

**Serveur Académique Lausannois SERVAL [serval.unil.ch](http://serval.unil.ch)**

## **Author Manuscript**

**Faculty of Biology and Medicine Publication**

**This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.**

Published in final edited form as:

**Title:** More is not always better-comparison of three instruments measuring volume of drinking in a sample of young men and their association with consequences.

**Authors:** Gmel G, Studer J, Deline S, Baggio S, N'Goran A, Mohler-Kuo M, Daeppen JB

**Journal:** Journal of studies on alcohol and drugs

**Year:** 2014 Sep

**Volume:** 75

**Issue:** 5

**Pages:** 880-8

**DOI:** <http://dx.doi.org/10.15288/jsad.2014.75.880>

In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.

**More is not always better – comparison of three instruments measuring volume of drinking in a sample of young men and their association with consequences**

*Gerhard Gmel<sup>1,2,3,4</sup>, Joseph Studer<sup>1</sup>, Stephane Deline<sup>1</sup>, Stephanie Baggio<sup>1</sup>, Alexandra N’Goran<sup>1</sup>, Meichun Mohler-Kuo<sup>5</sup>, Jean-Bernard Daeppen<sup>1</sup>*

<sup>1</sup>Alcohol Treatment Centre, Lausanne University Hospital CHUV, Av. Beaumont 21 bis,

Pavillon 2, CH - 1011 Lausanne, Switzerland, Stephanie.Baggio@chuv.ch,

Joseph.Studer@chuv.ch, Jean-Bernard.Daeppen@chuv.ch; Gerhard.Gmel@chuv.ch; Adjuda-

Alexandra.NGoran@chuv.ch; stephane.deline@chuv.ch.

<sup>2</sup>Swiss Institute for the Prevention of Alcohol and Drug Problems, Case postale 870, CH - 1001

Lausanne, Switzerland

<sup>3</sup>Centre for Addiction and Mental Health, 250 College St, Toronto, Ontario, M5T 1R8, Canada

<sup>4</sup>University of the West of England, Frenchay Campus Coldharbour Lane, Bristol BS16 1QY,

United Kingdom

<sup>5</sup>Institute of Social- and Preventive Medicine, University of Zurich, Hirschengraben 84, 8001

Zurich, Switzerland, meichun.mohler-kuo@uzh.ch

**Grant:** the Swiss National Science Foundation (33CSC0-122679).

**Corresponding author and guarantor:** G. Gmel; Alcohol Treatment Centre, Lausanne

University Hospital CHUV, Av. Beaumont 21 bis, Pavillon 2, CH-1011 Lausanne, Switzerland,

email: [gerhard.gmel@chuv.ch](mailto:gerhard.gmel@chuv.ch); Tel.:+41 21 3147352; Fax: +41 21 3140562

## **Abstract**

**Objective:** In general population survey instruments that measure volume of drinking, additional questions and shorter reference periods yield higher volumes. Comparison studies have focused on volume, but not on associations between volume and consequences.

**Method:** From a cohort study on substance use risk factors (C-SURF), baseline data were analyzed for 5,074 young (~ 20 years) males who were drinkers in the past 12 months. Volume of drinking was measured by a generic Quantity-Frequency (QF) instrument, an extended QF (separately for weekends and weekdays) instrument with 12-months recall, and a retrospective past week diary. Associations of consequences with and without attribution of alcohol as a cause, DSM-IV criteria for dependence, and DSM-5 alcohol use disorder in the past 12 months were analyzed.

**Results:** The generic QF resulted in lower volume compared with either the extended QF (more questions) or the retrospective diary (most questions and shortest recall period). For outcomes, however, the extended QF performed the best and the diary the worst.

**Conclusions:** Higher volume yields are not always better regarding associations with outcomes. The extended QF better captures the variability of drinking. The retrospective diary performs poorly for associations because of the mismatch with the recall period for past 12-months consequences and the potential for misclassification of past week abstainers and heavy drinkers due to an uncommon past week. Diaries are not recommended for research investigating individual associations between exposure and outcomes in young populations if consequences are measured with a sufficiently long interval to capture rare consequences.

## **Introduction**

Alcohol use is a serious public health concern and is a major risk factor for diseases, and injuries. Among adolescents and young adults alcohol use is the major risk factor for mortality and morbidity (Rehm et al., 2006). All over the world, adolescents and young adults are affected by a multitude of consequences. Besides the risk related to drinking and driving (Delcher et al., 2013), consequences include assaults and violence, academic problems, delinquency or other judicial consequences, vandalism and property damage, sexual assaults or unintended pregnancy, mental health condition, physical and medical health problems (Assanangkornchai et al., 2009; Diep et al., 2013; Hingson et al., 2009; Labhart et al., 2013; Wicki et al., 2010). Alcohol use is often measured in general population surveys. It is important to establish the best link between behaviors and consequences; accomplishing this depends on accurately measuring exposure to alcohol, consequences, and the time frames of exposure and outcome measures. Assessment methods for alcohol consumption at the individual level include e.g. self-reports, blood and breath analyses. Among other reasons, the advantages of self-reports are that they can be obtained relatively inexpensively and non-invasively from a large number of individuals compared with, e.g. biological markers, and they have shown reasonable levels of reliability and validity (Del Boca and Darkes, 2003). Compared with aggregate measures such as sales data it is possible to measure alcohol intake according to different characteristics of users and to link individual drinking behavior with other behaviors and consequences (Ekholm, 2004). However, it is important to measure alcohol use as accurately as possible (Dawson, 2003). There are a number of method comparison studies and reviews of alcohol measurement issues (Alanko, 1984; Bloomfield et al., 2013; Dawson, 2003; Del Boca and Darkes, 2003; Ekholm et al., 2008; Feunekes et al., 1999; Gmel and Rehm, 2004; Greenfield and Kerr, 2008; Lemmens et al., 1992; Rehm, 1998; Room, 1990). Very few of them have focused on adolescents and young adults (for

an exception see Townshend and Duka, 2002) and even fewer have attempted to validate various measurement approaches through their associations with potential consequences. The present study seeks to fill this gap.

