COMPARISON OF A FOOD-FREQUENCY QUESTIONNAIRE METHOD AND A QUANTITY-FREQUENCY METHOD TO CLASSIFY RISKY ALCOHOL CONSUMPTION IN WOMEN

SUSAN L. CLEMENS* and SHARON MATTHEWS

Turning Point Alcohol and Drug Centre, 54-62 Gertrude St, Fitzroy, VIC 3065, Australia

(Received 30 April 2007; first review notified 24 July 2007; in revised form 31 August 2007; accepted 31 August 2007; advance access publication 17 October 2007)

Abstract — Aims: Population surveys use a variety of methods to collect data on alcohol consumption. Comparability of results across methods is a prime consideration. Different methods have been demonstrated to be robust in terms of ranking individuals’ alcohol use, while results have been mixed regarding comparability in terms of volume of consumption. In Australia, evidence-based guidelines have been developed that identify critical thresholds of consumption that are associated with increased risk of alcohol-related morbidity. This study investigated whether the identification of individuals consuming alcohol above these thresholds was consistent across two methods used to collect data on consumption. Methods: The Australian Longitudinal Study of Women’s Health (ALSWH) incorporated both a quantity-frequency (QF) method and a food-frequency questionnaire (FFQ) to collect data on alcohol consumption. Comparisons were made between these two methods on the ability to classify women consuming alcohol as risky (between 176 and 350 ml of pure alcohol weekly) and at high risk (greater than 350 ml of pure alcohol weekly) levels. Results: The ranking of individuals was robust across methods. However, concordance in identifying risky/high-risk drinkers varied considerably based on the assumptions underlying the different methods used to calculate drinking volume using the FFQ. Similarly, the sensitivity and specificity of the FFQ methods compared to QF in terms of identifying risky/high-risk consumers were high but variable. Conclusions: This study indicated that the proportion of respondents exceeding consumption thresholds was sensitive to the instrument used to collect data on alcohol intake. Quantifying such differences is important when making comparisons between surveys that use different methodologies.

INTRODUCTION

Alcohol consumption is associated with a large proportion of the global burden of disease for both acute and chronic conditions (Mathers et al., 2001; Rehm and Monteiro, 2005; Boffetta et al., 2006; Rehm et al., 2006a,b). Significant effort is invested in the accurate collection of data on alcohol consumption worldwide. Several formats are available for the collection of data on alcohol consumption (for reviews see World Health Organization, 2000; Dawson, 2003; Del Boca and Darkes, 2003). Each has specific strengths and weaknesses, for example, in terms of validity, reliability, or the time and resources required to collect the data. This has led to the common usage of a variety of data collection techniques with comparability between methods becoming a prime consideration.

The quantity-frequency (QF) and graduated quantity-frequency (GF) methods are often utilized for large population surveys due to their brevity and ease of analysis compared to other methods. The QF asks respondents to report their usual drinking quantity and frequency over a reference period, typically the past year. The GF captures more detailed information on the pattern of drinking, with respondents reporting how often during the reference period they drank various quantities of alcohol. This ability to capture more detailed information on infrequent heavy drinking occasions results in typically higher volumes being recorded using the GF relative to the QF (Room, 1991; Rehm et al., 1999). However, difficulties are encountered when the frequency response items are summed, with some respondents reporting consumption for more than 365 days per annum (World Health Organization, 2000).

Relative to the QF and GF, food-frequency questionnaires (FFQs) require a large amount of time and resources to collect and analyse. FFQs are used in intensive studies of diet and chronic disease and are a powerful tool for examining total nutrient intake and the effects on morbidity and mortality. FFQs are extensively tested for reliability and validity, typically undertaken using a food intake diary for comparison (Ocke et al., 1997; Johansson et al., 2002; Jain et al., 2003; Nath and Huffman, 2005; Lee et al., 2006). Other methods commonly used include prospective or retrospective drinking diaries, however, these methods require larger amounts of resources to collect and analyse, and only capture drinking history over a short time period, typically 1 week. This shorter reference period can lead to erroneous classification of very infrequent drinkers as abstainers, due to the fact that alcohol wasn’t consumed during the reference period.

Each technique used to collect data on alcohol consumption has a correspondingly different approach and set of assumptions to convert data on consumption to volume of alcohol consumed, therefore, it is unlikely that different methods will record identical volumes of alcohol. One strategy to overcome discrepancies in terms of absolute volume of alcohol consumed, is to translate millilitres or grams of alcohol into the number of standard drinks. Alternately, comparisons may be based on relative ranks, which have been shown to be robust across methods (Feunekes et al., 1999). Significant effort has been invested into examining the comparability of various methods of collecting data on alcohol consumption in terms of absolute volume of alcohol intake, number of standard drinks consumed, and ranking of individuals.

