Published in final edited form as:

Contemp Drug Probl. 2009 October ; 36(3-4): .

Graduated Frequencies alcohol measures for monitoring consumption patterns: Results from an Australian national survey and a US diary validity study

Thomas K. Greenfield^{1,2}, William C. Kerr^{1,2}, Jason Bond¹, Yu Ye¹, and Tim Stockwell³

¹Alcohol Research Group, Public Health Institute, Emeryville, CA

²University of California San Francisco

³Centre for Addictions Research of BC, University of Victoria, BC, Canada

Abstract

We investigate several types of graduated frequency (GF) instruments for monitoring drinking patterns. Two studies with 12-month GF measures and daily data were used: (i) the Australian 2004 National Drug Strategy Household Survey ($n = 24,109$ aged 12+; 22,546 with GF and over 8000 with yesterday data) and (ii) a US methodological study involving a 28-day daily diary plus GF summary measures drawn from the National Alcohol Survey ($n = 3,025$ screened, 119 eligible study completers). The NDSHS involved (i) "drop and collect" self-completed forms with random sampling methods; the Measurement study (ii) screened 3+ drinkers by telephone and collected 28-day drinking diaries and pre- and post-diary 28-day GFs. We compared mean values for the GF quantity ranges from yesterday's drinks (study i) and 28-day diaries (study ii), also examining volume influence. Using Yesterday's drinking, Australian results showed GF quantity range means close to arithmetic midpoints and volume effects only for the lowest two levels (1–2, and 3–4 drinks; $p < .001$). U.S. calibration results on the GF using 28-day diaries were similar, with a volume effect only at these low quantity levels ($p < .001$). Means for the highest quantity thresholds were 23.5 drinks for the 20+ (10 gram) drink level (Australia) and 15.5 drinks for the 12+ (14 g) drink level (US). In the US study, summary GF frequency and volume were highly consistent with diary-based counterparts. A conclusion is that algorithms for computing volume may be refined using validation data. We suggest measurement methods may be improved by taking better account of empirical drink ethanol content.

Corresponding author: Tom K. Greenfield, PhD, Alcohol Research Group, Public Health Institute, 6475 Christie Ave., Ste 400, Emeryville, CA 94608, USA, tgreenfield@arg.org.

Notes on Contributors:

THOMAS K. GREENFIELD is Center Director and Senior Scientist at the Alcohol Research Group, Public Health Institute (PHI), Emeryville, California, USA and on the faculty of the Clinical Services Research Training Program in the Department of Psychiatry, University of California San Francisco (UCSF). A clinical psychologist, his grants and research interest include self-report alcohol measurement, national and comparative surveys, alcohol and mortality, and health services research and alcohol policy. **WILLIAM C. KERR** is an economist and senior scientist at the Alcohol Research Group, PHI with an adjunct position at UCSF. Interests and grants include methodological studies on drink ethanol content, time series analyses of alcohol consumption and mortality, analyses of national survey data, and alcohol policy. **JASON BOND** is a senior biostatistician at the Alcohol Research Group, PHI with interests in multivariate methods, measurement, longitudinal analysis and analyzing randomized and quasi-experimental health services data. **YU YE** is a biostatistician at the Alcohol Research Group, PHI with a background in economics and interests in cross-national studies, alcohol measurement, national alcohol survey and time series analysis, and emergency department studies. **TIM STOCKWELL** directs the Centre for Addictions Research of British Columbia and is professor in the Department of Psychology, University of Victoria, where he also co-leads the BC Mental Health and Addictions Research Network. A clinical psychologist he formerly directed Australia's National Drug Research Institute in Perth and has interests in numerous topics including monitoring drug and alcohol use and problems and alcohol measurement.

Introduction

For assessing volume and heavy drinking, graduated frequency (GF) measures (Greenfield 2000, Rehm et al. 1999) serve well in countries where “drinks” have some meaning. Validity studies suggest good correspondence with drinking diaries in special populations (Hilton 1989). With some inherent limitations (Midanik, 1994, 1998), such alcohol measures show reliable problem relationships (Greenfield 1998, Greenfield et al. 2006) suggesting they at least order individuals on intake fairly effectively. Hazardous drinking patterns identified by GF measures (Rehm et al. 1996, Rogers & Greenfield 1999) reveal plausible risk relationships (Greenfield et al. 2004, Greenfield & Rogers 1999). However, methodological studies of alcohol intake measures like the GF are still needed (Midanik & Greenfield 2004; Midanik, 1998). Because of the combined-beverage GF measure’s complexity, some see it as prone to errors (Gmel et al. 2006, Graham et al. 2004), in part due to inadequate training and implementation (Greenfield & Kerr, 2008). There remain questions about accuracy of the information derived from the GF at each of the quantity levels, and how best to combine these data in volume algorithms. The present paper aims to provide new information relevant to these issues.

Retrospective Recall and Reference Period

A drinking pattern with the same amount on every occasion is very rare (Midanik et al. 1999) but in principle easier to recall than a pattern with varying amounts. Cognitive studies about how people answer drinking questions (Greenfield 2000, Midanik 2003, Midanik et al. 1999) show the importance of providing a reference period like 12 months or 30 days; absent these, respondents assumed periods ranging from a week to several years for “recent” drinking (Greenfield 2000, Midanik & Greenfield 2003). The US “Improving Self-Report Alcohol Measurement” validity study used here involved 28-day drinking diaries (completed daily), comparing entries to pre- and post GF summary measures covering the same period. (A 28-day period includes 4 weekly drinking cycles while 30-days can have 5 weekends; although untested, 30-day and 28-day measures likely perform similarly.) Even if recall might be more accurate for periods like the prior week, short-duration measures are very prone to happenstance variability, seasonality, omission of occasional heavy episodes, and so may not yield “average” drinking (Greenfield & Kerr, 2008), resulting in poor estimates of harmful and hazardous drinking compared to 12-month GF measures (Rehm et al. 1999). Nonetheless, reports of drinking on the previous day (a still shorter period) address specific recent events and so may be particularly well recalled (Stockwell et al, 2004). The Australian NDSHS asked in detail questions about beverages (and sizes) drunk yesterday (Stockwell et al., 2008) and also included a 12-month GF measure.

This study therefore uses two types of daily data: yesterday’s drinks in the Australian case and prospective drinking diaries in the US one, comparing each to GF measures in a within-subjects design. Plausibly, reporting drinks consumed yesterday, or keeping a daily drinking diary (Greenfield & Kerr, 2008) can yield detailed information likely to be based on actual recall of particular drinking events and so be quite accurate (Stockwell et al. 2004). As Stockwell et al (2008) showed, yesterday reports are less downwardly biased and yield higher coverage of sales-based consumption. Thus, we take yesterday and daily diary measures as a “silver standard” (not quite a “gold standard”) to serve as criterion measures. Diaries have been validated with other health behaviors (Leigh et al. 1992, Verbrugge 1980) but previously, for alcohol use (Hilton 1989, Leigh et al. 1998), the samples have not been representative of a broad population. Both yesterday and diary measures used here, gather detailed information on the prior day’s beverages: type, brand and serving/container sizes (Stockwell et al. 2004, 2008), to assess (typical) drink ethanol (see Greenfield & Kerr, 2008, for more details). The diaries used in the present US validity study also requested drinking context and start/stop times for up to four occasions in each day, in effect using a very short

duration Time Line Follow Back approach (Sobell & Sobell 1992), expected to improve recall still further (Greenfield et al. 2007).

