ARTICLE IN PRESS

Addictive Behaviors xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Addictive Behaviors

journal homepage: www.elsevier.com/locate/addictbeh

Assessment of alcohol intake: Retrospective measures versus a smartphone application

Antoinette Poulton[a,*], Jason Pan[a], Loren Richard Bruns Jr[b], Richard O. Sinnott[b], Robert Hester[a]

[a] Melbourne School of Psychological Sciences, University of Melbourne, Parkville, VIC 3010, Australia
[b] Computing and Information Systems, University of Melbourne, Parkville, VIC 3010, Australia

HIGHLIGHTS

- Real-time assessment of drinking behaviour using a smartphone app was explored.
- Participants logged more drinking days via the app compared to Timeline Followback.
- Total intake was higher when recorded using the app relative to Timeline Followback.
- The app captured a greater number of high intake episodes than Timeline Followback.
- The app showed faster rate of consumption than the Alcohol Use Questionnaire.

ARTICLE INFO

Keywords:
Alcohol
Smartphone applications
Ambulatory assessment
Timeline Followback
Alcohol use questionnaire

ABSTRACT

Introduction: Research investigating problem drinking often relies on retrospective measures to assess alcohol consumption behaviour. Limitations associated with such instruments can, however, distort actual consumption levels and patterns. We developed the smartphone application (app), CNLab-A, to assess alcohol intake behaviour in real-time.

Methods: Healthy individuals ($N = 671$, M age 23.12) completed demographic questions plus the Alcohol Use Questionnaire and a 21-day Timeline Followback before using CNLab-A for 21 days. The app asked participants to record alcohol consumption details in real time. We compared data reported via retrospective measures with that captured using CNLab-A.

Results: On average, participants submitted data on 20.27 days using CNLab-A. Compared to Timeline Followback, a significantly greater percentage of drinking days (24.79% vs. 26.44%) and significantly higher total intake (20.30 vs. 24.26 standard drinks) was recorded via the app. CNLab-A captured a substantially greater number of high intake occasions at all levels from 8 or more drinks than Timeline Followback. Additionally, relative to the Alcohol Use Questionnaire, a significantly faster rate of consumption was recorded via the app.

Conclusions: CNLab-A provided more nuanced information regarding quantity and pattern of alcohol intake than the retrospective measures. In particular, it revealed higher levels of drinking than retrospective reporting. This will have implications for how particular at-risk alcohol consumption patterns are identified in future and might enable a more sophisticated exploration of the causes and consequences of drinking behaviour.

1. Introduction

In research focusing on the causes and consequences of problem drinking, accurate assessment of alcohol use – both in terms of volume and pattern – is vital. Currently, such research relies primarily on data collected using well-validated and reliable retrospective self-report measures administered in the laboratory. Retrospective measures can be broadly categorised into two main types: quantity-frequency surveys and retrospective diaries. In their simplest form, quantity-frequency surveys ask participants to specify how much they usually drink and how often (Utpala-Kumar & Deane, 2010). This permits average daily consumption to be determined rapidly and efficiently, but does not allow for any investigation of pattern of intake (Del Boca & Darkes, 2003). Moreover, such surveys have been found to under-estimate total consumption by almost 30% when compared to prospective daily assessments (Heeb & Gmel, 2005) and up to 50% when compared to per

* Corresponding author.
E-mail address: antoinette.poulton@unimelb.edu.au (A. Poulton).

http://dx.doi.org/10.1016/j.addbeh.2017.11.003
Received 28 June 2017; Received in revised form 1 November 2017; Accepted 3 November 2017
0306-4603/

capita sales of alcohol (Stockwell et al., 2004; Stockwell, Zhao, Chikritzhs, & Greenfield, 2008). This discrepancy is attributed, at least in part, to difficulties participants have conceptualizing their usual consumption, especially if their intake is highly variable across drinking occasions, a tendency to overlook occasional high intake sessions, and recall bias (Del Boca & Darkes, 2003; Stockwell et al., 2004). Participants are consequently thought to report modal rather than average consumption in response to quantity-frequency surveys (Utpala-Kumar & Deane, 2010).

The accuracy and detail of data obtained from quantity-frequency surveys can be enhanced somewhat by the inclusion of additional components. Beverage-specific questions, an explicit reference period, and including items that assess high intake behaviour have all been found to generate increased estimates of consumption and more comprehensive information about pattern of intake (Dawson, 2003; Del Boca & Darkes, 2003; Rehm, 1998). A number of measures adopt this type of approach. The Alcohol Use Questionnaire (AUQ), for instance, combines beverage-specific quantity-frequency questions about weekly intake over a six-month period with items that assess drinking behaviour and subjective effects (Mehrabian & Russell, 1978). Weekly intake, rate of consumption, and a composite binge score index can be derived from this measure (Townshend & Duka, 2002). Although this provides some insight into total intake and pattern of consumption, it nonetheless has still been found to under-estimate total intake when compared to prospective daily assessments (Townshend & Duka, 2002). Furthermore, binge behaviour is frequently determined based on tertile or median splits of the sample (Townshend & Duka, 2005; Townshend, Kambouropoulos, Griffin, Hunt, & Milani, 2014). Results are therefore inextricably bound to the sample studied (Bø, Billieux, & Landrø, 2016).

