Journal of Official Statistics, Vol. 12, No. 3, 1996, pp. 253–273

Response Effects in Consumption Surveys: An Application of the Beta-Binomial Model to Self-Reported Drinking Frequencies

Timo Alanko[1] *and Paul H. Lemmens*[2]

Prospective diaries, seven-day recall interviews, and a self-assessed summary measure are compared as methods of measuring drinking frequency. The distribution of the number of drinking days in diaries is modelled by the beta-binomial model and applied to a Dutch 1985 general population one-week diary and to a 1983 two-week diary. The model-based comparisons show that the deviances between the methods are most pronounced in respondents with intermediate drinking rates. Memory effects are suggested as an interpretation of the deviances.

Key words: Beta-binomial; surveys; respondent error; recall error; memory effects; drinking frequency; sensitive topics.

1. Introduction

Surveys often rely on reports of events or self-reports of behaviours that occurred in the past. The validity of retrospective self-reports is negatively affected by omissions in reporting events that occurred within a specific reference period. Omissions in surveys have been found to increase exponentially with time, to be related to the subjective importance of the event (saliency), to the characteristics of the person recalling, the interview technique, and the social meaning of the event that is being queried (Sudman and Bradburn 1982).

Specifically, in alcohol consumption surveys held in general populations, estimates of total alcohol consumption are systematically lower than what should be expected on the basis of official sales and tax data. This so-called undercoverage is considered indicative of the limited validity of self-reports of drinking. There is reason to believe that forgetting

[1] Technical Research Centre of Finland, Computing Services, P.O. Box 2103 FIN-02044 VTT, Finland. Internet: Timo.Alanko@VTT.FI
[2] Department of Medical Sociology, University of Limburg, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.
Acknowledgements: The work of Timo Alanko was supported by a grant from the Finnish Foundation for Alcohol Studies. The data have been gathered by the University of Limburg, Department of Medical Sociology, under a grant from the Dutch Foundation of the Moderate Use of Alcohol (STIVA). The authors wish to thank professor Anders Ekholm of the University of Helsinki and two anonymous referees for their valuable suggestions.
Correspondence to T. Alanko at the above address.

of drinking events is an important factor causing underestimation of total consumption (Alanko 1984; Lemmens 1991).

Alcohol consumption frequency is customarily measured by one of two types of questions: (1) asking the respondent to recall actual drinking events in the near past (often the past week), (2) requiring the subject to make a subjective summary of the typical or usual drinking frequency, taken over a longer period (ranging from one month to one year). Both methods have their relative merits and drawbacks (Duffy 1985; Duffy and Alanko 1992). As memory errors increase with time, limiting the length of recall period logically would reduce measurement error. Empirical findings by Mäkelä (1971), Wilson (1981), Philipsen, Knibbe, and van Reek (1983) and Poikolainen (1985) support this hypothesis.

In the actual consumption approach both a prospective diary and a retrospective recall have been used. Poikolainen and Kärkkäinen (1983) reported that a prospective diary technique produced higher consumption estimates than did retrospective methods. Lemmens, Tan, and Knibbe (1992) compared diary records kept over one week, a seven-day retrospective recall, and a typical summary measure and concluded that the discrepancies were mainly due to differences in reported frequency of drinking rather than the number of drinks on the recalled occasions. This result corroborated an earlier empirical finding by Alanko (1985) who reported that intra-individual variation of weekly consumption is caused mainly by variation in drinking frequency over weeks rather than by variation in the quantities consumed. Sobell, Bogardis, Schuller, Leo, and Sobell (1989) found that diaries yielded higher estimates of drinking than the weekly recall method and that the difference between recall and diary was due entirely to the number of drinking occasions reported.

Given the differences between the results obtained by differerent measurement techniques, we expect that a comparison of drinking frequency from prospective diaries on the one hand, and retrospective recall and reports on typical frequency on the other will also yield information on response failures. Furthermore, contrasting the diary with other methods of measurement permits an analysis of the interaction between recall failure and drinking pattern.

The aim of the study is to develop and apply statistical tools for a comparison of the retrospective recall and the typical frequency method against the prospective diary method. Further, the comparison focuses on whether the potential discrepancies are due to respondents with low, intermediate or high drinking frequencies or whether there are no significant differences between these consumer categories.

We first model drinking behaviour statistically in a manner that explicitly accounts for the individual level time-variation. Such a statistical model is necessary for a comparison where predictions based on an observed diary week are compared with observed responses of the same respondent in a different week or other reference period. Similarly, an explicit statistical model is required for the inferences drawn from these comparisons.

We then examine the statistical properties of drinking frequencies in terms of drinking days distributions. These are developed in Section 3. Some of the properties are then used to introduce inference tools for selected comparisons, given in Section 4. Finally, results related to model validation and the comparisons are applied to Dutch general population survey data in Section 5.

2. Data

We use two data sets; both are from general population surveys conducted in the Netherlands in 1983 and in 1985. In the 1983 survey (Lemmens et al. 1988) consumption data were collected as part of a larger survey among 20–64 year olds on life-style and health. The first part of the survey consisted of a personal interview on health problems, smoking, and alcohol consumption (response rate 71%). Alcohol consumption was assessed using a 7-day recall method, detailing location of drinking. In the second part of the survey a randomly selected subsample of 599 respondents was asked to keep a diary for 14 consecutive days. Alcohol consumption was recorded for each third part of the day. The completion rate was 83% ($N = 496$). The drinking behaviour in the diary subsample, as assessed by the 7-day recall, did not differ significantly from the original sample. The 1985 survey had a similar design, a personal interview and a diary, to be kept for 7 days (response rate 70%; Lemmens et al. 1988). In the interview, alcohol consumption was measured with several distinct indices of past consumption, among which a 7-day recall and a summary measure. In the latter, the respondent was asked to summarise his/her drinking frequency in the past six months, using eight options (Lemmens et al. 1992). The diary ($N = 918$) was completed in the week subsequent to the interview and inquired about consumption of all beverages, medicines and smoking, to be specified by every quarter of an hour. For additional details concerning the surveys the reader is referred to the works cited above.

For the ease of definition, a drinking occasion is defined in both data sets as any consumption during a day. Subjects that claimed not to have had any drink in the past year (the 1983 survey) or in the past six months (the 1985 survey) were considered abstainers and were excluded from the analysis.

The outcome measures considered here are the number of drinking days in the prospective diary, the number of drinking days in the 7-day retrospective recall, and a self-assessed typical frequency measure, all three recorded for each respondent in the 1985 survey, but only the first two in the 1983 survey.

3. Statistical Models

Like other types of purchasing and consumption behaviour (e.g., Sichel 1982), the variability of drinking in time has been tentatively modelled by assuming that the number of drinking occasions of each individual in a fixed time-period varies according to the Poisson distribution (Ekholm 1968; Alanko 1984). Nevertheless, the Dutch studies record data on the basis of daily totals of consumption. This gives the natural upper limit of drinking frequency as every day, i.e., seven days a week for the weekly recall and the one-week diary. In this context it is natural to define the individual's long-term drinking rate by his/her propensity to consume any alcoholic beverage on a given day.

