After how many drinks someone experiences acute consequences? Determining thresholds for binge drinking based on two event-level studies

Running head: Optimal thresholds for binge drinking

Florian Labhart[1,2], *Michael Livingston*[3,4], *Rutger Engels*[2,5], *Emmanuel Kuntsche*[3,6,7]

[1]Addiction Switzerland, Research Institute, Lausanne, Switzerland;
[2]Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands;
[3]Centre for Alcohol Policy Research, La Trobe University, Australia;
[4]Department of Clinical Neuroscience, Karolinska Institutet, Sweden;
[5]Trimbos-institute, Netherlands Intitute of Mental Health and Addiction, Utrecht, The Netherlands;
[6]Behavioural Science Institute, Radboud University, Nijmegen, the Netherlands;
[7]Faculty of Education and Psychology, Eötvös Loránd University, Budapest, Hungary

Correspondence should be addressed to Florian Labhart, Addiction Switzerland, Research Department, PO Box 870, CH 1001 Lausanne, SWITZERLAND; Phone: +41 21 321 29 51; Fax: +41 21 321 29 40; E-Mail: flabhart@addictionsuisse.ch

The authors do not have any conflicts of interest. The study was funded by the Swiss National Science Foundation (grant no. CR1I1_120181/1) and approved by the Lausanne and Zurich Cantonal Ethics Committees for Research on Human Beings (protocol 145/14). M. Livingston is supported by an NHMRC Career Development Fellowship (1123840). The Centre for Alcohol Policy Research is funded by the Foundation for Alcohol Research and Education, an independent, charitable organization working to prevent the harmful use of alcohol in Australia http://www.fare.org.au.

Word count: 3,446

Abstract

Aims: The threshold of 4+/5+ drinks per occasion has been used in alcohol research to distinguish between non-risky vs. risky episodic drinking for decades. Yet, no study has assessed the validity of this threshold using event-level data. This study aims to determine the optimal thresholds for the detection of five acute alcohol-related consequences (hangover, blackout, risky sex, fights and injury) using data from two event-level studies.

Methods: On 3,554 weekend nights, 369 participants, aged 16 to 25 years, documented their alcohol consumption and the occurrence of consequences the next morning. Separately for gender and age groups (16-17 vs. 18-25), the ability of number of drinks consumed to discriminate nights with and without consequence was measured using the area under the receiver operating characteristic (AUROC) curve. Optimal thresholds were determined using Youden Index and the shortest distance from perfect discrimination based on sensitivity and specificity.

Results: Hangover was the most frequently reported consequence and injury the least for both genders. Across age groups and studies, optimal thresholds for hangover only were equal to 4+/5+ while those for blackouts, risky sex, fights and injuries were up to 3 drinks higher. Adolescent men experienced consequences more often and at slightly lower drinking levels than adult men. For all consequences but injuries, the optimal thresholds were 1 to 2 drinks lower for women than for men.

Conclusions: Event-level data collection techniques appear particularly suitable to estimate thresholds at which acute alcohol-related consequences occur. In an epidemiological perspective, the 4+/5+ thresholds (corresponding to 40+/50+ grams of pure alcohol in these studies) appear slightly too low to optimally predict acute consequences of binge drinking. However, if the overall aim is simply to reduce negative consequences, using the standard 4+/5+ thresholds for drinking guidelines and advice to the public may still be warranted.

Keywords:

Alcohol-related consequences, Binge drinking, Optimal threshold, Event-level study

Citable statements:

- Binge drinking thresholds of 4+/5+ (women/men) drinks accurately predict the occurrence of hangovers but are too low for more severe acute alcohol-related consequences.
- Women and adolescents tend to experience acute alcohol-related consequences at slightly lower drinking levels (1-2 drinks less) than men and adults, respectively.
- Event-level data collection is particularly suitable to estimate thresholds at which acute alcohol-related consequences occur.

Introduction

Binge drinking is one of the most important concepts used in alcohol epidemiology to discriminate non-risky drinking from risky drinking and to determine the burden resulting from alcohol use [1]. This behavior has been linked to a wide range of acute (short-term; e.g. blackouts, injuries) and chronic (long-term; e.g. liver and cardiovascular diseases) consequences [2]. Interest in binge drinking has increased tremendously since the early nineties [3, 4], resulting in an immense body of almost 8,000 papers and 400 literature reviews published in the last three decades [5].

The most widely used definition of binge drinking is the consumption of 4 or more standards drinks among women, and 5 or more among men, within two hours or at a given drinking occasion, corresponding to 40 grams of pure alcohol for women and 50 for men [5, 6]. In practice, however, important variations are seen across countries, both in terms of number of drinks per gender (e.g. Australia: 5+ for both genders) and of alcohol content per drink (e.g. Australia, Switzerland: 10 grams of pure alcohol; the UK: 8g; the US: 12-14g; Canada: 13.5g) [7-9]. Despite regional variations in definitions, most studies still find that a 'binge' drinking occasion is associated with an increased risk of negative consequences. This convergence of findings appears largely due to the event-level nature of binge drinking in the sense that the high blood alcohol concentration achieved with the consumption of some drinks in a short period of time invariably increases the risk of consequences shortly after the drinking [10, 11].