Challenges for the Graduated Frequencies Approach

GF measures often rely on categories, both of the stated quantity ranges (e.g., 12+ drinks, 8–11 drinks, 5–7 drinks, 3–4 drinks, and 1–2 drinks, for the US NAS) and for the frequency of drinking each amount. Usually the arithmetic mean of the quantity range is used for volume calculations, e.g., 3.5 for 3–4 drinks. Choosing a value for the highest number-of-drinks listed is usually arbitrary (e.g., 13 for 12+ drinks). Empirically based estimates of the means and top, unbounded category would be better. The GF frequencies are sometimes number of days but usually categories like “every day or nearly every day”, “3–4 times a week”, etc. down to “Never”. Arithmetic mid-points of the implied day ranges are often used for volume calculations. These values may not well reflect the drinking distribution. When cumulative frequency exceeds 365 days, volume calculations may need capping (Greenfield 2000) or prorating (Gmel et al. 2006) of the summed quantity $\times$ frequency GF elements. Downward capping (Greenfield 2000) counts only the quantities until 365 total days is reached. In prorating; the volume yielded by the full summation of QF products is multiplied by (365/total N days) (Gmel et al. 2006). Each approach has its rationale and conceptual merits but neither has been empirically tested yet.

Aims of the Paper

Although GF methods have been recommended for use in international alcohol monitoring surveys (World Health Organization 2000), choice of optimal algorithms for calculating volume (or frequency of hazardous drinking) has not been evidence based. Comparing GF item responses to Yesterday or 28-day Diary data can help address some of the issues raised in recent critiques of the GF approach, e.g., for both developed countries like Canada (Graham et al. 2004) or multiple countries in cross-national studies (Gmel et al. 2006). We consider specifically what the data can tell us about: (a) for volume calculations, best values to use for quantity ranges (e.g., arithmetic midpoint, arbitrary top value, or empirically derived means); (b) whether the mean values differ between groups with lower vs. higher volume (since the underlying drinking distribution might alter the distributions in specific quantity ranges); (c) how best to handle frequencies of drinking at each specific GF quantity when the individual's cumulative frequency exceeds 365 days (12-month summary GF); last, (d) whether empirical estimates of an individual's preferences for specific beverages, brands and container/pour sizes can be used to improve GF measures.

Methods

Data Sources

Two main data sources are used, both providing versions of the 12-month GF and detailed data on daily drinking: (i) the Australian 2004 National Drug Strategy Household Survey ($n = 24,109$ aged 12+; 22,546 with drinkers completing the GF measure); and (ii) early results from ARG's “Improving Self-Report Alcohol Consumption Measurement” study involving daily diary measures, an initial 12-month GF, and two 28-day GF measures completed before and after filling out the diaries for 28 consecutive days ($n=3,025$ screened, 147 eligible study completers, with 119 who completed at least 26 days of diaries, analyzed here).

The Australian 2004 NDSHS

This is a large national survey conducted by the Australian Institute of Health and Welfare (AIHW) from June to November 2004. The sample was based on households selected by a

multi-stage, stratified-area random sample design including Australia's six states and two territories. In all 29,245 Australians aged 12 years and older participated, responding to alcohol questions as well as drug knowledge, attitudes and consumption (Australian Institute of Health and Welfare 2005). Only the randomized "drop and collect" household survey ($n=24,109$, response rate 46.7%) asked about alcohol, so remaining surveys completed by CATI methods are omitted. Respondents were household members aged at least 12 with the chronologically next birthday. Children aged 12–15 completed the survey with consent of parent or guardian, with those 12–13 administered a shorter questionnaire. For further details see Stockwell et al. (2008). Present analyses use 22,546 individuals with complete 12-month GF information. Mean age for this analysis group was 44.4 (SD 17.7; 5.2% 12–17, 9.7% 70+); 46.2% were males, 59.7% were married or cohabiting (24.3% never married) and 1.3% were aboriginal/Torres Strait Islanders. Some analyses are limited to days involving at least 1 full 10 g drink ($n=22,188$).

The Improving Self-Report Alcohol Measurement Study

This involved an experimental design with three groups recruited from an RDD CATI survey of the San Francisco area. Following screening for weekly drinking and drinking at least 3 drinks in a day sometime in the prior year using the 12-month combined-beverage GF included in the 2005 NAS (see below); eligibility also involved ability and willingness to undertake study tasks. Eligible volunteers were randomized into one of three groups. Group 1, 2 and 3 completed by telephone pre- and a post-retrospective summary measures—the 28-day GF. In addition to these surveys, Group 2 completed in-home drinking diaries each day, collected or mailed weekly for 4 consecutive weeks (28 days) giving detailed information about up to 4 drinking occasions/day (see Measures). In addition to the pre- and post-surveys given all groups, Group 3 completed the same diaries as Group 2 but also wore on their wrist a physiological measuring device, a non-invasive Wris-TAS-V® Trans-dermal alcohol sensor for 2 of the 4 weeks. For this study we use preliminary data only from Groups 2 and 3 ($n=147$, derived from 3,025 individuals originally screened) who completed diaries and both 28-day and 12-month GF summaries. Most analyses include only those providing at least 26 days of complete drinking diaries ($n=119$).

Measures

Graduated Frequencies

GF measures have been detailed (Greenfield 2000, Room 1990; Dawson & Room; World Health Organization, 2000) and so will be summarized as used here.

2004 NDSHS GF

This involved a self-completed form beginning with the instruction "Please record how often in the last 12 months you have had each of the following number of standard drinks in a day (Mark one response for each row below)". This NDSHS GF involves 8 rows of boxes with headings on the left indicating amounts in the descending ranges: "20 or more standard drinks in a day", followed by 11–19, 7–10, 5–6, 3–4, 1–2 (each level repeating "...standard drinks in a day"), ending with "Less than 1", and "None". For each row there were 8 frequency boxes (which respondents checked), with columns headed, left to right, "Every day" followed by 5–6, 3–4 and 1–2 times per week, "2–3 days a month, About 1 day a month, Less often, and Never". Numbers of days assigned to these respective frequencies (approximate midpoints) were: 365, 286, 182, 78, 30, 12, 6 and 0. For standard Volume calculations the values used for quantities were the arithmetic mid-points of the drink-range categories, in standard drinks (10 g. in Australia). A detailed standard drink guide preceded the GF matrix. The top level (20+) was assigned 23.5 drinks based on inspection of the mean of Yesterday results in this range ($n=64$; see Table 1, Total column, last row). The

quantity range arithmetic midpoints used in the Volume calculation are given in Table 1 under the “GF Quantity Range” column (standard drinks). The Volume calculation summed the multiplicative $Q \times F$ products in standard drinks and days, respectively, yielding 12-month total standard drinks.

Both a Capped- and a Prorated-Volume was computed and the respectively adjusted Frequencies involved in downward capping and prorating were also output. In addition, for the NDSHS a Maximum level (the highest Quantity for which the response was not None) was identified (see Table 2 which also provides the percentages and *ns* capped at various levels).

Self-report Measurement GFs

These involve (a) an initial NAS-based 12-month GF (Greenfield 2000), but with separate frequencies for 1 and 2 drinks (see below); (b) equivalent Pre-and (c) Post-28-Day GF measures which ask the number of days (of 28) that each quantity (12+, 8–11, 5–7, 3–4, and 1–2 (not separated) drinks was consumed. All GFs were based on telephone CATI surveys, to best inform the NAS, a survey also administered by telephone.