The individual level propensity is summarized in the sequel by a single parameter π, the probability to consume alcoholic beverages on a randomly chosen day. This definition leads to the analysis of time variation through the binomial distribution, instead of the Poisson. Naturally, the propensity to consume varies from consumer to consumer. The between individuals variability of π can be modelled by assuming that the corresponding random variate Π has a continuous density $f_{\Pi}(\pi)$ defined for $0 < \pi < 1$. A natural first choice for $f_\Pi(\pi)$ is to assume that Π follows the beta distribution. It is well known that, under the assumptions described below, this leads to the beta-binomial distribution which offers mathematical convenience and has been used in several applications (e.g., Hald 1968; Chatfield and Goodhardt 1970; Williams 1975). The reasons for using the beta assumption in the present context are purely empirical. As shown below in Section 5, both the univariate and the bivariate beta-binomial distributions provide a good description of Dutch drinking days data as measured by a prospective diary. For alternatives to the beta assumption in a similar context the reader is referred to Alanko and Duffy (1996).

In this section a beta-binomial model is specified for the distribution of self-reported drinking days as measured by the prospective diary. First we review some statistical aspects of the beta-binomial model.

3.1. *The univariate beta-binomial model for the number of drinking days in the diary*

Let the conditional random variable $X|(\Pi = \pi)$ denote the number of drinking days out of N days (e.g., $N = 7$) in the diary for an individual, given that the individual has the probability π of drinking on any particular day. Assuming independence between days, it follows that the probability function $f_{X|\Pi}(\cdot|\pi)$ of $X|(\Pi = \pi)$ for $x = 0, 1, \ldots, N$ follows the binomial distribution

$$f_{X|\Pi}(x|\pi) = \binom{N}{x} \pi^x (1 - \pi)^{N-x} \tag{1}$$

denoted by $X|(\Pi = \pi) \sim Bin(N, \pi)$.

The unobservable random variable Π describes the variation of the individual drinking probabilities among the individuals of the study population. We assume that $\Pi \sim Beta(\alpha, \beta)$, i.e., that the density function $f_\Pi(\cdot)$ of Π over the individuals of the population is given by the beta distribution with parameters α and β

$$f_\Pi(\pi) = \frac{\pi^{\alpha-1}(1 - \pi)^{\beta-1}}{B(\alpha, \beta)} \qquad \alpha, \beta > 0; \quad 0 < \pi < 1 \tag{2}$$

where $B(\alpha, \beta)$ is the beta function with arguments α and β. The assumption (2) is not very restrictive. The beta-distribution is extremely flexible and admits a large number of shapes for different values of α and β as illustrated by, e.g., Johnson and Kotz (1970, pp. 42–43).

From assumptions (1) and (2) it follows by the rule of total probability that X, the number of drinking days out of N days recorded, has the beta-binomial distribution with index N and parameters α and β denoted as $X \sim BB(N, \alpha, \beta)$. The probability function $f_X(\cdot)$ of the unconditional random variate X is of the form

$$f_X(x) = \binom{N}{x} \frac{B(\alpha + x, \beta + N - x)}{B(\alpha, \beta)} \qquad \alpha, \beta > 0; \qquad x = 0, 1, \ldots, N. \tag{3}$$

One should note that only the population parameters α and β have an effect on this distribution. Thus (3) describes the population under study but is not concerned with individual behaviour. But the population level (or marginal) X is observable and the distributional result (3) can be checked against cross-sectional diary data, whereas the individual level assumption (1) and the population level assumption (2) cannot be directly verified.

Using general rules for conditional moments the expectation $E(X)$ and the variance $var(X)$ of the beta-binomial distribution can be calculated as

$$E(X) = E_{\Pi}[E(X|\Pi)] = \frac{N\alpha}{\alpha+\beta} \tag{4}$$

$$\begin{aligned} var(X) &= var_{\Pi}[E(X|\Pi)] + E_{\Pi}[var(X|\Pi)] \\ &= \frac{N^2\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2} + \frac{N\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)} \\ &= \frac{N\alpha\beta(N+\alpha+\beta)}{(\alpha+\beta+1)(\alpha+\beta)^2} \end{aligned} \tag{5}$$

For higher moments and other properties of the beta-binomial see Johnson, Kotz, and Kemp (1992, pp. 239–242, 264–266). The expression (5) is of interpretational interest in that the variance of the observable variable $var(X)$ is partitioned into two parts: the between individuals and the within individuals components. The first term on the right hand side of (5) is the variance of the regression of X on Π, i.e., of $h(\pi) = E(X|\Pi = \pi) = N\pi$ with respect to the distribution of Π, given above in (2). This is the between individuals component. The second term on the right hand side of (5) is the expected variance with respect to Π of $X|(\Pi = \pi)$. This is the within individuals component.

3.1.1. The conditional distribution of the propensity Π, given a beta-binomial observation

Although the propensity to drink, Π, is not directly observable it is easy to derive the conditional distribution of Π, given an observed number of drinking days $X = x$. Denote the density function by $f_{\Pi|X}(\cdot|x)$. The interpretation of $f_{\Pi|X}(\cdot|x)$ is that it expresses the density function of the propensity to consume in the subpopulation having observed consumption on x days out of N. By Bayes theorem and substitution of (1), (2), and (3), we obtain

$$f_{\Pi|X}(\pi|x) = \frac{f_{\Pi}(\pi)f_{X|\Pi}(x|\pi)}{f_X(x)} = \frac{\pi^{\alpha+x-1}(1-\pi)^{\beta+N-x}}{B(\alpha+x,\ \beta+N-x)} \tag{6}$$

i.e., $f_{\Pi|X}(\pi|x) \sim Beta(\alpha+x,\ \beta+N-x)$. Consequently, results based on the distribution $f_{\Pi|X}(\pi|x)$, such as

$$E(\Pi|X=x) = \frac{\alpha+x}{\beta+\alpha+N} \tag{7}$$

follow from (6). Note that in (7) regression of Π on x is a linear function of x.

3.2. *The bivariate beta-binomial model*

Suppose now that two non-overlapping time periods of lengths N_1 and N_2 days are observed. Denote by X_1, X_2, the random variables for the number of drinking days in the two periods. Assume further local independence $(X_1 \perp\!\!\!\perp X_2)|\Pi$, that is, for a given individual with a stable $\Pi = \pi$ in the two periods, X_1 and X_2 are independent. The bivariate conditional probability function $f_{X_1,X_2|\Pi}$ of observing him/her consume on x_1 days in the first period and on x_2 days in the second is given by the double binomial

$$f_{X_1,X_2|\Pi}(x_1,x_2|\pi) = \binom{N_1}{x_1}\binom{N_2}{x_2}\pi^{x_1+x_2}(1-\pi)^{N_1+N_2-x_1-x_2} \tag{8}$$

$0 < \pi < 1$; $x_1 = 0, 1, \ldots, N_1$; $x_2 = 0, 1, \ldots, N_2$. On the basis of (8) and the distribution of Π from (2) it follows without further assumptions that the probability function $f_{X_1,X_2}(\cdot)$ of the unconditional (i.e., population level) bivariate random variable (X_1, X_2) is of the bivariate beta-binomial form

$$f_{X_1,X_2}(x_1,x_2) = \binom{N_1}{x_1}\binom{N_2}{x_2}\frac{B(\alpha + x_1 + x_2,\ N_1 + N_2 + \beta - x_1 - x_2)}{B(\alpha,\beta)} \tag{9}$$

for $\alpha, \beta > 0$; $x_1 = 0, 1, \ldots, N_1$; $x_2 = 0, 1, \ldots, N_2$.