The operationalization of any given threshold (e.g., 4+/5+ drinks) as the most appropriate threshold to predict the occurrence of alcohol-related consequences appears nevertheless questionable given the general collinearity between alcohol use and consequences (i.e. the higher the number of drinks, the more likely the consequence), suggesting that any threshold might discriminate a lower from a higher risk of consequence

[10, 12], and the fact that some consequences might occur at different drinking levels (e.g., a lower intake is usually required for a hangover than for blacking-out). Additionally, the few event-level studies available suggest that higher thresholds might be more appropriate than 4+/5+. Summarizing about 30 emergency room studies, Taylor and colleagues [13] showed that the risk of non-motor vehicle injury continuously increased with drinking levels, from an odds ratio of 1.79 at 24 grams of pure alcohol, up to a maximum odds ratio of 24.2 at 140 g. Moreover, Jackson [14] found that, using diaries completed by U.S. psychology students, thresholds of 10+ to 14+ drinks had the strongest prediction of different symptoms of hangover, compare to lower levels.

Researchers have recently attempted to determine the optimal number of drinks consumed to predict different adverse consequences using a well-established robust approach, namely by maximizing the sensitivity (i.e. the correct classification of the occurrence of a consequence at or above a certain number of drinks) and the specificity (i.e. the correct classification of the absence of a consequence at or below a certain number of drinks) of the drinking-consequence relationship [15, 16]. These studies produced mixed results. Dawson and colleagues [17] reported that thresholds that best discriminated between U.S. adults with and without severe concurrent alcohol-related harms (past-year alcohol dependence, alcohol abuse, injury, job loss, and hypertension) consisted of 4+ drinks for men and 3+ drinks for women. Among the general Australian population, Livingston [18] showed that optimal thresholds varied by the kind of outcome (past-year injury, hazardous behaviors and delinquent behaviors), while noting that optimal thresholds for all outcomes were 7 drinks or fewer. Finally, among adults from two US clinical trials, Pearson and colleagues [19] found no clear threshold and concluded that any binary classification of alcohol consumption levels performs poorly in predicting past-year occurrence of 45 different consequences from the Drinker Inventory of Consequences [20].

Strikingly, all three studies used retrospective self-reports of alcohol use and consequences over the past year. Such data might be adequate for the investigation of chronic alcohol-related consequences but appear inappropriate for the investigation of acute consequences. First, by using cross-sectional surveys, individuals, instead of drinking events, are compared in terms of frequency and intensity of drinking and of consequences and no temporal association can be established at the drinking occasion level. Second, retrospective self-reports of alcohol use are subject to recall bias due to memory deficits after even a few days [21, 22], which make such assessments particularly at risk of underestimation, both in terms of frequency of drinking occasions and of drinking levels per occasion [23].

The aim of this study is to determine the optimal drinking thresholds for the detection of several acute adverse alcohol-related consequences using sensitivity/specificity analyses of event-level data. As the highest rates for binge drinking are found among adolescents and young adults [24], we will first focus on people aged 16 to 25 years and investigate whether different thresholds apply to adolescents and young adults. Second, we will investigate whether different thresholds apply to women and men. Third, because different thresholds might apply to different types of consequences, we will conduct separate analyses for five acute adverse consequences (hangover, blackout, risky sex, involvement in fights and injury). Fourth, because the assessment method might alter the way the alcohol-consequence relationship is captured, we will use the data from two event-level studies (Table 1) using different assessment schedules of alcohol use (6 night-level questionnaires vs. 1 questionnaire the next morning) and different assessment modes of consequences (with and without explicit attribution to alcohol as the cause [25]).

--Table 1--

Method

The analysis was conducted on similar datasets from two studies. An overview is presented in Table 1.

Youth@Night study

Design and participants. The Youth@Night (Y@N) study aimed to document young people's behaviors on Friday and Saturday nights using a specifically developed smartphone application collecting event-level data (e.g. questionnaires, pictures, videos, GPS coordinates) repeatedly over the course of the night [see 26, 27 for full details of the study].

Participants were recruited in two major nightlife hubs in Switzerland, Lausanne and Zurich on Friday and Saturday nights of September 2014 from 9 to 12 p.m. [28]. Eligible volunteers (i.e. aged 16 to 25 and owning an Android smartphone) had to confirm their participation by entering their mobile phone number in the online consent form. After completion of a baseline questionnaire, participants were asked to document up to 10 Friday and Saturday nights over 7 consecutive weekends.

Of the 241 participants who used the app [28], 234 (97.1%) documented their previous nights' drinking and related consequences at least once. To ensure consistency with the selection procedure of the ICAT sample (described below), 17 participants (7.3%) who never reported any alcohol use during the study were excluded. The final dataset thus comprises 2,345 nights from 217 participants.

Measures.

Gender and age were recorded in the baseline questionnaire.

On Saturday and Sunday morning at 10 a.m., the app prompted participants to indicate the *total number of alcoholic drinks* they had consumed the previous night using a slider ranging from 0 to 30 drinks. Each drink corresponded to approximately 10 grams of pure ethanol [29].