2005 NAS 12-Month GF

This measure was adapted for use in the Self-Report Measurement Study (as noted above). The initiating question (maximum) begins by asking: “Think of all kinds of alcoholic beverages combined, that is any combination of bottles or cans of beer, glasses of wine, drinks containing liquor of any kind, or coolers, flavored malt beverages or p-made cocktails. In his question, one drink is equal to a 12 ounce bottle of beer or cooler, a four to five ounce glass of wine, or one shot of liquor (1.5 ounces). During the last 12 months, what is the largest number of drinks you had on any single day? Was it...” Categories begin with 24 or more drinks, then 12–23, 8–11, 5–7, 4, 3, 2 drinks and finally 1 drink. (Prior implementations of maximum/day used 3–4 (combined) and 1–2 drink categories (Greenfield et al. 2006)). Drinks, so defined are the US standard, variously taken as equivalent to 12 to 14 g (Kerr et al. 2005, Turner 1990). The response to the maximum question determines which GF quantity-level frequencies will be asked, the highest one being 12+ drinks (not 24+), with lower ones 8–11, 5–7, 3–4, 2 drinks, and 1 drink (previous GF forms combined the last as 1–2 drinks). For each of the quantity ranges a categorical frequency is asked: “Every day or nearly every day” (coded as 360 days), “3–4 times a week, 1–2 times a week, 1–3 times a month, Less than once a month (coded 6), Once in those 12 months and Never in those 12 months (i.e., 7 categories). Using category midpoints (or values noted), the same variables were computed as in the NDSHS, viz, Maximum, Capped volume, number of drinking days, and respective capped or prorated frequencies of the GF ranges.

Daily measures: 2004 NDSHS Yesterday

A simple standard-drink Yesterday question and a more elaborate matrix including items on serving size and beverage strength were both included. After indicating the day of the week, Monday through Sunday, “that is today,” respondents were first asked “How many standard alcoholic drinks did you have yesterday?” with spaces to write the number. An additional statement read: “If less than 1, please indicate to the nearest fraction”: with boxes for $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$ and ‘Never’ provided. This simple standard drinks Yesterday question forms the basis of most analyses here. However, estimates of the drink ethanol used to adjust the GF means in the rightmost column of Table 1 were derived from the much more detailed Yesterday data reported in response to a further question: “How many bottles, glasses, cans or nips of alcohol did you drink yesterday?” followed by a detailed matrix of response options laid out on a full page. For each of 6 major beverage types (Beer, Wine, Pre-Mixed Spirits, Straight

Spirits (not pre-mixed), Alcoholic Cider, and Other, including fortified wines), with subsets totaling to 12 beverage varieties listed downward, a range of descriptions of common glass and container sizes for the beverage variety was provided in a given row. Responses to this series were adjusted with typical serving size and beverage ethanol content, updated from Stockwell et al. (2004) using published industry and other sources; the multiple adjusted entries were summed to yield the day's ethanol intake (Stockwell et al., 2008).

Self-Report Measurement Study 28-Day Diaries

The drinking diary consisted of one page per day (Day Date and ID Number were pre-entered), divided down the page into sections for up to 4 "Drinking Sessions" with data on each entered in three rows with 8 columns. Column headings from left to right were Type of alcohol (W B S); # of Drinks (always record a whole number); Location (H=Home, R=Rest., B=Bar, P=Party, O=Other); Brand (Examples: Bud Lite, Mondavi Chardonnay, Meyers Dark Rum); Drink Size in Ounces ($\frac{1}{2}$, 1, 1 $\frac{1}{2}$ 2 4 6 8 12 16 24 32 40); Finished Drink? (# of dinks not finished, % of drink left); Meal within 1 hr? (S=Snack, M=Meal, N=No). For each session, respondents recorded Time began and Time ended. A 30-minute training session introduced the diary and explained the meaning of the columns (e.g., how to figure out drink sizes, meaning of "session", use of shorthand, etc.). The Letters/numbers given above were pre-printed to allow circling. Data were mailed weekly in pre-paid envelopes, or collected. A significant monetary incentive was associated with completion of diaries and surveys with a bonus for full completion (up to a total \$175 for all tasks). Among those completing diaries and GF measures, half (Group 2) in addition wore the WrisTAS for 2 weeks. TAS data are not used here, but a report on pilot data ($n=33$) indicated that when physiological measures agreed on a diary drinking event, "peak height measurements appeared to correspond with small or large [diary quantities]" (Greenfield et al. 2005).

Results

Considering first the Australian 2004 NDSHS, Table 1 gives results of analyses using the distributions of the Yesterday amounts, in standard drinks for corresponding GF quantity ranges. For example, for the GF 3–4 drink level, the Yesterday quantities selected are either 3 or 4 drinks (only). The mean of Yesterday amounts in this range is given in the column headed "Total Sample". Of those drinking yesterday, 2446 reported either 3 or 4 drinks on the simple Yesterday measure, with the resultant mean of 3.40 drinks (unadjusted for alcohol content) so estimated with some confidence (unweighted data). The three columns of Table 1, grouped under the heading Volume Tertiles, decompose the full drinking sample into those with high, medium and low Volumes (approximate thirds, based on the capped GF volume). For each GF level, the means of the high, medium and low volume subgroups are given and the significance of the difference tested with a one-way ANOVA). The means are ordered as expected a priori. There is an overall difference in the 3–4 means, by volume ($F_{(2, 2443)} = 8.46, p < .001$). Pairwise Scheffe tests indicate that the low and medium thirds differ from the high third but not from one another, but differences are not very great, only about a 0.1 drink difference from the lower two to the high volumes. The results for the 1–2 drink level are stronger overall ($F_{(2, 5067)} = 145.76, p < .001$), with means varying between a low of 1.33 drinks and a high of 1.64 drinks (reflecting a preponderance of 1 drink for lower- and of 2 drinks for higher-volume drinkers). The overall mean (unweighted) was exactly the arithmetic mid point (1.50). At higher GF levels than 3–4, differences are not significant. In some (but not all) cases they are ordered appropriately but analyses suffer because the volume thirds do not well subdivide the groups (e.g., even at 5–6 drinks there are only 26 cases in the low volume group of intermittent heavy drinkers). Importantly, overall means mostly lie very close to the arithmetic midpoints, exactly so for the 1–2 drink range, just below for 3–4, and 7–10 drink ranges, and just above for the 5–6 drink range.

The mean for the 11–19 drink range (on the descending limb of the drinking quantity distribution, is considerably (1.4 drinks) below the arithmetic mean of 15 drinks; conversely, the 20+ drinks (unweighted) mean is 23.5 drinks, an important value since otherwise undefined. For any given individual the value may be imprecise but for the population this value should not be biased, although even in this large sample only a relatively small number reported 20 or more drinks yesterday ($n=64$). Without the yesterday measure (or diaries in the US instance), one could only guess at a value to assign the top category (20+ drinks) in a volume calculation.

The final column in Table 1 provides the weighted estimate, in standard 10g ethanol drink equivalents, of the means for GF levels when the “empirical” estimate is instead based on the beverage type and drink size information in the detailed Yesterday assessment (Stockwell et al. 2008). When adjusted for reported drink ethanol (rightmost “Ethanol Adjusted” column), the levels are all substantially above the total group unadjusted drink means based on the simple Yesterday measure (“Total Sample” second to last column). All but one (for the 11–19 drink level) are well above the arithmetic mean. The highest GF-category mean of 20+ becomes a fully 36.3 drinks vs. 23.5, unweighted (or 24.2 weighted), assuming standard drinks

Table 2 shows the distribution on the maximum level of responses to the GF. Unlike the NAS GF measure, which for efficiency begins with a maximum question, allowing unnecessary GF quantity-levels to be skipped, the paper-and-pencil format used in the Australian NDSHS requires derivation of the maximum based on the pattern of completion of the GF matrix. We see that 11 percent of drinkers indicate 20+ drinks ($n=2452$) as their maximum. It is worth noting that of the 8364 who drank yesterday, in the group also completing the GF, only 64 said they drank 20 or more drinks (around 0.8%). This serves to illustrate forcefully how longer reference periods (such as 12 months) are needed to capture adequately intermittent heavy (especially very heavy) drinking episodes, which are easily missed in the shorter reference period measures, tending to probabilistically reflect more usual amounts.

