An online daily diary study of alcohol use using Amazon's Mechanical Turk

Dr Marcella H. Boynton and
University of North Carolina at Chapel Hill, Lineberger Comprehensive Cancer Center, Chapel Hill, North Carolina, USA

Laura Smart Richman
Assistant Professor, Duke University, Durham, North Carolina, USA

Abstract

Introduction and Aims—In recent years, unprecedented levels of internet access and the widespread growth of emergent communication technologies have resulted in significantly greater population access for substance use researchers. Despite the research potential of such technologies, the use of the internet to recruit individuals for participation in event-level research has been limited. The purpose of this paper is to provide a brief account of the methods and results from an online daily diary study of alcohol use.

Design and Methods—Participants were recruited using Amazon's Mechanical Turk (MTurk). Eligible participants completed a brief screener assessing demographics and health behaviours, with a subset of individuals subsequently recruited to participate in a two week daily diary study of alcohol use.

Results—Multilevel models of the daily alcohol data derived from the MTurk sample ($n = 369$) replicated several findings commonly reported in daily diary studies of alcohol use.

Discussion and Conclusions—Results demonstrate that online participant recruitment and survey administration can be a fruitful method for conducting daily diary alcohol research.

Obtaining accurate measures of alcohol consumption using traditional survey measures can be challenging due to retrospective biases as well as to substantial within- and between-person variations in consumption patterns [1–2]. Daily diary survey methodology, defined as the administration of a brief daily survey for a series of weeks, offers great potential for assessing alcohol use proximate to its real-time occurrence [3–4]. Past research has employed a variety of technologies to collect daily measures of alcohol [5–9]; however, to date, there have been a dearth of studies leveraging online recruitment platforms for the purposes of daily diary research, in general, and alcohol-related research, specifically.

Correspondence concerning this report should be addressed to Dr Marcella H. Boynton, Lineberger Comprehensive Cancer Center, 450 West Drive, Chapel Hill, NC 27514; mhb23@unc.edu.

Amazon's Mechanical Turk (MTurk) is an online crowdsourcing tool that allows "workers" to complete online tasks or "human intelligence tasks" for relatively small amounts of remuneration. MTurk has become an increasingly popular tool for social science research, with multiple experimental and survey studies consistently replicating findings from prior research [10–24]. The purpose of this paper is to illustrate the utility of MTurk to recruit a diverse sample of adults for participation in an online daily diary study of alcohol use.

Method

Participants

Five hundred and eighteen participants enrolled in the daily diary study and completed at least one daily survey. Of those, 130 were removed because they completed fewer than four daily surveys, and an additional 19 were removed because they did not have at least two consecutive daily surveys. The final sample was comprised of 369 adults contributing 3145 daily observations. The average number of completed daily measures was 8.5 (SD = 3.9). Table 1 presents general sample demographics.

Procedure

Participants were recruited as part of a larger online survey study of personality and health that was comprised of three separate recruitment phases. Initially, we posted a human intelligence task on MTurk inviting interested individuals to complete a brief screener assessing basic demographic and health factors. Inclusion criteria for the follow-up survey were a primary racial/ethnic identity of White, Black/African American, Latino or Asian/Pacific Islander, age 21 to 65, currently residing in the US, and ability to speak and read English. Because one of the broader aims of the study was focused on issues particularly relevant to US ethnic minority groups, an over-sampling approach was employed such that enrolment for each of the four targeted ethnic groups was kept relatively balanced. Individuals invited to complete the follow-up survey were sent a unique link to a secure website [25] where they completed a series of social-personality and health-related measures. Participants recruited for the daily diary study used the same secure website to complete a five minute daily survey for up to 14 consecutive days. Each daily survey assessed cognitions and behaviours for the prior evening (5 PM to 6 AM) and the current day (6 AM to 5 PM) and could be completed between 5:00 PM (at which time an email reminder was sent to participants) and 6:00 AM the following morning. Participants were compensated for completing both the screener/baseline survey ($0.85 USD) and daily diary surveys (up to $5.00 USD for perfect adherence).

Measures

Possible history of alcohol misuse was assessed using the 4-item CAGE Alcohol Questionnaire [26–27]; concordant with past research [28–29] the CAGE was coded 0 for zero item endorsement and 1 for ≥ 1 items endorsed. Depression was assessed using the 8-item Center for Epidemiologic Studies Depression (CES-D) Scale [30], which is a validated short version of the 20-item CES-D [31]. Both the 8-item CES-D ($\alpha = 0.78$) and the 4-item CAGE ($\alpha = 0.74$) exhibited acceptable scale reliability in our sample.

