

Measurement and distribution of alcohol consumption

Citation for published version (APA):

Lemmens, P. H. H. M. (1991). *Measurement and distribution of alcohol consumption*. [Doctoral Thesis, Maastricht University]. Rijksuniversiteit Limburg. <https://doi.org/10.26481/dis.19910208pl>

Document status and date:

Published: 01/01/1991

DOI:

[10.26481/dis.19910208pl](https://doi.org/10.26481/dis.19910208pl)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Measurement and Distribution of Alcohol Consumption

18 2

Measurement and Distribution of Alcohol Consumption

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Rijksuniversiteit Limburg te Maastricht,
op gezag van de Rector Magnificus, Prof. Mr. M.J. Cohen,
volgens het besluit van het College van Dekanen,
in het openbaar te verdedigen
op vrijdag, 8 februari 1991 om 16.00 uur

door

Paulus Hermanus Hubertus Marie Lemmens

geboren te Beek in 1955

Promotor:

Mevr. Prof. Dr. M.J. Drop

Co-promotor:

Dr. R.A. Knibbe

Beoordelingscommissie:

Prof. Dr. R.J.J. Hermus (voorzitter)

Prof. Dr. P.G. Knipschild

Prof. Dr. Ir. D. Kromhout (RIVM, Bilthoven)

Prof. Dr. P.J. van der Maas (Erasmus Universiteit Rotterdam)

Prof. Dr. Ir. W.H.M. Saris

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Lemmens, Paulus Hermanus Hubertus Marie

Measurement and distribution of alcohol consumption /

Paulus Hermanus Hubertus Marie Lemmens. -[S.l. : s.n.]. - I11.

Proefschrift Maastricht. -Met lit. opg. - Met samenvatting in het Nederlands.

ISBN 90-9003877-9

SISO 614.72 UDC 613.81.000.316(043.3) NUGI 652

trefw.: alcoholgebruik; medische sociologie.

Research has been supported by a grant (# 900.568.406) of the Netherlands Organization for Scientific Research (NWO) and the Foundation for the Moderate Use of Alcohol (STIVA).

Publication of this book has been supported by Stichting Alcohol Fonds and the Jan Dekkerstichting and the Ludgardine Bouwmanstichting.

text design

cover design

printed by

Drukwerkvormgeving Bert Lemmens Schimmert

Mieke Derickx en Brian Swan

Datawyse Maastricht / Krips Repro Meppel

¿LE GUSTA ESTE JARDIN?

¿QUE ES SUYO?

¡EVITE QUE SUS HIJOS LO DESTRUYAN!

The consul stared back at the black words on the sign without moving. You like this garden? Why is it yours? We evict those who destroy! Simple words, simple and terrible words, words which one took to the bottom of one's being, words which, perhaps a final judgement on one, were nevertheless unproductive of any emotion whatsoever, unless a kind of colourless cold, a white agony, an agony chill as that iced mescal drunk in the Hotel Canada on the morning of Yvonne's departure.

UNDER THE VOLCANO, Malcolm Lowry.

dedicated to my parents

VOORWOORD

Een voorwoord bij een proefschrift bestaat meestal uit een opsomming van personen aan wie je het allemaal te danken hebt. Nu, het eerste wat mij in gedachten kwam toen ik me aan het schrijven van deze woorden zette, waren een paar regels uit Simon Carmiggelt's "Tekst voor een wijnkaart", uit een in de ramsj geplukt boekje "Vreugden en verschrikkingen van de dronkenschap". Het begint aldus: "O, drank, je hebt zoveel verpest./ Toch ben je in mijn dorstig leven/ altijd die ene hoer gebleven,/ die mij het diepste heeft gelest,". Ik geef toe dat het wat overtrokken overkomt, maar toch lijkt me de geschetste ambiguiteit tegenover de drank hier op zijn plaats: ook ik ben veel aan de drank verschuldigd. Maar wellicht moet nog meer water bij de wijn, want het zijn niet zozeer de drank en zijn verpesting, alswel de drinkers die ik hier zou moeten bedanken. Het drinken van velen en het gezuip van die enkeling vormen namelijk de grondslag van dit boek.