The reviews cited above provide a wealth of information on differences, advantages and shortcomings of various instruments, but there is still no widely accepted "optimal" way to measure alcohol use. The information contained in these reviews can be summarized as follows. There are three main approaches to measuring alcohol intake in survey research: quantity–frequency (QF) measures with extended variants; graduated–frequency (GF) measures; and short-term recall measures (diaries). Generic QF assesses usual quantities per drinking days and multiplies them by usual frequencies of drinking to yield volume. Extensions of generic QF ask about frequency and quantity for different beverages, drinking on workdays versus weekend days, or in different contexts, e.g. at home, in bars or pubs, at someone else's home. The graduated frequency approach looks at the frequencies of different amounts, such as number of days that one consumes 1-2 drinks, 3-4 drinks, 5-7 drinks, etc. Finally, short-term recall measures, e.g. seven-day diaries, try to measure alcohol use as accurately as possible over a short period. This can be accomplished retrospectively or prospectively. A "prospective" measurement commonly consists of a series of subsequent 24-hour recalls in which respondents are typically asked to self-report at the end of each day before going to bed. The "yesterday" approach with the shortest recall period typically asks about consumption on the prior day only (Stockwell et al., 2008).

One of the clearest findings from method comparison studies is that the more questions asked, the higher the reported alcohol use (Dawson, 2003; Feunekes et al., 1999; Gmel and Rehm, 2004; Rehm, 1998). For example, Kühlnhorn & Leifman (1993) divided the days of the

week into Monday-Thursday and Friday-Sunday; this yielded higher volume than did a generic QF. Similarly, beverage-specific QF measures, where quantities and frequencies are asked separately for different alcoholic beverages (wine, beer, etc.) commonly yield higher volumes than do generic QF, as do instruments asking about consumption in different contexts, settings and locations (Casswell et al., 2002; Single and Wortley, 1993).

The recall period also has an influence on reporting, e.g. because of recall biases such as forgetting certain drinking occasions. Instruments with shorter recall periods, such as retrospective diaries over one week, on average yield higher volumes than do instruments with recall periods of 12 or 6 months, although such recall biases can occur within a week (Ekholm, 2004; Gmel and Daeppen, 2007). Thus, prospective diaries consistently yield higher volumes than do retrospective measures with longer recall periods (e.g. Leigh, 2000).

Coupled with the observation that survey measurements often result in estimates that are lower than sales data (Dawson, 2000; Del Boca and Darkes, 2003; Knibbe and Bloomfield, 2001; Midanik, 1982) is a common assumption regarding the validity of instruments that “the more the better” (Del Boca and Darkes, 2003; Leigh, 2000). This view has been criticized (Dawson, 1998; Del Boca and Noll, 2000; Rehm, 1998), because with more questions there is also the risk of double counting, e.g. when contexts (consumption at a friend’s home or at a party) or use of different beverages are overlapping. On some occasions as with the GF, instances of heavy consumption (e.g. 8-10 drinks) may be double counted when included with occasions of fewer drinks (e.g. 4-6 drinks), because respondents may have misunderstood the concept of non-overlapping occasions (Dawson, 1998; Del Boca and Noll, 2000; Rehm, 1998). The GF approach has also been criticized because it sometimes results in more than 365 drinking days a year for some respondents (Gmel et al., 2006; Graham et al., 2004; Poikolainen et al.,

2002). On the other hand, the GF approach is clearly one of the best measure to capture heavy drinking days (Stockwell et al., 2004), and there have been several suggestions how to fix the problem of overcounting drinking days (Greenfield et al., 2009; Stockwell et al., 2008). Nevertheless, it is not always certain that higher volumes automatically mean higher validity.

Most measurement comparison studies base their analyses on the aspect of which instrument yields the highest overall consumption (or, which is related, the most heavy drinkers), despite the claim that the most important aspect in epidemiology is the link between use as exposure and consequences (Dawson, 2003; Ekholm et al., 2008). The link with consequences may favor instruments other than those yielding the highest population volume estimates. As outlined by Gmel & Rehm (2004) it makes a difference whether the focus of research is on the most valid consumption levels aggregated for population segments (men, women, or different age groups, etc.) or on correlational analyses at the individual level (e.g. the link between exposure and outcome). For example, short recall measures such as the yesterday method may capture aggregated consumption better than would a 12-months generic volume measure (because the risk of not remembering drinking occasions is reduced). The advantage of the yesterday method is the very detailed assessment of alcohol use. The method is excellent - if aggregated over sufficiently large population subgroups – for several aspects such as the beverage-specific contribution to heavy drinking days, the impact of different strengths of the same type of alcoholic beverages (e.g. beers), or the compliance of population subgroups with drinking guidelines (Stockwell et al., 2008). However, short recall measures may not be representative of a respondent's general drinking pattern, particularly when alcohol use is variable (e.g. no daily use, or different phases with high and low volumes over a year). Someone who did not drink yesterday or during the past week is not necessarily an abstainer, while