In order to maximise reliability of self-reported daily alcohol consumption, a figure with accompanying text depicting a "standard drink" for each major type of alcohol was presented immediately prior to the daily alcohol measures (see Figure 1). When reporting daily alcohol consumption, individuals were asked to specify type and number of drinks consumed, again with accompanying imagery (see Figure 2). Alcohol use intentions were measured using the item "Approximately how many standard alcoholic drinks do you intend to consume between 5 PM today and 5 PM tomorrow?" Alcohol salience was measured with the item "How much have you thought about alcohol or alcohol-related places and activities since waking up this morning?"

Results

Of the 369 participants, 87% ($n = 322$) reported some level of past alcohol consumption. The average number of drinking days and heavy drinking days across the entire sample was 1.6 (SD = 2.5) and 0.45 (SD = 1.3), respectively. Those with a positive CAGE score had a significantly higher number of drinking days ($M = 2.7$; SD = 2.9) and heavy drinking days ($M = 1.2$; SD = 2.2) compared to those without a positive CAGE score ($M = 1.4$; SD = 2.4; $M = 0.3$; SD = 0.9, respectively), $F(1, 366) = 15.9$, $p < .001$ for drinking days and $F(1, 366) = 30.7$, $P < 0.001$ for binge days.

Of the 3145 reported days, approximately nineteen percent ($n = 600$) included consumption of one or more standard alcoholic drinks. Of those, 36% ($n = 214$) were beer only, 20% were wine only ($n = 122$), 9% were shots only (n = 54), 11% were mixed drinks only ($n = 64$), 3% were some other form of liquor (e.g. alcopop, malt liquor, $n = 18$), and 21% were two or more different types of alcoholic beverages ($n = 128$). Of the 600 reported drinking days, 27% ($n = 162$) met the National Institute on Alcohol Abuse and Alcoholism criteria for a heavy drinking episode, which is 4 drinks for females and 5 for males [32]. The average number of drinks consumed when drinking occurred was 3.6 (SD = 3.5). Those who reported at least one drinking day during the study, had a slightly higher measures completion rate ($M = 9.0$, SD = 3.4) compared to those who reported no drinking ($M = 8.5$, SD = 3.9), $F(1, 367) = 4.45$, $P = 0.01$. However, those with a history of problematic drinking, as measured by the CAGE, had a slightly lower measures completion rate ($M = 7.2$, SD = 3.8) compared to those who reported no problematic history drinking ($M = 8.8$, SD = 3.8), $F(1, 366) = 9.74$, $P = 0.002$.

Multilevel modelling was used to analyse the daily diary data. Number of drinks consumed (0–15+) and heavy drinking episode (0 = *no*; 1 = *yes*) were the two daily alcohol outcomes of interest. Random intercept models were estimated using the PROC GLIMMIX procedure in SAS 9.3 [33]. Table 2 details all model results. In the final combined effects model for number of drinks consumed, being male, having a history of problematic drinking, daily alcohol intentions, daily alcohol salience, and weekend day predicted a higher number of drinks consumed. For the heavy drinking episode final combined effects model, odds of having a heavy drinking day was greater for those individuals who were Latino (as compared to White), had a history of problematic drinking, or reported alcohol intentions or salience at the daily level.

Discussion

One of the primary purposes of this study was to develop and successfully implement a daily diary study of alcohol use using MTurk with a diverse US sample. Examination of the measures adherence rates and demographic factors indicates that we were successful in this regard. Further, our alcohol-related findings replicated a number of associations noted in previous research. In particular, daily alcohol consumption was greater on weekends, for those with a positive CAGE score, and when alcohol salience and intentions were present. Reflective of typical American drinking preferences and behaviours [34], beer was the most common type of beverage consumed, followed by various forms of spirits and liquors, and then wine as the least common type of alcohol consumed.