Het proefschrift komt voort uit en is een afronding van, wat men is gaan noemen, het Ledermann-project. Als je naar de auteurs van de artikelen kijkt waaruit dit boek is opgebouwd wordt duidelijk dat het niet het werk van mij alleen is. De basis voor het Ledermann-project is gelegd door Ronald Knibbe. Hij heeft destijds het stramien bepaald waarin ik de afgelopen 5 jaar heb gewerkt. Als ik wel eens niet meer wist hoe nou verder, keerde ik altijd terug naar het begin, Ronald's projectvoorstel. Hij heeft ook de nodige fondsen geworven bij de Stiva en, samen met Riet Drop, bij ZWO. Bovendien komt Ronald de eer toe de eerste praktische voorbereidingen te hebben getroffen, zoals ontwerp van vragenlijst en de onderhandelingen met het Instituut voor Toegepaste Sociologie, dat het veldwerk heeft verricht. Als supervisor én kamergenoot in de eerste jaren heb ik Ronald leren respecteren door zijn inzet, inzicht en geduld in de vele discussies die we hebben gevoerd over dit onderwerp. Hij is een collega, in de ware zin des woords, waar ik goed en graag mee samenwerk. En dat ondanks dat hij mijn "slapie" was op menig buitenlandse studiereis, zo heb ik boze tongen horen beweren (het begint nu wel erg veel op een wielereportage te lijken).

Een tweede hoofdrolspeler in het Ledermann drama is Frans Tan, waar ik vanaf het begin van het project vele uren van intensieve discussie mee heb doorgebracht. Ook van hem heb ik veel geleerd. Zijn inbreng in het project is zeer effectief geweest. Zonder hem zou dit proefschrift er geheel anders hebben uitgezien, als het er al ooit van gekomen was. Op deze plaats verontschuldigd ik me nogmaals dat hij niet als auteur genoemd staat bij het artikel dat in Gezondheid en Samenleving is verschenen (zie hoofdstuk 1). Het is ter leniging van de grote schuld die ik bij hem heb dat ik hem altijd de overwinning gun bij het badminton. Een voorname plaats in het Ledermannse neemt Riet Drop in, promotor en deelproject-leider. Ik heb haar nooit als

een strenge beoordelaar van mijn werk ervaren maar altijd als gesprekspartner en raadgever. Hoewel we uit verschillende disciplines komen heeft dit nooit tot controversen geleid. Haar visie en ideeën over wetenschappelijk onderzoek bestrijken een groot gebied en zijn algemeen geldend, en haar autoriteit blijft geenszins beperkt tot het vakterrein. Als ik bedenken dat ze wel 35 versies heeft gelezen van artikelen, papers en hoofdstukken die aan dit boekje vooraf zijn gegaan, realiseer ik me dat frustratie-tolerantie en taai zitvlees de voornaamste eigenschappen van een hoogleraar moeten zijn. Ik bedank haar voor het vertrouwen dat ze van het begin in mij heeft gehad en, misschien wel net zo belangrijk, duidelijk heeft laten blijken. Ze heeft hiermee de voorwaarden geschapen waardoor ik altijd met plezier heb gewerkt. Van mijn andere collegae bij MedSoc wil ik vooral Renier noemen. De omgang met deze "jonge wilde" werkt zeer verfrissend en raad ik alle ouwe lulen aan. Ik meen veel gemeenschappelijks in ons beider "Weltanschauung" te hebben ontdekt. Ik hoop je daarmee nog vaak te mogen vervelen. Verder noem ik Jos Diederiks, die ik erkentelijk ben voor de moeite die hij zich getroost heeft om samen het modelletje in hoofdstuk 4 te ontvouwen en voor de vele grappen en grollen tussendoor. Voorts mag Bob Wilkinson (Vertalersopleiding) niet onvermeld blijven die een deel van de Engelse tekst heeft gecorrigeerd als onderdeel van de cursussen wetenschappelijk Engels. Lex Volovics en Marion de Leeuw hebben geassisteerd bij het uitvoeren van de multivariate analyse in hoofdstuk 6, waarvoor dank.