Retrospective diary methods of collecting alcohol consumption information ask participants to recall day-to-day intake over a preceding designated time period (Del Boca & Darkes, 2003). Alcohol Timeline Followback (TLFB; Sobell & Sobel, 1992) is a commonly employed, well-validated example of this technique. Such diaries furnish researchers with considerable information pertaining to volume consumed and pattern of intake (Kypri, Langley, & Stephenson, 2005). In a review of more than thirty papers, however, Feunekes, van't Veer, van Staveren, and Kok (1999) found these diaries significantly under-estimate consumption levels – by about 20% – when compared to quantity-frequency surveys and prospective assessments. A study comparing repeated 7-day TLFB with 30-day TLFB revealed how more frequent assessments identified higher volume of intake, greater frequency of binge episodes and fewer abstinent days, with the absolute value of volumetric discrepancies between the two measures increasing as a function of length of recall (Hoeppner, Stout, Jackson, & Barnett, 2010). Longer assessment time periods yield less precise data (Ekholm, 2004); however, collecting data over narrow time frames distorts alcohol consumption information because drinking behaviour has been found to vary considerably from week to week (Del Boca, Darkes, Greenbaum, & Goldman, 2004; Goldman, Greenbaum, Darkes, Brandon, & Del Boca, 2011).

Real-time assessment of alcohol intake and pattern of consumption potentially overcome disadvantages associated with retrospective measures. Such assessment enables drinking to be recorded repeatedly, in close proximity to the event, often in the natural environment, and in the absence of the researcher (Trull & Ebner-Priemer, 2013). To date, studies comparing daily intake recorded via hand-held electronic devices and interactive voice response systems with that captured using retrospective methods have yielded varied results. In some cases, real-time assessments have revealed significantly higher consumption (Searles, Helzer, Rose, & Badger, 2002) or greater variability of intake (Carney, Tennen, Affleck, Del Boca, & Kranzler, 1998), while in others no significant differences have been identified (Bernhardt et al., 2009). Moreover, such devices pose a significant cost to researchers, potentially limiting the scope of the research (Kuntsche & Labhart, 2013; Trull & Ebner-Priemer, 2013). They additionally place considerable burden on participants, possibly diminishing compliance (Kuntsche & Labhart, 2013; Shiffman, Stone, & Hufford, 2008). With the advent of smartphones and application technologies (apps), real-time assessment limitations can be reduced. Researchers can take advantage of high smartphone ownership (Poushter, 2016), for instance, by using apps on participants' own devices to collect data. Apps enable considerable information about drink type, size, and ethanol content to be quickly logged, reducing the burden on participants in terms of time required to record information. Their capacity to compute rate of consumption and standard drink calculations reduce the potential for inaccurate reporting.

Few studies to date have validated app assessment of alcohol intake and pattern of consumption with retrospective measures. Monk, Heim, Qureshi, and Price (2015) found healthy participants ($n = 51$) recorded greater consumption when using an app for 7 days as compared to when such information was gathered retrospectively using researcher-generated beverage-specific surveys. Similarly, alcohol intake recorded daily via an app over a six-week period was shown to be significantly higher than when reported using TLFB; indeed, discrepancies between the measures increased over time (Dulin, Alvarado, Fitterling, & Gonzalez, 2017). Participants ($n = 25$) in this study, however, were diagnosed as alcohol dependent and were simultaneously undertaking treatment modules deployed by the app (Dulin et al., 2017). In both studies, sample size was relatively modest and retrospective data were collected after real-time recording, potentially enhancing participants' recall of drinking information and thus underestimating differences between real-time and retrospective data. This later point is a limitation often cited by researchers in this area (Carney et al., 1998; Perrine, Mundt, Searles, & Lester, 1995; Toll, Cooney, McKee, & O'Malley, 2006). Moreover, neither study examined differences in pattern of alcohol intake across measures.

The aim of this study was to examine differences between alcohol consumption information captured using an app for 21 days with data recorded via traditional, well-validated retrospective measures – namely, 21-day TLFB and the AUQ – in a large healthy sample. We chose a 21-day time frame in order to capture drinking variability; this appears to be particularly important when assessing binge patterns of alcohol intake (Courtney & Polich, 2009). We hypothesised that, relative to retrospective reports, indices related to alcohol intake and pattern of consumption would be greater when recorded via CNLab-A. Further, we expected app data to better accord with Australian Bureau of Statistics (ABS) apparent per capita alcohol consumption.

2. Methods

2.1. Participants

The present study consists of data from 671 participants (*M* age 23.12, *SD* = 7.24, range: 16–56, 70% female) that form a subset of an ongoing project – entitled CheckMyControl – investigating the relationship between alcohol use and various social and cognitive factors in the healthy population (see Fig. 1). Participants were recruited via adverts posted in and around the University of Melbourne, researcher networks, and social media posts. The University of Melbourne Human Ethics Committee approved the study in accordance with the standards for ethical research of the National Health and Medical Research Council.

2.2. Procedure

After reading a plain language statement and providing informed consent, participants answered a brief online researcher-devised demographic survey. They then completed the Alcohol, Smoking and Substance Involvement Screen (ASSIST), Alcohol Use Disorders Identification Test (AUDIT), AUQ, and a 21-day TLFB. Finally, participants were required to download and use a smartphone app for