Using the identity $E(X_1X_2) = E_{X_1}[E(X_1X_2|X_1)]$ the covariance between X_1 and X_2 can be derived. In fact, it is of the form

$$Cov(X_1,X_2) = \frac{N_1N_2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{10}$$

from which it is seen that $Cov(X_1,X_2)$ is always positive. Similarly, the correlation between X_1 and X_2 is

$$Corr(X_1,X_2) = \sqrt{\frac{N_1N_2}{(N_1+\alpha+\beta)(N_2+\alpha+\beta)}}. \tag{11}$$

When X_1 and X_2 have the same index, say N, the correlation coefficient is particularly simple: $Corr(X_1,X_2) = N/(N+\alpha+\beta)$.

On the basis of (9), the distribution of the conditional variate $X_2|(X_1 = x_1)$, i.e., the number of drinking days in the second period given that x_1 days in the first period have been observed, is obtained as being also beta-binomial

$$X_2|(X_1 = x_1) \sim BB(N_2, \alpha + x_1, \beta + N_1 - x_1). \tag{12}$$

From (12) and (4) it follows that the expected number of drinking days in the second period, for the persons having had $X_1 = x_1$ drinking days in the first period is

$$E(X_2|X_1 = x_1) = \frac{N_2(\alpha + x_1)}{\alpha+\beta+N_1} = \frac{N_2\alpha}{\alpha+\beta+N_1} + \frac{N_2}{\alpha+\beta+N_1}\,x_1 \qquad x_1 = 0, 1 \ldots N_1. \tag{13}$$

Note that in (13) the regression is linear with respect to x_1.

3.3. *Estimation of beta-binomial parameters*

Maximum likelihood estimation of the parameters of the univariate beta-binomial distribution is described in the literature. The log likelihood and the likelihood equations are given by, e.g., Griffiths (1973) and will not be repeated here. Asymptotic variances and covariances of the ML-estimates are readily obtained from the Fisher information matrix of the beta-binomial. The ML-estimation of the bivariate beta-binomial involves only a slight and straightforward modification of the univariate case. In the sequel we assume that the maximum likelihood estimates $\widehat{\alpha}$ and $\widehat{\beta}$ together with their asymptotic variance estimates $\widehat{var}(\widehat{\alpha})$, $\widehat{var}(\widehat{\beta})$, and their asymptotic covariance estimate $\widehat{Cov}(\widehat{\alpha},\widehat{\beta})$ have been obtained.

4. Differences Between Drinking Frequency Measures

Assume now that the diary records of the respondents are the gold standard with which to compare statistically the corresponding records obtained by the retrospective recall or the typical frequency measure. It is natural to base the statistical comparison on the null hypothesis that the three ways of measuring drinking frequency are equivalent. Thus, we will use the following null model in the comparison: The beta-binomial model holds for the diary records and the derived properties of the model hold also for the other two drinking frequency measures. For instance, under the null model assumption, the retrospective recall records follow the beta-binomial distribution and the typical frequency records follow the beta distribution with the same parameters α and β as in the diary beta-binomial.

4.1. Comparing the retrospective recall measure with the diary

Denote by X the random variate describing the number of drinking days out of N_1 days as recorded by the diary and let Y denote the random variate for the number of drinking days out of N_2 (non-overlapping) days recorded by the same respondents with the retrospective recall method. Assume that $X \sim BB(N_1, \alpha, \beta)$. Under the null model also $Y \sim BB(N_2, \alpha, \beta)$ and further, (Y, X) will have the bivariate beta-binomial distribution given above in (9).

There are numerous ways of using the beta-binomial null model and we will note only a few of them. First, the observed marginal distribution of Y can be compared with the predicted marginal $\widehat{Y} \sim BB(N_2, \widehat{\alpha}, \widehat{\beta})$ where $\widehat{\alpha}$ and $\widehat{\beta}$ have been estimated from the observed distribution of X. Under the null model, the distributions of Y and $\widehat{Y}$ should not differ significantly and neither should their means (or higher moments). Yet another possibility is to use (13), i.e., that the regression of Y on X is linear and of the form $E(Y|X = x) = \beta_0 + \beta_1 x$ where $\beta_0 = (N_2\alpha)/(\alpha + \beta + N_1)$ and $\beta_1 = N_2/(\alpha + \beta + N_1)$. Substituting $\widehat{\alpha}$ and $\widehat{\beta}$ as estimated from the observed distribution of X, the null hypothesis can be represented as a fully specified linear hypothesis. Under the assumption of normally distributed errors (an assumption which in the present context may be rather crude) this leads to a standard F-test.

4.1.1. A local comparison based on prediction intervals

In the case that the marginal distribution of Y is shown empirically to differ significantly from $\widehat{Y}$, we are interested in locating the respondent categories causing the difference. For that purpose, denote by n_x the number of respondents having recorded x drinking days in their diaries, $x = 0, 1, \ldots, N_1$ and forming thus a subsample S_x of size n_x. The entire sample of respondents is thus partitioned into $N_1 + 1$ classes by the diary records, i.e., $S = \cup_{x=0}^{N_1} S_x$. Each of the n_x respondents belonging to subclass S_x also have a number of drinking days, y out of N_2 recorded by the retrospective recall. In the S_x-group, the expected value of number of drinking days in the retrospective recall is as given above in (13), and the distribution of Y under the null model is $Y|(X = x) \sim BB(N_2, \alpha + x, \beta + N_1 - x)$ as seen from (12). Both aspects are useful for a local comparison.

With $\widehat{\alpha}$ and $\widehat{\beta}$ estimated from the diary records of the entire sample S, the maximum likelihood estimate of the conditional expected value is

$$\widehat{E}(Y|X=x)=\widehat{\mu}_x=\frac{N_2(\widehat{\alpha}+x)}{\widehat{\alpha}+\widehat{\beta}+N_1}. \tag{14}$$

An estimate for the asymptotic variance of $\widehat{\mu}_x$, $\widehat{var}(\widehat{\mu}_x)$ can be derived by the standard δ-method (e.g., Bishop, Fienberg, and Holland 1975, pp. 486–502). Denote by $\bar{y}_x$ the subgroup average of the number of drinking days from the retrospective recall records in subgroup S_x. Under the null model $E(\bar{y}_x)=\mu_x$. Similarly, under the null model

$$var(\bar{y}_x)=\frac{1}{n_x}var(Y|X=x)=\frac{1}{n_x}\frac{N_2(\alpha+x)(\beta+N_1-x)(N_1+N_2+\alpha+\beta)}{(\alpha+\beta+N_1+1)(\alpha+\beta+N_1)^2}. \tag{15}$$

Substituting $\widehat{\alpha}$ and $\widehat{\beta}$ for α and β in (15) results in the maximum likelihood estimate $\widehat{var}(\bar{y}_x)$. Consider the deviation between $\bar{y}_x$ and $\widehat{\mu}_x$. For a reasonable n_x, $\widehat{\mu}_x-\bar{y}_x$ approaches the normal distribution with expected value 0 and variance estimated by $\widehat{var}(\widehat{\mu}_x)+\widehat{var}(\bar{y}_x)$. Thus the statistic

$$z=\frac{\widehat{\mu}_x-\bar{y}_x}{\sqrt{\widehat{var}(\widehat{\mu}_x)+\widehat{var}(\bar{y}_x)}} \tag{16}$$

is approximately $N(0,1)$ distributed and the corresponding asymptotic prediction interval for $\bar{y}_x$ is

$$\left[\widehat{\mu}_x-u_{1-\alpha/2}\cdot\sqrt{\widehat{var}(\bar{y}_x)+\widehat{var}(\widehat{\mu}_x)},\quad \widehat{\mu}_x+u_{1-\alpha/2}\cdot\sqrt{\widehat{var}(\bar{y}_x)+\widehat{var}(\widehat{\mu}_x)}\right] \tag{17}$$

where $u_{1-\alpha/2}$ refers to the $(1-\alpha/2)$-quantile of the standard normal distribution. The expression (17) is suitable for a quick screening of local deviances of the recall measure from the diary measure. If one wants to use prediction intervals for many S_x's simultaneously, the significance levels can be adjusted accordingly, for instance by using the Bonferroni inequality. In a simultaneous graphical comparison (see Figure 1) one compares a diary based linear prediction line and prediction intervals with the observed means of the retrospective recall measure.