Limitations and Future Directions

Although our findings were very similar to those described in past daily dairy studies of alcohol, it should be noted that were several key demographic differences, with our sample substantially older, more affluent, and more ethnically diverse than a typical college student sample. Another major difference between this and past samples is that we did not require a certain level of past alcohol use for participation. Researchers interested in studying patterns of alcohol use in groups more similar to those in past studies could readily adapt MTurk screening criteria to select for such individuals, and future online daily diary alcohol studies would benefit from this approach. Also of note, daily measures adherence in our study was approximately 60%, which is somewhat lower than the typical 75–85% rates found in many college student daily diary studies of alcohol use. One contributing factor may be the level of compensation, which was a small fraction of what is typically offered. Future online daily diary research adherence rates would likely benefit from a higher incentive schedule. In summary, this research demonstrates the potential value of using online participant recruitment for daily diary alcohol research.

Acknowledgments

References

1. Fitzgerald JL, Mulford HA. Self-report validity issues. J Stud Alc. 1987; 48:207–11.

2. Stockwell T, Donath S, Cooper-Stanbury M, Chikritzhs T, Catalano P, Mateo C. Under-reporting of alcohol consumption in household surveys: a comparison of quantity-frequency, graduated-frequency and recent recall. Addiction (Abingdon, England). 2004; 99:1024–33.

3. Armeli S, Todd M, Mohr C. A Daily Process Approach to Individual Differences in Stress-Related Alcohol Use. J Pers. 2005; 73:1657–86. [PubMed: 16274449]

4. Conner TS, Tennen H, Fleeson W, Barrett LF. Experience Sampling Methods: A Modern Idiographic Approach to Personality Research. Soc Personal Psychol Compass. 2009; 3:292–313. [PubMed: 19898679]

5. Armeli S, Carney MA, Tennen H, Affleck G, O'Neil TP. Stress and alcohol use: a daily process examination of the stressor-vulnerability model. J Pers Soc Psychol. 2000; 78(5):979–94. [PubMed: 10821203]

6. Mohr CD, Armeli S, Tennen H, Carney MA, Affleck G, Hromi A. Daily interpersonal experiences, context, and alcohol consumption: crying in your beer and toasting good times. J Pers Soc Psychol. 2001; 80:489–500. [PubMed: 11300581]

7. Barrett LF, Barrett DJ. An Introduction to Computerized Experience Sampling in Psychology. Soc Sci Comp Rev. 2001; 19:175–85.

8. Sacco P, Smith CA, Harrington D, Svoboda DV, Resnick B. Feasibility and Utility of Experience Sampling to Assess Alcohol Consumption Among Older Adults. J Appl Gerontol. 2014 Epub ahead of print.

9. O'Hara RE, Boynton MH, Scott D, Armeli S, Tennen H, Williams C, Covault J. Drinking to cope among African-American college students: An assessment of episode-specific motives. Psychol Addict Behav. in press.

10. Azzam T, Jacobson MR. Finding a Comparison Group: Is Online Crowdsourcing a Viable Option? Amer Jour Eval. 2013; 34:372–84.

11. Amir O, Rand DG, Gal YK. Economic Games on the Internet: The Effect of $1 Stakes. Plos One. 2012; 7:e31461. [PubMed: 22363651]

12. Casler K, Bickel L, Hackett E. Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. Comput Human Behav. 2013; 29:2156–60.

13. Gardner RM, Brown DL, Boice R. Using Amazon's Mechanical Turk website to measure accuracy of body size estimation and body dissatisfaction. Body Image. 2012; 9:532–4. [PubMed: 22832086]

14. Holden CJ, Dennie T, Hicks AD. Assessing the reliability of the M5-120 on Amazon's mechanical Turk. Comput Human Behav. 2013; 29:1749–54.

15. Horton J, Rand D, Zeckhauser R. The online laboratory: conducting experiments in a real labor market. Exp Econ. 2011; 14:399–425.

16. Rand DG. The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. J Theor Biol. 2012; 299:172–9. [PubMed: 21402081]

17. Simons DJ, Chabris CF. Common (Mis)Beliefs about Memory: A Replication and Comparison of Telephone and Mechanical Turk Survey Methods. Plos One. 2012; 7:e51876. [PubMed: 23272183]

18. Summerville A, Chartier CR. Pseudo-dyadic "interaction" on Amazon's Mechanical Turk. Behav Res Methods. 2013; 45:116–24. [PubMed: 23055159]

19. Joinson A. Social desirability, anonymity, and internet-based questionnaires. Behav Res Meth Instrum Comput. 1999; 31:433–8.

20. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. Am Psychol. 2004; 59:93–104. [PubMed: 14992636]

21. Berinsky AJ, Huber GA, Lenz GS. Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. Pol Anal. 2012; 20:351–68.

22. Paolacci G, Chandler J, Ipeirotis PG. Running experiments on Amazon Mechanical Turk. Judgm Decis Mak. 2010; 5:411–9.

23. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspect Psychol Sci. 2011; 6:3–5.

24. Mason W, Suri S, Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods. 2012; 44:1–23. [PubMed: 21717266]

25. Qualtrics. Qualtrics; Provo, Utah: 2012. http://www.qualtrics.com

26. Ewing JA. Detecting alcoholism: The CAGE questionnaire. JAMA. 1984; 252:1905–7. [PubMed: 6471323]

27. O'Brien CP. The CAGE questionnaire for detection of alcoholism: a remarkably useful but simple tool. JAMA. 2008; 300:2054–6. [PubMed: 18984895]

28. Bush B, Shaw S, Cleary P, Delbanco TL, Aronson MD. Screening for alcohol abuse using the CAGE questionnaire. Am J Med. 1987; 82:231–5. [PubMed: 2880504]

29. Girela E, Villanueva E, Hernandez-Cueto C, Luna JD. Comparison of the CAGE questionnaire versus some biochemical markers in the diagnosis of alcoholism. Alcohol Alcohol. 1994; 29:337–43. [PubMed: 7945575]

30. Andresen EM, Byers K, Friary J, Kosloski K, Montgomery R. Performance of the 10-item Center for Epidemiologic Studies Depression scale for caregiving research. SAGE Open Med. 2013:1.

31. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Appl Psych Meas. 1977; 1:385–401.

32. National Institute on Alcohol Abuse and Alcoholism, NIAAA. NIAAA council approves definition of binge drinking. Washington, DC: Office of Research Translation and Communication; 2004 Winter. NIAAA Newsletter, 3, 3

33. SAS Institute Inc. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc; 2011.

34. National Institute on Alcohol Abuse and Alcoholism, NIAAA. Apparent per capita alcohol consumption: National, state, and regional trends, 1977–2009. 2009. p. 35http://pubs.niaaa.nih.gov/publications/Surveillance92/CONS09.pdf

Figure 1.
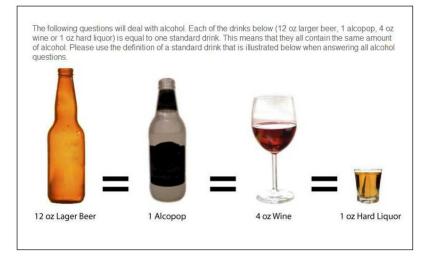Alcohol measures introduction and standard drink definition (Note: 1 oz is equal to 28.4 grams)

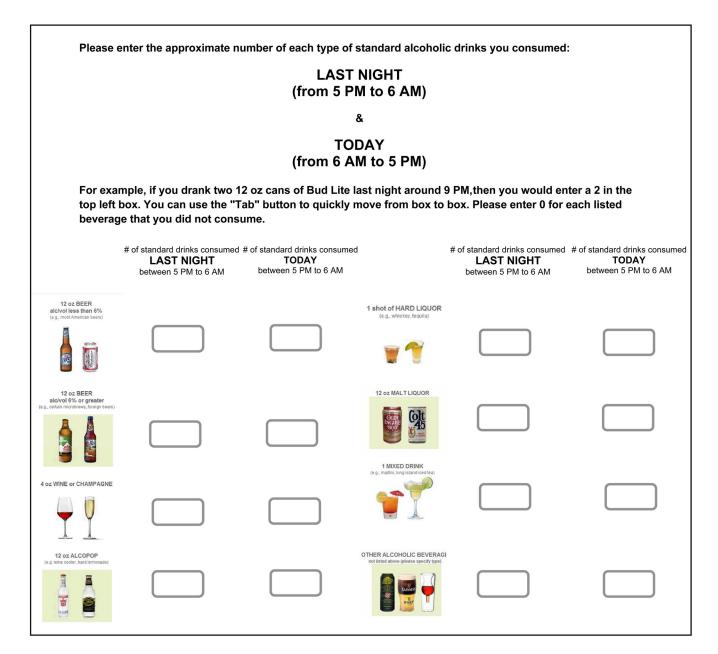Please enter the approximate number of each type of standard alcoholic drinks you consumed:

LAST NIGHT
(from 5 PM to 6 AM)

&

TODAY
(from 6 AM to 5 PM)

For example, if you drank two 12 oz cans of Bud Lite last night around 9 PM, then you would enter a 2 in the top left box. You can use the "Tab" button to quickly move from box to box. Please enter 0 for each listed beverage that you did not consume.