Table 2 also provides descriptive information on those whose responses exactly account for 365 days because they indicated drinking ‘every day’ at a single (top) level, totaling 2.7% ($n=550$) of drinkers, mostly indicating consuming lower quantities such as 1–2 ($n=305$) and 3–4 drinks ($n=146$). On the 12-month GF, a further 13.2% ($n=2,662$) of the 20,188 drinkers consuming at least a full standard drink provided frequencies such that the implied days’ summation exceeded 365 days. Most of the (downward) capping occurs at lower quantities 1–2 (8.0%, $n=1,605$) and 3–4 drink (3.4%, $n=686$), with smaller numbers at higher levels. We consider later the effects of downward capping (eliminating or truncating excess days at lower amounts, by design) versus prorating the frequencies by adjusting them downward at *all* levels.

Moving to the US Self-Report Measurement data, the 28-day sample of daily drinking overcomes some difficulties of having only a one day sample (yesterday) in the Australian data. Focusing first on the distribution of drinking days at the GF-quantity levels, Table 3 provides the number of days in the two summary instruments, the Pre-28-day GF and the Post-28-day GF measures. Because the CATI system created a flag if 28 days were exceeded, the interviewer could reconcile responses with the respondent, so no cases involved > 28 days of data and no capping or prorating was needed. An open-ended response format is possible and preferred for a 4 week period, as previously used in paper-and-pencil formats (Greenfield 2000).

First, the reported number of days drinking any alcohol on the Diary and the Post-GF were extremely similar, 61.6 and 61.7 percent of days, respectively (see Table 3). The Table shows that the number of 12+ days on the GF (Pre and Post) was inflated compared to the diary data, but this was offset by lower numbers of days on the GF for all other levels excepting 1–2 days, also slightly inflated. Despite these slight variations across the levels, the overall correspondence of sample distributions was extremely high (Pearson's $r = .996$). The similarity of the distributions seen may not be due only to better recall after completing diaries because the relationship is also very high ($r = .987$) between the drinking diary and the Pre-GF (reporting drinking in the prior 28-day period before diary keeping). The Pre- and Post-GFs distributions are also highly similar ($r = 0.992$) suggesting consistency of estimates and little reactivity of measurement due to diary record keeping. Figure 2 displays results in a parallel bar chart.

Such sample-level aggregate distributions should show greater agreement than individual-level comparisons, but on an individual basis there is also good agreement between 28-day Diary and Post-GF data (results not shown). Correlations between an individual's diary and GF summary values were higher for larger drink amounts (5–7 = .73; 8–11 = .80; 12+ = .76) than for smaller ones (1–2 = .65; 3–4 = .55). Overall, drinking frequency and volume agreed closely between diary and the summary measure: both yielded 17.3 drinking days; and volumes of 59.1 (SD = 4.4) vs 60.4 (SD = 6.7) drinks, respectively ($r = 0.86$ and 0.87 , on a within-subjects basis). Sample statistics (volume and frequency of drinking), and sample distributions agree more closely than within-subjects' recall of days drinking at each GF quantity.

Next we consider, for the US data, how arithmetic GF quantity means compare with empirical means, and how volume affects these, using a parallel analysis to that for the Australian yesterday data. Volume was calculated from the 12-month GF administered prior to the 28-day Pre-GF and drinking diaries. Diary data were divided into amounts in a day with the same ranges as the GF quantities (1–2 drinks, 3–4, etc.) and their means derived. Results are in Table 4; with Table 5 providing mean values of the volume tertile subgroups and the overall sample for the 28-day Pre- and Post-GFs and 12-month GF (capped). Table 4 results were very similar to those found for the NDSHS (see Table 1), although the upper levels of the NAS GF quantity ranges (13 g US standard drinks) differ from those used in the Australian survey (10 g standard drinks). Empirical means for each quantity level are slightly below the respective arithmetic means—almost all are .96 to .97 of arithmetic values. The mean for the 12+ group is 15.5 drinks, larger than the value typically used (13 drinks in the standard volume algorithm, shown to be too conservative). The means for the volume tertile groups, as in the Australian case, differ significantly only for the 1–2 and 3–4 drink levels. At other levels the order consistently supports the notion that higher volume individuals' distributions are shifted upward, affecting each empirical range mean slightly (but not significantly). With the exception of the high volume in the 1–2 drink range, resultant means remain below the arithmetic mean. Many people in the US drink small amounts and even in this slightly heavier drinking methodological sample, about 30% of diary days (or 50% drinking days) involved 1–2 drinks. Just under a quarter of these days involved drinking by the top tertile volume group, averaging 1.65 drinks per drinking day when drinking 1 or 2 drinks (the lowest tertile volume group, with 37% of 1–2 drink events, averaged only 1.32 drinks).

Finally, we used the distributions of the daily data from both the Yesterday measure (Australian NDSHS) and the 28-day Diaries (Self-Report Measurement Study), taken as the criteria distribution, in each case, to examine whether adjusting the 12-month GF levels by Downward Capping or Prorating more closely resembled the criterion distribution (data not shown; tables available from the first author). We focus on 12-month current drinkers

drinking a least 1 drink, omitting partial drinks from consideration. In both cases we consider 1–2 drink levels combined (the NDSHS and standard NAS format) combining distinct 1 drink and 2 drink levels in the US 12-month GF. For the NDSHS data, the Capped GF, compared to the Yesterday distribution as the criterion, slightly over-represents the 20+ drinks, the 5–6 and the 1–2 drinks level, is extremely close on the 3–4 drinks level, and is somewhat depleted for the remaining levels (11–19 and 7–10 drinks). Prorating GF data leads to lower levels for the top 5 GF levels (20+, 11–19, 7–10, 5–6 and even 3–4 drinks) and a greatly increased proportion at 1–2 drinks, compared to the criterion (42.5% compared to 30.4% for the criterion). A general indication of the extent to which the two methods of handling excess GF days match the criterion is in the correlation between the respective (matched) levels. The Capped GF's distribution is highly similar overall to the Yesterday criterion ($r = 0.99$) while the corresponding correlation for the Prorated GF distribution is 0.94. Note that partial drinks are not considered here, largely because when given this option is given on the GF, large numbers, particularly very light drinkers, opt for this response. Over 10% ($n = 2,358$ of 23,161) identified amounts less than 1 drink as their maximum on the GF (Table 2) while only 3.6% of GF drinkers with 1 or more drinks who also reported drinking Yesterday ($n = 360$ of 9,980) stated they had fractional amounts (see Table 1). The partial drinks level appears to introduce an unnecessary distortion into the GF distributions, has never been the practice in the US, so this level was omitted for better performance of the Australian GF and better comparability with US data.

As seen in Table 2, 13.2% report frequencies such that total drinking days > 365 requiring recourse to either prorating or capping (21.8% if partial drinks are included, again suggesting that including this level is problematic, perhaps adding an opportunity to minimize one's drinking). For those who drank yesterday whose cumulative GF drinking days totaled $\# 365$, the mean of average quantity on each drinking day is 3.2 drinks (capped and prorated give the same result since no adjustment is required), while yesterday's drinking mean is 3.0 drinks, agreeing closely. Among those who drank yesterday with total GF drinking days > 365 , the mean of the average quantity per drinking day is 3.8 drinks from prorated approach, 4.5 from the downward capped approach and 4.8 from yesterday's drinking ($n=2,260$). Thus, when adjustment is required due to excess GF drinking days, the Capping algorithm more closely reflects the "silver standard" than the Prorating algorithm.