Acute consequences. Participants were also requested to report whether or not each of "the following situations occurred during or since last night" (answer categories: 'yes' or 'no'): 'hangover (headache, upset stomach, etc.),' 'inability to remember what happened (even for a short period of time),' 'unintended or unprotected sex,' 'involvement in a fight or a quarrel',' and 'injury to yourself or someone else'.

ICAT study

Design and participants. The ICAT study aimed at documenting young adults' behaviors on Thursday, Friday and Saturday nights with hourly questionnaires completed on the smartphone browser using the Internet-based Cellphone-optimized Assessment Technique [ICAT: 30].

Participants were recruited from three higher education institutions in French-speaking Switzerland in April 2010. An invitation email was sent to all students, including detailed information about the study and a link to the registration webpage. Volunteers had to confirm a unique code sent by text message (SMS) to validate the online consent form and access the baseline internet questionnaire [23]. On Thursday, Friday and Saturday nights over five consecutive weeks, participants were prompted by SMS to complete six assessments (at 8, 9, 10 and 11 p.m., and 12 and 11 a.m.) about their alcohol use covering the timeframe from 5 p.m. to the end of the night.

Of the 183 participants who fully documented their nights and reported the consumption of at least one alcoholic drink over the study [23], 31 participants aged above 25 were excluded to obtain a sample comparable to the Youth@Night study (see above). The final dataset comprises 1,209 nights from 152 participants.

Measures.

Gender and age were recorded in the baseline questionnaire.

Each assessment across the night asked: "how many of the following alcoholic drinks did you have between…?" with the timeframes being '5-8 p.m.,' '8-9 p.m.,' '9-10 p.m.,' '10-11 p.m.,' '11 p.m.-midnight' and 'midnight-end of the night'. For each drink types – 'beer,' 'wine or champagne,' 'aperitifs or liqueurs,' 'spirits,' 'self-mixed drinks (e.g., whiskey-coke) or cocktails,' and 'pre-mixed drinks)' – six answer categories were provided, ranging from '0' to 'five or more' (coded as 5.5). A standard drink was defined as 10 g of pure ethanol [31]. The *total number of drinks* consumed was obtained by summing up the drinks reported per type and assessment over the entire night.

Acute consequences. At 11 a.m. the next morning, participants were asked whether "any of the following occurred last night as a result of your drinking" (i.e. explicitly with mentioning alcohol as being the cause; answer categories: 'yes' or 'no'): 'hangover (headache, upset stomach, etc.),' 'unable to remember what has happened (even for a short period of time),' 'unintended or unprotected sex,' 'involved in fight or quarrel,' and 'injured yourself or someone else.'

Analytic strategy

The analyses were conducted separately for each study, consequence and gender. Additionally, age differences between *adolescents*, aged 16 to 17 (i.e. unlike adults, they can legally purchase beer and wine but not distilled alcoholic beverages and cannot usually enter nightclubs), and *adults*, aged 18 to 25, were investigated among the participants in the Y@N study.

First, in order to assess whether the consequences had a significant dose-response relationship to drinking levels [12], the number of drinks consumed was compared between nights without and with each consequence. Using the STATA 14 [32] statistical software, differences were tested using *t*-tests, adjusted for the observations being nested within individuals.

Second, the area under the receiver operating characteristic (AUROC) curve [15, 33] was calculated to measure the ability of an increasing number of drinks consumed to correctly discriminate nights with from nights without each consequence. AUROC values range from 0 to 1, with 0.5 indicating no better discrimination than random chance and 1.0 indicating perfect discrimination [33]. AUROCs were estimated using bootstrapping to account for the nested structure of the data.

Thirdly, two methods maximizing sensitivity and specificity were used to determine the optimal threshold discriminating nights with from nights without each consequence. First, the Youden Index [34] identifies the threshold that maximizes the sum of sensitivity + specificity. The second method identifies the threshold that minimizes the distance between the receiver operating characteristic (ROC) curve and perfect discrimination (sensitivity = 1, specificity = 1) using Pythagorean Theorem [16]. Usually the two methods provide the same threshold [35]. In case of disagreement, both thresholds were reported.

Finally, in order to obtain summary scores of AUROCs and optimal thresholds for the different sample characteristics, respective values were averaged across genders, consequences, age groups and studies.

Results

Hangover was the most commonly reported consequence both on the person level (e.g. at least one hangover was reported by 38.3% to 60.3% of women; Table 2) and on the night level (e.g. 5.9% to 14.0% of all women's nights) and injury was the least common for both genders. Except for risky sex among women, consequences tended to be more often reported by adolescents than older participants, both at the individual and the night levels.

--Table 2--

Overall, an average of 2 to 3 drinks were consumed by men (Table 3) and 3 to 4 by women (Table 4) on nights without consequences. About 3 to 4 times more drinks were consumed on nights with hangovers and blackouts for both genders and all age groups, as well as for all consequences in the ICAT samples. Numbers of drinks consumed were also much higher (about 2 to 3 times higher) on nights with risky sex, fights and injuries in the Y@N samples, although the difference failed to reach 5%-significance level in some cases.