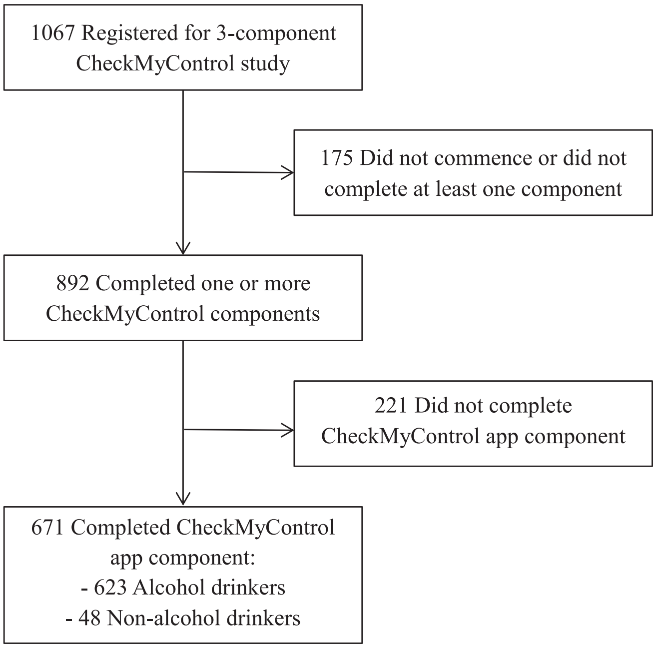

Fig. 1. Flow diagram following Consolidated Standards of Reporting Trials (CONSORT) guidelines of study participation.

recording alcohol use over a period of 21 days. They were compensated AU$10 for time spent completing online surveys and AU$0.50 each day information about alcohol consumption was submitted via the app (regardless of whether alcohol had been consumed or not). Participants received a bonus AU$9.50 if app data were submitted on all 21 days. The maximum participants could be reimbursed was AU$30.

2.3. Measures

2.3.1. Drug and alcohol screens

The ASSIST is designed to identify harmful use of alcohol, tobacco, and illicit drugs (WHO ASSIST Working Group, 2002). It comprises eight questions covering 10 substances and assesses frequency of use and associated problems over the preceding three months.

The AUDIT (Saunders, Aasland, Babor, de la Fuente, & Grant, 1993) is a 10-item screening measure that asks participants to respond to questions assessing alcohol intake, problems, and dependence with reference to the preceding six months. Scores of eight or more indicate hazardous alcohol consumption (Babor, Higgins-Biddle, Saunders, & Monteiro, 2001).

2.3.2. Retrospective alcohol consumption

The AUQ asks quantity/frequency questions pertaining to alcohol consumption as well as questions related to speed of intake and drunkenness over the preceding six months. Weekly use, average rate of consumption, and binge score can be derived from this measure (Mehrabian & Russell, 1978; Townshend & Duka, 2002). Tertile splits of the binge scores of regular drinkers can be used to assign binge and non-binge group membership (e.g., Townshend & Duka, 2005; non-binge ≤ 16, binge ≥ 24). As applied to the drinkers in this sample ($n = 623$), scores of 10 or less denote non-bingers while scores of 24 or more suggest binge behaviour.

TLFB is a retrospective diary method of collecting alcohol consumption information (Sobell & Sobel, 1992). Participants are presented with a visual calendar and are required to record the number of standard drinks consumed each day over a set number of preceding days. They are encouraged to use cues – such as holidays and special events – to assist their memory. TLFB enables variables – such as the percentage of drinking days, total number of standard drinks consumed, average standard drinks consumed per day and/or per drinking day, and number of occasions 4 (or more) drinks are consumed in 1 day – to be quantified.

A visual reference, which detailed how many standard drinks are contained in common serving sizes of various alcoholic beverages in Australia, was available to participants as they responded to retrospective surveys. This visual reference utilised the same images and information about alcohol content, drink size, and standard drinks as used in the app.

2.3.3. Smartphone app

CNLab-A is a freely available app that can be used on iPhone (iOS 8.4 +) or android (Kitkat 4.1 +) smartphones to record alcohol intake. Once downloaded, CNLab-A requires participants to allow it to send them notifications. One notification is pre-set to 8 am while the other can be set to suit the user. While participants are directed at the outset to record alcohol consumption as it happens (or as soon thereafter as possible), notifications serve to prompt individuals to input information twice daily in case they neglect to do so when drinking. Thus, alcohol intake data can be submitted at any time, either in response to notifications or while drinking. A unique ID code, provided to participants during the online component of the study, is also required before the app opens.

The app is programmed to include commonly available alcoholic drinks. For each beverage, users can choose from a range of sizes corresponding to those sold in licensed premises and retail outlets across Australia. Participants are instructed that in some cases they may need to use the closest approximation to their drink size. Submitted data, including all "No" responses, upload to a server and are automatically date and time stamped. An instructional video explaining how to use CNLab-A is available to participants (https://youtu.be/WNqr-otRsTM). See Fig. 2 for further details.

2.4. Data analysis

For TLFB and the app, data related to the number of days drinking, total standard drinks, and occasions where four or more/six or more (and so forth) drinks were consumed in one episode were aggregated across days for each individual. Average drinks per day and per drinking day were calculated by dividing total standard drinks consumed by 21 and number of days drinking respectively. Average weekly consumption was determined by dividing total standard drinks consumed by three. Where participants uploaded less than 21 days of app-based data, daily consumption was calculated as a function of the number of submission days; average weekly intake was similarly determined. Each time drinking was submitted via the app, an hourly rate of consumption was computed based on the start/end time recorded by participants. This allowed average daily rate of consumption to be computed for each individual across the experimental period.

Repeated measures analysis of variance (ANOVA) and paired-samples *t*-tests were conducted to assess differences between TLFB, CNLab-A, and AUQ data. Sphericity was assessed using Mauchly's test; where this assumption was violated, Greenhouse-Geisser corrections were applied. Paired comparisons were Bonferroni-corrected for multiple comparisons. Effect sizes were computed for *t*-tests using *r*-values and for one-way repeated measures ANOVA using partial eta squared (η_p^2); they were interpreted according to Cohen's guidelines: 0.01 = small, 0.06 = moderate, and 0.14 = large effect (Cohen, 1988).