Het wordt persoonlijker als ik Roelien noem. Mijn beslissing om in Maastricht te gaan werken heeft ook haar leven sterk beïnvloed. Niet zonder spijt en pijn zijn we uit elkaar gegaan. Onze relatie was wellicht, in figuurlijke zin dan, een van de slachtoffers van den drankduivel, zo treffend uitgebeeld in het Postbus 51 spotje. Ik ben blij dat ze nog steeds mijn beste Amsterdams maatje is.

Persoonlijk is ook mijn dank aan Mieke Derickx, voor haar steun, vriendschap, en wat niet al. Met haar computerkennis en vaardigheden heeft ze een groot aandeel gehad in het wassen en gladstrijken van de ruwe data. Bij al dat ingewikkelde, maar ook saaie werk wist ze toch altijd de lach te laten overheersen, waarmee zelfs een dooie tot leven kan worden gewekt. Met haar directheid en enthousiasme heeft ze me vaak mijn eigen cynische spiegel voorgehouden.

Als laatste noem ik mijn broer en, vooral, mijn ouders, aan wie ik dit boek opdraag. Ze hebben zonder enig eigenbelang ontzettend veel voor me gedaan en gelaten. Gewillig hebben ze zich altijd veel van me laten welgevalen. Zij zijn voor mij de enige echte rotsen in het roerend bestaan.

Contents

- 1 The Distribution of Alcohol Consumption in the General Population. 1**
 - Summary 1
 - 1.1 Introduction 2
 - 1.2 Description of the Ledermann model 3
 - 1.3 Model estimates of prevalence of consumption 6
 - 1.4 Empirical studies and criticism 8
 - 1.5 The theory of the collectivity of drinking cultures 12
 - 1.6 A distribution-free approach 16
 - 1.7 Threats to validity of data on alcohol consumption 17
- 2 Bias due to Non-response in a Dutch Survey on Alcohol Consumption 25**
 - Summary 25
 - 2.1 Introduction 26
 - 2.2 Methods 28
 - 2.3 Results 30
 - 2.4 Discussion 37
- 3 Weekly Recall and Diary Estimates of Alcohol Consumption in a General Population Survey 41**
 - Summary 41
 - 3.1 Introduction 42
 - 3.2 Method 43
 - 3.3 Results 44
 - 3.4 Discussion 47
- 4 Measuring Quantity and Frequency of Drinking in a General Population Survey. A Comparison of 5 Indices. 51**
 - Summary 51
 - 4.1 Introduction 52
 - 4.2 Methods 53
 - 4.3 Results 56
 - 4.4 Discussion 66
- 5 Seasonal Variation in Survey and Sales Estimates of Alcohol Consumption 71**
 - Summary 71
 - 5.1 Introduction 72
 - 5.2 Methods 73
 - 5.3 Results 75
 - 5.4 Discussion 80

6	Agreement between Respondent and Partner on Reports of Drinking and Drinking Problems in a General Population Survey.	83
	Summary	83
6.1	Introduction	84
6.2	Methods	86
6.3	Results	89
6.4	Discussion	100
7	Comparing Distributions of Alcohol Consumption : Empirical Probability Plots.	105
	Summary	105
7.1	Introduction	106
7.2	Empirical Probability Plots	108
7.3	Data	110
7.4	Results	111
7.5	Discussion	114
8	Comparisons of Distributions of Male and Female Alcohol Consumption from Surveys in 5 Western Countries.	119
	Summary	119
8.1	Introduction	120
8.2	Data	121
8.3	Assessment of the relation between different types of methods	124
8.4	Comparisons of distributions from different countries	125
8.5	Longitudinal comparisons between distributions from the USA and the UK	129
8.6	Discussion	129
9	A recapitulation and extension.	131
	Summary	131
9.1	Introduction	132
9.2	Recapitulation	132
9.3	Region estimates of the a-and b-parameter of the linear shift function	143
9.4	Impact and limitations of the notion of a "single"- distribution of alcohol consumption for a primary prevention model	144
9.5	Longitudinal stability of drinking status	151
9.6	Final comment	152
	Samenvatting (summary in Dutch)	157
	Appendix 1.1 & 1.2	

The Distribution of Alcohol Consumption in the General Population.

