Published in final edited form as:

Psychol Addict Behav. 2008 March ; 22(1): 68–77. doi:10.1037/0893-164X.22.1.68.

Heavy Episodic Drinking: Determining the Predictive Utility of Five or More Drinks

Kristina M. Jackson

Brown University

Abstract

Although the heavy episodic drinking (HED) measure of 5+ drinks (sometimes 4+ for women) is used extensively, there is no empirical basis for the designation of 5 drinks as the threshold (vs. another threshold that may perform equally). The present study sought to determine the threshold for HED that maximally predicts proximal and distal adverse-drinking-related outcomes. Participants included 115 young adults (57% female; 96% Caucasian) who partook in an 8-week Internet survey that assessed daily drinking as well as next-day hangover; 10 months later, adverse outcomes (problem drinking, alcohol-related problems, maximum number of drinks, and drug use) were surveyed. Thresholds were computed, with a range from 1+ drinks to 15+ drinks, and outcomes were predicted from each threshold. Findings for hangover measures showed relatively good convergence across multiple indicators, with greatest prediction occurring at a threshold of 10+ drinks per occasion. Different thresholds were observed for long-term outcomes, with higher thresholds indicative of outcomes with greater severity. Although alternatives to HED, such as subjective effects and blood alcohol concentration, can indicate risky drinking, a threshold measure of HED may have advantages in terms of prevention and of intervention efforts.

Keywords

heavy episodic drinking; binge drinking; threshold; alcohol; measurement

One of the goals of public health research is understanding the risks and the benefits associated with alcohol consumption (Gunzerath, Faden, Zakhari, & Warren, 2004). Commensurate with the recommendations of the U.S. Department of Agriculture (U.S. Department of Agriculture, 2007), consumption of one to two drinks per day is associated with the lowest mortality risk for the adult U.S. population (Gunzerath et al., 2004). Research is increasingly showing that with regard to alcohol-associated risk, particularly among youths, drinking patterns are of greater importance than is volume consumed (generally measured as the product of frequency of consumption and of quantity consumed on a “typical” occasion). Indeed, at a given volume, less frequent heavy use is associated with greater risk of experiencing a negative outcome than is frequent light drinking (Midanik, Tam, Greenfield, & Caetano, 1996; Wechsler & Nelson, 2001). Occurrence or frequency of heavy episodic drinking (HED), which gauges heavy consumption in a short time frame (Wechsler & Nelson, 2001), has become a key metric in estimations of the prevalence of problematic alcohol involvement. Those who drink heavily on a given occasion are more likely to experience negative outcomes (e.g., injury, driving after drinking, alcohol dependence) compared with those who do not drink heavily on that occasion (e.g., Knight et al., 2003; Midanik et al., 1996; Wechsler, Davenport, Dowdall, Moeykens, &

Castillo, 1994; Wechsler, Dowdall, Maenner, Gledhill-Hoyt, & Lee, 1998; Wechsler & Nelson, 2001).

Although other thresholds have been proposed (e.g., 8+ drinks, 12+ drinks; Conigrave, Saunders, & Reznik, 1995; Knupfer, 1984; Nadeau, Guyon, & Bourgault, 1998; White, Kraus, & Swartzwelder, 2006), 5 or more drinks on a single occasion has by far been the standard definition of HED. The origin of this criterion is unclear. The 5+ threshold was used by Room (1972) and was based on items by Cahalan, Cisin, and Crossley (1969) that assessed proportion of occasions on which a given number of drinks was consumed. Since then, Midanik (1999) has noted that “the 5+ measure has generally maintained its status of being the standard for measuring heavy episodic drinking in general population alcohol surveys” (p. 888). An exception to this threshold is the gender-specific measure (5+ for men; 4+ for women; Wechsler & Austin, 1998; Wechsler et al., 1994; Wechsler, Dowdall, Davenport, & Rimm, 1995), which has been widely adopted in the field.

Yet, there is no empirical basis for the designation of 5 drinks (vs. some other criterion that performs equally well) as the HED threshold (DeJong, 2001). As noted by Perkins, DeJong, and Linkenbach (2001), “Research literature does not provide a justification for a singular focus on a cutoff point of 5+/4+ drinks” (p. 317). Risk-function analyses, which use “typical” volume rather than heavy consumption on a single occasion, have failed to pinpoint the best threshold for prediction of negative consequences (Midanik et al., 1996; Room, Bondy, & Ferris, 1995; Wechsler et al., 1995; Weitzman & Nelson, 2004). They suggest that risk increases monotonically with volume, although considerable risk emerges at just 1 drink per day (Weitzman & Nelson, 2004).

The World Health Organization (2000) has suggested a threshold of 60 g as a cutoff for risky drinking; this volume corresponds to roughly 5 or more drinks per occasion. The Advisory Council Task Force of the National Institute on Alcohol Abuse and Alcoholism (NIAAA) recommended in 2004 that the definition of HED be based on drinking behaviors that elevate an individual’s blood alcohol concentration (BAC) up to or above the level of 0.08% (NIAAA, 2004). This level corresponds to 5 drinks in 2 hr for men and to 4 drinks in 2 hr for women, although clearly there is variability due to factors such as body mass, age, and recent drug and food ingestion.