someone who drank heavily yesterday or during the past week may have had a rather “uncommon” yesterday or past week and may be a moderate drinking otherwise. For example, Midanik and colleagues (2013) showed that the prevalence rate of current drinking was 47.3% with a 30 days measure, but was 20% higher (67.5%) with the 12 months measure. Additionally, short recalls of consumption may not coincide with the occurrence of consequences, where longer reference periods (e.g. 12 months) may be needed to capture those events that occur rarely (Dawson, 1998; Dawson and Room, 2000). In one of the rare studies comparing alcohol measurement instruments as regards their potential to predict consequences, Keogh et al. (2012) showed that a Food Frequency Questionnaire (FFQ, in principle comparable to a 12 months QF measurement) yielded a lower volume measure of alcohol use compared with a 7-day retrospective diary for the total sample. The association with breast cancer, however, was higher with the FFQ compared with the retrospective diary.

When measuring associations with individual data, a main assumption for a reliable instrument is that rank order should be preserved (Willett, 1998), i.e. despite potential biases, a "true" heavy drinker reports more alcohol use than a "true" moderate drinker. As stated by Greenfield and Kerr (2008) regarding self-reports, the glass is half full rather than half empty, because many instruments do perform fairly well in ranking individual intake. A second aspect when measuring associations with consequences is the matching of reference periods for both alcohol use and consequences. As alcohol-related consequences commonly have a low prevalence, long reference periods of e.g. 12 months have been recommended for the measurement of consequences (Dawson and Room, 2000).

To our knowledge, most studies use older subjects or general population data; there are only a handful that compare instruments within younger samples (Townshend and Duka, 2002)



where alcohol use is especially variable (e.g. heavy drinking on weekends, but no alcohol use on weekdays). We found no research among this age group that attempted to validate instruments with respect to convergent validity, namely, associations with consequences. The present study compares a generic QF with an extended QF measuring quantity and frequency separately on weekdays and weekends, and with a retrospective seven-day diary, but does not include a GF measure.

We hypothesize that the extended QF and the retrospective diary will yield higher volumes than the generic QF, and the retrospective diary will have the lowest associations with consequences for two reasons: First, it may not capture individual usual drinking patterns well, since drinking among youth is highly variable and a given past week often represents an uncommon week. Second, the mismatch between the time frame for the reporting of consequences and alcohol consumption may penalize the retrospective diary method.

## **Methods**

### *Sample*

The data are part of the Cohort Study on Substance Use Risk Factors (C-SURF), a longitudinal project designed to assess substance use patterns and related consequences in young Swiss men. Enrolment took place between August 2010 and November 2011 in three of the six Swiss army recruitment centers located in Lausanne (French-speaking), Windisch and Mels (German-speaking). These three centers cover 21 of the 26 cantons in the country, including all French-speaking cantons. In Switzerland, army recruitment is mandatory, so virtually all young men around 20 years old were eligible for study inclusion.

Of the 7,563 who gave written consent to participate, 5,990 (79.2%) completed a baseline questionnaire. Questionnaires were sent to consenters' home around two weeks later. So responses were given outside the army environment. We excluded lifetime abstainers, former drinkers, and those who have started drinking but have not yet consumed 12 drinks in their lifetime (n=900). Sixteen participants had insufficient alcohol use data and had to be dropped, leaving 5,074 past year drinkers in the final sample.

### *Alcohol use measures*

Three different measures assessed alcohol use. First, a generic QF measure asked for usual drinking frequency in the past 12 months (answer categories: 7 days per week down to 1 day per week, 2-3 times a month, once a month or less often) plus an open-ended question for the usual number of standard drinks per drinking day. Second, an extended QF measure asked separately for usual number of drinking days on weekends (Friday, Saturday and Sunday) in the past 12 months. Choices were 3 days per weekend, 2 days, 1 day, 2-3 days on weekends per month, 1 day or less per month. Quantities per drinking days on weekends were closed-ended (answer categories: 12 drinks or more, coded 13, 9-11 drinks, 7-8 drinks, 5-6 drinks, 3-4 drinks and 1-2 drinks). Midpoints of the categories were chosen. The same was done for weekdays (Monday to Thursday) with the highest frequency category of 4 days per week. Third, a retrospective diary asked for number of drinks during the past week on each day separately for beer, wine, liquor (whisky, vodka, etc), aperitifs (suze, martini), alcopops (ready-to-drink bottles containing spirits with around 5% alcohol by volume (ABV)), beer- and winepops, chillers, coolers (ready-to-drink bottles containing beer or wine, ABV around 3%), and mixed drinks. Illustrations of standard drinks were provided for each beverage containing approximately 10-12 grams of pure ethanol.

All three measures were converted into number of drinks per week. For method comparison the instruments with open-ended choices (retrospective diary and generic QF) were capped at a maximum of 13 drinks per drinking day (equating to the maximum number of drinks category on the extended QF).

We also measured risky single occasion drinking (RSOD) with the standard question of the Alcohol Use Disorder Identification Test (AUDIT) as frequency of 6 drinks or more on an occasion in the past 12 months (never, less than monthly, monthly, weekly daily or almost daily) whereby the last two categories were collapsed, because only 1.4% (n=72) drank such an amount daily or almost daily.