BASED ON "DE VERDELING VAN ALCOHOLCONSUMPTIE IN DE BEVOLKING",
PUBLISHED IN GEZONDHEID EN SAMENLEVING, 8, PP. 23-36, 1987.
P.H.H.M. LEMMENS, R.A. KNIBBE & M.J. DROP.

Summary

According to the French epidemiologist S. Ledermann, the distribution of alcohol consumption in a homogeneous population is best described by a one-parameter lognormal distribution model. This model implies a fixed relationship between mean consumption and the prevalence of drinkers at certain levels of consumption. In this introductory chapter the theoretical and empirical foundations of the Ledermann-model are discussed, as well as its practical implications for an alcohol prevention policy. A review of the extensive literature that deals with the subject is given. Alternative approaches are discussed, such as the collectivity theory of O.-J. Skog, Ledermann's main critic. Several problems are noticed. The main difficulties are 1) inadequacy of statistical procedures to test distributional form, 2) considerable undercoverage of sales statistics with survey data, and, hence, doubts about the validity of empirical distribution data, 3) scarcity of longitudinal studies, and 4) the absence of a theoretical basis for the lognormal hypothesis. These criticisms have been the point of departure for the present study. In this introduction, an outline of the basic hypotheses and designs of the study is presented.

1.1 Introduction

The large increase in alcohol consumption in the Netherlands since the early fifties has raised public awareness of the negative effects such a development can have on society. Even though per capita consumption in the Netherlands has been fairly stable for the last 12-15 years (Produktschap voor Gedistilleerde Dranken, 1988), public concern about ways to control consumption of alcohol has grown particularly in the last decade. This has resulted, for example, in a new alcohol policy formulated by the Dutch government, along lines and goals set by the WHO. The renewed moderation spirit is accompanied by a nationwide campaign aimed at the general public, aimed at a reduction of consumption by all members of society. This generalistic policy is clearly inspired by, what has been named, the single-distribution model for the prevention of alcohol problems (see Bruun *et al.*, 1975). The model consists of arguments, hypotheses and empirical data which statistically link, at the aggregate level, per capita consumption to several harmful consequences of excessive alcohol use. The model does not pretend, as one might expect, to offer substantial evidence as to why people drink (too much) or what makes people increase their drinking at a certain time. So it is not a deterministic, social theory of drinking. Basic argument of the single distribution model is that a strong, relatively invariant relationship exists between excessive use and per capita consumption. This relationship determines the "*degrees of freedom*" of an effective alcohol control policy. In generalistic terms, the model asserts that without a reduction of consumption by all members of a community, one cannot expect alcohol abuse to decline in that community, and without a decline in abuse, there will be no decline in negative effects. In an extreme sense, the idea is reminiscent of pre-war prohibitionism. In contrast with prohibitionism, however, the model does not have a moral base but a scientific one.

The idea of a strong, fixed relation between mean and excessive use was first introduced by the French epidemiologist Sully Ledermann (1956), in an effort to explain regional variation in alcohol related mortality in France. In essence, he argued that the distribution of alcohol consumption in a homogeneous population could be described by a mathematically fixed, one-parameter lognormal distribution. A description of the literature in this chapter starts with, what has been named, the Ledermann-model. In the paragraphs that follow, Ledermann's mathematical model will be introduced, together with a critical review of the literature on this subject, an assessment of the shortcomings that have emerged over the years, and the implications these might have for the formulation of a primary prevention model. One of the problems that are noted, and, for that matter, a basic question in all scientific work not only in this area, concerns the methodology of empirical research, such as issues of reliability and validity of measurement. Special

attention in this thesis will be given to several possible threats to a valid estimation of alcohol consumption in the general population (ch2 thru 6). In two chapters changes in the distribution of consumption over time (ch7) and across several European countries (ch8) are studied with the use of the technique of probability plots.