However, studies examining BAC and the 5+/4+ measure are open to interpretation. Mean BAC for heavy episodic drinkers hovers around .08% (Beirness, Foss, & Vogel-Sprott, 2004; Lange & Voas, 2001), which is consistent with the NIAAA’s definition and with the legal limit for driving (Insurance Institute for Highway Safety, 2005). Yet, Beirness et al. (2004) observed that only 18% of those who met the 5+/4+ criteria exceeded a BAC of .08%, and Lange and Voas (2001) found that over half of drinkers returning from Tijuana who had consumed 5+ drinks had BACs < .06%. Lange and Voas calculated that it would take 8.2 drinks for men and 6.7 drinks for women to reach a BAC of .08% and that 5+ drinks corresponds to a BAC of .048% for men and of .072% for women. Using survey data, Perkins et al. (2001) estimated maximum BAC on the basis of last and of typical drinking episode and of sex, body weight, and drinking habits. Only 52% and 63% of those who drank heavily reached .08% BAC for last and for typical drinking episode, respectively. On the basis of this work, the 5+/4+ measure does not reliably identify drinkers with high BACs, and consumption of 5+/4+ drinks on an occasion is a poor marker of acute impairment that leads to adverse outcomes (Beirness, Foss, & Vogel-Sprott, 2005). Borsari, Neal, Collins, and Carey (2001) recommended further examination of the relationship between different cutoffs for HED and alcohol-related consequences. This is the goal in the present study.

The present study took an empirical approach to determination of the threshold for HED that maximally predicted both proximal and distal adverse-drinking-related outcomes. With data from a daily survey, thresholds were computed that ranged from 1 or more drinks to 15+ drinks. Outcome variables included next-day hangover as well as more distal outcomes, such as problem drinking, alcohol-related problems, maximum number of drinks, and drug use. These outcomes were selected because they reflect problematic substance involvement of varying degrees. Also, preliminary evidence suggests that maximum number of drinks consumed on a given occasion is a genetically influenced endophenotype (Bierut et al., 2002). In addition, because a criticism of HED research is the failure to account for body mass, duration of drinking episode, and sex (DeJong, 2001), these factors are controlled. Given the recommendations for a sex-specific HED measure, findings are presented separately for men and for women.

Method

Participants and Procedure

Data were taken from a larger study that examined the association between daily recordings of alcohol and tobacco use with a nonclinical young adult sample. A sample of 115 college students in three cohorts (Cohort 1 $N = 18$, Cohort 2 $N = 33$, Cohort 3 $N = 64$) undertook an 8-week Internet survey that assessed daily drinking and smoking as well as positive and negative mood and positive and negative stressors. Participants were recruited from a pool of introductory psychology students who attended a large Midwestern university. In all, 939 participants were contacted, either by telephone ($n = 336$) or by e-mail ($n = 603$), and were assessed for eligibility; participants were eligible if they (a) had smoked at least 100 lifetime cigarettes, (b) had reported past-month smoking and past-month drinking, and (c) had consumed more than five lifetime drinks. Because the original study was focused on tobacco use as well as on drinking, smokers were oversampled.

The first 115 (12.2% of $N = 939$) respondents who met eligibility criteria and were scheduled successfully participated in the study. Of these, 57% were female, 96% were Caucasian, and 90% were either 18 or 19 years old (primarily freshmen or sophomores). In addition, 100% of the respondents had consumed alcohol in the past 30 days, 88.7% had drunk five or more drinks in the prior 2 weeks, 52.2% were daily smokers, and 28.7% reported smoking at least one half pack per day. There were no cohort differences in sex, number of drinks consumed, or any of the outcome variables with the exception of hangover, which showed slightly higher levels in the second cohort. However, this is not a meaningful effect, and cohorts were collapsed for the purpose of analysis.

Prior to survey commencement, participants underwent a training session that described study procedure and use of the Web survey as well as the definition of a standard drink. During the training session, participants completed a paper-and-pencil survey that assessed substance use, personality, motivations for substance use, family history of substance use, mood, and other psychosocial constructs. After 1 week (Cohort 1), 2 weeks (Cohort 2), or 3 weeks (Cohort 3), participants began the daily Web-based, 26-item survey. Approximately 10 months following the Web survey, the paper-and-pencil survey was readministered. Compensation (course credit or monetary) was based on weekly participation, with bonuses for completing assessments on time.

Measures

Drinking (daily survey)—Prior-day alcohol consumption was assessed with an item that indexed quantity of drinks consumed, with a range from 0 to 25 or more ($M = 2.55$ drinks, $SD = 4.74$). To create alternate HED thresholds, this quantity variable was categorized 15

different ways: 1 drink versus no drinks, 2 drinks versus 0 or 1 drink, 3 drinks versus 2 or fewer drinks, and so on, up to 15 or more (vs. 14 or fewer) drinks.

Proximal outcome (daily survey)—Current-day hangover symptoms were taken from Slutske, Piasecki, and Hunt-Carter (2003) and included five items: had a headache, felt more tired than usual, had difficulty concentrating on things, felt nauseated, and felt very weak. Item responses ranged from 1 = *not at all* to 7 = *extremely*. A hangover scale was formed by taking the mean across the items ($M = 1.30$, $SD = 0.84$; Cronbach's α across time = .92; α aggregated across individuals = .94).