#### *Alcohol-related consequences with and without causal attribution*

Ten consequences from Wechsler et al. (1994) mentioning alcohol as a cause in the past 12 months included having problems with the police, drinking alcohol to get over the effects, blackouts, something done regretted later, unplanned sex, sex without condoms, accident or injury, argument or fight, damage to property, missed school or work or neglected family obligations. Each could be answered 'yes' or 'no'. A summary score from zero to 10 was constructed.

Gmel et al. (2010) have shown that due to criterion contamination, attributing alcohol as the cause of consequences and linking this to alcohol use may result in different associations with consequences than without mentioning alcohol as the cause. Therefore, sixteen consequences were selected from standard instruments (Bucholz et al., 1994; Hesselbrock et al., 1999; Hibell et al., 2012; Wechsler et al., 1994), without explicit mention of being alcohol-related. The items dealt with social problems (physical fights, problems with family, problems

with friends, performed poorly at school or work, theft, trouble with police, regretted sexual intercourse, sexual intercourse without condoms, and damaged property) and health consequences (accident/injury, admitted to an emergency room, attempted suicide, required medical treatment, spent a night in hospital, outpatient surgery, and treated in an emergency room because of an accident/injury). Each consequence was coded '0' if it did not occur in the past 12 months, or '1' if it did. A summary score was computed ranging from zero to 16.

#### *DSM-IV dependence and DSM-5 alcohol use disorder*

The questionnaire also contained items (Knight et al., 2002) that were originally adapted from the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz et al., 1994; Hesselbrock et al., 1999). They were designed to elicit self-reporting of DSM-IV dependence during the year preceding the survey, i.e. 12-month diagnoses, according to the American Psychiatric Association (APA, 1994).

The seven dependence criteria were: 1) tolerance; 2) withdrawal symptoms or withdrawal relief/avoidance; 3) drinking larger amounts or for longer periods than intended; 4) persistent desire or unsuccessful efforts to cut down or control drinking; 5) great deal of time spent on activities to obtain alcohol or to recover from its effects; 6) giving up or reducing important social, occupational or recreational activities in favor of drinking; and 7) continued drinking despite knowledge of a physical or psychological problem caused or exacerbated by drinking. The number of positive responses to the items was then summed.

For DSM-5 alcohol use disorders, a summary score of the seven dependence criteria plus three DSM-IV abuse criteria that were retained in DSM-5 (drinking in hazardous situations,

failure to fulfill major role obligations at work/school/home, and continued use despite persistent or recurrent social or interpersonal problems due to drinking) plus an item on craving was used.

Both the DSM-IV dependence and the DSM-5 alcohol use disorder were assessed on a continuum (i.e. number of criteria met) rather than having a dichotomous cut-off. This concurs with the current assumed dimensional structure of these concepts (see e.g. Kerridge et al., 2013).

### *Statistical analysis*

Descriptive statistics for mean weekly number of drinks as well as number of drinks at the 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup>, 95<sup>th</sup> and 97.5<sup>th</sup> percentiles were reported for the original and for the capped generic QF and retrospective diary measures. Pearson correlations were calculated for the capped and uncapped measures to measure the rank invariance across instruments. We also used natural logarithmic transformations to account for the skewness in the measures; a constant of one drink was added before applying the natural logarithmic transformations in order to avoid an undefined value of zero drinks. We used a proportional Venn diagram to determine the overlap of the top 25% drinkers on each measure. The diagram illustrates the proportion of those classified as top 25%-drinkers on all three measures, on each of two measures, and on a single measure only.

To analyze associations with outcomes, the number of consequences (with and without causal attribution to alcohol) and the number of criteria (dependence and AUD) were regressed on log-transformed alcohol use measures. Count models (i.e. Poisson models or negative binomial models) are appropriate for the number of consequences and the number of criteria. Vuong-Tests in Stata12 were used to check for over-dispersion in the Poisson models and for zero-inflation. In all cases, zero-inflated negative binomial models (zinb) were optimal since

over-dispersion in the Poisson models favored negative binomial models and zero-inflated negative binomial models were preferable to standard negative binomial models. Only the zinb-model results are reported. Differences in associations among the three instruments were conservatively tested by means of non-overlapping 95% confidence intervals of the coefficients of the zinb models. The regression coefficients are reported, and incident rate ratios (IRR) can be obtained by exponentiation of Euler's number  $e$  with the regression coefficient ( $e^{\text{coeff}}$ ). Models were run with and without adjusting for RSOD, because RSOD is an important dimension of alcohol use particularly among young men.

## Results

As expected (Table 1), the generic QF yielded the lowest mean consumption, the extended QF (with more questions) the second highest mean consumption, and the past week retrospective diary (with the most questions and a shorter recall period of one week) the highest mean consumption. This ranking was independent of whether the maximum number of drinks indicated on the open-ended question of the retrospective diary and the generic QF were capped or not. The corresponding values for comparable capped measures would be 7.92 drinks per week with the generic QF, 9.59 drinks per week with the extended QF (1.48 per week on week days and 8.11 per week on weekends, not presented in Table 1), and 10.02 drinks per week with the retrospective diary.

Open-ended quantity questions, particularly when capped, mainly had an impact on the last 5% of the drinking distribution. Thus, although closed-ended questions affected the overall drinking mean by about 6% (generic) to 10% (retrospective diary), this did not affect the distribution of drinkers or their rank order within an instrument generally, but only at the highest level. It should also be noted that the retrospective diary yielded not only the heaviest

consumption among the heavier drinkers, but also the lowest consumption in the first 25% of drinkers, which included many “abstainers” (drinkers who did not use alcohol in the past week).