Recently, investigation of alcohol consumption at the population level has focused on patterns of consumption...
associated with alcohol-related harm, frequently defined as exceeding thresholds of low-risk drinking. This has arisen out of the recognition that the omission of a consumption criterion in alcohol use disorder classification in psychiatry was based on methodological feasibility in a rather limited study (Saha et al., 2007). Typically, 60 g of alcohol per occasion is often used as a harmful consumption threshold in the USA (Midanik et al., 1996; Dawson, 2003; Greenfield and Kerr, 2003). In Australia, risky drinking thresholds were established in 2001 by the National Health and Medical Research Council (NHMRC). Risk levels were differentiated as those associated with acute harm in the short term, and those representing long-term chronic risk, and different thresholds were set for various sub-populations (e.g., women, the elderly). The rationale for different thresholds for women has been detailed elsewhere (Graham et al., 1998; Graham et al., 2004).

As interest in the use of consumption thresholds continues to increase, the ability to correctly classify individuals on the basis of such thresholds across methods used to collect data on alcohol consumption becomes increasingly important. The comparability on the basis of thresholds is not as well investigated as comparability on the basis of absolute amount of alcohol consumed, number of drinks consumed, or rankings. Further, only rarely are results from a FFQs compared to those from methods commonly used to collect data on alcohol consumption, such as the QF or GF (for a review see Feunekes et al., 1999). One study compared a FFQ to a more detailed risk-factor questionnaire using thresholds of consumption and reported that 62% of subjects were classified into the same risk category (Witte and Haile, 1996). To the authors’ knowledge, comparisons between a FFQ and the QF method using thresholds for increased risk of experiencing adverse alcohol-related consequences have not previously been reported.

This study aims to address this issue by comparing results from a QF and an FFQ method of collecting data on alcohol consumption. Methods are compared in terms of overall ranking and in terms of identifying individuals that consume alcohol above thresholds associated with long-term alcohol-related risk.

MATERIALS AND METHODS

Data
The Australian Longitudinal Study of Women’s Health (ALSWH) consists of three cohorts of women recruited from the Medicare database, which includes all Australian citizens and permanent residents, regardless of age or income. The cohorts were aged 18 to 23 years (younger cohort; \( N = 14,247 \)), 45 to 50 years (mid-aged cohort; \( N = 13,716 \)), and 70 to 75 years (older cohort; \( N = 12,432 \)) at initial data collection, and all cohorts completed the initial survey in 1996. Initial response rates for the mailed questionnaire were 41, 54, and 36% for the younger, mid-aged, and older cohorts, respectively. After 1996, one cohort was surveyed each year on a rotational basis such that data were collected for each cohort approximately every 3 years. Data are collected by means of a mailed self-completed survey. Detailed methods of this study have been reported elsewhere (Brown et al., 1998).

During the approximately 10 years that this survey has been conducted, measures of alcohol consumption have been collected using two different methods. From 1996 to 2000, data were collected using a QF method. In 2001, respondents to Survey 3 for the mid-aged cohort reported data on alcohol consumption as part of the Dietary Questionnaire for Epidemiological Studies (DQES) Version 2, a FFQ developed by The Cancer Council of Victoria that collects information regarding usual eating habits, including alcohol intake, over the previous 12 months (Hodge et al., 2000). In 2003, the younger cohort received both the FFQ and the QF questions used in previous surveys in separate sections of the questionnaire. The fact that this cohort received both question formats simultaneously enables the comparison of classification based on thresholds across instruments. Details of the questions for both formats are presented in Table 1.

Alcohol consumption calculations

Consumption recorded by the QF method. Calculating the number of drinks using the basic QF method is straightforward (Dawson, 2003). The quantity question instructed respondents to report their consumption in ‘standard drinks’. A standard drink has been defined by the NHMRC as containing approximately 12.5 ml of alcohol. Usual weekly alcohol consumption was calculated by multiplying the midpoints of the response categories for the quantity, in terms of drinks per occasion, and frequency, in terms of days per week. From these weekly consumption volumes, respondents were categorized according to NHMRC drinking guidelines as abstaining from alcohol consumption, low-risk drinkers (up to 175 ml of pure alcohol per week), risky drinkers (between 175.1 and 350 ml per week) and high-risk drinkers (over 350 ml per week).