1.2 Description of the Ledermann model

Ledermann's model of the distribution of alcohol consumption is built upon two hypotheses concerning the underlying (viz., lognormal) model (Ledermann, 1956). The first hypothesis states that, in homogeneous populations, the distribution of drinkers according to their average daily consumption is unimodal and positively skewed. Unimodality means that the distribution has only one peak (viz., the mode). Ledermann asserts that visual inspection of several empirical distributions did not reveal a second peak at the higher levels, and that there is no sharp distinction between normal, socially acceptable consumption levels and excessive levels. In addition, Ledermann argues that the distribution should be skewed because symmetry would imply a maximum consumption of twice the mean (since consumption can vary between 0 and MAX, mean consumption is $\frac{1}{2}$ MAX). Since, for instance, per capita consumption in countries in 1987 did not exceed 13 liter 100% alcohol (which is 3.5 glasses per day; corrected for a abstainers fraction of .15), it is clear that such a maximum is indeed not realistic.

In the second hypothesis Ledermann further argues that it is not consumption itself which is normally distributed but the logarithm of the consumption. Ledermann's choice of, particularly, a logarithmic transformation of the raw data does not become very clear, and other models are theoretically possible. He only mentions that it is not uncommon in the case the phenomenon develops according to a "boule de neige" or snowball mechanism. Ledermann uses the example of contagion (ibid, p. 126). In the case of drinking, contagion means that one could distinguish many, mostly, socio-cultural factors that determine a person's consumption. Skog (1980a) has added that lognormality can often be observed if there are many variables that each contribute (only) a small fraction to the resulting variate (see also footnote 2, ch7). Aitchison and Brown (1969) note that lognormal distributions are not uncommon for certain variables in economics (e.g. income) or in sociology (e.g. certain types of deviant behavior).

Ledermann proceeded by specifying the lognormal distribution. In particular, he assumed that it is a one-parameter lognormal distribution, one in which only the mean varies (and, in other words, in which there is a relationship between the mean and dispersion parameter). Usually, lognormal distributions are determined by the mean and the dispersion parameter. Populations that have identical means but which differ in dispersion will have different distributions and, hence, different rates of excessive use. According to Ledermann, however, the differences in dispersion between homogeneous populations with identical means are negligible. The justification for this stability are to be found in the following three assumptions, with which the relation between the two parameters are determined:

- a. there is not only a minimum to consumption (0 glasses) but also a physiological maximum, set at an average of 80 glasses per day, above which death would be imminent and the incidence of drinkers over this level is equal to the mortality in this category;
- b. the fraction of drinkers over 80 glasses per day was assumed equal for all populations, regardless its mean consumption;
- c. using 5 empirical studies, Ledermann estimated (by means of extrapolation) this fraction to be about 0.03% of the drinking population (1956).

The assumption of a limited range of consumption is appealing and seems plausible. The estimated maximum average daily consumption of 80 glasses, however, is quite extreme, at least for some sub-populations (for instance, there are good reasons to expect women to have a lower physiological maximum than men). One can, even more strongly, doubt the estimated fraction of 0.03% and its assumed constancy over populations.

With the reduction of the variation between the mean and the dispersion parameter, knowledge of the mean is sufficient for the calculation of the proportion of drinkers at the various consumption levels. In practice, the number of alcoholics or excessive users are estimated with per capita sales of alcoholic beverages which are often taken as estimates of mean consumption (per capita means per head of the D.A.P., the Drinking Age Population, consisting of all people over 15 years of age). It is a clear advantage over the time- and money-consuming general survey approach. Still, however, estimates of the number of abstainers are needed, data that can be obtained only by means of a survey.