--Tables 3 and 4--

Across age groups and studies, the highest AUROC values (indicating higher accuracy in the discrimination of nights with from nights without consequences) were generally found for hangovers, injuries and blackouts among men and for hangovers and blackouts among women, while the lowest were found for fights and risky sex (see also Table 5 for a summary). Regarding age groups and studies, AUROC values were the lowest among Y@N adolescents and the highest among adults in the ICAT sample.

--Table 5--

Almost no thresholds were below the commonly used binge threshold, namely 4 drinks for women and 5 drinks for men, and these concerned only adolescents (Table 2 and 3). The methods for determining the optimal thresholds provided identical results in 60% (9 out of 15) of the cases for men and 86.7% (13 out of 15) for women. In the discordant cases, the difference exceeded two drinks only for fights among adolescents of both genders.

On average per consequence, the optimal thresholds ranged from 4.2 to 7.7 drinks among women and 5.2 to 8.8 drinks among men (Table 5). For all consequences but injuries, the thresholds were 1.0 to 2.6 drinks lower for women than for men. Lower thresholds were found on average across consequences for adolescent men (around 5.1) compared to adult men (7.3) whereas no major difference was found for women (around 5.2 for both age

groups). Finally, regarding study differences among adults, optimal thresholds in the ICAT sample were about 1.2 drinks higher than in Y@N sample.

Discussion

Using data from two event-level studies, the aim of this study was to determine the optimal number of alcoholic drinks consumed to discriminate nights with from those without acute adverse consequences. For all investigated consequences, the association between heavier drinking levels and increased risks of consequences was evident. Hangovers and blackouts were strongly and consistently related to drinking levels in all age groups, genders and studies, whereas the dose-response effect appeared slightly less consistent for risky sex, fights and injuries. This difference might be explained by the differential contribution of alcohol to these consequences. In contrast to risky sex and fights and injuries, hangovers and blackouts are unlikely to occur without alcohol use, which explains to a large extent the generally lower AUROCs of the former than the latter.

In the large majority of cases, the two methods for determining the optimal thresholds converged or diverged by only one drink, supporting the reliability of the thresholds found. The only notable exception was fights among adolescents of both genders. In line with particularly low AUROC values, these results suggest that, during adolescence, inebriation is just one of many causes of fights experienced during a night out.

Overall, the 4+/5+ thresholds for binge drinking, representing consumption of 40 grams of alcohol for women and 50 for men in the present studies, appeared only optimal for predicting hangovers in both studies and across consequences among adolescent men. However, higher optimal thresholds were found for blackouts, risky sex, fights and injuries, as well as among adolescent women and adults of both genders. The different body composition (total body water) and accelerated rate of ethanol elimination of adolescents mean that high blood alcohol concentration is reached more quickly [36] and their lower experience in

identifying consequences may be among the reasons that male adolescents report a hangover at lower drinking levels than men. With regard to genders, higher thresholds were found for men than for women with few exceptions. These results confirm the general use of different thresholds to account for differences in body constitution and alcoholism metabolism of men and women [37, 38].

Among adults, higher drinking levels on nights with consequences and higher AUROCs were consistently found among the ICAT samples compared to the Y@N samples. These differences might be explained by alcohol use being assessed using six assessments per night in the ICAT study while only one assessment was used in the Y@N study. Thus, as evidenced in previous studies [39], a higher number of assessments and short recall periods, such as used in the ICAT study, were likely to result in higher reported drinking levels. With regard to AUROC values, consequences were assessed using explicit attribution to alcohol use as the cause in the ICAT study and without attribution in the Y@N study. Thus, it is not surprising to observe closer associations of the number of drinks consumed and the occurrence of consequences in the ICAT study since only consequences with an obvious relationship to previous alcohol use for the participants were reported. Such findings should nevertheless be interpreted with caution since the use of alcohol-attributed items in questionnaires is prone to underestimate the effective number of experienced consequences and to bias the measurement of the alcohol-consequence relationship [25].

Several limitations need to be acknowledged. First, with the exception of hangover, the investigated consequences did not occur frequently, despite having sampled thousands of nights. Considering the relatively low number of reports of risky sex and injuries, related results might be affected by reporting errors of participants and levels of sensitivity and specificity might be slightly overestimated [40]. Second, the present findings were obtained among two samples of young people in Switzerland. Different dose-response relationships

between drinking levels and the occurrence of adverse consequences might be expected in other drinking cultures as well as among older populations. Third, no information was available on the actual size of the drinks consumed, making the results dependent on the participants' self-estimation of standard drinks. This approach has nevertheless the advantage that participants reported their drinks the same way they would understand drinking guidelines, making these findings a reliable basis for health communication. Fourth, no information was collected on participants' height and weight and on the time and size of each drink consumed. It was therefore not possible to adjust the analyses for the participants' body mass index [42] or estimate blood alcohol concentration. Future studies using event-based measures of drink and alcohol biosensors are recommended to validate the present findings. Finally, the present findings relate to standard drinks containing about 10 grams of pure alcohol. The thresholds might need to be recalculated in countries with different standard drink sizes. Although thresholds for blackouts, risky sex, fights and injuries among adults (i.e., approximately 5+/7+ among adults) might match the US and Canadian definition of binge drinking, namely 4+/5+ drinks containing 12 to 14g of alcohol, more research is needed to validate the equivalence.