Consumption recorded by the FFQ method. Calculating weekly alcohol consumption using the beverage-specific DQES FFQ is more complex for several reasons. Unlike the QF calculation with an overall quantity and an overall frequency of consumption, the DQES FFQ collected an overall quantity but collected frequency based on beverage type. The omission of an overall frequency required that two assumptions be applied for individuals that drank more than one type of alcohol, as there is no means to determine whether consumption took place on the same or different occasions. One, that consumption of all types of alcohol occurred on the same days and that the maximum frequency across drink types was the true weekly frequency (‘max of days’ assumption). Conversely, it may be assumed that individuals only drank one type of alcohol per occasion and therefore the total number of drinking days was the sum of the drink-specific frequencies (‘sum of days’ assumption). Under this assumption, it was possible for the sum to total more than 7 days. In these cases, the drink-specific frequencies were proportionately reduced to total 7 days.

The different assumptions underlying the calculation of consumption frequency (‘sum of days’ (vs) ‘maximum of days’) impacts the derivation of weekly consumption in two ways. When weekly consumption frequency is assumed to be the sum of individual drink-type frequencies, weekly consumption becomes the sum of the consumption of each beverage calculated using the basic QF calculation (the product of
Table 1. Alcohol consumption questions in the 2003 ALSWA survey

<table>
<thead>
<tr>
<th>Question</th>
<th>Response options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity–frequency method</td>
<td>I never drink alcohol</td>
</tr>
<tr>
<td>How often do you usually drink alcohol?</td>
<td>Less than once a month</td>
</tr>
<tr>
<td></td>
<td>Less than once a week</td>
</tr>
<tr>
<td></td>
<td>On 1 or 2 days a week</td>
</tr>
<tr>
<td></td>
<td>On 3 or 4 days a week</td>
</tr>
<tr>
<td></td>
<td>On 5 or 6 days a week</td>
</tr>
<tr>
<td></td>
<td>Every day</td>
</tr>
<tr>
<td>On a day when you drink alcohol, how many standard drinks do you usually have?</td>
<td>1 or 2 drinks per day</td>
</tr>
<tr>
<td></td>
<td>3 or 4 drinks per day</td>
</tr>
<tr>
<td></td>
<td>5 to 8 drinks per day</td>
</tr>
<tr>
<td></td>
<td>9 or more drinks per day</td>
</tr>
</tbody>
</table>

Food frequency method

Over the last 12 months, how often did you drink beer, wine and/or spirits? (Note: Separate questions for each type of alcohol.)

(a) Beer (low alcohol) Never
(b) Beer (full strength) Less than once per month
(c) Red wine 1–3 days per month
(d) White wine (including sparkling wines) 1 day per week
(e) Fortified wines, port, sherry, etc. 2 days per week
(f) Spirits, liqueurs, etc 3 days per week

Over the last 12 months, on days when you were drinking, how many glasses of beer, wine and/or spirits altogether did you usually drink?
(Note: One question for all types of alcohol combined.)

One
Two
Three
Four
Five
Six
Seven
Eight
Nine
Ten or more

Each beverage-specific frequency by the overall quantity per occasion. However, when weekly frequency is the maximum of drink-type frequencies, this approach would overestimate weekly intake. Under the ‘maximum of days’ assumption, each drink-specific frequency must be proportionately reduced such that the sum of the pro rata frequencies will equal the maximum number of drinking days reported. Weekly consumption was calculated as the sum of the pro rata frequencies multiplied by the overall quantity.

A second difficulty with the FFQ was that the quantity of consumption was collected in terms of ‘glasses’ rather than a standard based on alcohol content regardless of beverage type. The survey provided some guidance in recording glasses of different beverages, however, these examples did not conform to the Australian standard drink guidelines nor was alcohol content equalized across drink types. Conversion from survey description of glasses to standard drinks was performed for each of the drink types used in the examples (see Table 2). Spirits were not included as an example in the survey and were interpreted with a glass equalling a standard drink. This conversion enabled calculation of consumption in terms of standard drinks across drink types. It may be argued that respondents do not convert usual beverage portions to standard drinks when reporting, and therefore that glasses are an appropriate surrogate measure of standard drinks without conversion. A second means of calculating quantity, with glasses analysed as a standard drink, was therefore included.

Lastly, the documentation provided for the DQES FFQ suggests a weighted method combining the results from the ‘max of days’ and ‘sum of days’ methods (The Cancer Council of Victoria, 2005). From the National Health Survey 3-day alcohol consumption diary, it was determined that individuals who drink more than one type of alcohol were about twice as likely to do so on different days. The weighted average was calculated by giving the results from the ‘sum of days’ method two times the weight as the ‘max of days’ method.