Maybe even more impressive than estimates of excessive use as such are the expected changes in rates of excessive use the model predicts when per capita consumption rises. Figure 1.1 shows the non-linear change in the proportion of drinkers at two heavy drinking levels (an average over 8 and 16 glasses per day) with increasing mean consumption. A doubling of mean consumption from, for instance, 5 to 10 liters pure alcohol corresponds with an increase of 150% of the fraction drinking more than 8 glasses per day and

an increase of 100% in the fraction drinking over 16 glasses per day. Another striking effect is shown in figure 1.2., which depicts low mean consumption levels. A change in mean consumption from 2 to 7 glasses raises the percentage of drinkers who drink more than the three depicted levels (an average over 21, 49 and 70 glasses per week) with 5.4%, 2.1%, and 0.9%, respectively, whereas an equal change in mean levels of 20 to 25 glasses raises the respective percentages with 8.9%, 3.4% and 1.8%. Given a fixed relation between consumption at certain levels, and various risks (health or otherwise), it is apparent that according to the model, a steady growth of mean consumption will result in an increasingly larger growth of the population-at-risk. If the *"quasi-mathematical connection between reasonable and unreasonable consumption cannot be broken.."*, the consequences of this model for a primary prevention policy are a *"suppression of alcohol in all the forms in which it is consumed.."* since in a permissive, tolerant attitude toward alcohol *"society discharges its responsibility of the individual granting him liberty of choice.."* (Ledermann, 1964a, p.8). In Ledermann's view, this liberty of choice is what causes the distribution law to be in effect.

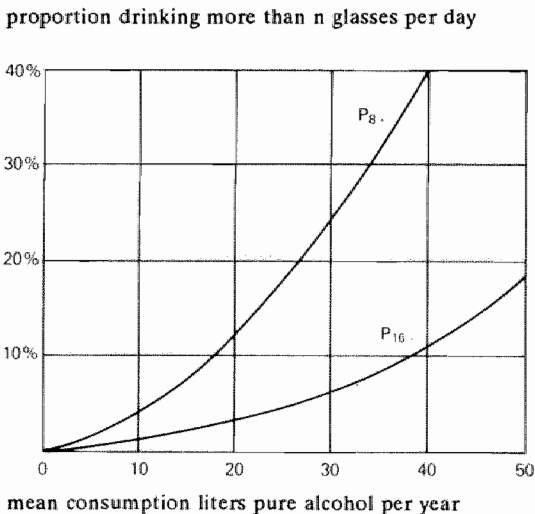


Figure 1.1. Relationship between average consumption (per drinker) and the proportion of the drinking population consuming more than 8 and 16 glasses per day, according to the Ledermann formula.

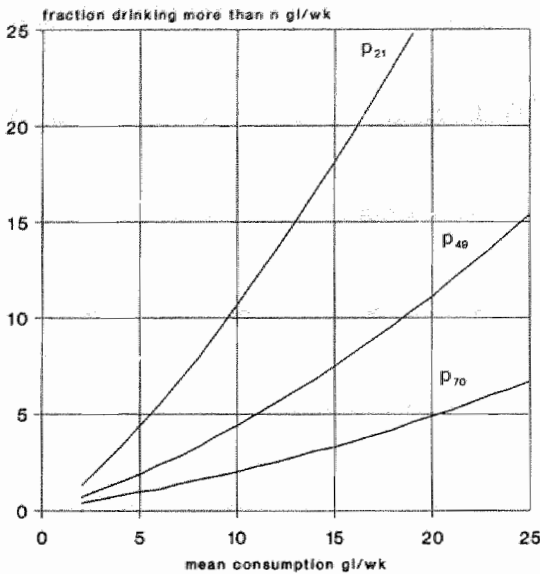


Figure 1.2. Relationship between average consumption (per drinker) and the proportion of the drinking population consuming more than 21, 49 and 70 glasses per week