The main strength of the present paper is the use of event-level studies to determine the most appropriate thresholds for risky drinking. In contrast to previous studies based on yearly retrospective assessments, the present thresholds were supported by higher levels of sensitivity and specificity. Hence, all but three Youden Index values in this study were higher than the highest value (1.34) reported by Pearson and colleagues [19] and almost two third of the shortest distance from the perfect discriminating test were lower than the shortest distance (0.351) reported by Livingston [18]. These comparisons confirm the higher accuracy of event-level studies to determine relevant thresholds with regard to acute alcohol-related consequences [10] because of minimized recall bias and the longitudinal assessment of the

drinking-consequence relationship. Another strength is the assessment of five acute consequences on about eight to ten nights among the same participants in their natural environment. Each participant serves thus as its own control, across nights and consequences, and the obtained data are characterized by a high ecological validity.

Conclusions

Relying on event-level evidence, results of this study showed that the consensual 4+/5+ threshold for binge drinking [3, 5], corresponding to 40+/50+ grams of pure alcohol in this study, appeared appropriate to predict hangovers in both genders and acute consequences overall among for adolescent men. Thresholds to predict more severe consequences and those for adults were however higher. These findings suggest revising the thresholds used to measure the real health-related burden resulting from acute consequences of binge drinking. In terms of reducing the overall number of negative consequences, it might however be recommended to keep communicating thresholds corresponding to the more conservative (and widely used) 4+/5+ drinks per occasion, especially for adolescent drinkers.

Reference

1. World Health Organization. *Global Status Report on Alcohol and Health* [online]. 2014. Available at: http://www.who.int/substance_abuse/publications/global_alcohol_report/en/ (accessed 12.1. 2015).
2. Gmel G., Rehm J., Kuntsche E. Binge drinking in Europe: Definitions, epidemiology, and consequences. *Sucht*. 2003; **49**: 105-16.
3. Wechsler H., Davenport A., Dowdall G. W., Moeykens B., Castillio S. Health and behavioural consequences of binge drinking in college - a national survey of students at 140 campuses. *JAMA*. 1994; **272**: 1671-7.
4. Wechsler H., Isaac N. E. 'Binge' drinkers at Massachusetts colleges. Prevalence, drinking style, time trends, and associated problems. *JAMA*. 1992; **267**: 2929-31.
5. Kuntsche E., Kuntsche S., Thrul J., Gmel G. Binge Drinking: Definition, Impact on Health, Demographic Correlates, Psychosocial Determinants and Intervention approaches. *Psychology & Health*. in press.
6. Wechsler H., Nelson T. F. Binge drinking and the American college student: What's five drinks? *Psychol Addict Behav*. 2001; **15**: 287-91.
7. World Health Organization. *International guide for monitoring alcohol consumption and related harm* [online]. 2000. Geneva: WHO, Department of Mental Health and Substance Dependence, Noncommunicable Diseases and Mental Health Cluster. Available at: http://whqlibdoc.who.int/hq/2000/WHO_MSD_MSB_00.4.pdf (accessed 13.08. 2015).
8. National Health and Medical Research Council. *Australian guidelines to reduce health risks from drinking alcohol*2009.

9. Butt P., Beirness D., Gliksman L., Paradis C., Stockwell T. *Alcohol and health in Canada: A summary of evidence and guidelines for low-risk drinking*. Ottawa, ON: Canadian Centre on Substance Abuse2011.

10. Pearson M. R., Kirouac M., Witkiewitz K. Questioning the validity of the 4+/5+ binge or heavy drinking criterion in college and clinical populations. *Addiction*. 2016; **111**: 1720-6.

11. Gmel G., Kuntsche E., Rehm J. Risky single-occasion drinking: Bingeing is not bingeing. *Addiction*. 2011; **106**: 1037-45.

12. Gruenewald P. J., Johnson F. W., Light J. M., Lipton R., Saltz R. F. Understanding college drinking: Assessing dose response from survey self-reports. *J Stud Alcohol*. 2003; **64**: 500-14.

13. Taylor B., Irving H., Kanteres F., Room R., Borges G., Cherpitel C., *et al.* The more you drink, the harder you fall: A systematic review and meta-analysis of how acute alcohol consumption and injury or collision risk increase together. *Drug Alcohol Depend*. 2010; **110**: 108-16.

14. Jackson K. M. Heavy episodic drinking: Determining the predictive utility of five or more drinks. *Psychol Addict Behav*. 2008; **22**: 68-77.

15. Metz C. E. Basic principles of ROC analysis. *Semin Nucl Med*. 1978; **8**: 283-98.

16. Kumar R., Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr*. 2011; **48**: 277-87.

17. Dawson D. A., Smith S. M., Pickering R. P., Grant B. F. An empirical approach to evaluating the validity of alternative low-risk drinking guidelines. *Drug Alcohol Rev*. 2012; **31**: 141-50.