Distal outcomes (paper-and-pencil survey)—Problem drinking was measured with a single item (i.e., “How would you best describe yourself in terms of your current use of alcohol?”), which was scored on a scale with the following response options: 1 = *abstainer*, 2 = *abstainer—former problem drinker and/or in recovery*, 3 = *infrequent drinker*, 4 = *light drinker*, 5 = *moderate drinker*, 6 = *heavy drinker*, and 7 = *problem drinker*. Endorsement of either *heavy drinker* or *problem drinker* was coded as “problem drinker.”

Past-year alcohol-related problems were assessed with the 37-item Young Adult Alcohol Problems Screening Test (YAAPST; Hurlbut & Sher, 1992). Items include problems relevant to college student populations (e.g., missing class, getting involved in regrettable sexual situations) as well as more general problems (hangovers, blackouts, driving while intoxicated). Response options were 1 = *Not in the past year*, 2 = *Once in the past year*, 3 = *Twice in the past year*, 4 = *3 times in the past year*, and 5 = *4+ times in the past year*.

Maximum number of drinks was assessed by a single open-ended item asking, “What is the maximum number of drinks you have had in one sitting in the past 30 days?” Responses ranged from 0 to 23 ($M = 10.32$ drinks, $SD = 4.91$).

Number of drugs used was a mean across nine classes of drugs, including (a) marijuana, (b) ecstasy, (c) other club drugs, (d) inhalants, (e) stimulants or amphetamines, (f) crack or other forms of cocaine, (g) psychedelics or hallucinogens, (h) barbiturates or tranquilizers, and (i) heroin or other opiate drugs. Response options were 1 = *Not in the past year*, 2 = *Once in the past year*, 3 = *2–5 times in the past year*, 4 = *6–10 times in the past year*, 5 = *11–20 times in the past year*, 6 = *21–40 times in the past year*, and 7 = *41+ times in the past year*. Given the skew that resulted from conversion of responses to days per year, the ordinal nature of the scales was retained.

Attrition

Of the 115 respondents, 81 (70%) completed the Wave 2 (distal outcome) survey. Of the 34 who did not participate in Phase 2, 3 refused further participation in the study, 5 refused Phase 2 (because of poor computer access or limited time), 9 were contacted but failed to complete or to return the survey and consent form, and 17 could not be located or reached. Three participants (4%) who had relocated returned the Wave 2 survey via mail. Attrition analyses determined whether those lost to follow-up differed from those who were retained on relevant variables. Those lost to follow-up did not differ from those assessed at Wave 2 on age, race, alcohol consumption (age of onset, maximum number of drinks, past-year and past 30-day frequency and quantity, and past 30-day frequency of HED based on 5+ drinks or on 12+ drinks), problem drinker status, alcohol-related consequences, or drug use. The only significant effect was that men were less likely to be retained than were women, $\chi^2(1, N = 115) = 8.08$, $p < .01$.

Data Analysis

For each set of analyses, a series of 15 models was conducted, each with an alternate definition of heavy drinking. That is, for a given outcome, 15 models were estimated, one with heavy drinking defined as one or more drinks, one with heavy drinking defined as two or more drinks, one with heavy drinking defined as three or more drinks, and so on.¹ This approach enabled comparison of parameters across the 15 models and determination of which of the models was maximally predictive. Consistent with Borsari et al. (2001), the best HED threshold was based on the largest regression parameter value. Although goodness-of-fit indices could be used in pinpointing the best fitting model, these indices reflect both fixed and random effects. The goal in the current study was determination of the best fixed effect of the binary threshold variable, which is indicated by the largest parameter. Because the parameter for analyses with high thresholds (which have low base rates) may be a function of a few influential observations (events) that have high leverage values, the goal was to look for consistent patterns across thresholds and variables, rather than to base conclusions on a single parameter from a given analysis.

Because hangover outcomes were collected on a daily level, the model was estimated with PROC MIXED in SAS, with the hangover item as the outcome and the binary heavy drinking threshold variable as the randomly varying predictor. Mixed models were selected because they account for clustering of observations within individuals. Models controlled for sex, body mass, and duration of drinking episode as covariates. The measure of body mass was computed on the basis of weight and of height (weight in pounds divided by height in inches squared, multiplied by 703). However, body mass was unrelated to any outcome variables, and parameter values were unchanged when the variable was removed from the model. Therefore, the analyses described below did not control for body mass. Sex was modeled as a fixed effect, and duration of drinking episode was modeled as a random effect. Because there was little or no random variability in the intercept once duration was controlled, intercept was modeled as a fixed effect.² When no drinking was reported on a given day, duration of drinking episode was set to the mean value for the full sample (or within gender for the subgroup analyses). That is, this covariate was a constant for nondrinkers. Setting these values to missing would have eliminated these events from analyses, which would have markedly reduced sample size and generalizability.

The analytic method for the distal outcomes was an attempt to conduct a parallel analysis to the multilevel modeling analysis for prediction of hangover. It is not possible to predict a fixed outcome from a random effect in a single model, so the analysis had two steps.³ An intercept-only model was estimated for each binary heavy drinking threshold with PROC NL MIXED in SAS, and predicted values were computed as a function of the parameter estimates and of the empirical Bayes estimates of the random effects. These empirical Bayes estimates of the random effects were output. Each individual, then, had a single empirical Bayes estimate that was a measure of that HED threshold aggregated across the 56 time points. In the second step, the empirical Bayes estimates were predictors in a regression or a logistic regression, with the

¹In the creation of the thresholds, those who reported zero drinks were included, because the HED measure typically compares five or more drinks with four or fewer drinks (vs. consumption of some number between one and four drinks). This article sought to be consistent with the extant measure.