Table 1 about here

As can be expected for instruments designed to measure the same parameters, and a sample size of over 5,000 individuals, all correlations among the different instruments were significant (Table 2). Although significant, the convergent reliability (correlation of one instrument with another) was rather weak. For example, the correlation between the original measures of the retrospective diary and the generic measure was 0.421, meaning that less than 20% of the variance of one instrument was explained by the other. The correlation increased to  $r(5072) = 0.589$  ( $p < 0.001$ ) when measures were capped and natural logarithms were computed. This means that the correlations were influenced by high consumption values, since capping and logarithmic transformations both reduce the impact of high values. However, the correlations between the retrospective diary and the other two measures were generally weaker, suggesting that other influences are affecting these associations. The highest correlation was found between the generic and extended QF ( $r(5072) = 0.837$ ,  $p < 0.001$ ) with the capped and log-transformed generic QF.

Insert Table 2 about here

When looking at the concordance between the instruments and focusing on the 25% of heaviest drinkers identified by any of the three instruments, only one third were consistently identified by all three instruments (see Figure 1). Each of the instruments identified approximately two thirds of all top 25% drinkers (generic QF: 62%; extended QF: 65%; retrospective diary: 67%). The highest proportion identified by only a single instrument was that

of the retrospective diary (19%), which was about twice as high as for the generic or the extended QF (10% each).

Please insert Figure 1 about here

Zinb-models for consequences with and without causal attribution of alcohol, DSM-IV dependence criteria, and DSM-5 alcohol use disorder criteria consistently revealed the lowest association with the retrospective diary (Table 3). In terms of overlapping confidence intervals, both the generic and the extended QF performed significantly better than the retrospective diary. Although the extended QF consistently yielded higher associations than the generic QF, these differences were not statistically significant. When adjusting for RSOD the order of effects remained the same. The retrospective diary had significant lower effects compared with the generic and the extended QF with one exception, namely for consequences without causal attribution where confidence intervals were overlapping between the retrospective diary method and the generic QF.

Insert Table 3 about here

## **Discussion**

To our knowledge, this is one of the first studies looking at various instruments that measure volume of drinking among a population of adolescents and young adults. It shows that the choice of an instrument depends on the aim of the study. If only different aggregated usage levels within subpopulations are of interest, then short recall measures are useful since they often



yield higher consumption than measures with longer reference periods. If the instruments are used to establish associations with consequences or other behaviors, then short recall periods may not be the best choice.

The present study confirmed the general literature finding that instruments asking more alcohol use questions yield higher volumes compared with instruments using fewer questions (Alanko, 1984; Bloomfield et al., 2013; Dawson, 2003; Del Boca and Darkes, 2003; Ekholm et al., 2008; Feunekes et al., 1999; Gmel and Rehm, 2004; Greenfield and Kerr, 2008; Rehm, 1998; Room, 1990). The present study adds to this literature using a sample of young men with highly variable drinking pattern. In addition, instruments using shorter recall periods may reduce recall errors and yield higher drinking volumes compared with instruments using longer periods, which was also confirmed in the present study of young men. The present retrospective diary with a short recall period of one week and beverage-specific questions yielded the highest volume, followed by the extended QF with alcohol use assessed separately for weekend and weekdays, followed by the generic QF consisting of only two questions. When measures were made comparable by capping the generic QF and the retrospective diary at a maximum of 13 drinks per drinking day, the generic QF yielded less than 8 drinks per week, the extended QF about 9.5 drinks and the retrospective diary 10 drinks, on average. When looking at associations with consequences the extended QF always yielded the highest association, but was not significantly different from the generic QF, while both were significantly different from the retrospective diary having the lowest associations. The extended QF is a good compromise between measuring usage levels and establishing associations with outcomes, since usage levels were similar to the retrospective diary, but the associations were stronger.

The reason for the rather poor associations of the retrospective diary with consequences may lie in one of its advantages, namely the short recall period. Drinking among young individuals is typically highly variable and a recall period of one week may not always reflect usual drinking patterns (Dawson and Room, 2000; Del Boca and Darkes, 2003). On the one hand there were many past year drinkers with zero consumption in the corresponding week when assessed by the retrospective diary. On the other hand, the retrospective diary also found the largest proportion of heavy drinkers that were not heavy drinkers according to the other two instruments. The latter may reflect a particularly heavy week of drinking which is not one's usual pattern. Thus, regarding the link with consequences, the retrospective diary "misclassifies" the usual patterns of drinkers through their consumption in the past week only, which biases associations with consequences toward the Null. A common recommendation for measuring consequences is to use a longer recall period (Dawson and Room, 2000). The retrospective diary method may not necessarily preserve the rank order of usual low to heavy drinking. This is inferred from the correlation of the retrospective diary with the other two instruments which was as low as 0.421 (i.e. less than 20% common variance) for the uncapped correlation between the generic QF and the retrospective diary. Part of the reason for low correlations may be the presence of high consumption values for uncapped measures that includes outliers. However, this does not explain why the retrospective diary generally resulted in lower correlations with the other two instruments, which were also lower when the open-ended question for quantities in the retrospective diary and the generic QF were taken into account by capping.

The generic QF and the extended QF with log-capped values for the generic QF resulted in a correlation of  $r = 0.837$ . As a side note, when association with outcomes are the focus, extreme values at the upper end may downbias associations, thus capping and logarithmic

transformations may be useful. As shown in table 2, uncapped measures affect only the very high end of the distribution and capping will not grossly affect the overall distribution and rank order of drinkers.