4.1.2. A local comparison of conditional distributions

To supplement the prediction interval comparison, the empirical frequency distribution of retrospective recall records can be compared in the subgroup S_x with the distribution predicted by the null model. Plotting the empirical frequency distribution of Y restricted to the subgroup S_x and the predicted probability distribution $BB(N_2,\widehat{\alpha}+y,\widehat{\beta}+N_1-x)$ in the same graph is usually sufficient for a simple graphical comparison (see Figures 2 and 3). The χ^2 test of goodness of fit, for example, can be used to supplement the graphical comparison.

4.2. *Comparing the typical frequency measure with the diary*

The "summary" or typical frequency method is very often used in surveys and it is of interest also in the present context. The "typical frequency" question asks the respondent to place his/her drinking frequency on a classification scale with a limited number of classes. Lemmens et al. (1992) interpreted the answers to these questions as subjective estimates of the probability to drink on a randomly chosen day. In Table 1 we have transformed the questionnaire items to an appropriate probability-to-consume scale.

The last column of Table 1 contains M alternatives $\pi^{(j)}, j = 1, \ldots, M$ which we will interpret as direct observations of the propensity to consume. The fact that the $\pi^{(j)}$ are actually class midpoints will be ignored in the sequel. Denote by $p_{ij}, j = 1, \ldots, M$ that respondent i in the sample has assessed his/her score to be $\pi^{(j)}$. Assume again that X, the random variate representing diary records, is distributed $X \sim BB(N, \alpha, \beta)$. Under the null model, the propensity to consume, Π, is beta-distributed with parameters α and β. With ML-estimates $\widehat{\alpha}$ and $\widehat{\beta}$, the sample distribution of the p_{ij} over the entire sample S can be compared with the $Beta(\widehat{\alpha}, \widehat{\beta})$ distribution. A first comparison can be based on the average of the empirical distribution, $\overline{p}$, and the beta mean $\widehat{\alpha}/(\widehat{\alpha} + \widehat{\beta})$. A graphical comparison is obtained by plotting the histogram of the questionnaire records p_{ij} and the null model probability distribution $Beta(\widehat{\alpha}, \widehat{\beta})$ on the same scale.

4.2.1. A local prediction interval comparison

If the empirical distribution of the p_{ij} is shown to differ from $Beta(\widehat{\alpha}, \widehat{\beta})$ we will be interested in locating the respondent categories causing the difference. Under the null model the conditional distribution of Π has the distribution $\Pi|(X = x) \sim Beta(\alpha + x,\ \beta + N - x)$. Given ML-estimates $\widehat{\alpha}$ and $\widehat{\beta}$, and formula (7), the ML-estimate of the conditional mean of Π is given by

$$\widehat{E}(\Pi|X = x) = \widehat{\pi}_x = \frac{\widehat{\alpha}+x}{\widehat{\beta}+\widehat{\alpha}+N}. \tag{18}$$

The asymptotic variance estimate $\widehat{var}(\widehat{\pi}_x)$ of $\widehat{\pi}_x$ is obtained by the δ-method.

Again, let there be n_x respondents having recorded x drinking days in their diaries and forming thus a subsample S_x. Now use the scores p_{ij} of the n_x respondents in subclass S_x to calculate the subsample average $\overline{p}_x = (1/n_x)\Sigma_{i \in S_x} p_{ij}$. Obviously, under the null model $E(\overline{p}_x) = (\alpha + x)/(\beta + \alpha + N)$. Equally, under the null model, the variance of $\overline{p}_x$ is given by

$$var(\overline{p}_x) = \frac{1}{n_x} \cdot \frac{(\alpha + x)(\beta + N - x)}{(\alpha + \beta + N)^2(\alpha + \beta + N + 1)}. \tag{19}$$

Table 1. The typical frequency classification scale and the transformation from drinking frequency into probability to consume on a given day

Questionnaire Item	Lower class limit per half year	Upper class limit per half year	Lower score limit per day	Upper score limit per day	Probability score per day
1–2 times in 6 months	0	2.5	0.00	0.014	0.008
3–5 times in 6 months	2.5	5.5	0.014	0.03	0.02
1–3 times a month	5.5	22	0.03	0.12	0.07
1–2 times a week	22	65	0.12	0.36	0.21
3–4 times a week	65	117	0.36	0.64	0.5
5–6 times a week	117	169	0.64	0.93	0.78
Every day	169	182	0.93	1.0	0.965

Table 2. Distribution of 7-day prospective diary drinking-days of Dutch males and females fitted by the beta-binomial distribution. Drinking days data from the 1985 general population survey in the Netherlands

Number of drinking days	Observed frequency: All	Expected beta-binomial frequency: All	Observed frequency: males	Expected beta-binomial frequency: males	Observed frequency: females	Expected beta-binomial frequency: females
0	114	116.3	50	48.2	64	67.8
1	87	86.4	39	42.0	48	43.6
2	85	78.4	40	41.7	45	36.8
3	77	76.3	44	42.9	33	33.7
4	77	77.9	50	45.9	27	32.5
5	81	83.8	50	51.5	31	32.6
6	93	98.0	63	63.2	30	34.7
7	153	149.9	106	106.2	47	43.2
Totals:	767	767	442	442	325	325
G^2	–	1.03	–	0.86	–	4.45
d.f	–	5	–	5	–	5
$\Pr(G^2)$	–	0.96	–	0.97	–	0.48
ML estimates		$\hat{\alpha} = 0.702(0.050)$ $\hat{\beta} = 0.625(0.044)$		$\hat{\alpha} = 0.823(0.079)$ $\hat{\beta} = 0.581(0.054)$		$\hat{\alpha} = 0.621(0.067)$ $\hat{\beta} = 0.761(0.084)$

Denote by $\widehat{var}(\overline{p}_x)$ the corresponding maximum likelihood estimate. Asymptotically, $\widehat{\pi}_x - \overline{p}_x$ is under the null model normally distributed with expectation 0 and variance estimated by $\widehat{var}(\widehat{\pi}_x) + \widehat{var}(\overline{p}_x)$. A z-statistic which can be used for constructing prediction intervals is essentially the same as in (16)

$$z = \frac{\widehat{\pi}_x - \overline{p}_x}{\sqrt{\widehat{var}(\widehat{\pi}_x) + \widehat{var}(\overline{p}_x)}}. \tag{20}$$

The prediction intervals and their use are analogous to (17).

4.2.2. A local comparison of conditional distributions

To supplement the local prediction interval comparison, the empirical distribution of typical frequency records can be compared in the subgroup S_x with the distribution predicted by the null model. Plotting the histogram of the questionnaire records $p_{ij}, i \in S_x$ and the predicted probability density $Beta(\widehat{\alpha} + x, \widehat{\beta} + N - x)$ on the same scale is usually sufficient for a graphical comparison. For an example see Figures 5 and 6. The χ^2 test of goodness of fit can be used to supplement the graphical comparison.