18. Livingston M. Testing different thresholds for risky episodic drinking--what's so special about five drinks? *Drug Alcohol Depend*. 2013; **133**: 248-53.

19. Pearson M. R., Bravo A. J., Kirouac M., Witkiewitz K. The search for an elusive cutoff remains: Problems of binary classification of heavy drinking as an endpoint for alcohol clinical trials. *Drug Alcohol Depend*. 2017; **171**: 91-6.

20. Miller W. R., Tonigan J. S., Longabaugh R. The drinker inventory of consequences (DrInC). *Project MATCH Monograph Series*. 1995; **4**.

21. Ekholm O. Influence of the recall period on self-reported alcohol intake. *Eur J Clin Nutr*. 2004; **58**: 60-3.

22. Gmel G., Daeppen J. B. Recall bias for seven-day recall measurement of alcohol consumption among emergency department patients: implications for case-crossover designs. *J Stud Alcohol Drugs*. 2007; **68**: 303-10.

23. Kuntsche E., Labhart F. Investigating the drinking patterns of young people over the course of the evening at weekends. *Drug Alcohol Depend*. 2012; **124**: 319-24.

24. Kuntsche E., Rehm J., Gmel G. Characteristics of binge drinkers in Europe. *Soc Sci Med*. 2004; **59**: 113-27.

25. Gmel G., Kuntsche E., Wicki M., Labhart F. Measuring alcohol-related consequences in school surveys: Alcohol-attributable consequences or consequences with students' alcohol attribution. *Am J Epidemiol*. 2010; **171**: 93-104.

26. Santani D., Biel J., Labhart F., Truong J., Landolt S., Kuntsche E., *et al.* The night is young: Urban crowdsourcing of nightlife patterns. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* ACM; 2016. p. 427-38.

27. Labhart F., Kuntsche E. Development and validation of the predrinking motives questionnaire. *J Appl Soc Psychol*. 2016; **47**: 136-47.

28. Labhart F., Tarsetti F., Bornet O., Santani D., Truong J., Landolt S., *et al.* Development and evaluation of an innovative method to recruit representative samples in the urban

nightlife. *International Journal of Social Research Methodology*. 2017; **Epub ahead of print**.

29. Labhart F., Anderson K., Kuntsche E. The spirit is willing, but the flesh is weak: Why young people drink more than intended on weekend nights - an event-level study. *Alcohol Clin Exp Res*. in press.
30. Kuntsche E., Labhart F. ICAT: Development of an Internet-Based Data Collection Method for Ecological Momentary Assessment Using Personal Cell Phones. *Eur J Psychol Assessm*. 2013; **29**: 140-8.
31. Labhart F., Graham K., Wells S., Kuntsche E. Drinking before going to licensed premises: An event-level analysis of predrinking, alcohol consumption, and adverse outcomes. *Alcohol Clin Exp Res*. 2013; **37**: 284-91.
32. StataCorp. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP; 2015.
33. Hanley J. A., McNeil B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; **143**: 29-36.
34. Youden W. J. Index for rating diagnostic tests. *Cancer*. 1950; **3**: 32-5.
35. Perkins N. J., Schisterman E. F. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 2006; **163**: 670-5.
36. Donovan J. E. Estimated blood alcohol concentrations for child and adolescent drinking and their implications for screening instruments. *Pediatrics*. 2009; **123**: e975-81.
37. Wechsler H., Dowdall G. W., Davenport A., Rimm E. B. A gender-specific measure of binge drinking among college students. *Am J Public Health*. 1995; **85**: 982-5.
38. Lange J. E., Voas R. B. Defining binge drinking quantities through resulting blood alcohol concentrations. *Psychol Addict Behav*. 2001; **15**: 310-6.

39. Monk R. L., Heim D., Qureshi A., Price A. "I have no clue what I drunk last night" using Smartphone technology to compare in-vivo and retrospective self-reports of alcohol consumption. *PLoS One*. 2015; **10**: e0126209.

40. Leeflang M. M. G., Moons K. G. M., Reitsma J. B., Zwinderman A. H. Bias in Sensitivity and Specificity Caused by Data-Driven Selection of Optimal Cutoff Values: Mechanisms, Magnitude, and Solutions. *Clin Chem*. 2008; **54**: 729-37.

41. LaBrie J. W., Hummer J. F., Kenney S. R., Lac A., Pedersen E. R. Identifying factors that increase the likelihood for alcohol-induced blackouts in the prepartying context. *Subst Use Misuse*. 2011; **46**: 992-1002.

42. Courtney K. E., Polich J. Binge drinking in young adults: Data, definitions, and determinants. *Psychol Bull*. 2009; **135**: 142-56.