The present study did not address other measures such as beverage-specific QF or the Graduated-Frequency approach. Additional research would be needed to investigate those issues. The sample consists of young men only; it should be extended to include women. However, we think the data do show that a more detailed measurement is preferable, and that the use of short-recall measures for assessing associations with outcomes is questionable in populations with highly variable drinking. There is high variability in our sample, as demonstrated by the fact that (on average) young men consumed only one-and-a-half drinks per week during four workdays, but have more than eight drinks per week on weekends.

C-SURF also measures risky single occasion drinking (RSOD), or the frequency of drinking 6+ drinks on an occasion, but we did not use this measure in the present study directly because it cannot be compared with the other instruments since the scale is different. The coefficients in zlnb-models, as well as the variable means, would be different and not comparable to volume measures. RSOD measures occasions of heavy use and is therefore a frequency measure, whereas our volume measures assess number of drinks per week. There is little doubt, particularly in Switzerland (Gmel et al., 2008), that RSOD is one dimension of heavy alcohol use that needs to be measured in surveys, especially among younger individuals where RSOD is practically the norm. The present study focuses on "volume" of drinking that is a dimension that also needs to be included in on-going research. In addition, young men often exhibit a pattern of only drinking heavily on each drinking occasion, and therefore volume of drinking may strongly reflect RSOD. We included, however, RSOD as control variable in our

models. When adjusting for RSOD, the effect sizes of volume of drinking went down, supporting the assumption that volume of drinking in this age group is often made up by RSOD. The decrease in effect size, however, was found for all volume measures and therefore the rank order of different volume measures as regards associations with consequences was preserved.

A caveat of the present study was that we could not look at the association with consequences measured with a shorter reporting interval, i.e. matching the reporting period of alcohol use and consequences with respect to the retrospective diary. A retrospective diary may show the highest association when consequences were also measured with the same reference period. We would hypothesize, however, that a stronger association may then mainly be due to negative matches, i.e. non-drinkers having no consequences. However, such a speculative statement would certainly need empirical validation. We also believe that many rare consequences would be missed, and there is – to our knowledge – no recommendation for measuring alcohol-related consequences in a short time frame.

A recommendation from the present study for measuring drinking volume in epidemiological research to study associations with consequences is not to use retrospective diaries with short recall periods and to use extended QF measures that can distinguish between weekends (Friday-Sunday) and workdays (Monday-Thursday), in order to capture the consumption patterns of young drinkers. This is particularly true if consequences are measured with a longer time frame to capture rare alcohol related consequences (Dawson & Room, 2000). If space is limited in the questionnaire and the main aim is to measure associations with outcomes (consequences), then a generic QF should be sufficient. Of course, if the aim is to measure drinking levels or the composition of drinking levels by various beverages for aggregated subgroups of the population, diaries may be more precise.

## References

- Alanko, T. (1984). Overview of techniques and problems in the measurement of alcohol consumption. *Research Advances in Alcohol and Drug Problems*, 8, 209-226.
- American Psychiatric Association (APA). (1994). *Diagnostic and Statistical Manual of Mental Disorders: Diagnostic Criteria From DSM-IV*. Washington, DC: APA.
- Assanangkornchai, S., Mukthong, A., & Intanont, T. (2009). Prevalence and patterns of alcohol consumption and health-risk behaviors among high school students in Thailand. *Alcoholism, Clinical and Experimental Research*, 33, 2037-2046.
- Bloomfield, K., Hope, A., & Kraus, L. (2013). Alcohol survey measures for Europe: A literature review. *Drugs: Education, Prevention and Policy*, 20, 348-360.
- Bucholz, K. K., Cadoret, R., Cloninger, C. R., Dinwiddie, S. H., Hesselbrock, V. M., Nurnberger, J. I., Jr., . . . Schuckit, M. A. (1994). A new, semi-structured psychiatric interview for use in genetic linkage studies: A report on the reliability of the SSAGA. *Journal of Studies on Alcohol*, 55, 149-158.
- Casswell, S., Huckle, T., & Pledger, M. (2002). Survey data need not underestimate alcohol consumption. *Alcoholism: Clinical and Experimental Research*, 26, 1561-1567.
- Dawson, D. A. (1998). Measuring alcohol consumption: Limitations and prospects for improvement. *Addiction*, 93, 965-968; discussion 969-977.
- Dawson, D. A. (2000). Alcohol consumption, alcohol dependence, and all-cause mortality. *Alcoholism: Clinical and Experimental Research*, 24, 72-81.
- Dawson, D. A. (2003). Methodological issues in measuring alcohol use. *Alcohol Research and Health*, 27, 18-29.