²Although an unconditional model suggested there is random variance in the intercept, when number of drinks was controlled, the variance of the intercept was close to zero. This is because the level of hangover is relatively constant when number of drinks is controlled (i.e., if there is no drinking, there is no hangover).

³An alternative to this two-step analysis is an analysis where distal outcomes are predicted from a sum of the number of occasions a given threshold was reached over the 8-week interval. However, findings showed that in all cases, the 15+ threshold was consistently the strongest predictor of the outcome. This analysis apparently differentiates the very heavy drinkers from the rest, consistent with much of the risk function work on the relationship between consumption and harm (e.g., Wechsler & Nelson, 2006; Weitzman & Nelson, 2004), and is less informative with regard to the best HED threshold.

distal outcomes as the dependent variable. As the focus was on prediction of a given behavior (rather than on change in behavior), corresponding Year 1 behaviors were not controlled.

Results

Descriptive Information

Out of 6,440 (56 days $\times$ 115 participants) possible person days, 5,930 (92%) were obtained. Daily retention ranged from 71% (the last survey day) to 100%; the median retention rate was 93%. Retention dropped monotonically over the course of study. Over the 56-day interval, individuals engaged in drinking on 29.4% of the days. When data were aggregated over occasion, 29.3% of participants drank (rates from 0% to 71.0%). Considering aggregated weekly drinking rates, drinking decreased over the course of the study, multivariate $F(7, 87) = 4.44, p < .001$. There was no sex difference in drinking over time, multivariate $F(7, 87) = 1.59, ns$.

Prevalence of days that individuals consumed alcohol at each of the HED thresholds ranged from 29.4% of the days (1+ drink vs. 0 drinks) to 3.3% of the days (15+ drinks vs. 14 or fewer drinks). Rates are presented separately for men and for women in Figure 1. Table 1 presents descriptive information for the hangover outcome measures aggregated across days ($N = 5,892$ – $5,906$) and aggregated across individuals ($N = 115$) as well as information for the distal outcomes (aggregated across individuals; $N = 81$).

Comparing Predictive Utility of Different Thresholds

Proximal outcome—Multilevel models were used for prediction of next-day hangover from each of the 15 HED thresholds. Separate models were run for the five hangover symptoms and for the full scale; Figure 2 presents unstandardized coefficients for each threshold. Confidence intervals are shown for the full sample. Bearing in mind that there are no formal tests for determination of the maximally predictive outcome, it appears that there is a trend for the greatest regression coefficient to hover around 10+ drinks for the full sample, although there was some predictive utility at the high end for some of the outcomes, particularly for having trouble concentrating.

For men, the maximally predictive threshold was generally 13+ drinks; again, high thresholds showed the strongest prediction for having trouble concentrating. For women, the maximally predictive threshold again hovered around 10+, the same value that was observed for the full sample, but there was also high predictability at 14+ drinks.

Distal outcomes—Next, the relation between HED and distal outcomes was explored with a two-part analysis designed for comparability with a multilevel model that was used in the proximal outcome analyses. The first step calculates the predicted random effects for each participant, and the second step predicts each distal outcome from these values. Figure 3 shows regression parameters for each outcome for the full sample, for men, and for women. For problem drinking, a low threshold (1–3 drinks) was most predictive. For men, the threshold of 15+ drinks showed the best prediction, with some prediction by 10+ drinks. For women, low levels were most predictive; in fact, the effect at higher thresholds was in the opposite direction.

For maximum number of drinks in the past 30 days, the most predictive threshold tended to be at the low-to-middle values, although this was true to a lesser extent for men. For past-year number of drugs, very low and very high thresholds showed maximum prediction, although for men, thresholds in the range of 8+ to 10+ tended to also predict drug use.

For alcohol-related consequences (YAAPST), low-to-middle values were consistently the most predictive. Prediction of YAAPST by lower thresholds may occur in part because some

of the consequences are mild in nature, and these consequences are in essence driving the analysis. This idea was further examined by formation of a three-item scale that included mild consequences (passed out, acted obnoxious or rude, been a passenger in a car with a drunk driver). The figure suggests that the predictability by low thresholds was even stronger, although this finding appeared to be primarily driven by women. All thresholds seemed to perform equally well for men.

For all distal outcomes, there was an elevation at the highest threshold (15+). This finding suggests the presence of a few outliers who drank at very high levels and who experienced adverse distal outcomes (although the data suggest that consumption of 15 or more drinks was relatively spread out among 34 individuals).

Discussion

The present study examined the predictive utility of different thresholds for HED for both proximal and distal outcomes. For proximal hangover measures, findings showed relatively good convergence across multiple indicators, with the greatest prediction occurring at a threshold of 10 or more drinks (vs. 9 or fewer) per occasion, with low thresholds revealing poorer predictive utility. This was true for the full sample and for women; the threshold was higher for men, consistent with research advocating a sex-specific HED threshold (Wechsler et al., 1994, 1995; Wechsler & Austin, 1998). This finding suggests that heavy drinking college students may not be very sensitive to the physical symptoms of hangover, which can be an important indicator to the drinker that his or her heavy drinking is excessive and is potentially harmful. This lack of sensitivity can be misleading and, in fact, might inadvertently suggest to young drinkers that what is clearly heavy drinking (up to 9 drinks per occasion) is not harmful, although drinking at this level can result in a BAC at a very dangerous level, depending on the time frame. This issue may serve as a barrier to prevention efforts among college student drinkers. It is important to note, however, that the standard 5+ threshold was still highly predictive of hangover, and thresholds other than 10+ may show greater prediction of proximal outcomes not assessed here (e.g., immediate consequences such as missing class or a sexual situation that was later regretted).