Conversion to NHMRC risky drinking categories. The different assumptions regarding weekly frequency and conversion to standard drinks resulted in six different calculations of weekly consumption based on the FFQ (see Table 3). The resulting volume of weekly consumption from the QF and FFQ methods was translated to consumption
above the thresholds defined by the NHMRC low-risk drinking guidelines. Specifically, four thresholds were compared: those that abstain from alcohol consumption, those that drink within low-risk guidelines, risky drinkers, and high-risk drinkers. Nineteen cases with contradictory information regarding drinking status within a given method were eliminated from analysis; 197 cases reported contradictory drinking status across the QF and FFQ methods and were retained. One inconsistency observed on the FFQ involved individuals who reported that they never consumed any type of alcohol based upon their frequency of drinking, yet reported that they consumed one drink in the quantity question and one to two drinks in the maximum quantity question (the lowest categories available for both questions). These individuals are assumed to be non-drinkers that erroneously answered the remainder of the alcohol questions because ‘non-drinker’ was not an available option for these items.

Statistical analysis
Analysis consisted of Spearman’s rank correlation coefficients to assess the comparability in overall ranking in terms of drinks per week between the QF and the results based on each assumption using the FFQ. Spearman’s rank correlations were chosen because the results from the QF and FFQ approaches were not jointly normally distributed. Unweighted kappa statistics were used to investigate agreement based upon the NHMRC risky drinking threshold classifications. Tests for sensitivity, the ability to correctly classify cases that consume alcohol above NHMRC guidelines, and specificity, the ability to correctly classify cases that do not consume alcohol above NHMRC guidelines, were performed. In all cases, the results from FFQ methods were compared to the QF results using the QF as the benchmark. The ability to correctly classify any consumption above the guidelines was tested by combining the risky and high-risk categories, and comparing the combined abstainers and low-risk drinking categories.

RESULTS
Mean weekly consumption was estimated as 3.4 standard drinks per week by the QF method, approximately 4 drinks per week by the FFQ maximum of drinking days assumption, approximately 6.4 standard drinks by the weighted method, and approximately 7.5 drinks per week based upon the FFQ sum of drinking days approach (Table 4). Spearman’s correlation coefficients indicate high comparability in ranking individuals between the QF and FFQ methods, with coefficients exceeding 0.8.

The ability to correctly predict high-risk consumption was tested by comparing the classification of high-risk drinkers to all other categories combined. Analysis was performed in Stata V9.2.
method (8.5 compared to 8.1%, respectively). The QF method also resulted in the highest proportion of low-risk drinkers (88.9%) and correspondingly lowest proportions of risky and high-risk consumers (2.4 and 0.6%, respectively). Estimates of the proportion of risky drinkers under the ‘max of days’ assumption were about two times greater, and estimates under the ‘sum of days’ assumption were approximately five times greater than corresponding estimates from the QF method. Similar figures were observed for the proportion of high-risk drinkers (two times and eight times greater for ‘max of days’ and ‘sum of days’, respectively). A comparison of unadjusted (glasses) and standard drink adjusted (standard) estimates indicated that such adjustment had little effect on proportions of risky or high-risk drinkers.

Concordance between the QF and the FFQ methods are presented in Table 6. Using the criteria outlined by Landis and Koch (1977) where values greater than 0.75 indicate excellent agreement beyond chance, values below 0.40 indicate poor agreement, and values between 0.40 and 0.75 indicate fair to good agreement, concordance between methods was very high for the crude abstainer/drinker dichotomy (K = 0.857). Examining all NHMRC categories, kappa statistics indicated fair to good agreement, with the ‘max of days’ assumption performing better than either the ‘sum of days’ or weighted methods. Combining the risky and high-risk categories improves kappa results with the ‘max of days’ assumption almost reaching excellent agreement. Concordance in agreement for high-risk drinkers only was basically poor across all methods. The results of all kappa analyses are significant at the P < 0.001 level.

Table 7 presents the sensitivity and specificity analysis for the FFQ assumptions using the QF results as the benchmark. Sensitivity, in terms of the FFQ to correctly classify individuals consuming alcohol at NHMRC risky/high-risk levels relative to the QF, is highest for calculations based on the ‘sum of days’ assumption although this technique incorrectly classifies approximately 14% of cases as risky/high-risk. Calculations based on the ‘max of days’ assumption performed much better in terms of these ‘false positives’ at the expense of reduced sensitivity to detect risky/high-risk consumers relative to the QF method, and misclassified 24 and 29% of cases (max-standard and max-glasses, respectively). Corresponding figures for the weighted method were 11% false positives and 9% false negatives. Sensitivity to correctly classify high-risk drinkers was lower although specificity was improved.