5. Results

5.1. Validation of the beta-binomial model

Our estimation of the drinking days distributions was performed in two stages (for the univariate cases). In practice, ML-estimation requires good starting values and the moment estimates for α and β were used. The ML-estimates were calculated using the

Table 3. Distributions from 14-day prospective diary drinking-days fitted by the beta-binomial distributions. Drinking days data are from the 1983 general population survey in the Netherlands (both females and males) and the two diary weeks are presented separately

Number of drinking days	Observed frequency: Diary week 1	Expected beta-binomial frequency: week 1	Observed frequency: diary week 2	Expected beta-binomial frequency: week 2
0	47	54.6	42	47.9
1	54	42.0	47	42.9
2	43	38.9	54	41.9
3	40	38.5	40	42.5
4	40	40.1	49	44.3
5	41	44.0	40	47.8
6	39	53.1	43	54.9
7	95	87.8	84	76.7
Totals:	399	399	399	399
G^2	–	9.6	–	9.7
d.f	–	5	–	5
$\Pr(G^2)$	–	0.086	–	0.082
ML estimates		$\widehat{\alpha} = 0.722(0.072)$ $\widehat{\beta} = 0.581(0.057)$		$\widehat{\alpha} = 0.857(0.086)$ $\widehat{\beta} = 0.700(0.069)$

Gauss–Newton method with the moment estimates as starting values and with constraints $\alpha > 0, \beta > 0$. The diary distributions of drinking days are presented below by comparing the fitted beta-binomial distributions with the observed ones. Only the final ML-estimates are reported in the tables. The G^2 goodness of fit measure, appropriate for maximum likelihood estimates (Bishop et al. 1975, p. 125) is used throughout the tables.

As seen from Table 2, the beta-binomial model provides an excellent fit to the univariate distributions of the Dutch 1985 diary data. With the data presented in Table 2, also the proportions of between individuals and within individuals variance as given in (5) were calculated. For the data of Table 2 the proportion of the between individuals variance of the total variance was about 94%. One should note that the ratio of the between individuals variance to the within individuals variance is a linear function of N, so even for a relatively modest N, the between individuals part of the variance dominates.

As shown by the G^2 statistics in Table 3, the beta-binomial model cannot be rejected at the 5% level for the 1983 data sets although the fit is worse than that of 1985. Similar results are obtained when females and males are treated separately (result not shown).

Table 4. Bivariate distribution of drinking days (both females and males) from the 1983 Netherlands diary data fitted by a bivariate beta-binomial distribution. In each cell, the upper number is the observed frequency and the lower is the bivariate beta-binomial expected frequency. The ML-estimates of the bivariate beta-binomial parameters are $\hat{\alpha} = 0.762$ (0.065) and $\hat{\beta} = 0.608$ (0.051), $G^2 = 32.7$ with 27 degrees of freedom. ($Pr(G^2_{27} > 32.7) = 0.20$). The cell grouping for the G^2 test statistic is available from the authors on request

	0	1	2	3	4	5	6	7	
0	26	12	6	3	–	–	–	–	47
	31.6	12.4	5.2	2.0	0.7	0.2	0.1	0.0	52.2
1	13	16	15	5	4	1	–	–	54
	12.4	12.1	8.6	5.1	2.5	1.0	0.3	0.1	42.0
2	3	10	15	4	7	3	1	–	43
	5.2	8.6	9.2	7.7	5.1	2.7	1.1	0.2	39.6
3	–	7	10	9	9	3	1	1	40
	2.0	5.1	7.7	8.4	7.5	5.3	2.8	0.8	39.5
4	–	1	6	13	9	6	3	2	40
	0.7	2.5	5.1	7.5	8.8	8.3	5.8	2.5	41.2
5	–	–	1	2	13	11	9	5	41
	0.2	1.0	2.7	5.3	8.3	10.5	10.4	6.7	45.1
6	–	1	–	4	3	11	9	11	39
	0.1	0.3	1.1	2.8	5.8	10.4	15.7	17.8	53.8
7	–	–	1	–	4	5	20	65	95
	0.0	0.1	0.2	0.8	2.5	6.7	17.8	57.3	85.4
	42	47	54	40	49	40	43	84	399
	52.2	42.0	39.6	39.5	41.2	45.1	53.8	85.4	

The bivariate beta-binomial distribution was also fitted. This was done for the 1983 two-week diary, considered as a bivariate distribution of one-week drinking days in Table 4. The table also shows the week to week variation in the observed number of drinking days. Unfortunately, similar data are not available from 1985.

By Table 4, also the fit to the bivariate distribution remains good although the margins are far from perfect. A similar conclusion was reached when males and females were fitted separately (result not shown).

We may conclude that the beta-binomial models proposed for diary records cannot be rejected by the data. On the contrary, they appear to provide a fairly accurate description of the diary observations.

5.2. *Comparison results*

Due to lack of space, a selection of the comparison possibilities introduced in Section 4 are reported here. Because most overall comparisons (although not based on the beta-binomial statistical modelling) have been reported by Lemmens et al. (1992) and Lemmens et al. (1988), we concentrate on the conditional (groupwise) comparisons.

5.2.1. Comparison of the retrospective recall measure with the diary

The sample averages of drinking frequencies for different measures were compared by Lemmens et al. (1992). They showed that in the 1985 Dutch survey the average drinking frequencies obtained through the diary are significantly higher than average drinking frequencies obtained from the seven-day recall for both men and women. Our model-based comparisons show further that the observed distributions are skewed to the left (results not shown).

A similar result is obtained by comparing the average frequencies from the diary and the seven-day recall in the Dutch 1983 survey (Lemmens et al. 1988).

Because the comparisons based on the entire sample S showed systematic differences we turn to screening the respondent categories S_x. Table 5 shows the results in the local comparison between the diary and the seven-day recall. In Figure 1 the same information is conveyed graphically by using the 95% Bonferroni-adjusted prediction limit approach.

Table 5. Predicted and observed average numbers of drinking days from the Dutch 1985 study. Predicted values are predictions from the one-week diary, fitted with the beta-binomial model (see Table 2 above). Observed values are averages of the observed numbers of drinking days in the retrospective seven-day recall measure. The z-statistic is calculated as in (16)

Number of drinking days in diary	Number of respondents	Predicted number of drinking days from the diary	Observed average number of drinking days in 7-day recall	z-statistic
0	114	0.59	0.61	−0.16
1	87	1.43	1.13	2.36
2	85	2.27	1.58	4.40
3	77	3.11	2.45	3.61
4	77	3.95	3.08	4.35
5	81	4.79	4.07	3.35
6	93	5.63	5.24	1.76
7	153	6.47	6.15	1.67

It is clear from Table 5 and Figure 1 that the discrepancies are concentrated in the intermediate drinking frequencies as measured by the diary. Judged by prediction limits in Figure 1, the observed retrospective recall means are outside the prediction limits for days between 2 and 5, forming a seemingly smooth (perhaps quadratic) pattern of deviances.

As an illustration of the observed differences between distributions, Figures 2 and 3 show the predicted and observed distributions of drinking days for $x = 2$ and $x = 7$, respectively. The discrepancy in distributions is clearly observable in Figure 2 and the lack of it in Figure 3.