For distal outcomes, the maximal threshold differed from outcome to outcome, and findings varied widely as a function of outcome and of subgroup. For problem drinking, low thresholds (including the traditional 5+ HED measure) showed greatest prediction, although sex differences were observed, with a much higher threshold for men. Women who drink in very large quantities over a short interval are less likely to label themselves as heavy drinkers, perhaps because they perceive higher consumption rates by their male peers as normative. Alternately, low prevalence rates at high thresholds for women may have led to unreliable estimates.

For alcohol-related consequences, a low threshold was most predictive, a finding consistent with Dawson (2000) and Borsari et al. (2001), who showed that volume predicted alcohol-related problems as strongly as did HED. For examination of the extent to which findings were due to the severity of the consequences, three minor consequences were selected and the model was reestimated. Consistent with expectation, the tendency for low thresholds (vs. high thresholds) to predict mild consequences was even more apparent. The sex difference whereby larger thresholds were more predictive for men in particular may be in part because more severe consequences were more likely to be experienced by men (22%) than by women (10%).

A low-to-moderate threshold (including the traditional 5+ threshold) consistently predicted maximum drinks in the past month. This finding suggests that what may constitute a single bout of drinking is not necessarily predicted by drinking pattern so much as it is predicted by

being a drinker. That is, an individual who generally drinks few drinks is just as likely to have a drinking episode marked by a large quantity as is an individual who tends to frequently drink relatively large numbers of drinks—in fact, the low-quantity drinker is slightly more likely to do so. It may be that high-quantity drinkers have learned to minimize negative consequences (e.g., by not drinking after driving), consistent with work that uses a harm reduction approach to college student drinking, but low-quantity drinkers have much more variability in their drinking and may actually account for the majority of risky alcohol behaviors (the preventive paradox; Kreitman, 1986). In fact, the preventive paradox disappears when low-quantity drinkers report their maximum drinking level (Stockwell, Hawks, Lang, & Rydon, 1996). In short, for maximum number of drinks in the past 30 days, the traditional 5+ threshold fares no better or worse than do the surrounding thresholds.

Very low and very high thresholds were most predictive of prior-year drug use. In general, low thresholds were more predictive for women than for men, although the highest threshold (15+) was also strong for women. There was also evidence that for men, thresholds surrounding 9+ drinks discriminated drug use well. As with alcohol-related consequences, women had less severe scores on the outcome measure (mean number of illicit drugs taken was $M = 1.51$ for women and $M = 1.85$ for men), which suggests that the most predictive threshold may vary along a severity gradient.

One final finding of interest was that for the proximal hangover outcome, consumption generally tended to be a stronger predictor for women than for men; conversely, distal outcomes were more strongly predicted by drinking for men than for women. This relation may be affected by sex differences in body mass that are associated with acute but not with chronic alcohol outcomes.

Utility of a Measure of HED

Concerns have been raised that HED is a term that represents “normative” drinking (Wechsler & Kuo, 2000) and that it connotes that there exists a level of “safe” drinking (DeJong, 2001), and the present study suggests that there is no clear-cut threshold for HED. What, then, are other options? Volume is a valid indicator of risky drinking, but detailed retrospective assessment is limited by recall bias. One measure that assesses both pattern and volume, the graduated frequency approach (Greenfield, 2000), determines frequencies of consumption at a series of levels. Unfortunately, problems in reducing all of the categories at the point of analysis are commonly encountered (Dawson, 2003). Another alternative is to focus on subjective “effect” drinking (e.g., getting high or drunk; Midanik, 1999), which is associated with HED (Wechsler & Nelson, 2001) as well as with alcohol-related harm and alcohol dependence symptoms (Midanik, 1999) but which avoids problems associated with variability in body mass, metabolism, and tolerance. Many individuals who engage in HED report drinking with the intention to get drunk (Naimi et al., 2003). Knupfer (1984) argued that frequency of intoxication is of greater importance than is number of drinks; she found that only a quarter of individuals who reported drinking 8+ drinks three times per week reported being drunk three times per week.

Yet another option would be to use an objective measure such as BAC (Beirness et al., 2005), which takes into account duration of drinking episode, metabolism, and body mass. BAC is difficult to assess in survey research, although it can be calculated from volume, duration, body mass, and sex (Borsari et al., 2001), and widely available BAC estimators can be personalized based on weight and on sex. Given that Borsari et al. failed to find unique predictive utility of BAC over weekly drinking when predicting alcohol-related problems, however, it is still unclear whether this measure warrants the cost.