- Dawson, D. A., & Room, R. (2000). Towards agreement on ways to measure and report drinking patterns and alcohol-related problems in adult general population surveys: The Skarpö conference overview. *Journal of Substance Abuse, 12*, 1-21.
- 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., & Noll, J. A. (2000). Truth or consequences: the validity of self-report data in health services research on addictions. *Addiction, 95*, S347-S360.
- Delcher, C., Johnson, R., & Maldonado-Molina, M. M. (2013). Driving after drinking among young adults of different race/ethnicities in the United States: Unique risk factors in early adolescence? *Journal of Adolescent Health, 52*, 584-591.
- Diep, P. B., Knibbe, R. A., Giang, K. B., & De Vries, N. (2013). Alcohol-related harm among university students in Hanoi, Vietnam. *Global Health Action, 6*, 1-10.
- Ekholm, O. (2004). Influence of the recall period on self-reported alcohol intake. *European Journal of Clinical Nutrition, 58*, 60-63.
- Ekholm, O., Strandberg-Larsen, K., Christensen, K., & Gronbaek, M. (2008). Comparison of assessment methods for self-reported alcohol consumption in health interview surveys. *European Journal of Clinical Nutrition, 62*, 286-291.
- 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.
- Gmel, G., & Daeppen, J. B. (2007). Recall bias for seven-day recall measurement of alcohol consumption among emergency department patients: Implications for case-crossover designs. *Journal of Studies on Alcohol and Drugs, 68*, 303-310.

- Gmel, G., Gaume, J., Faouzi, M., Kulling, J.-P., & Daeppen, J.-B. (2008). Who drinks most of the total alcohol in young men - risky single occasion drinking as normative behaviour. *Alcohol and Alcoholism, 43*, 692-697.
- Gmel, G., Graham, K., Kuendig, H., & Kuntsche, S. (2006). Measuring alcohol consumption - should the "graduated frequency" approach become the norm in survey research? *Addiction, 101*, 16-30.
- Gmel, G., Kuntsche, E., Wicki, M., & Labhart, F. (2010). Measuring alcohol-related consequences in school surveys: Alcohol-attributable consequences or consequences with students' alcohol attribution. *American Journal of Epidemiology, 171*, 93-104.
- Gmel, G., & Rehm, J. (2004). Measuring alcohol consumption. *Contemporary Drug Problems, 31*, 467-540.
- Graham, K., Demers, A., Rehm, J., & Gmel, G. (2004). Problems with the graduated frequency approach to measuring alcohol consumption: Results from a pilot study in Toronto, Canada. *Alcohol and Alcoholism, 39*, 455-462.
- Greenfield, T. K., & Kerr, W. C. (2008). Alcohol measurement methodology in epidemiology: Recent advances and opportunities. *Addiction, 103*, 1082-1099.
- Greenfield, T. K., Kerr, W. C., Bond, J., Ye, Y., & Stockwell, T. (2009). Graduated Frequencies alcohol measures for monitoring consumption patterns: Results from an Australian national survey and a US diary validity study. *Contemporary Drug Problems, 36*.
- Hesselbrock, M., Easton, C., Bucholz, K. K., Schuckit, M., & Hesselbrock, V. (1999). A validity study of the SSAGA--a comparison with the SCAN. *Addiction, 94*, 1361-1370.

- Hibell, B., Guttermsson, U., Ahlström, S., Balakireva, O., Bjanason, T., Kokkevi, A., & Kraus, L. (2012). The 2011 ESPAD report - Substance use among students in 36 European countries. Stockholm: ESPAD.
- Hingson, R. W., Zha, W., & Weitzman, E. R. (2009). Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18-24, 1998-2005. *Journal of Studies on Alcohol and Drugs. Supplement*, 12-20.
- Keogh, R. H., Park, J. Y., White, I. R., Lentjes, M. A., McTaggart, A., Bhaniani, A., . . . Khaw, K. T. (2012). Estimating the alcohol-breast cancer association: A comparison of diet diaries, FFQs and combined measurements. *European Journal of Epidemiology*, 27, 547-559.
- Kerridge, B. T., Saha, T. D., Gmel, G., & Rehm, J. (2013). Taxometric analysis of DSM-IV and DSM-5 alcohol use disorders. *Drug and Alcohol Dependence*, 129, 60-69.
- Knibbe, R. A., & Bloomfield, K. (2001). Alcohol consumption estimates in survey in Europe: Comparability and sensitivity for gender differences. *Substance Abuse*, 22, 23-38.
- Knight, J. R., Wechsler, H., Kuo, M., Seibring, M., Weitzman, E. R., & Schuckit, M. A. (2002). Alcohol abuse and dependence among U.S. college students. *Journal of Studies on Alcohol*, 63, 263-270.
- Kühlhorn, E., & Leifman, H. (1993). Alcohol surveys with high and low coverage rate: A comparative analysis of survey strategies in the alcohol field. *Journal of Studies on Alcohol*, 54, 542-554.
- Labhart, F., Graham, K., Wells, S., & Kuntsche, E. (2013). Drinking before going to licensed premises: An event-level analysis of predrinking, alcohol consumption, and adverse outcomes. *Alcoholism, Clinical and Experimental Research*, 37, 284-291.



- Leigh, B. C. (2000). Using daily reports to measure drinking and drinking patterns. *Journal of Substance Abuse, 12*, 51-65.
- Lemmens, P. H., Tan, E. S., & Knibbe, R. A. (1992). Measuring quantity and frequency of drinking in a general population survey: A comparison of five indices. *Journal of Studies on Alcohol, 53*, 476-486.
- Midanik, L. T. (1982). The validity of self-reported alcohol consumption and alcohol problems: A literature review. *British Journal of Addiction, 77*, 357-382.
- Midanik, L. T., Ye, Y., Greenfield, T. K., & Kerr, W. (2013). Missed and inconsistent classification of current drinkers: Results from the 2005 US National Alcohol Survey. *Addiction, 108*, 348-355.
- Poikolainen, K., Podkletnova, I., & Alho, H. (2002). Accuracy of quantity-frequency and graduated frequency questionnaires in measuring alcohol intake: Comparison with daily diary and commonly used laboratory markers. *Alcohol and Alcoholism, 37*, 573-576.
- Rehm, J. (1998). Measuring quantity, frequency, and volume of drinking. *Alcoholism: Clinical and Experimental Research, 22*, 4S-14S.
- Rehm, J., Taylor, B., & Room, R. (2006). Global burden of disease from alcohol, illicit drugs and tobacco. *Drug and Alcohol Review, 25*, 503-513.
- Room, R. (1990). Measuring alcohol consumption in the United States: Methods and rationales. In L. T. Kozlowski, H. M. Annis, H. D. Cappell, F. B. Glaser, M. S. Goodstadt, Y. Israel, H. Kalant, E. M. Seller & E. R. Vingilis (Eds.), *Research Advances in Alcohol and Drug Problems* (Vol. 10, pp. 39-80). New York, NY: Plenum Press.