Turning to the algorithm comparison in the Self-Report Measurement study (data not shown; table available from first author), considering current drinkers only, the Capped GF percentage at 12+ drinks (3.1%) exceeds that for Diary 12+ drink days (2.7%) as did the 1–2 drinks level (62.4% vs. 50%, respectively). For intermediate GF levels (8–11, 5–7, and 3–4) the GF under-represents the equivalent Diary percentages, although the 8–11 drink proportion is fairly close (5.8 vs. 6.3%). Prorating comes closer to the criterion on 12+ drink percentages (2.6% vs. 2.7%) but all the other percentages are further from the criterion than the Capped GF percentages. As a result, the Capped GF to Diary correlation is very high ($r = 0.97$) and somewhat higher than that for the Prorated GF ($r = 0.94$). For those requiring adjustment (> 365 days, $n = 36$) the 28-day average was 3.70 drinks/day ($SD = 1.96$) while the 12-Month Capped GF drinks/day was 3.65 ($SD = 1.48$) and the Prorated GF was 2.90 ($SD = 1.87$) drinks. Both GFs are adjusted to 365 days drinking while not all diary days involved drinking (mean drinks per drinking day 4.24). Thus, as with the NDSHS, for the US sample requiring adjustment, the mean volume in drinks per day given by the capping procedure is closer to the diary-based mean volume than the equivalent result based on prorating.

The Self-Report Measurement Study permits an analysis not possible in the case of the NDSHS Yesterday measure since 28 days diary data provides a *distribution* across drinking quantity levels to compare with the 12-Month GF levels on an individual (not only a group)

basis. Recall that only the 12-Month GF required adjustment in 30% of the cases for excess total days (> 365). To conduct this analysis, marginals from cross-tabulations of the 12-Month GF frequency categories (expressed as days drinking at each quantity level) with the Diary days (at equivalent levels) can be individually correlated. A summary measure of fit is taken as the mean correlation of all individuals. Using this method we compared the 12-Month GF distribution of frequencies across quantity levels (either Capped or Prorated) to that found in the Diary data (and to each other). We conducted this analysis using the diary days sample with > 26 valid days ($n = 119$) both for the sample requiring adjustment because they had cumulative total GF days > 365 ($n = 36$) and the full sample, yielding mean correlations for pairs of the three measures (Capped GF, Prorated GF and Diaries). Considering first the 36 cases requiring adjustment, means of individual distributions' correlations for the pairs were as follows: GF-Capped to GF-Prorated = 0.84; GF-Capped to Diary = 0.52; and GF-Prorated to Diary = 0.47. This selected sample focuses on those requiring adjustment (because for the others Capped and Prorated do not occur and so results are identical). However, to get a "population" picture (based on the full sample) of the impacts of capping or prorating, the equivalent mean individually-based distribution correlations are: GF-Capped to GF-Prorated = 0.95; GF-Capped to Diary = 0.70; and GF-Prorated to Diary = 0.69. Therefore, when all cases are included, capping or prorating will not make much difference. However, capping performs somewhat better for the affected cases (no doubt not significantly, although the difference in r s was not tested).

Discussion

The Australian survey's Yesterday and the US methodological study's Diary measures appear to have been valuable for validating and potentially refining the Graduated Frequency (GF) measures also included in these studies. They have helped answer some practical questions regarding optimal methods of computing volume from the GF measures. Perhaps the most important finding borne out in both studies is that use of the arithmetic mid points in standard algorithms is not very far off, involving very little overall bias. More important is the use of these data to estimate the mean drinks value to use for the top GF threshold, which is otherwise a complete guess because responses can range up to the maximum amount which is not typically assessed in a continuous fashion. Even if maximum is measured, this part of the distribution is particularly sensitive to the heavy drinking individuals' actual amounts. The mean value for the highest level of the GF has typically been set too low in volume algorithms, previously 13 drinks for the 12+ drinks level in the NAS implementation, where Table 4 suggests 15.5 drinks would be a better value (although based on samples of 28 days' not 12 months' drinking). While as hypothetically reasonable, given the shapes of the drinking distribution underlying GF responses, the adjustments to quantity means introduced by volume differences, though generally in the right direction, were only significant for the lower (1–2 and 3–4 drink) levels and small in any case. This finding was replicated in both Australian and Self-Report Measurement study data. An attempt to use maximum amount to predict mean quantity level was largely unsuccessful so volume seems to better characterize the "drinking profile".

The impact of reported drink ethanol on the volume calculation is considerably greater than adjustments to GF quantity means from an individual's volume. As noted by Stockwell et al. (2008) for the GF, the estimated coverage for this survey measure (survey-based mean consumption compared to Australian 12-and-older per capita alcohol consumption), was increased from 52.41% (assuming standard drinks) to 69.17% (taking full account of drink ethanol based on the detailed Yesterday estimate)—a 32.0% increase in coverage. The widening gap between the adjusted and unadjusted values at the two highest levels (11–19 and 20+ drinks) suggests that the riskiest drinkers have more ethanol-rich drinks, given greater straight spirits (80%) and regular-strength beer (73%), as opposed to low-alcohol

beer (26%) consumed on risky drinking days, defined as days drinking more than 60 g ethanol or 40 g for women on any single day (Stockwell et al. 2008). The Australian empirical adjustment for yesterday's drink ethanol boosts all levels beyond their arithmetic means and particularly dramatically raises the heaviest drinking level's mean in volume and hazardous drinking algorithms. Heaviest drinkers have been predicted on conceptual grounds to be the likely source for most of the alcohol missed by surveys, or undercoverage (Greenfield & Kerr, 2008). The Australian data seem to confirm findings from a methodological study at ARG in which typical drink ethanol content was assessed in the home using measuring beakers, showing heaviest drinkers tend to have larger drinks (Greenfield & Kerr 2008; Kerr et al, 2005). The Diary data also has detailed brand, drink type and drink size data which will be analyzed later to inform this issue.

Regarding the full-sample similarity between drinkers' frequency distributions on the GF quantities and the equivalently partitioned daily data "silver standard," this seems generally to support the usefulness of GF-style combined-alcohol measures in capturing some genuine underlying pattern of drinking—the "drinking profile". It is clear that on an individual basis agreement is far from perfect, as has also been seen also for different ways of assessing the maximum (Greenfield et al. 2006). The GF and Diary-based individual distributions (days at each quantity level) are moderately correlated on average, at about the 0.70 level, meaning about 50% of the variance is in common between diary and GF-amount drinking profiles. It is good news that in spite of these individual errors, the sample distributions match much better (see Table 3 and Figure 2), meaning that the GF summary measures do assess population drinking patterns quite accurately. Although there may be some biases at particular levels (e.g., higher GF summary frequencies at 12+ and 1–2 drink levels, and lower ones for the intermediate levels, seen in Figure 2), as was found by Hilton's (1989) earlier GF-diary study, in calculating volume the effects largely appear to cancel out. In the case of agreement between diary data and the 28-day Post GF, it could be argued that the task of completing daily diaries makes it more likely that the numbers of days drinking the various amounts will be recalled reasonably accurately. This may well be so, but an indication that this source of agreement is unlikely to completely account for the similarity of the sample distributions seen is that the relationship between the distributions based on the drinking diary and the Pre-GF (which reports drinking during the *previous* 28-day period and completed before diary keeping) is *also* very high ($r = .99$). The Pre- and Post-GFs distributions are extremely similar ($r = 0.99$) indicating overall consistency of estimation over a two month period with little or no measurement reactivity appearing to stem from the record keeping involved in the diary tasks.