3. Results

3.1. Sample demographics

At the time of testing, 39.9% of the sample was aged under 20, 46.2% was aged 20 to 29, and 13.9% was aged 30 or over. Participants had completed, on average, 14.80 years of education ($SD = 2.31$).

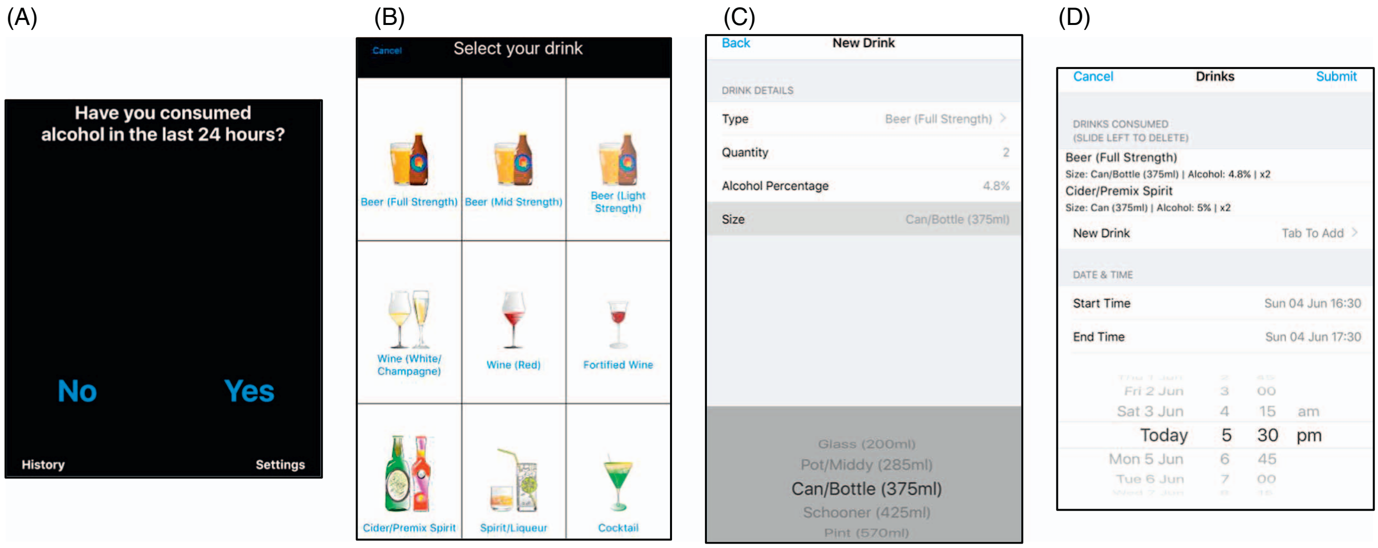

Fig. 2. Screenshots from the CNLab-A app. On opening, CNLab-A asks users if alcohol has been consumed in the last 24 h (A). Thereafter, participants are asked if they have consumed alcohol since their last submission. If they indicate – by pressing "No" – that no drinking has occurred, the app can be closed. This serves to differentiate between participants who have forgotten to upload data and those who have not been drinking. If participants indicate drinking has occurred – by pressing "Yes" – images of common alcoholic beverages (including beer, wine, cider/premix, spirit/liqueur, and cocktail) are displayed (B). Participants then select the type of beverage consumed by touching the appropriate image on the screen. They are required to indicate quantity and size consumed for each beverage via a simple scroll option menu (C). Alcohol content as a function of beverage type is prefilled. Participants are able to repeat this process by tapping "Back" in order to add as many drink types as required. Erroneously entered data can be deleted by swiping left. Prior to submitting data, participants must also specify the start and end time of their drinking, again using a scroll option menu (D). Data cannot be submitted more than 15 min ahead of the current time or after more than 24 h have elapsed in order to circumvent potential forward and/or back filling. Participants are able to either report their drinking in separate sessions or they can leave the app open to the 'drinks' screen so as to record beverages as they are consumed. The later option still allows participants to use other features on their phone. Participants can access a history of their submission dates and times (but not their drinking data) via the "History" button. At the conclusion of the experimental period, an automated message thanks participants and gives them simple feedback regarding the number of days they consumed alcohol, total standard drinks consumed, and average daily consumption.

Table 1
Alcohol and other drug use characteristics as a percentage of the sample (N = 671).

	%
Age at first drink	
Never	2.20
≤ 12	3.40
13–15	36.70
16–17	37.10
≥ 18	20.60
AUD diagnosis	
Self	0.30
1st degree relative	3.90
SUD diagnosis	
Self	0.40
1st degree relative	2.70
ASSIST moderate/high risk drug use	
Tobacco	20.90/1.80
Alcohol	33.20/3.10
Cannabis	14.30/0.40
Cocaine	3.30/0.00
Amphetamines	7.90/0.00
Inhalants	2.10/0.00
Sedatives	4.30/0.30
Hallucinogens	4.20/0.00
Opiates	2.10/0.00
Other	0.40/0.00

Note. AUD = alcohol use disorder. SUD = substance use disorder. All AUD/SUD frequencies are derived from self-report data. ASSIST = alcohol, Smoking and Substance Involvement Screen. Drug involvement assessed via the ASSIST is categorised as low (≤ 10 for alcohol; ≤ 3 for all other drugs), moderate (11–26 for alcohol; 4–26 all other drugs) or high risk (≥ 27).