The beta-binomial model was also fitted to the 14-day diary of 1983 and the observed average seven-day recall results were compared with predictions from the 14-day diary. Here the discrepancies were also concentrated in the intermediate drinking frequencies (result not shown).

5.2.2. Comparison of the typical frequency measure with the diary

For an overall or marginal comparison we refer again to Lemmens et al. (1992) who showed that the average frequency obtained by the diary was significantly higher than that obtained by the typical frequency questionnaire.

Given the above overall difference, we turn again to the diary respondent classes. The results are given in Table 6 and Figure 4.

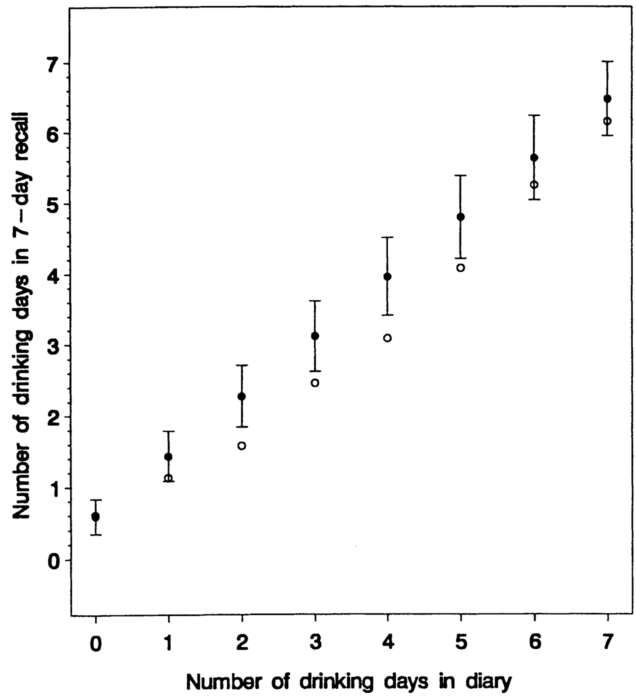

Fig. 1. Predicted (dots) and observed (circles) denote numbers of drinking days in the seven-day recall by the number of observed days in the 1985 diary. Dots are predicted numbers of drinking days on the basis of the 1985 one-week diary. Circles are averages of the corresponding numbers of drinking days in the retrospective seven-day recall. The 95% Bonferroni prediction limits for observed means are drawn around the predicted means

Table 6 and Figure 4 show that the pattern of discrepancies between the predicted and observed π's is very similar to that of the retrospective recall method. Here the days between 2 and 6 are clearly outside the prediction limits.

As an illustration of the observed differences between distributions, Figures 5 and 6 show the predicted and observed distributions of drinking propensity for $x = 0$ and $x = 6$, respectively. The similarity between observed and predicted distributions is reasonable in Figure 5 while in Figure 6 the observed distribution is clearly to the left of the predicted one.

6. Discussion

6.1. On the interpretation of the results

According to Sudman and Bradburn (1982) diaries are used for assessment of "frequent, nonsalient events that are difficult to recall accurately." They are to be used when it is important to limit "reliance on recall" (p. 47) and thus limit the time period. Silberstein and Scott (1991) review the literature on expenditure surveys using diaries and conclude that "Recall errors ... are produced by partial recall of past events, increasing with longer reference periods and less salient events."

Our results show a method effect in inquiring about consumption frequency. Summing

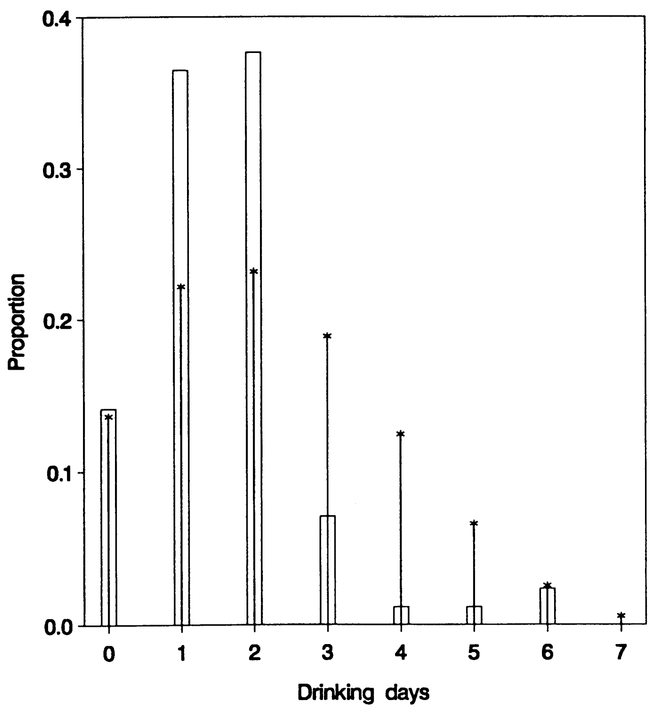

Fig. 2. Predicted and observed conditional distributions of the number of drinking days in 1985 given two drinking days in the diary. The predicted distribution (spikes) is the predicted beta-binomial from the 1985 one-week diary for the subgroup with two drinking days in the diary. The observed distribution (bars) represents the number of drinking days in the same subgroup as measured by the seven-day retrospective recall

up, it seems that response discrepancies depend on the drinking frequency of the respondent. By contrasting diaries with frequencies from the retrospective recall method, it was shown that recall errors occur mostly with respondents drinking at intermediate frequencies. Thus it seems we are dealing with partial recall of nonsalient events: For most Dutch respondents, drinking occasions are fairly frequent and nonsalient. Quite possibly, stability of drinking behaviour over time is a factor influencing recall, with daily and very rare drinking events being easiest to reconstruct accurately. As the prospective diary is kept daily the crucial difference between the diary and the retrospective recall method is that in the diary the recall reference period is limited to 24 hours. For these reasons we suggest that faulty memory is the main explanation for the method differences found.

Comparison of typical frequency scores with the diary showed again that the most pronounced assessment errors occur with respondents drinking at intermediate frequencies. The pattern of discrepancies is surprisingly similar to that of the retrospective recall result. In the comparison between typical frequency and diary, things are admittedly conceptually more complex and cognitive processes other than memory are also involved.

As an alternative to the cognitive interpretation social desirability might be suggested. Dillman and Tarnai (1991) report on social desirability, and refer to Bradburn (1983): A more anonymous method of question administration appears to work somewhat better by lowering the degree of underreporting or overreporting. The diary can surely be seen as

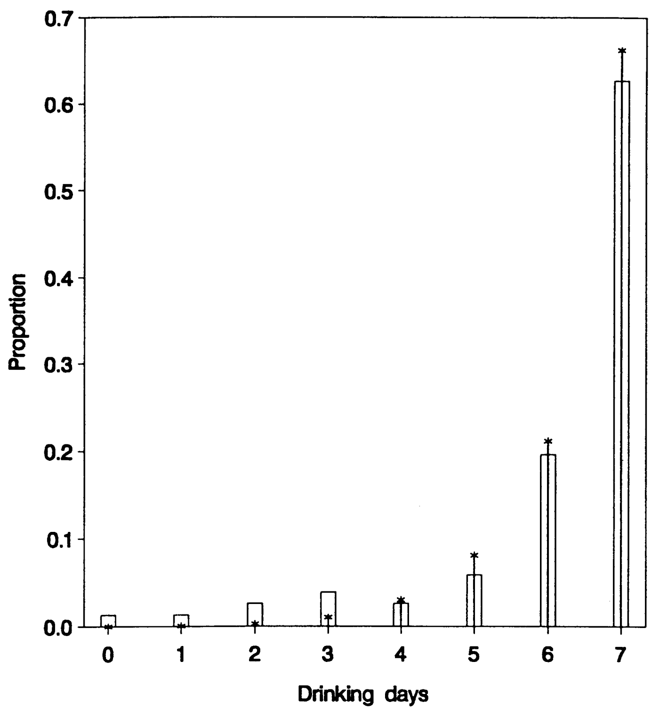

Fig. 3. Predicted and observed conditional distributions of the number of drinking days in 1985 given seven drinking days in the diary. The predicted distribution (spikes) is the predicted beta-binomial from the 1985 one-week diary for the subgroup with seven drinking days in the diary. The observed distribution (bars) represents the number of drinking days in the same subgroup as measured by the seven-day retrospective recall