The Self-Report Measurement study includes other measures such as QF and beverage-specific GF (Knupfer Series) so it will be interesting to examine in detail the biases in the different types of measures in within-subjects designs, and identify strengths and weaknesses implied by each type of measurement (Midanik 1994). We have begun to work on combining data from several alcohol measures, beverage-specific and combined alcohol GFs for example (Greenfield et al. 2003, Kerr & Greenfield 2007). It is hoped that ultimately some of the errors in any one measure may be reduced by use of several, as would be an implication of standard psychometric theory (Green et al. 1979, Greenfield et al. 2006).

For helping decide how to resolve the GF problem of respondents who provide too many total days on the 12-Month GF, the daily data was also informative. Overall, the results from both studies suggest that downward capping may be slightly preferred to prorating, particularly for those cases where adjustment is necessary. Why might downward capping be slightly more reflective of daily distributions? For one thing, it has been pointed out by Schwarz (1999) that up to 30% of any representative sample offer opinions on fictitious

issues, which he sees as indicative of the conversational norm in interviews. If you ask the question, responses will follow, whether they make sense or not! GF's with many levels, as used in the 1979 NAS for example, appear to give too much opportunity to proliferate spurious days of drinking, which is why ARG simplified the GF after that survey and reduced its number of quantity levels. We believe asking about partial drinks in the GF series may have the same problem, generating spurious answers. In telephone interviews a CATI system can be programmed to stop asking questions when cumulative days exceed 365, or to promote interviewer checks and data reconciliation efforts at that point. This appears to have worked well in the methodological diary study (Figure 2). We intend to develop ways of efficiently doing this in 12-month GFs as well. Yet despite the capping versus prorating differences seen, choice of one or the other strategy may not greatly affect population study results when taken together with the much larger number of cases not exceeding the 365 day reference period and so not needing adjustment.

As noted, in the Self-Report Measurement Study, within subject GF- and Diary-based frequency distributions (days drinking at each quantity level) do not correspond as well as the distributions found in the aggregated sample, i.e., when population rather than individual-level comparisons are made (as in Table 3 and Figure 2). Such a difference of results would be expected on the basis of individual measurement errors and is why large samples are needed for population estimates in monitoring surveys. The GF population estimates and the drinking distribution show very little bias overall (compared to the silver standard), suggesting GF reporting errors on an individual level may be relatively random. Nonetheless, an important future agenda for the Self-Report Measurement study is to examine individual differences such as drinking pattern, drinking problems, and demographics like educational level that have been hypothesized to influence accuracy of reporting. Doing this was beyond the scope of the present paper and awaits the completion of the data collection. If ways can be found of developing instruction sets or question series that work better for those prone to make reporting errors, measurement should be improved.

Acknowledgments

Work was supported by the U.S. National Institute on Alcohol Abuse and Alcoholism (NIAAA) Center Grant P30 AA05595 and grants R01 AA013309 and R21 AA014773 (all T. Greenfield, PI) to the Alcohol Research Group, Public Health Institute and by the Centre for Addictions Research of British Columbia, University of Victoria. Data collections were funded from various agencies: Australian Government's Department of Health and Aging to the Australian Institute of Health and Welfare (2004 NDSHS) and the NIAAA grants just mentioned. Opinions are those of the authors and not necessarily of the sponsoring institutions. We wish to thank Jinhui Zhao for drink ethanol calculations for the Australian NDSHS.

References

- Dawson, Deborah A.; Room, Robin. Toward 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* 2000;12:1–21. [PubMed: 11288465]
- Gaunekar, Gaurish; Patel, Vikram; Rane, Anil. The impact and patterns of hazardous drinking amongst male industrial workers in Goa, India. *Social Psychiatry and Psychiatric Epidemiology* 2005;40:267–275. [PubMed: 15834777]
- Gmel, Gerhard; Graham, Kathryn; Kuendig, Hervé; Kuntsche, Sandra. Measuring alcohol consumption – should the 'graduated frequency' approach become the norm in survey research? *Addiction* 2006;101(1):16–30. [PubMed: 16393189]
- Graham, Kathryn; Demers, André; Rehm, Jürgen; Gmel, Gerhard. Problems with the graduated frequency approach to measuring alcohol consumption: results from a pilot study in Toronto, Canada. *Alcohol & Alcoholism* 2004;39(5):455–462. [PubMed: 15289203]
- Green RS, Nguyen TD, Attkisson CC. Harnessing the reliability of outcome measures. *Evaluation and Program Planning* 1979;2:137–142. [PubMed: 10244559]

- Greenfield, Thomas K. Evaluating competing models of alcohol-related harm. *Alcoholism: Clinical and Experimental Research* 1998;22(2):52s–62s.
- Greenfield, Thomas K. Ways of measuring drinking patterns and the difference they make: experience with graduated frequencies. *Journal of Substance Abuse* 2000;12(1):33–49. [PubMed: 11288473]
- Greenfield, Thomas K.; Bond, Jason; Kerr, William C.; Ye, Yu. Using Daily Drinking Diaries to Validate 28-Day and 12-month Graduated Frequencies Measures: A self-reported alcohol intake measurement study. The 33rd Annual Meeting of the Kettil Bruun Society for Social and Epidemiological Research on Alcohol; June 4–8; Budapest, Hungary. 2007.
- Greenfield, Thomas K.; Bond, Jason; Ye, Yu; Kerr, William C. Deriving heavy-per-day (5+) drinking measures for trend surveys from beverage-specific consumption series: possible or Quixotic?. The 29th Annual Meeting of the Kettil Bruun Society for Social and Epidemiological Research on Alcohol; June 2–6; Krakow, Poland. 2003.
- Greenfield, Thomas K.; Bond, Jason; Ye, Yu; Kerr, William C.; Anton, Raymond; Kranzler, Henry R. Research Society on Alcoholism. Victoria B.C.: 2004 Jun 26–30. Improving outcome measures for trials of new compounds for treating alcohol dependence. [Abstract: *Alcoholism: Clinical and Experimental Research* 28(5) Supplement 33A]
- Greenfield, Thomas K.; Kerr, William C. Alcohol measurement methodology in epidemiology: recent advances and opportunities. *Addiction* 2008;103(7):1082–1099. [PubMed: 18422826]
- Greenfield, Thomas K.; Nayak, Madhabika; Bond, Jason; Ye, Yu; Midanik, Lorraine T. Maximum quantity consumed and alcohol-related problems: assessing the most alcohol drunk with two measures. *Alcoholism: Clinical & Experimental Research* 2006;30(9):1576–1582.
- Greenfield, Thomas K.; Rogers, John D. Alcoholic beverage choice, risk perception, and self-reported drunk driving: effects of measurement on risk analysis. *Addiction* 1999;94(11):1735–1743. [PubMed: 10892011]
- Greenfield, Thomas K.; Tujague, Jennifer; Bond, Jason; Kerr, William C. Research Society on Alcoholism. Santa Barbara California: 2005 Jun 25–30. A pilot study of self-report alcohol consumption: diary versus transdermal alcohol sensor accuracy. [Abstract: *Alcoholism: Clinical and Experimental Research* 29(5) Supplement 130A]
- Hilton, Michael E. A comparison of a prospective diary and two summary recall techniques for recording alcohol consumption. *British Journal of Addiction* 1989;84:1085–1092. [PubMed: 2790272]
- Kerr, William C.; Greenfield, Thomas K. Distribution of alcohol consumption and expenditures and the impact of improved measurement on coverage of alcohol sales in the 2000 National Alcohol Survey. *Alcoholism: Clinical & Experimental Research* 2007;31(10):1714–1722.
- Kerr, William C.; Greenfield, Thomas K.; Tujague, Jennifer; Brown, Stephan. A drink is a drink? Variation in the alcohol content of beer, wine, and spirits drinks in a U.S. methodological sample. *Alcoholism: Clinical & Experimental Research* 2005;29(11):2015–2021.
- Leigh, Barbara C.; Aramburu, Beatriz; Norris, Jeanette. The morning after: Gender differences in attributions about alcohol-related sexual encounters. *Journal of Applied Social Psychology* 1992;22(5):343–357.
- Leigh, Barbara C.; Gillmore, Mary R.; Morrison, Diane M. Comparison of diary and retrospective measures for recording alcohol consumption and sexual activity. *Journal of Clinical Epidemiology* 1998;51(2):119–127. [PubMed: 9474072]
- Midanik, Lorraine T. Comparing usual quantity/frequency and graduated frequency scales to assess yearly alcohol consumption: Results from the 1990 U.S. National Alcohol Survey. *Addiction* 1994;89(4):407–412. [PubMed: 8025493]
- Midanik, Lorraine T. Comments on Dawson's Measuring alcohol consumption: limitations and prospects for improvement—Improving alcohol consumption measures: not doing badly. *Addiction* 1998;93(7):969–970.
- Midanik, Lorraine T. Definitions of drunkenness. *Substance Use & Misuse* 2003;38(9):1285–1303. [PubMed: 12908812]
- Midanik, Lorraine T.; Greenfield, Thomas K. Defining “current drinkers” in national surveys: results of the 2000 National Alcohol Survey. *Addiction* 2003;98(4):517–522. [PubMed: 12653821]