Their mean AUDIT and AUQ binge scores were 7.97 (SD = 5.49) and 20.09 (SD = 19.74) respectively. Alcohol and other drug use characteristics are presented in Table 1.

3.2. App compliance and adherence

On average, participants used CNLab-A 20.27 (SD = 1.88) days out of 21. As data submission was either event- or notification-contingent, there was no upper limit to the number of drinking sessions participants could report using the app. Participants received a maximum of 42 notifications asking them to record information about their drinking. They submitted data, on average, 2.00 (SD = 0.41) times per day. There were 27,355 data points captured via the app in total. A one-way repeated measures ANOVA showed significant differences between weekly totals of app data, $F(1.97, 1321.07) = 38.67$, $p < 0.001$, $\eta_p^2 = 0.06$. Pairwise comparisons revealed average total intake recorded via CNLab-A during the first week (M = 9.14, SD = 10.90) was significantly higher than that recorded in either of the subsequent weeks ($p < 0.001$); there was no difference between average total intake recorded in the second (M = 8.02, SD = 9.70) and third weeks (M = 7.11, SD = 9.48, p = 0.449).

3.3. Comparisons between retrospective and CNLab-A data

Average alcohol consumption indices, as captured via TLFB and CNLab-A, and results of paired samples t-tests are presented in Table 2. A repeated measures ANOVA showed significant differences in average weekly intake as recorded via TLFB (M = 6.77, SD = 7.34), CNLab-A (M = 8.42, SD = 8.75), and the AUQ (M = 10.18, SD = 22.53), $F(1.13, 758.29) = 11.53$, $p < 0.001$, $\eta_p^2 = 0.02$. Pairwise comparisons revealed average weekly consumption was significantly higher when captured using either CNLab-A ($p < 0.001$) or the AUQ ($p < 0.001$), as compared to TLFB; there was no difference in average weekly consumption recorded via CNLab-A or the AUQ (p = 0.128). A paired samples t-test indicated average hourly rate of consumption was significantly higher when measured via CNLab-A (M = 2.20, SD = 2.09), as compared to the AUQ (M = 1.94, SD = 1.30), $t(670) = 3.14$, p = 0.002, r = 0.12.

Table 2
Average alcohol intake indices as recorded via 21-day TLFB and the CNLab-A app (21 days).

	TLFB	CNLab-A			95% CI		
	M (SD)	*M (SD)*	*t*(670)	*p*	LL	UL	*r*
Percent days drinking	24.79 (21.00)	26.44 (20.75)	2.73	0.007	0.46	2.84	0.10
Total drinks	20.30 (22.02)	24.26 (25.41)	5.37	< 0.001	2.51	5.41	0.20
Drinks per day	0.97 (1.05)	1.20 (1.25)	6.58	< 0.001	0.17	0.31	0.25
Drinks per drinking day	3.35 (2.77)	3.98 (3.02)	6.72	< 0.001	0.45	0.81	0.25
4/4 + intake	2.09 (2.63)	2.16 (2.58)	0.88	0.381	− 0.09	0.24	0.03
6/6 + intake	1.26 (1.90)	1.31 (1.92)	0.74	0.458	− 0.08	0.18	0.03
8/8 + intake	0.72 (1.39)	0.85 (1.44)	2.47	0.014	0.03	0.23	0.09
12/12 + intake	0.21 (0.71)	0.31 (0.80)	3.18	0.002	0.04	0.16	0.12
20/20 + intake	0.02 (0.25)	0.07 (0.31)	3.22	0.001	0.02	0.07	0.12

Note. Drinks refer to self-reported alcohol consumption in Australian standard drinks (1 drink = 10 g alcohol). CI = confidence interval; LL = lower limit; UL = upper limit. TLFB = Timeline Followback. 4/4 + (and so forth) intake refers to occasions where four or more drinks were consumed in one episode.

3.4. Comparisons between CNLab-A data and national alcohol sales

Across the 21-day experimental period, participants indicated consuming, on average, 24.26 standard drinks via CNLab-A and 20.30 via TLFB. In Australia, each standard drink contains 12.7 ml of pure alcohol. Thus, 308.10 ml of alcohol was consumed according to CNLab-A and 257.81 ml according to TLFB. In a year, this equates to 5340 ml and 4469 ml respectively. Annual apparent alcohol consumption by Australians 15 years and older in 2013–2014 was 9700 ml (ABS, 2017).

4. Discussion

This study investigated differences between alcohol intake and pattern of consumption information collected via a smartphone app with data gathered using traditional, well-validated, retrospective measures in a large healthy sample. We hypothesised that, relative to both a 21-day TLFB and the AUQ, indices related to quantity of intake and pattern of consumption would be greater when recorded via the 21-day CNLab-A app. Results revealed participants recorded drinking on a significantly greater percentage of days when using the app, as compared to TLFB. Results also indicated significantly higher intake – whether expressed in total volume, average drinks per day or average drinks per drinking day – when using CNLab-A. Additionally, the app captured a significantly greater number of high intake occasions at all levels from 8 or more drinks than TLFB. With regard to the AUQ, while participants recorded significantly greater average weekly intake using this measure than TLFB, there was no difference on this index between the AUQ and CNLab-A. This is an interesting finding given weekly alcohol intake derived from the AUQ is rarely employed in research assessing consumption (but see Townshend & Duka, 2002 as an exception). The app recorded a significantly higher hourly rate of intake, however, than the AUQ. Thus, while the AUQ might be an efficient method of assessing weekly consumption, it may be less useful as a means of investigating pattern of intake.