Figure 2.
Daily alcohol consumption measures (Note: 1 oz is equal to 28.4 grams) 0251677696

Table 1

Descriptive statistics

Baseline Measures	No history of alcohol misuse CAGE 1; *n* = 297	Possible history of alcohol misuse CAGE 2; *n* = 71
	M (*SD*) or %	*M* (*SD*) or %
Gender		
% male	42.8	49.3
% female	57.2	50.7
Age	31.5 (9.4)	31.8 (10.1)
Race		
% White/Caucasian	28.0	35.2
% Black/African American	27.6	31.0
% Latino/Hispanic	17.5	16.9
% Asian/Pacific Islander	26.9	16.9
Income bracket	$55,000–$59,999 ($40,000)	$50,000–$54,999 ($40,000)
Education		
% high school diploma or less	11.1	9.9
% some college or associate's degree	33.0	42.3
% college degree	40.4	35.2
% Master's or doctoral-level degree	15.5	12.6
Center for Epidemiologic Studies Depression (CES-D) Scale	6.8 (4.4)	7.7 (4.3)
Daily Measures		
# of completed daily surveys	7.24 (3.77)	8.81[*] (3.83)
# of reported drinking days	1.38 (2.34)	2.68[*] (2.91)
# of reported heavy drinking days	0.26 (0.91)	1.18[*] (2.18)

[*] significantly higher mean or median, *P* < 0.05

Table 2

Predicting number of drinks and heavy drinking episodes on the daily level

Predictor	Number of drinks				Heavy drinking episode			
	Intercept-only and Individual Main Effects Models Estimates[1]		Final Model		Intercept-only and Individual Main Effects Model Estimates[1]		Final Model	
	b^2	(SE)	b^2	(SE)	b^2	(SE)	b^2	(SE)
Intercept	0.18 **	0.11	0.05 **	0.21	0.03 **	0.14	0.00 **	0.40
Level 1 (within person/daily effects)								
Daily alcohol intentions (0=no intention drink any alcohol; 1=intends to consume 1 standard alcoholic drink)	4.85 **	0.06	3.60 **	0.06	21.1 **	0.06	10.52 **	0.28
Daily alcohol salience (0=did not think about alcohol or alcohol venues; 1=thought about alcohol or alcohol-related places and activities)	5.26 **	0.09	3.35 **	0.09	1.58 **	0.05	6.79 **	0.41
Weekend	1.47 **	0.05	1.24 **	0.05	1.79 *	0.20	1.46†	0.22
Level 2 (between person effects)								
Gender (1=male;0=female)	2.24 *	0.23	1.64 *	0.16	1.85 *	0.28	1.38	0.30
Age grand mean centred	0.99	0.01	1.00	0.01	1.01	.014	1.01	0.02
Race Reference group is White								
African American	0.58 †	.31	0.74	0.21	0.95	0.37	1.11	0.40
Latino/Hispanic	0.89	.35	1.16	0.23	1.74	0.40	3.25 *	0.42
Asian/Pacific Islander	0.77	.31	0.99	0.21	0.81	0.40	1.36	0.43
Education (0=high school or less; 1=associate's degree or higher)	1.22	0.24	1.15	0.17			0.95	0.33
Income grand mean centred	1.03 †	0.01	1.01	0.01			1.00	0.02
CAGE Questionnaire (0=endorsement of < 2 items; 1=endorsement of 2 items)	2.08 **	0.26	2.58 **	0.18	6.97 **	0.29	4.16 **	0.31
Center for Epidemiologic Studies Depression Scale grand mean centred	0.98	0.03	1.03		1.03	0.03		
Intraclass correlation	0.78				0.43			

[1] Each estimate in this column is unadjusted for all other predictors; the final model column presents the adjusted estimates from the combined effects model.

[2] Estimates provided are exponentiated betas to reflect the true drink number estimate or odds ratio

[†] $P <0.10$ (marginally significant),

[*] $P <0.05$,

[**] $P <0.0001$.