Table 6. Predicted and observed mean values of the typical frequency measure on the probability score scale in the Dutch 1985 study. Predicted values of π are predictions from the one-week diary, fitted with the beta–binomial model, for each class. Observed values are averages of the observed typical frequency probability scores. The z-statistic is calculated as in (20)

Number of drinking days in diary	Number of respondents	Predicted π from the diary	Observed average π	z-statistic
0	113	0.084	0.091	−0.36
1	86	0.204	0.133	2.46
2	83	0.324	0.189	4.04
3	77	0.444	0.282	4.61
4	76	0.565	0.387	5.04
5	80	0.685	0.508	5.32
6	93	0.805	0.694	3.90
7	150	0.925	0.873	2.63

more anonymous than the interview. The differences between the methods in the present article could therefore be mode differences caused by social desirability reactions. However, it is not the most likely explanation – why would respondents at both ends of the frequency scale have weaker reactions to social desirability?

Although the diary coverage of aggregate sales and tax figures (an external measure of consumption) is relatively high, the diary data necessarily include recall errors and other

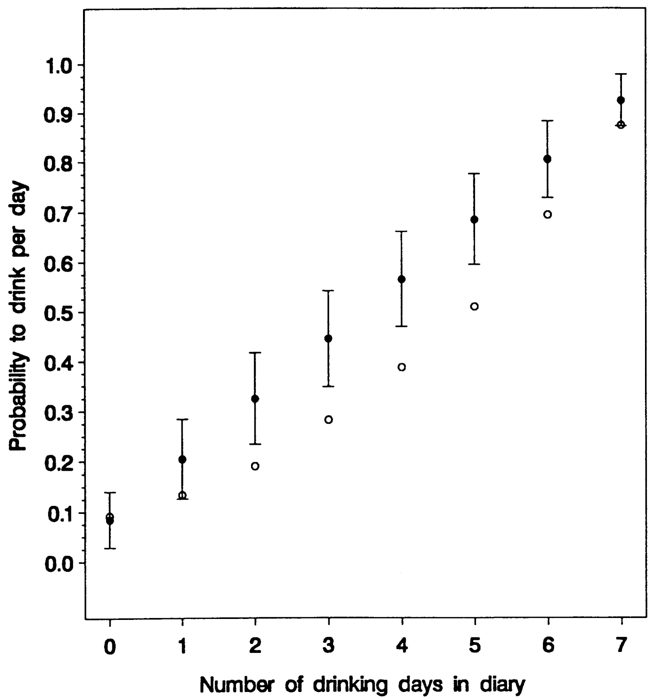

Fig. 4. Predicted (dots) and observed (circles) average probabilities to consume by the number of days in the 1985 diary. Dots are predicted values of probability to consume per day on the basis of the 1985 one-week diary. Circles are averages of corresponding typical frequency records. The 95% Bonferroni prediction limits for observed means are drawn around the predicted means

measurement effects of various sorts. In the absence of a better gold standard, we have chosen the diary scores as the best approximations of the drinking frequency and based the modelling approach on the diary. Thus the results given are relative to the consumption frequency scores obtained by the diary method.

The ultimate goal in method effect studies is, however, to treat the discrepancy between "true" consumption and that measured by instruments prone to be affected by response errors. Large errors in diary reports may of course affect the beta-binomial drinking model. Strong assumptions are required for generalising from the diary to the "true" drinking frequency. One possibility is opened by assuming that the recall error mechanisms observed here follow a similar pattern as those between true frequency and the diary. Evidently no empirical basis is at present, if ever, available for testing this generalisation in general population surveys. We are left with assumptions that are merely given a degree of plausibility by the present results.

6.2. *Modelling issues*

The days of the week are, of course, not uniform with regard to the probability of a drinking event to occur. It can be argued that drinking behaviour in general populations conforms to a weekly cycle. Proportionally, drinking occurs more frequently on weekends (67% of all drinkers in diary 1985 report drinking on Saturday; 61% on Sunday; 59% on Friday; normal weekdays between 42% and 49%). It is thus likely that there is day to day

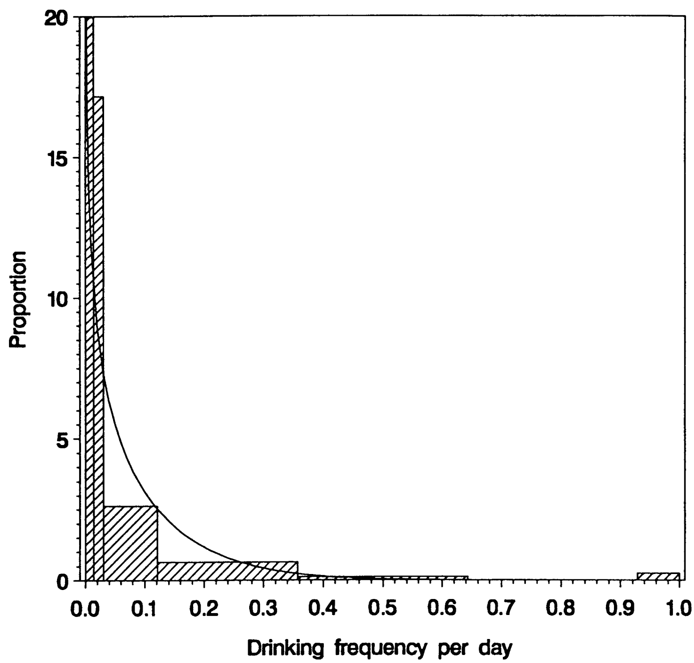

Fig. 5. Predicted and observed conditional distributions of the propensity to consume in 1985, given zero drinking days in the diary. The predicted distribution (continuous density) is the predicted beta density from the 1985 one-week diary for the subgroup with zero drinking days in the diary. The observed distribution (histogram) is from the questionnaire records in the same subgroup

heterogeneity of π at the individual level, contrary to the assumption underlying (1). The day to day heterogeneity would in theory be expected to affect the individual level variance but it is not possible to assess the extent of this empirically. Similarly there appears to be no way to study the possible effect of dependencies between drinking on successive days on the basis of the present data. We must be content to assume that the effects of potential deviations from the binomial assumption are not very prononounced in practice.

The amounts consumed have not been considered here. Lemmens et al. (1992) noted no significant differences in the mean amounts per day between the measures in the 1985 data. We conjecture that on average, amounts play a minimal role in drinking frequency response effects.

Finally, the present results for drinking frequency might be useful in an attempt to model response errors as a function of the gold standard frequency. The pattern of differences between the diary prediction and the seven-day retrospective recall and the typical frequency averages suggest a quadratic model for the effect of recall/assessment error as a function of diary frequency.