We also compared CNLab-A data with ABS apparent per capita alcohol consumption statistics. Calculations based on 21 days of data showed the app captured 55% of per capita consumption. It should be noted apparent per capita alcohol consumption, which is derived from national customs and excise data, is thought to over-estimate the true level of alcohol intake in Australia, as it does not adjust for wastage or for alcohol used in cooking and for other purposes (ABS, 2017). There is also some evidence to suggest variation in consumption as a function of Australian state or territory, though figures for Victoria are not yet available (Loxley, Gilmore, Catalano, & Chikritzhs, 2016). Additionally, while the sample in this study was relatively large, undergraduate students formed a substantial sub-group (85.7%) and their alcohol intake may not be representative of the general population. For instance, only 0.3% of participants indicated any diagnosis of alcohol use disorder, yet current data suggest the Australian prevalence rate of this disorder stands at 1.4% (Slade et al., 2009); high intake drinkers may therefore have been under-represented in our study. As > 50% of all alcohol consumed is drunk by the top 10% of high intake drinkers (Livingston, 2013), future research should seek to validate the use of apps to assess real-time alcohol consumption in high intake samples.

Our results accord with those from other studies examining app-based real-time assessment of alcohol consumption. Both Monk et al. (2015) and Dulin et al. (2017) found apps reveal greater levels of intake than retrospective surveys. In the case of the former, investigators utilised researcher-generated measures, whereas TLFB – often considered the gold standard in assessing alcohol consumption (Leeman et al., 2010) – was employed in the later. In both studies, real-time and retrospective assessment covered the same time periods. In this study, we sought to validate app-based real-time assessment against both the gold standard and the AUQ, but – in order to reduce the possibility real-time reporting enhanced recall of drinking on retrospective measures – we assessed drinking over two different time periods. While it is therefore possible differences between TLFB and the app reflect actual differences in drinking between the two time periods, we feel this is unlikely, particularly as total intake averaged across the two measures still only captures 50% of apparent alcohol consumption in Australia. Regardless, our findings support the notion that real-time app-based assessment of alcohol intake is at least an equivalent method of assessing drinking as TLFB.

While compliance with app protocols was high – as indicated by a 96.5% response rate – quantity of intake reduced significantly from the first to the second experimental week (but not from the second to the third week). It is possible that for periods greater than 7 days, participants found the app burdensome and so recorded “No” responses even when alcohol was consumed. It is equally probable, however, that participants curbed their drinking in response to using CNLab-A, even though no feedback pertaining to intake or national drinking guidelines was provided. Indeed, reductions in alcohol consumption due only to measurement have been reported in other studies (McCambridge & Kypri, 2011; Smith, Dash, Johnstone, Houben, & Field, 2017). Future studies should seek to examine further how drinking and/or responding changes over time in response to app-based measurement; a multi-level modeling approach, as opposed to the repeated measures ANOVA employed in this study, might be a more pertinent means of shedding light on this phenomenon.

With regard to the CNLab-A app in particular, several limitations must be noted. Firstly, any recording of drinks was limited by the options available in the app. While we ensured all drinks sizes (sold in Australian licensed and retail outlets) and associated alcohol content for beer (all strengths), wine, champagne, fortified wine, spirits, ciders/premix drinks, and cocktails were programmed into the app, participants may have consumed other types of alcoholic beverages or those with non-standard alcohol content. Such drinks may have been recorded as some other drink or they might not have been reported at all.

Similarly, although licensed premises and retail outlets in Australia are required by law to sell types of alcohol by reference to volume (National Measurement Institute, 2009), research has shown standard serve sizes are rarely adhered to when consumption takes place in non-licensed locations (Kerr & Stockwell, 2012). Inaccuracies related to serving size would, however, likely be reported across both retrospective measures and the CNLab-A app, especially as the visual reference supplied for the retrospective measures contained the same images and information about alcohol content and drink size as the app. Finally, although participants were encouraged to input drinking information as they consumed alcohol or very shortly thereafter, there is no guarantee they did so. Participants may have submitted data a number of hours after drinking depending on the timing of notifications. In this case, the data would be more akin to that derived from daily interviews and may still therefore under-estimate actual consumption.

In sum, limitations associated with retrospective methods of assessing alcohol intake impact the accuracy and detail of drinking behaviour information (Del Boca & Darkes, 2003; Feunekes et al., 1999; Heeb & Gmel, 2005; Hoeppner et al., 2010; Stockwell et al., 2004, 2008; Townshend & Duka, 2002; Utpala-Kumar & Deane, 2010). While real-time assessment methods involving hand-held electronic devices or interactive voice response systems have shown some promise in terms of overcoming drawbacks associated with retrospective measures, they can be expensive and burdensome (Kuntsche & Labhart, 2013; Shiffman et al., 2008; Trull & Ebner-Priemer, 2013). Apps, by contrast, can be downloaded to participants' own smartphones and allow for considerable information to be easily and quickly recorded and/or computed. Previous research has established participants log higher alcohol consumption when using a smartphone app, as compared to intake recorded via retrospective surveys (Dulin et al., 2017; Monk et al., 2015). We replicate and extend these findings by showing participants indicated higher total consumption, greater number of high intake episodes, and increased rate of consumption when using an app for 21 days than when assessed using traditional retrospective measures, such as 21-day TLFB and the AUQ. This will have implications for how particular alcohol consumption patterns are identified in future and might enable a more detailed exploration of the causes and consequences of drinking behaviour.

Role of funding source

This research was supported by an Australian National Health and Medical Research Council grant (1050766), and an Australian Research Council fellowship (FT110100088). The funding bodies had no role in the study design, collection, analysis or interpretation of data, writing the manuscript, or in the decision to submit the paper for publication.