- Midanik, Lorraine T.; Greenfield, Thomas K. [Editorial] Methodological articles in addiction journals: We need more, less, or about the same amount? *Addiction* 2004;99:145–147. [PubMed: 14756697]
- Midanik, Lorraine T.; Hines, Alice M.; Greenfield, Thomas K.; Rogers, John D. Face-to-face versus telephone interviews: using cognitive methods to assess alcohol survey questions. *Contemporary Drug Problems* 1999;26:673–693.
- Rehm, J Jürgen; Ashley, Mary J.; Room, Robin; Single, Eric; Bondy, Susan; Ferrence, Roberta; Giesbrecht, Norman. On the emerging paradigm of drinking patterns and their social and health consequences. *Addiction* 1996;91(11):1615–1621. [PubMed: 8972920]
- Rehm, Jürgen; Greenfield, Thomas K.; Walsh, Gordon; Xie, Xiaodi; Robson, Lynda; Single, Eric. Assessment methods for alcohol consumption, prevalence of high risk drinking and harm: a sensitivity analysis. *International Journal of Epidemiology* 1999;28:219–224. [PubMed: 10342682]
- Rogers, John D.; Greenfield, Thomas K. Beer drinking accounts for most of the hazardous alcohol consumption reported in the United States. *Journal of Studies on Alcohol* 1999;60(6):732–739. [PubMed: 10606483]
- Room, Robin. Measuring alcohol consumption in the United States: methods and rationales. In: Lynn, Kozlowski; Annis, Helen M.; Cappell, Howard D.; Glaser, Frederick B.; Goodstadt, Michael S.; Israel, Yedi; Kalant, Harold; Sellers, Edward M.; Vingilis, Evelyn R., editors. *Research advances in alcohol and drug problems*. Plenum Press; 1990. p. 39-80.
- Schwarz, Norbert. Self-reports how the questions shape the answers. *American Psychologist* 1999;54(2):93–105.
- Stockwell, Tim; Donath, S.; Cooper-Stanbury, M.; Chikritzhs, Tanya; Catalano, P.; Mateo, C. Under-reporting of alcohol consumption in household surveys: a comparison of quantity-frequency, graduated-frequency and recent recall. *Addiction* 2004;99(8):1024–1033. [PubMed: 15265099]
- Stockwell, Tim; Zhao, Jinhui; Chikritzhs, Tanya; Greenfield, Thomas K. What did you drink yesterday? Public health relevance of a recent recall method used in the 2004 Australian National Drug Strategy Household Survey. *Addiction* 2008;103:919–928. [PubMed: 18482414]
- Turner, Charlotte. How much alcohol is in a ‘standard drink’? An analysis of 125 studies. *British Journal of Addiction* 1990;85:1171–1175. [PubMed: 2224197]
- Verbrugge, Lois M. Health diaries. *Medical Care* 1980;18(1):73–95. [PubMed: 6986517]
- World Health Organization. *International Guide for Monitoring Alcohol Consumption and Related Harm*. World Health Organization, Department of Mental Health and Substance Dependence, Noncommunicable Diseases and Mental Health Cluster. 2000.

Figure 1.
Effect on GF-quantity means of accounting for Yesterday's drink ethanol (NDSHS) (Dashed line = unadjusted means; solid line = ethanol-adjusted means (both 10 g AU drinks)).

Figure 2.
Distributions of 28-day Pre-GF, 28-day Post-GF and 28-day Diary quantities

Table 1

Australian 2004 NDSHS Means for GF-equivalent ranges of Yesterday's combined drinks for low, medium and high volume tertiles, for all cases (Total, unweighted)^a and ethanol-adjusted^b

GF-Quantity Range (10g drinks)	Range Mid-point	Volume Tertile			Total Sample (10 g drink	Ethanol Adjusted equivalents)
		Low	Medium	High		
Less than 1 drink	0.5	0.43(150)	0.48(132)	0.44(44)	0.45(326)	0.77
1 – 2	1.5	1.33 (864)	1.44 (2301)	1.64*** (1905)	1.50 (5070)	1.92
3 – 4	3.5	3.32 (112)	3.34 (527)	3.43*** (1807)	3.40 (2446)	3.90
5 – 6	5.5	5.50 (26)	5.49 (134)	5.54 (887)	5.53 (1047)	6.02
7 – 10	8.5	8.43 (7)	8.51 (72)	8.47 (667)	8.47 (746)	8.67
11 – 19	15	-	13.00 (9)	13.65 (272)	13.63 (281)	14.41
20+	un-defined	-	21.00 (2)	23.58 (62)	23.50 (64)	36.32

Significant difference between volume tertile groups: * p < .05

** p < .01

*** p < .001

^aVolume was computed using the downward capped GF algorithm (see Methods) and divided in approximate thirds, for current drinkers (at least 1 drink in prior 12 months). Sample n is given below each estimate in parentheses (N=9980).

^bEthanol-adjusted estimates based on beverage-specific drink size and ethanol content "empirical" adjustments to number of 10 g standard drinks based on Yesterday data (courtesy of JinHui Zhao and Tim Stockwell, see also Stockwell et al., in press, N=8364).

Table 2

Australian 2004 NDSHS Maximum based on the 12-month GF data ^a and number of cases (a) capped by “Top Level = Every Day” Rule and (b) by Summed Days > 365 Rule ^b

GF Range (drinks)	Maximum		Capping (of all drinkers)	
	Frequency	Percent	Percent 1+ Drinkers	% (n) Summed Days > 365
No Drinks	615	2.7	-	-
Less than 1 drink	2358	10.2	-	-
1 – 2	4720	20.4	21.3	1.51 (305)
3 – 4	4145	17.9	18.7	0.72 (146)
5 – 6	3590	15.5	16.2	0.34 (67)
7 – 10	3294	14.2	14.8	0.09 (18)
11 – 19	1987	8.6	9.0	0.07 (14)
20+	2452	10.6	11.1	-
Total	23161	100.0 (23161)	100.0 (20188)	2.72 (550)
				13.19 (2662)

^aResults for Maximum shown including and excluding partial and zero drinks.