6.3. A note on implications

What can be said about the effect of the results for social and medical epidemiological studies using any of these methods, given that the above generalisation to true drinking frequency is assumed?

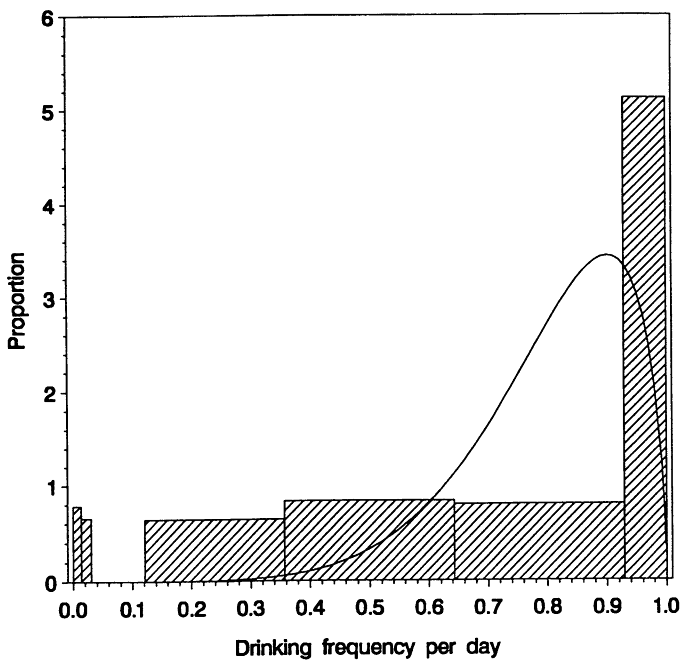

Fig. 6. Predicted and observed conditional distributions of the propensity to consume in 1985, given six drinking days in the diary. The predicted distribution (continuous density) is the predicted beta density from the 1985 one-week diary for the subgroup with six drinking days in the diary. The observed distribution (histogram) is from the questionnaire records in the same subgroup

The extra drinking hidden by response effects is not equally distributed over all drinker categories. With retrospective recall of actual consumption, an adjustment would have to be made in the Dutch data for the intermediate categories, with values between two and five drinking days per week. The effect on assessed risk (for medical outcome or other) will depend on the particular risk function. The fact that the people drinking at the six– or seven-day per week level are relatively correct in their reports implies that shape of the curve is affected by the use of a seven-day retrospective recall. For the typical frequency method (most often used in studies linking consumption to medical outcomes), similar adjustments are necessary. This implies that in the Dutch population risk thresholds at the frequency levels in question are in fact higher than estimated by the present measures. A systematic exploration of the effect of response errors as assessed here on several empirical risk curves would be the logical next step.

7. References

Alanko, T. (1984). An Overview of Techniques and Problems in the Measurement of Alcohol Consumption. In R. Smart, H. Capell, F. Glaser, Y. Israel, H. Kalant, R. Popham, W. Schmidt and E. Sellers (eds.), Research Advances in Alcohol and Drug Problems, 8, 209–226. New York: Plenum Press.

Alanko, T. (1985). Effects of Time-Variation of Drinking on the Validity of Consumption Surveys: An Empirical Investigation Using Finnish Drinking Rhythm Data. Paper presented at the ICAA/Epidemiology Section meeting in Rome, June 3–7, (unpublished).

Alanko, T. and Duffy, J.C. (1996). Compound Binomial Distributions for Modelling Consumption Data. The Statistician, 45, 269–286.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambrigde, Massachusetts: MIT Press.

Bradburn, N.M. (1983). Response Effects. In P.H. Rossi, J.D. Wright, and A.B. Anderson, (eds.), Handbook of Survey Research. New York: Academic Press.

Duffy, J.C. (1985). Questionnaire Measurement of Drinking Behaviour in Sample Surveys. Journal of Official Statistics, 1, 229–234.

Duffy, J.C. and Alanko, T. (1992). Self-Reported Consumption Measures in Sample Surveys: A Simulation Study of Alcohol Consumption. Journal of Official Statistics, 8, 327–350.

Chatfield, C. and Goodhardt, G.J. (1970). The Beta-Binomial Model for Consumer Purchasing Behaviour. Applied Statistics, 19, 240–250.

Dillman, D.A. and Tarnai, J. (1991). Mode Effects of Cognitively Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds.), Measurement Error in Surveys, 73–94. New York: John Wiley and Sons.

Ekholm A. (1968). A Study of the Drinking Rhythm of Finnish Males. 28th International Congress on Alcohol and Alcoholism, September 15–20, Washington D.C.

Griffiths, D.A. (1973). Maximum Likelihood Estimation for the Beta-Binomial Distribution and an Application to the Household Distribution of the Total Number of Cases of Disease. Biometrics, 29, 637–648.

Hald, A. (1968). The Mixed Binomial Distribution and the Posterior Distribution of *p* for a

Continuous Prior Distribution. Journal of the Royal Statistical Society, Series B, 30, 358–367.
Johnson, N.L., Kotz, S., and Kemp, A.W. (1992). Univariate Discrete Distributions, 2nd edition. New York: John Wiley and Sons.
Johnson, N.L. and Kotz, S. (1970). Continuous Univariate Distributions – 2. New York: John Wiley and Sons.
Lemmens, P. (1991). Measurement and Distribution of Alcohol Consumption. Maastricht: University of Limburg.
Lemmens, P., Knibbe, R., and Tan, F. (1988). Bias Due to Nonresponse in a Dutch Survey on Alcohol Consumption. British Journal of Addiction, 83, 1069–1077.
Lemmens, P.H., Tan, F.E., and Knibbe, R. (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.
Mäkelä, K. (1971). Measuring the Consumption of Alcohol in the 1968-1969 Alcohol Consumption Study. Helsinki: Social Research Institute of Alcohol Studies.
Philipsen, H., Knibbe, R.A., and van Reek, J. (1983). Alcohol Consumption in The Netherlands as a Social Phenomenon. In R.J.J. Hermus (ed.), Alcohol, Health and Society, 53–70 (Zeist, TNO).
Poikolainen, K. and Kärkkäinen, P. (1983). Diary Gives More Accurate Information about Alcohol Consumption than Questionnaire. Drug and Alcohol Dependence, 11, 209–216.
Poikolainen, K. (1985). Measurement of Recalled Alcohol Intake in Relation to Actual Consumption. British Journal of Addiction, 80, 215–216.
Sichel, H.S. (1982). Repeat-Buying and the Generalized Inverse Gaussian-Poisson Distribution. Applied Statistics, 31, 193–204.
Silberstein, A.R. and Scott, S. (1991). Expenditure Surveys and Their Associated Errors. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds.), Measurement Error in Surveys (pp. 303–326), New York: John Wiley and Sons.
Sobell, M.B., Bogardis, J., Schuller, R., Leo, G.I., and Sobell, L.C. (1989). Is Self-Monitoring of Alcohol Consumption Reactive? Behavioral Assessment, 11, 447–458.
Sudman, S. and Bradburn, N.M. (1982). Asking Questions: A Practical Guide to Questionnaire Design. San Francisco: Jossey-Bass.
Williams, D. (1975). The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity. Biometrics, 31, 949–952.
Wilson, P. (1981). Improving the Reliability of Drinking Surveys. The Statistician, 30, 159–167.

Received July 1994
Revised September 1996