Contributors

AP and RH designed the study. JP and AP developed the smartphone application. LB and RS programmed electronic protocols and built the database. AP collected data and conducted the statistical analysis. Both AP and RH contributed to the writing and editing of this article. All authors have read and approved the final manuscript.

Conflict of interest

No conflict declared.

Acknowledgements

The authors wish to thank Jiajie Li and Jemie Effendy who assisted with upgrades to the smartphone application.

References

Australian Bureau of Statistics (2017). 4307.0.55.001 – Apparent consumption of alcohol, Australia, 2013–14. Retrieved November 1, 2017, from http://www.abs.gov.au/ausstats/abs@.nsf/mf/4307.0.55.001/.

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *AUDIT: The alcohol use disorders identification test: Guidelines for use in primary care.* Geneva: World Health Organisation. Retrieved from http://www.who.int/substance_abuse/publications/alcohol/en/.

Bernhardt, J. M., Usdan, S., Mays, D., Martin, R., Cremeens, J., & Arriola, K. J. (2009). Alcohol assessment among college students using wireless mobile technology. *Journal of Studies on Alcohol and Drugs, 70*, 771–775. http://dx.doi.org/10.15288/jsad.2009.70.771.

Bø, R., Billieux, J., & Landrø, N. I. (2016). Binge drinking is characterized by decisions favoring positive and discounting negative consequences. *Addiction Research & Theory,* 1–8. http://dx.doi.org/10.3109/16066359.2016.1174215.

Carney, M. A., Tennen, H., Affleck, G., Del Boca, F. K., & Kranzler, H. R. (1998). Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real time "electronic interviews". *Journal of Studies on Alcohol, 59*, 447–454.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NY: Lawrence Erlbaum Associateshttp://dx.doi.org/10.1234/12345678.

Courtney, K. E., & Polich, J. (2009). Binge drinking in young adults: Data, definitions, and determinants. *Psychological Bulletin, 135*, 142–156. http://dx.doi.org/10.1037/a0014414.

Dawson, D. A. (2003). Methodological issues in measuring alcohol use. *Alcohol Research and Health, 27*, 18–29.

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

Del Boca, F. K., Darkes, J., Greenbaum, P. E., & Goldman, M. S. (2004). Up close and personal: Temporal variability in the drinking of individual college students during their first year. *Journal of Consulting and Clinical Psychology, 72*, 155–164. http://dx.doi.org/10.1037/0022-006X.72.2.155.

Dulin, P. L., Alvarado, C. E., Fitterling, J. M., & Gonzalez, V. M. (2017). Comparisons of alcohol consumption by timeline follow back vs. smartphone-based daily interviews. *Addiction Research & Theory, 25*, 195–200. http://dx.doi.org/10.1080/16066359.2016.1239081.

Ekholm, O. (2004). Influence of the recall period on self-reported alcohol intake. *European Journal of Clinical Nutrition, 58*, 60–63. http://dx.doi.org/10.1038/sj.ejcn.1601746.

Feunekes, G. I. J., van't Veer, P., van Staveren, W. A., & Kok, F. J. (1999). Alcohol intake assessment: The sober facts. *American Journal of Epidemiology, 150*, 105–112.

Goldman, M. S., Greenbaum, P. E., Darkes, J., Brandon, K. O., & Del Boca, F. K. (2011). How many versus how much: 52 weeks of alcohol consumption in emerging adults. *Psychology of Addictive Behaviors, 25*, 16–27. http://dx.doi.org/10.1037/a0021744.

Heeb, J., & Gmel, G. (2005). Measuring alcohol consumption: A comparison of graduated frequency, quantity frequency, and weekly recall diary methods in a general population survey. *Addictive Behaviors, 30*, 403–413. http://dx.doi.org/10.1016/j.addbeh.2004.04.022.

Hoeppner, B. B., Stout, R. L., Jackson, K. M., & Barnett, N. P. (2010). How good is fine-grained timeline follow-back data? Comparing 30-day TLFB and repeated 7-day TLFB alcohol consumption reports on the person and daily level. *Addictive Behaviors, 35*, 1138–1143. http://dx.doi.org/10.1016/j.addbeh.2010.08.013.

Kerr, W. C., & Stockwell, T. (2012). Understanding standard drinks and drinking guidelines. *Drug and Alcohol Review, 31*, 200–205. http://dx.doi.org/10.1111/j.1465-3362.2011.00374.x.

Kuntsche, E., & Labhart, F. (2013). Using personal cell phones for ecological momentary assessment: An overview of current developments. *European Psychologist, 18*, 3–11. http://dx.doi.org/10.1027/1016-9040/a000127.

Kypri, K., Langley, J., & Stephenson, S. (2005). Episode-centred analysis of drinking to intoxication in university students. *Alcohol and Alcoholism, 40*, 447–452. http://dx.doi.org/10.1093/alcalc/agh178.

Leeman, R. F., Heilig, M., Cunningham, C. L., Stephens, D. N., Duka, T., & O'Malley, S. S. (2010). Ethanol consumption: How should we measure it? Achieving consilience between human and animal phenotypes. *Addiction Biology, 15*, 109–124. http://dx.doi.org/10.1111/j.1369-1600.2009.00192.x.

Livingston, M. (2013). *Measuring risky drinking: An examination of the validity of different episodic drinking thresholds in predicting alcohol-related harms.* Centre for Alcohol Policy Research. Retrieved from http://fare.org.au/2013/08/measuring-risky-drinking-an-examination-of-the-validity-of-different-episodic-drinking-thresholds-in-predicting-alcohol-related-harms/.