^bWhen top level was ‘Every Day’, respondent (or data edit) gave missing values below, termed “Top Level = 365” capping; when frequencies at several GF-levels total > 365 days, downward capping algorithm is used (“Summed Days > 365”), n for capping analyses = 20,188 drinkers.

Table 3

US Self-Report Measurement Validity Study: Number of days (and percentage distributions) of 28-day Pre-GF, 28-day Post-GF and 28-day Diary by GF-quantity levels, for all cases with at least 26 diary days ($n=119$)^a

GF-Level (drink ranges)	28-Day Pre-GF		28-Day Post-GF		28-Day Diary	
	Days	%	Days	%	Days	%
12+	101	3.0%	111	3.4%	56	1.7%
8-11	118	3.5%	95	2.9%	129	3.9%
5-7	242	7.3%	232	7.0%	286	8.6%
3-4	388	11.6%	522	15.8%	553	16.6%
1-2	1126	33.8%	1078	32.6%	1025	30.8%
0	1357	40.7%	1266	38.3%	1276	38.4%
Drinking Days	1975	59.3%	2038	61.7%	2049	61.6%
Total Days	3332	100.0%	3304	100.0%	3325	100.0%

^a Pearson Correlations between days-at-levels distributions: Post-GF to Diaries $r = 0.996$; Pre-GF to Diaries $r = 0.987$; Pre-GF to Post-GF $r = .992$.

US Self-Report Measurement Study: Means (*n*) for GF-equivalent ranges of drinks reported consumed in 28-day diaries (days) for Volume Tertiles (see text) and all diary-day cases (28-day Total)^a; percent (*n*) capped per GF-level for 12-month GF^b

Table 4

GF-Quantity Range (12-14 g drinks)	Range Arithmetic Mid-point	Volume Tertile			28-day Total Mean	12-month GF Capping ^c
		Low	Medium	High		
1 – 2	1.5	1.32 (408)	1.47 (428)	1.65*** (268)	1.46 (1104)	26.5 (39)
3 – 4	3.5	3.33 (81)	3.29 (206)	3.44*** (318)	3.38 (605)	3.4 (5)
5 – 7	6.0	5.52 (23)	5.74 (82)	5.86 (209)	5.80 (314)	2.7 (4)
8 – 11	9.5	8.00 (2)	9.00 (23)	9.21 (122)	9.16 (147)	0.7 (1)
12+	Un-defined	-	15.33 (6)	15.49 (57)	15.48 (63)	-

^a A “case” is a person-drinking day (data for estimates *n*=119 with 26+ days valid)

Significant difference between volume tertile groups: * *p* < .05

**

p <

p < .001

^b Capping is only relevant to 12-Month GF and data are based on *n* = 147 cases (36 cases of 119, or 30%, were capped in the sample with 26 or more valid diary days).

Table 5

Tertile and overall Mean Volumes for three Measures: 28-day Pre- and Post-GF and Prior 12-month GF (all in total drinks during respective reference period)

Volume Measure	Volume Tertile			Overall Mean Volume (n=119)
	Low (n=40) Mean Volume	Medium (n=39) Mean Volume	High (n=40) Mean Volume	
28-day Pre-GF	18.66	45.95	109.86	58.26
28-day Post-GF	21.46	48.76	110.54	60.35
Prior 12-month GF	180.05	514.74	1502.53	734.27

Table 6

2004 NHSDS Drinkers of at least 1 drink whose Cumulative 12-month GF Frequency Exceeded 365 days ($n = 2,662$): Comparison of Aggregated Drinking Days for (a) Downward Capped GF (Rows B, D); (b) 'Prorated GF' (Rows C, E); and Yesterday's number of drinks in GF Levels (Rows F to H)

^a

A	B	C	D	E	F	G	H	I
GF-Level (drink ranges)	Aggregated Days		Distribution (Percentages)					
	GF-Capped days	GF-Prorated days	GF-Capped	GF-Prorated	Yesterday (only 1+ drink days)	Yesterday (w/ partial and 0 drink days)	Yesterday days drinking in GF ranges	GF-Level (drink ranges)
20+	20802	12385.09	2.1%	1.3%	1.6%	1.4%	37	20+
11-19	41711	27184.12	4.3%	2.8%	5.7%	4.9%	129	11-19
7-10	101168	72688.44	10.4%	7.5%	12.4%	10.5%	278	7-10
5-6	189704	152222.1	19.5%	15.7%	17.2%	14.6%	386	5-6
3-4	314148	293774.5	32.3%	30.2%	32.7%	27.8%	736	3-4
1-2	304097	413375.8	31.3%	42.5%	30.4%	25.8%	683	1-2
<1	-	-	-	-	-	0.4%	11	<1
Zero drinks	-	-	-	-	-	14.7%	390	0
Total days 1+	971630	971630	100.0%	100.0%	100.0%	100.0%	2249	N 1+ days
Total Case N	2662						2650	N all days

^aPartial drink days are excluded in Rows B to F, but shown, together with zero drink days for Yesterday in columns G and H; Correlations between GF-Capped and Yesterday (Rows D & F) = .99; between GF-Prorated and Yesterday (Rows E & F) = .94

Table 7

Self-Report Measurement Study Data: All Current Drinkers (Prior 12-months; $n = 119$): Comparison of Aggregated Drinking Days for (a) Downward Capped GF (Rows B, D); (b) 'Prorated GF' (Rows C, E); and 28-day Diary number of drinks in a day for GF Levels (Rows F to H) ^a

A	B		C		D	E	Distribution (Percentages)			H	I
	Aggregated Days		GF-Prorated days		GF-Capped	GF-Prorated	Diary Days (drinking drinks)	Diary Days (with non drinking days)	Diary days in GF ranges	GF-Level (drink ranges)	
GF-Level (drink ranges)	GF-Capped days	GF-Prorated days	GF-Capped	GF-Prorated	GF-Capped	GF-Prorated	Diary Days (drinking drinks)	Diary Days (with non drinking days)	Diary days in GF ranges	GF-Level (drink ranges)	
12+	874	723.5	3.1%	2.7%	3.1%	2.7%	2.7%	1.7%	56	12+	
8-11	1623	1103.5	5.8%	4.0%	5.8%	4.0%	6.3%	3.9%	129	8-11	
5-7	1998	1882.6	7.2%	6.8%	7.2%	6.8%	14.0%	8.6%	286	5-7	
3-4	5999	4852.7	21.5%	17.4%	21.5%	17.4%	27.0%	16.6%	553	3-4	
(1-2)	(17385)	(19316.7)	(62.4%)	(69.3%)	(62.4%)	(69.3%)	(50.0%)	(30.8%)	(1025)	(1-2)	
2	10228	10087.9	35.7%	36.2%	35.7%	36.2%	23.2%	14.3%	475	2	
1	7157	9228.8	25.7%	33.1%	25.7%	33.1%	26.8%	16.5%	550	1	
Zero drinks	-	-	-	-	-	-	-	38.4%	1276	0	
Total days 1+	45264	47195.7	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	2049	N 1+ days	
Total Case N	119								3325	N all days	

^aNon-drinking days are excluded in Rows B to F, but zero drink days are shown for Diaries in columns G and H; Correlations between GF-Capped and Diary (Rows D & F) = .972; between GF-Prorated and Diary (Rows E & F) = .954, based on 1-2 drink levels combined (with 1 and 2 drink levels separated, respective r s are .844 and .832).