Abandoning the use of a threshold HED measure altogether seems premature. Inclusion criteria for entry into prevention and treatment programs are frequently based on HED rates. The best strategy may be to choose a consumption measure that is meaningful for the research question. Total volume may be a better determinant of chronic health problems and mortality than of social and acute consequences (Mäkelä, 1996; Midanik et al., 1996), whereas drinking and driving rates are affected by HED but not by chronic drinking (Duncan, 1997). Findings by others (Wechsler & Nelson, 2006) and in the present study support this: Outcomes with greater severity tended to be maximally predicted by a higher threshold than were outcomes with lower severity.⁴

A researcher using a HED measure must weigh the importance of sensitivity versus specificity. A measure with high specificity will be more stringent but will capture fewer at-risk drinkers (Type II error). In contrast, a measure with high sensitivity will capture more at-risk and problem drinkers but may overclassify individuals (Type I error). Evidence of thresholds greater than the traditional 5+ measure reflect high specificity, which is useful for accurate screening but less so for prevention of alcohol-related harms (LaBrie, Pedersen, & Tawalbeh, 2007; Wechsler & Nelson, 2006). On the other hand, our findings showing that the traditional 5+ HED threshold was generally a better predictor of distal outcomes than were the subsequent (higher) thresholds suggests that a sensitive measure might be most useful in indicating risk for alcohol-related harm. Ascertaining the maximally predictive threshold measure for different adverse outcomes also has implications for stepped care that provides varying degrees of intervention on the basis of level of risky alcohol use. Decision rules could be based on different thresholds, depending on the outcome of interest.

One final note concerns how the alcohol field might label a heavy drinking measure that reflects pattern of drinking. Because the term *binge drinking* was redefined by the NIAAA (NIAAA, 2004), a threshold measure ought to be given a distinct label, such as “heavy episodic drinking.” However, researchers should accept both terms with recognition that each has a different meaning.

Limitations

The present study benefited from a large number of repeated assessments, the capability to calculate different HED thresholds, and prospective assessment of distal outcomes. However, it is important to acknowledge the limitations of the sample, which consisted of a subset of college student smokers from a single university. The present sample was more likely to drink (88.7% vs. 41.7% reporting 5+ drinks in the past 2 weeks) and to smoke (52.2% vs. 13.8% reporting daily smoking) compared with a nationally representative sample of college students from *Monitoring the Future* (Johnston, O’Malley, Bachman, & Schulenberg, 2006) and compared with an independent sample drawn from the same university (88% of the student body; Sher & Rutledge, 2007), although these two university samples were roughly comparable in terms of sex and race (data not shown).

In addition, college students in general are more likely to drink heavily than are their noncollege student peers (Johnston et al., 2006; O’Malley & Johnston, 2002). However, having a heavily drinking sample was valuable for exploration of very high HED thresholds. Finally, the small sample size may have reduced variability in the outcome variables. Attrition at Wave 2, although not unreasonable for a complex study such as this, was somewhat high (30%), which

⁴As an ancillary analysis for exploration of the idea that outcomes with greater severity were better predicted by a higher threshold than were ones with lower severity, a variable was created that indicated whether a hangover was experienced (0 = *no hangover*, 1 = *some degree of hangover*). This variable is a less severe outcome than is the continuous measure that indicates degree of hangover. As expected, the most predictive threshold for whether a hangover was experienced was lower than was the best threshold for the continuous hangover measure.

may have led to unstable parameter estimates, particularly in subgroups ($n = 28$ for men). Yet, the random effects were calculated on approximately 1,490 observations; only the regression model was based on $n = 28$. Still, the small number of male Wave 2 participants is a limiting factor.

Conclusion

The ultimate aim in this study was not to recommend a new threshold for the HED measure. This would be confusing for researchers, clinicians, and the public (DeJong, 2001); indeed, the traditional 5+ HED measure was a very good indicator of alcohol-related risk. Rather, study findings suggest that no single definition of HED is sufficient to characterize different aspects of risky drinking. Although having a consistent measure enables comparison across studies (Wechsler & Nelson, 2006), having a standard measure may be at the expense of understanding the myriad adverse consequences of HED. It is critical that researchers be thoughtful about their research question when characterizing HED (e.g., desiring high specificity for screening vs. high sensitivity for prevention). The present study is a crucial first step toward identifying an empirically meaningful threshold of HED and, it is hoped, will inspire further research with larger and more representative samples.

Acknowledgments

Preparation of this article was supported by National Institute on Alcohol Abuse and Alcoholism Grant K01 AA13938 to Kristina M. Jackson. I thank Andrew Littlefield for his assistance with data collection and data management and Don Hedeker, Denis McCarthy, Ken Sher, and Lance Swenson for their assistance in analyses and their helpful comments on previous drafts of this article.