**DISCUSSION**

This study was conducted to assess concordance between alcohol-related risk thresholds established by the NHMRC derived from two different methods of measuring alcohol consumption. Such studies are valuable due to the variety of methods used to measure alcohol consumption and the need to investigate differences across surveys or across waves of longitudinal data collection when the instrument varied over time. The investigation of agreement across data collection methods in terms of consumption above low-risk drinking...
consumption as measured by two FFQs, and a more detailed study comparing the frequency of consumption from a FFQ format, resulting in higher consumption. The inclusion of an overall frequency question, as recommended by Dawson (2003), would resolve this problem. A second explanation may be that embedding alcohol-specific questionnaire, it was reported that differences in estimated alcohol consumption were larger for men than women (McCann et al., 1999). Lastly, the context of a FFQ supports more cues that may enhance accurate recall. It has been argued that a more intensive data collection process, such as a dietary history interview, is expected to have greater face validity that the QF method (Koppes et al., 2002). That assumption likely holds true in the current analysis of the FFQ, but to a lesser degree due to the self-administered mode of data collection. The present study used the QF results as the benchmark as it was the method employed in the majority of ALSWH surveys. In the absence of a more robust benchmark, such as a comprehensive dietary intake diary, it may be equally valid to interpret results as indicating that the QF under estimated harmful consumption in previous surveys.

That agreement on crude abstainer/drinker was not higher is troublesome. Cases where FFQ responses indicated an abstainer while QF respondents were in the low-risk category, and vice versa, might allow those respondents that very infrequently consume alcohol, for example only during holidays, to report that small quantities of alcohol, for example only during holidays, to report that small quantities of alcohol were consumed. This study is subject to several limitations. First, the cohort available for analysis consisted of young women aged 25 to 30 years. It has been reported that differences in recorded consumption across methods are higher among men (McCann et al., 1999) which indicates that the concordance rates found in this study may be overestimated compared to the general population. Further, as the cohort available consisted of women from a narrow age range, analysis of differences in reporting due to age were not pursued. Lastly, the alcohol measures used in the ALSWA study did not permit an analysis of patterns of consumption. In addition to the shortcomings of the FFQ discussed previously, it also does not contain a means to determine the frequency of heavy drinking episodes.

FFQs are a valuable source of alcohol consumption data and demonstrate high reliability when compared to standards such as prospective consumption diaries. It is important to quantify the comparability of results from an FFQ to those more commonly used in large population studies. Results from this study indicated that relative rankings are robust and that agreement on non-drinking status and low-risk drinking is high. Misclassification based on risky drinking thresholds was observed, however, demonstrating the importance of including more than one method to collect alcohol consumption data, and conducting appropriate sensitivity and specificity analysis, when the choice of instruments changes over time. Despite good agreement across methods, prevalence estimates based on risky and high-risk consumption thresholds under most assumptions. The significance of kappa results is likely inflated, however, due to the non-normal distribution of NHMRC risky drinking categories and due to the assumption regarding erroneous reporting on the FFQ. For example, it may also be possible that individuals that reported never drinking any type of alcohol were in actuality drinkers that consumed a type of alcohol not represented on the beverage-specific frequency FFQ question. For example, respondents may not be reporting pre-mixed spirits in the ‘spirits, liquors, etc.’ category and there are times when a respondent may not know the type of alcohol contained in pre-mixed beverages (Dawson, 2003). Conversely, for individuals classified as non-drinkers by the QF but consumers on the FFQ, the beverage-specific frequency FFQ question may allow those respondents that very infrequently consume alcohol, for example only during holidays, to report small quantity. Discordance in non-drinking status among individuals who consume small amounts of alcohol has been reported in other studies (Parker et al., 1996).
thresholds varied greatly and researchers should be cautious when comparing risky drinking prevalence estimates based on thresholds derived from different methods.

SOURCE OF SUPPORT

The Australian Longitudinal Study on Women’s Health is funded by the Australian Government Department of Health and Ageing. Funding for the alcohol analyses was provided by the Alcohol Education and Rehabilitation Foundation.

Acknowledgements — The Australian Longitudinal Study on Women’s Health is conducted by a team of researchers at the University of Newcastle and the University of Queensland, and funded by the Australian Government Department of Health and Ageing. The Dietary Questionnaire for Epidemiological Studies (Version 2) is used for the ALSWH with permission from Professor Graham Giles of the Cancer Epidemiology Centre of The Cancer Council of Victoria.

The authors would like to acknowledge the Alcohol Education and Rehabilitation Foundation for funds that supported analysis specifically related to alcohol consumption.

REFERENCES