Table 1: Characteristics of the ICAT and Youth@Night (Y@N) studies

	ICAT study	Y@N study
Recruitment		
Recruitment period:	April 2010	September 2014
Target population:	Students from three higher education institutions in French-speaking Switzerland	Nightlife goers, aged 16 to 25, from the two major nightlife hubs in Switzerland
Method:	Mass mail sent to all students	Street intercept using the Geographical Proportionate-to-size Street intercept sampling (GPSIS: Labhart et al, 2017)
Participants		
N:	152	217
Gender ratio (% men):	46.7	52.5
Age range (adolescents):	n.a.	16-17 years old
Age range (adults):	18-25 years old	18-25 years old
Event-level data collection		
Period:	5 consecutive weekends in May-June 2010	7 consecutive weekends September-November 2014
Nights of interest:	Thursday, Friday and Saturday	Friday and Saturday
Method:	Smartphone-optimized online questionnaires	Smartphone application
Nights (N, average per person in parentheses):	1,209 (8.0)	2,345 (10.8)
Measures		
Age and gender:	Baseline questionnaire before event-level data collection	Baseline questionnaire before event-level data collection
Number of drinks consumed per night:	Beverage-specific assessments submitted six times over the course of the night	Total night consumption assessed the next morning
Occurrence of acute consequences:	Assessment the next morning; with attribution to alcohol as the cause	Assessment the next morning; without attribution to alcohol as the cause
Ethics Review Board	Ethics commission of canton de Vaud (protocol 223/08)	Lausanne and Zurich cantonal ethics commissions for the Research on Human Beings (protocol 145/14)

Table 2: Number of participants, nights and prevalence of consequences, per age group, study and gender

	Men		Women	
Study	Persons[1]	Nights	Persons[1]	Nights
Participants				
Adolescents [Y@N] (n)	32	347	35	391
Adults [Y@N] (n)	82	857	68	750
Adults [ICAT] (n)	71	528	81	681
Consequences				
Hangover				
Adolescents [Y@N] (%)	81.3	18.7	57.1	12.3
Adults [Y@N] (%)	61.0	16.6	60.3	14.0
Adults [ICAT] (%)	35.2	8.0	38.3	5.9
Blackout				
Adolescents [Y@N] (%)	31.3	4.6	17.1	2.8
Adults [Y@N] (%)	13.4	2.2	16.2	2.3
Adults [ICAT] (%)	12.7	2.7	8.6	1.0
Risky sex				
Adolescents [Y@N] (%)	21.9	4.6	8.6	1.5
Adults [Y@N] (%)	11.0	2.3	10.3	1.5
Adults [ICAT] (%)	8.5	1.1	6.2	0.9
Fight				
Adolescents [Y@N] (%)	25.0	3.6	17.1	2.0
Adults [Y@N] (%)	13.4	4.6	5.9	0.7
Adults [ICAT] (%)	5.6	0.5	3.7	0.4
Injury				
Adolescents [Y@N] (%)	25.0	2.9	8.6	1.3
Adults [Y@N] (%)	8.5	0.8	4.4	0.4
Adults [ICAT] (%)	7.0	0.9	2.5	0.3

Note: (1) For consequences: at least once during the study

Table 3: Levels of alcohol use without and with consequences, AUROC and optimal thresholds, per age group and study (Women)

	Drinks per night			AUROC (95%-CI)		Optimal threshold			
	Without conseq.	With conseq.					Sensitivity/ Specificity	Method[2]	
	mean (SD)	mean (SD)	Test-value[1]					Youden Index	Shortest distance
Hangover									
Adol. [Y@N]	1.6 (2.2)	6.0 (4.6)	59.6***	0.85	(0.80-0.90)	**3**	93.8/74.9	168.7	0.26
Adults [Y@N]	2.1 (3.1)	7.9 (5.5)	52.9***	0.85	(0.80-0.90)	**4**	81.9/79.1	161.0	0.28
Adults [ICAT]	2.6 (4.0)	10.5 (5.1)	82.5***	0.90	(0.86-0.95)	**5**	90.0/80.3	170.3	(0.22)
						6	85.0/85.3	170.3	0.21
Blackout									
Adol. [Y@N]	2.0 (2.6)	7.4 (8.4)	14.9***	0.74	(0.58-0.90)	**3**	72.7/67.6	140.4	0.42
Adults [Y@N]	2.8 (4.0)	7.9 (3.0)	30.4***	0.87	(0.78-0.95)	**5**	94.1/77.9	172.0	0.23
Adults [ICAT]	2.9 (4.4)	11.6 (4.0)	36.1***	0.92	(0.87-0.98)	**6**	100.0/82.1	182.1	0.18
Risky sex									
Adol. [Y@N]	2.0 (2.7)	7.3 (11.3)	2.8	0.63	(0.49-0.76)	**4**	66.7/75.3	142.0	0.41
Adults [Y@N]	2.9 (4.0)	8.7 (6.4)	16.7***	0.77	(0.63-0.92)	**5**	72.7/77.0	149.7	0.36
Adults [ICAT]	3.0 (4.5)	8.7 (6.0)	8.3**	0.80	(0.62-0.93)	**4**	83.3/71.1	154.4	0.33
Fight									
Adol. [Y@N]	2.0 (2.6)	7.5 (10.4)	3.8	0.58	(0.34-0.81)	**2**	62.5/57.4	(119.9)	0.57
						12	37.5/99.2	136.7	(0.63)
Adults [Y@N]	2.9 (4.1)	5.6 (4.5)	1.6	0.64	(0.22-1.06)	**7**	60.0/87.0	147.0	0.42
Adults [ICAT]	3.0 (4.5)	10.0 (6.1)	5.9*	0.88	(0.78-0.98)	**6**	100.0/81.6	181.6	0.18
Injury									
Adol. [Y@N]	2.0 (2.6)	10.6 (12.1)	9.9**	0.68	(0.44-0.92)	**10**	60.0/98.7	158.7	0.40
Adults [Y@N]	2.9 (4.1)	6.3 (7.1)	1.0	0.58	(0.07-1.09)	**5**	66.7/76.4	143.1	0.41
Adults [ICAT]	3.0 (4.4)	17.0 (12.7)	4.8*	0.94	(0.85-1.03)	**8**	100.0/87.8	187.8	0.12