References

- Beirness DJ, Foss RD, Vogel-Sprott M. Drinking on campus: Self-reports and breath tests. *Journal of Studies on Alcohol* 2004;65:600–604. [PubMed: 15536769]
- Beirness DJ, Foss RD, Vogel-Sprott M. Drinking on campus: Self-reports and breath tests: Response. *Journal of Studies on Alcohol* 2005;66:439–440.
- Bierut LJ, Saccone NL, Rice JP, Goate A, Foroud T, Edenberg H, et al. Defining alcohol-related phenotypes in humans: The Collaborative Study on the Genetics of Alcoholism. *Alcohol Research and Health* 2002;26:208–213. [PubMed: 12875049]
- Borsari B, Neal DJ, Collins SE, Carey KB. Differential utility of three indexes of risky drinking for predicting alcohol problems in college students. *Psychology of Addictive Behaviors* 2001;15:321–324. [PubMed: 11767264]
- Cahalan, D.; Cisin, IH.; Crossley, HM. American drinking practices: A national study of drinking behavior and attitudes (Monograph No. 6). New Brunswick, NJ: Rutgers Center of Alcohol Studies; 1969.
- Conigrave KM, Saunders JB, Reznik RB. Predictive capacity of the AUDIT questionnaire for alcohol-related harm. *Addiction* 1995;90:1479–1485. [PubMed: 8528033]
- Dawson DA. Alternative measures and models of hazardous consumption. *Journal of Substance Abuse* 2000;12:79–91. [PubMed: 11288476]
- Dawson DA. Methodological issues in measuring alcohol use. *Alcohol Research and Health* 2003;27:18–29. [PubMed: 15301397]
- DeJong W. Finding common ground for effective campus-based prevention. *Psychology of Addictive Behaviors* 2001;15:292–296. [PubMed: 11767259]
- Duncan DF. Chronic drinking, binge drinking and drunk driving. *Psychological Reports* 1997;80:681–682. [PubMed: 9129384]
- Greenfield TK. Ways of measuring drinking patterns and the difference they make: Experience with graduated frequencies. *Journal of Substance Abuse* 2000;12:33–49. [PubMed: 11288473]
- Gunzerath L, Faden V, Zakhari S, Warren K. National Institute on Alcohol Abuse and Alcoholism report on moderate drinking. *Alcoholism: Clinical and Experimental Research* 2004;28:823–847.

- Hurlbut SC, Sher KJ. Assessing alcohol problems in college students. *Journal of American College Health* 1992;41:49–58. [PubMed: 1460173]
- Insurance Institute for Highway Safety, Highway Loss Data Institute. DUI/DWI laws. 2005. Retrieved April 21, 2006, from http://www.iihs.org/laws/state_laws/dui.html
- Johnston, LD.; O'Malley, PM.; Bachman, JG.; Schulenberg, JE. Secondary school students (Publication No. 06–5883). Vol. I. Bethesda, MD: National Institute on Drug Abuse; 2006. Monitoring the Future national survey results on drug use, 1975–2005.
- Knight JR, Wechsler H, Kuo M, Seibring M, Weitzman ER, Schuckit MA. Alcohol abuse and dependence among U.S. college students. *Journal of Studies on Alcohol* 2003;63:263–270. [PubMed: 12086126]
- Knupfer G. The risks of drunkenness (or, ebrietas resurrecta): A comparison of frequent intoxication indices and of population subgroups as to problem risks. *British Journal of Addiction* 1984;79:185–196. [PubMed: 6611166]
- Kreitman N. Alcohol consumption and the preventive paradox. *British Journal of Addiction* 1986;81:353–363. [PubMed: 3461846]
- LaBrie JW, Pedersen ER, Tawalbeh S. Classifying risky-drinking college students: Another look at the two-week drinker-type categorization. *Journal of Studies on Alcohol and Drugs* 2007;68:86–90. [PubMed: 17149521]
- Lange JE, Voas RB. Defining binge drinking quantities through resulting blood alcohol concentrations. *Psychology of Addictive Behaviors* 2001;15:310–316. [PubMed: 11767262]
- Mäkelä K. How to describe the domains of drinking and consequences. *Addiction* 1996;91:1447–1449.
- Midanik LT. Drunkenness, feeling the effects and 5+ measures. *Addiction* 1999;94:887–897. [PubMed: 10665077]
- Midanik LT, Tam TW, Greenfield TK, Caetano R. Risk functions for alcohol-related problems in a 1988 U.S. national sample. *Addiction* 1996;91:1427–1437. [PubMed: 8917925]
- Nadeau L, Guyon L, Bourgault C. Heavy drinkers in the general population: Comparison of two measures. *Addiction Research* 1998;6:165–187.
- Naimi TS, Brewer RD, Mokdad A, Denny C, Serdula M, Marks J. In reply: “Definitions of binge drinking. *Journal of the American Medical Association* 2003;289:1636. [PubMed: 12672729]
- National Institute on Alcohol Abuse and Alcoholism. NIAAA council approves definition of binge drinking. NIAAA Newsletter. 2004 Winter. Retrieved September 23, 2005, from http://pubs.niaaa.nih.gov/publications/Newsletter/winter2004/Newsletter_Number3.htm
- O'Malley PM, Johnston LD. Epidemiology of alcohol and other drug use among American college students. *Journal of Studies on Alcohol* 2002;(Suppl 14):23–39.
- Perkins HW, DeJong W, Linkenbach J. Estimated blood alcohol levels reached by “binge” and “nonbinge” drinkers: A survey of young adults in Montana. *Psychology of Addictive Behaviors* 2001;15:317–320. [PubMed: 11767263]
- Room R. Drinking patterns in large U.S. cities: A comparison of San Francisco and national samples. *Quarterly Journal of Studies on Alcohol* 1972;(Suppl 6):28–57. [PubMed: 5031415]
- Room R, Bondy S, Ferris J. The risk of harm to oneself from drinking, Canada 1989. *Addiction* 1995;90:499–513. [PubMed: 7773113]
- Sher KJ, Rutledge PC. Heavy drinking across the transition to college: Predicting first-semester heavy drinking from precollege variables. *Addictive Behaviors* 2007;32:819–835. [PubMed: 16860940]
- Slutske WS, Piasecki TM, Hunt-Carter EE. Development and initial validation of the Hangover Symptoms Scale: Prevalence and correlates of hangover symptoms in college students. *Alcoholism: Clinical and Experimental Research* 2003;27:1442–1450.
- Stockwell T, Hawks D, Lang E, Rydon P. Unravelling the preventive paradox for acute alcohol problems. *Drug and Alcohol Review* 1996;15:7–15. [PubMed: 16203346]
- U.S. Department of Agriculture. Dietary guidelines for Americans 2005: Alcoholic beverages. 2007. Retrieved April 26, 2007, from <http://www.health.gov/dietaryguidelines/dga2005/document/html/chapter9.htm>
- Wechsler H, Austin SB. Binge drinking: The five/four measure. *Journal of Studies on Alcohol* 1998;59:122–123. [PubMed: 9498324]