- Single, E., & Wortley, S. (1993). Drinking in various settings as it relates to demographic variables and levels of consumption: Findings from a national survey in Canada. *Journal of Studies on Alcohol, 54*, 590-599.
- 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.
- 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.
- 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.
- Wechsler, H., Davenport, A., Dowdall, G. W., Moeykens, B., & Castillio, S. (1994). Health and behavioural consequences of binge drinking in college - a national survey of students at 140 campuses. *JAMA, 272*, 1671-1677.
- Wicki, M., Kuntsche, E., & Gmel, G. (2010). Drinking at European universities? A review of students' alcohol use. *Addictive Behaviors, 35*, 913-924.
- Willett, W. C. (1998). *Nutrition Epidemiology* (2nd ed.). Oxford: Oxford University Press.

*Table 1: Statistics of the distribution of weekly alcohol use among past year drinkers (n = 5074)*

	Percentiles							Overall	
	10	25	50	75	90	95	97.5	Mean	SE
extended QF	0.63	2.02	6.00	12.50	23.50	29.50	39.17	9.59	0.15
generic QF	0.69	2.00	5.00	10.00	20.00	28.00	36.00	8.47	0.19
capped generic QF	0.69	2.00	5.00	10.00	20.00	26.00	30.25	7.92	0.13
past week diary	0.00	1.00	7.00	15.00	27.00	37.00	49.00	11.09	0.21
capped past week diary	0.00	1.00	7.00	15.00	25.00	31.25	41.00	10.02	0.16

**Remarks:** Extended QF asks Quantity/Frequency questions separately for workdays (Monday to Thursday) and weekend days (Friday to Sunday); past week diary is a retrospective diary asking consumption over the past 7-days

*Table 2: Correlations between uncapped, capped and log-capped measures*

	<b>extended QF</b>	<b>generic QF</b>	<b>past week diary</b>
extended QF	-	0.749 (.837)	0.608 (.622)
generic QF	0.607	-	0.539 (.589)
past week diary	0.585	0.421	-

**Remarks:** Below diagonal: uncapped diary and uncapped generic QF; above diagonal: capped (log-capped) diary and generic QF; all correlations are significant at  $p < 0.001$ ; past week diary is a retrospective diary asking consumption over the past 7-days

*Table 3: Associations of volume measures with consequences, DSM-IV dependence and DSM-5 alcohol use disorder, unadjusted and adjusted for frequency of RSOD*

		<b>Coeff.</b>	<b>SE</b>	<b>CI. 95%</b>	
				upper	lower
<b>Unadjusted for risky single occasion drinking</b>					
consequences without causal attribution (0-16)					
	generic QF	0.150	0.012	0.126	0.174
	past week diary	0.098	0.009	0.080	0.116
	extended QF	0.154	0.011	0.131	0.176
causally attributed consequences (0-10)					
	generic QF	0.320	0.020	0.281	0.360
	past week diary	0.176	0.015	0.147	0.205
	extended QF	0.373	0.019	0.335	0.410
DSM-IV dependence (0-7)					
	generic QF	0.354	0.026	0.303	0.405
	past week diary	0.172	0.020	0.134	0.211
	extended QF	0.387	0.025	0.337	0.436
DSM-5 alcohol use disorder(0-11)					
	generic QF	0.415	0.023	0.371	0.459
	past week diary	0.223	0.017	0.190	0.256
	extended QF	0.449	0.022	0.407	0.492
<b>Adjusted for risky single occasion drinking</b>					
consequences without causal attribution (0-16)					
	generic QF	0.079	0.018	0.045	0.114
	past week diary	0.046	0.011	0.024	0.067
	extended QF	0.107	0.017	0.073	0.141
causally attributed consequences (0-10)					
	generic QF	0.142	0.026	0.091	0.193
	past week diary	0.051	0.015	0.020	0.081
	extended QF	0.256	0.026	0.206	0.306
DSM-IV dependence (0-7)					
	generic QF	0.272	0.034	0.206	0.338
	past week diary	0.060	0.021	0.019	0.100
	extended QF	0.340	0.033	0.275	0.404
DSM-5 alcohol use disorder(0-11)					
	generic QF	0.265	0.028	0.210	0.321
	past week diary	0.092	0.017	0.058	0.126
	extended QF	0.344	0.028	0.289	0.399

Remark: RSOD = risky single occasion drinking defined as 5+ drinks on an occasion; past week diary is a retrospective diary asking consumption over the past 7-days

Figure Caption

*Figure 1: Proportional Venn diagram of the 25% heaviest drinkers with each of the three volume measures*

Figure 1