Notes: 1) $F(1, n-1)$: adjusted *t*-tests for the design effect of cluster on individuals; 2) Values in brackets and grey: less optimal threshold.

Table 4: Levels of alcohol use without and with consequences, AUROC and optimal thresholds, per age group and study (Men)

	Drinks per night			AUROC (95%-CI)		Optimal threshold			
	Without conseq.	With conseq.					Sensitivity/ Specificity	Method[2]	
	mean (SD)	mean (SD)	Test-value[1]					Youden Index	Shortest distance
Hangover									
Adol. [Y@N]	2.3 (3.5)	9.2 (6.5)	47.7***	0.87	(0.80-0.93)	**5**	86.2/81.2	167.4	0.23
Adults [Y@N]	3.2 (4.1)	9.1 (5.6)	87.5***	0.81	(0.76-0.86)	**5**	80.3/73.3	153.6	0.33
Adults [ICAT]	3.6 (4.7)	14.2 (8.3)	63.3***	0.89	(0.85-0.93)	**5**	97.6/71.8	169.4	(0.28)
						6	85.7/77.4	(163.1)	0.27
Blackout									
Adol. [Y@N]	3.3 (4.5)	9.6 (9.3)	8.8**	0.72	(0.54-0.89)	**4**	81.3/63.4	144.7	0.41
Adults [Y@N]	4.0 (4.8)	10.5 (4.3)	60.0***	0.85	(0.78-0.91)	**7**	84.2/76.6	160.8	0.28
Adults [ICAT]	4.0 (5.1)	19.9 (8.5)	34.1***	0.96	(0.93-0.99)	**11**	100.0/89.3	189.3	0.11
Risky sex									
Adol. [Y@N]	3.4 (4.6)	7.9 (9.5)	2.7	0.59	(0.37-0.80)	**5**	56.3/69.8	(126.0)	0.53
						6	50.0/79.2	129.2	(0.54)
Adults [Y@N]	4.1 (4.9)	7.4 (6.6)	3.9	0.63	(0.45-0.81)	**8**	50.0/79.0	129.0	0.54
Adults [ICAT]	4.3 (5.7)	12.3 (7.8)	7.5**	0.84	(0.74-0.95)	**5**	100.0/67.1	167.1	(0.33)
						7	83.3/76.4	(159.8)	0.29
Fight									
Adol. [Y@N]	3.4 (4.7)	7.7 (9.0)	1.7	0.56	(0.24-0.87)	**5**	53.9/69.5	(123.3)	0.55
						8	46.2/87.1	133.3	(0.55)
Adults [Y@N]	4.0 (4.8)	10.9 (5.5)	28.7***	0.83	(0.74-0.91)	**6**	87.5/71.0	158.5	0.32
Adults [ICAT]	4.3 (5.6)	20.0 (13.4)	7.4**	0.89	(0.75-1.04)	**14**	75.0/92.2	167.2	0.26
Injury									
Adol. [Y@N]	3.4 (4.7)	10.3 (9.0)	4.0	0.75	(0.54-0.96)	**4**	90.0/62.9	152.9	(0.38)
						5	80.0/70.0	(150.0)	0.36
Adults [Y@N]	4.1 (4.8)	15.7 (4.3)	53.6***	0.95	(0.90-0.99)	**10**	100.0/86.5	186.5	0.14
Adults [ICAT]	4.3 (5.6)	17.4 (13.0)	6.5*	0.87	(0.75-0.98)	**5**	100.0/66.9	166.9	(0.33)
						7	80.0/76.3	(156.3)	0.31

Notes: 1) $F(1, n-1)$: adjusted *t*-tests for the design effect of cluster on individuals; 2) Values in brackets and grey: less optimal threshold.

Table 5: Average AUROCs and optimal thresholds per consequences, age groups and gender

	AUROC		Optimal threshold	
	Women	Men	Women	Men
Consequences:				
Hangover	0.87	0.86	4.2	5.2
Blackout	0.84	0.84	4.7	7.3
Risky sex	0.73	0.69	4.3	6.5
Fight	0.70	0.76	6.7	8.8
Injury	0.73	0.85	7.7	6.8
Age groups:				
Adolescents	0.69	0.70	5.1	5.1
Adults	0.82	0.85	5.3	7.9
Y@N	0.74	0.81	4.7	7.2
ICAT	0.89	0.89	5.9	8.5