Loxley, W., Gilmore, W., Catalano, P., & Chikritzhs, T. (2016). *National Alcohol Sales Data Project (NASDP): Stage 5 report.* Perth, Western Australia: National Drug Research Institute, Curtin University. Retrieved from https://ndri.curtin.edu.au/local/docs/pdf/nasdp/nasdp005.pdf.

McCambridge, J., & Kypri, K. (2011). Can simply answering research questions change behaviour? Systematic review and meta analyses of brief alcohol intervention trials. *PLoS One, 6*, e23748. http://dx.doi.org/10.1371/journal.pone.0023748.

Mehrabian, A., & Russell, J. A. (1978). A questionnaire measure of habitual alcohol use. *Psychological Reports, 43*, 803–806.

Monk, R. L., Heim, D., Qureshi, A., & Price, A. (2015). "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, 10*, e0126209. http://dx.doi.org/10.1371/journal.pone.0126209.

National Measurement Institute (2009). Guide to the sale of alcohol. Retrieved from http://measurement.gov.au/Industry/Business/Pages/Alcohol.aspx.

Perrine, M. W., Mundt, J. C., Searles, J. S., & Lester, L. S. (1995). Validation of daily self-reported alcohol consumption using interactive voise response (IVR) technology. *Journal of Studies on Alcohol, 56*, 487–490.

Poushter, J. (2016). Smartphone ownership and internet usage continues to climb in emerging economies but advanced economies still have higher rates of technology use. Retrieved from http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/.

Rehm, J. (1998). Measuring quantity, frequency, and volume of drinking. *Alcoholism: Cinical and Experimental Research, 22*, 4–14. http://dx.doi.org/10.1111/j.1530-0277.1998.tb04368.x.

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction, 88*, 791–804. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8329970.

Searles, J. S., Helzer, J. E., Rose, G. L., & Badger, G. J. (2002). Concurrent and retrospective reports of alcohol consumption across 30, 90 and 366 days: Interactive voice response compared with the Timeline Follow Back. *Journal of Studies on Alcohol, 63*, 352–362.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32.

Slade, T., Johnston, A., Teesson, M., Whiteford, H., Burgess, P., Pirkis, J., & Saw, S. (2009). *The mental health of Australians 2: Report on the 2007 national survey of mental health and wellbeing.* Canberra: Department of Health and Ageing.

Smith, J. L., Dash, N. J., Johnstone, S. J., Houben, K., & Field, M. (2017). Current forms of inhibitory training produce no greater reduction in drinking than simple assessment: A preliminary study. *Drug and Alcohol Dependence, 173*, 47–58. http://dx.doi.org/10.1016/j.drugalcdep.2016.12.018.

Sobell, L. C., & Sobel, M. B. (1992). Timeline follow-back: A technique for assessing self-reported alcohol consumption. In R. Z. Litten, & J. P. Allen (Eds.). *Measuring alcohol consumption: Psychosocial and biochemical methods* (pp. 41–72). Totowa, NJ: Humana Press.

Stockwell, T., Donath, S., Cooper-Stanbury, M., Chikritzhs, T., Catalano, P., & Mateo, C. (2004). Under-reporting of alcohol consumption in household surveys: A comparison of quantity–frequency, graduated–frequency and recent recall. *Addiction, 99*, 1024–1033. http://dx.doi.org/10.1111/j.1360-0443.2004.00815.x.

Stockwell, T., Zhao, J., Chikritzhs, T., & Greenfield, T. K. (2008). What did you drink yesterday? Public health relevance of a recent recall method used in the 2004 Australian National Drug Strategy Household Survey. *Addiction, 103*, 919–928. http://dx.doi.org/10.1111/j.1360-0443.2008.02219.x.

Toll, B. A., Cooney, N. L., McKee, S. A., & O'Malley, S. S. (2006). Correspondence between Interactive Voice Response (IVR) and Timeline Followback (TLFB) reports of drinking behavior. *Addictive Behaviors, 31*, 726–731. http://dx.doi.org/10.1016/j.addbeh.2005.05.044.

Townshend, J. M., & Duka, T. (2002). Patterns of alcohol drinking in a population of young social drinkers: A comparison of questionnaire and diary measures. *Alcohol and Alcoholism, 37*, 187–192.

Townshend, J. M., & Duka, T. (2005). Binge drinking, cognitive performance and mood in a population of young social drinkers. *Alcoholism: Clinical and Experimental Research, 29*, 317–325. http://dx.doi.org/10.1097/01.ALC.0000156453.05028.F5.

Townshend, J. M., Kambouropoulos, N., Griffin, A., Hunt, F. J., & Milani, R. M. (2014). Binge drinking, reflection impulsivity, and unplanned sexual behavior: Impaired decision-making in young social drinkers. *Alcoholism: Clinical and Experimental Research, 38*, 1143–1150. http://dx.doi.org/10.1111/acer.12333.

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology, 9*, 151–176. http://dx.doi.org/10.1027/1016-9040.14.2.95.

Utpala-Kumar, R., & Deane, F. P. (2010). Rates of alcohol consumption and risk status among Australian university students vary by assessment questions. *Drug and Alcohol Review, 29*, 28–34. http://dx.doi.org/10.1111/j.1465-3362.2009.00082.x.

WHO ASSIST Working Group (2002). The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): Development, reliability and feasibility. *Addiction, 97*, 1183–1194. http://dx.doi.org/10.1046/j.1360-0443.2002.00185.x.