- Wechsler H, Davenport A, Dowdall G, Moeykens B, Castillo S. Health and behavioral consequences of binge drinking in college: A national survey of students at 140 campuses. *Journal of the American Medical Association* 1994;272:1672–1677. [PubMed: 7966895]
- Wechsler H, Dowdall GW, Davenport A, Rimm EB. A gender-specific measure of binge drinking among college students. *American Journal of Public Health* 1995;85:982–985. [PubMed: 7604925]
- Wechsler H, Dowdall GW, Maenner G, Gledhill-Hoyt J, Lee H. Changes in binge drinking and related problems among American college students between 1993 and 1997. *Journal of American College Health* 1998;47:57–68. [PubMed: 9782661]
- Wechsler H, Kuo M. College students define binge drinking and estimate its prevalence: Results of a national survey. *Journal of American College Health* 2000;49:57–64. [PubMed: 11016129]
- Wechsler H, Nelson TF. Relationship between level of consumption and harms in assessing drink cut-points for alcohol research: Commentary on “Many college freshman drink at levels far beyond the binge threshold” by White et al. *Psychology of Addictive Behaviors* 2001;15:287–291. [PubMed: 11767258]
- Wechsler H, Nelson TF. Binge drinking and the American college student: What’s five drinks? *Alcoholism: Clinical and Experimental Research* 2006;30:922–927.
- Weitzman ER, Nelson TF. College student binge drinking and the “prevention paradox”: Implications for prevention and harm reduction. *Journal of Drug Education* 2004;24:247–265. [PubMed: 15648886]
- White AM, Kraus CL, Swartzwelder HS. Many college freshman drink at levels far beyond the binge threshold. *Alcoholism: Clinical and Experimental Research* 2006;30:1006–1010.
- World Health Organization. WHO Publication No. WHO/MSD/MSB/00, 4, 2000. Geneva, Switzerland: Author, Department of Mental Health and Substance Abuse; 2000. International guide for monitoring alcohol consumption and related harm.

Figure 1.
Prevalence of days that men and women consumed alcohol at each of the heavy episodic drinking thresholds.

Figure 2. Unstandardized regression parameters for multilevel models predicting hangover symptomatology from heavy drinking comparisons, for full sample (square), for men (triangle), and for women (circle).

Figure 3.

Unstandardized regression parameters for models predicting distal outcomes 1 year later from heavy drinking comparisons, for full sample (square), for men (triangle), and for women (circle). For problem drinking, the parameter for the highest threshold (15+) for women was $\beta = -2.03$, but for the sake of presentation it was graphed at $\beta = -.38$, the value for the 14+ threshold.

Table 1

Descriptive Statistics for Proximal and Distal Outcome Measures, Aggregated Across Days (Proximal Outcomes) and Aggregated Across Individuals (Proximal and Distal Outcomes)

Outcome measure	Full sample				Men				Women			
	Aggregated across days		Aggregated across individuals		Aggregated across days		Aggregated across individuals		Aggregated across days		Aggregated across individuals	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
(N = 5,892–5,906)												
(N = 115)												
(N = 2,433–2,443)												
(N = 3,453–3,463)												
(N = 65)												
Proximal outcome												
Full hangover scale	1.30	0.84	1.31	0.25	1.29	0.79	1.30	0.27	1.32	0.87	1.31	0.29
Had a headache	1.26	0.90	1.26	0.24	1.22	0.82	1.23	0.25	1.28	0.95	1.28	0.24
Felt more tired than usual	1.50	1.26	1.51	0.38	1.51	1.25	1.53	0.44	1.50	1.26	1.50	0.32
Trouble concentrating on things	1.27	0.89	1.27	0.27	1.28	0.91	1.29	0.29	1.26	0.87	1.26	0.26
Felt nauseated	1.20	0.78	1.20	0.21	1.16	0.67	1.17	0.22	1.22	0.86	1.22	0.21
Felt very weak	1.29	0.89	1.29	0.27	1.27	0.85	1.28	0.28	1.30	0.92	1.30	0.26
(N = 81)												
(N = 28)												
(N = 53)												
Distal outcome												
YAAPST	35.57		19.32				41.79		23.13		32.28	
Maximum no. drinks, past 30 days	10.32		4.91				13.61		4.86		8.58	
No. times did illicit drugs, past year	1.63		0.79				1.85		0.88		1.51	

Note. YAAPST = Young Adult Alcohol Problems Screening Test.