1.3 Model estimates of prevalence of consumption

Compared with other methods for estimating prevalence of alcohol (ab)use in society, for example, the direct survey method or indirect methods based on differential mortality statistics (see, e.g. Wever, 1976), the Ledermann formula is relatively easy to employ. As was outlined above, the only crucial parameter is the average daily consumption in the population, which is usually readily available in official tax and sales statistics. For this reason, estimates of the number of excessive (and even alcoholic) users have been derived from the Ledermann model (e.g. in the Netherlands by Wever & Gips (1977), and De Zwart (1983); De Lint (1974) for several countries). However simple, three conditions should be met in the application of the Ledermann formula. The first concerns the use of sales and population statistics, the second concerns the proportion abstainers, and the third condition is that of homogeneity of the population under study.

Consumption statistics are usually gathered by agencies on behalf of producers or governments on the basis of sales or tax revenues (see, for an extensive list of agencies, Produktschap voor Gedistilleerde Dranken, 1988, appendix 1). Illegal production, home production, tourism, interstate or border-crossing purchases, effects of stockpiling, and certain losses (such as bottle breakage, unfinished drinks, beer spilled at the tap, etc.) constitute a threat to the validity of the implicate assumption that sales in a certain period are equivalent to consumption in that period (e.g. Mulford & Fitzgerald, 1981, Midanik, 1982, Watts *et al.*, 1981). The impact of these factors depends on their magnitude and direction. According to Wever & Gips (1977) illegal and home production in the Netherlands are relatively small. In Norway, however, survey-based estimates of home production (which is not recorded in official statistics) amount to nearly 30% of the total consumption (Nordlund, 1989). It is evident that negligence of these factors will result in biased estimates of excessive use in society.

Similarly, a bias in prevalence rates is introduced as a result of the variation in estimates of abstinence. It has been shown that in Dutch surveys these estimates vary over the years and over surveys (Lemmens, 1987). For example, de Zwart (1983) calculating prevalence(s) of heavy use in the Netherlands by means of the Ledermann formula, assumes that 13% of the D.A.P. abstains from alcohol in 1981, whereas van Reek *et al.* (1983) report an abstinence rate of 20% for the same year (among those over 20 years of age). Lemmens (1987) concluded that variations in the way abstinence is operationalized, and particularly the length of the reference period, will produce different results. The reader is also referred to chapter 5 where it is found that self-reported abstinence is also affected by the period in which the interview takes place.

Homogeneity in drinking behavior refers to the condition that all subjects should be exposed to the same factors that facilitate or hamper consumption. For example, it is important that the population is homogeneous with regard to the upper limit of consumption (the physiological maximum already mentioned), the metabolism of alcohol, and with regard to the various social constraints that are in effect. In a sense, the exclusion of abstainers can also be seen as an aspect of the homogeneity of the population. According to Ledermann, disregarding homogeneity in drinking pattern can lead to biased estimates. The populations for which sales and tax-based total consumption figures are available, however, are not very homogeneous in drinking behavior. For instance, men and women differ quite dramatically in drinking pattern, norms and physiological parameters. In the Netherlands age, educational level, religion and region are correlates of consumption (see van Reek *et al.*, 1983; Knibbe *et al.*, 1985).

The above restraints show that even when the the assumptions of the Ledermann model were justified, the estimates of excessive use based on per capita tax and sales figures should be interpreted with sufficiently large margins.

The Ledermann's theory of the distribution of alcohol consumption in the early sixties clearly was at odds with prevailing ideas about aetiology and prevention of alcohol abuse. Neither in the medical, individualistic view on alcoholism (e.g. Jellinek, 1960), nor in the social integration theory was mean level of consumption regarded as a crucial variable in the epidemiology of abuse. Ledermann has been the first to link, on a scientific level, "normal" drinking behavior with, from a public health point of view, hazardous consumption patterns.

1.4 Empirical studies and criticism

In appendix 1.1 an overview of the most important empirical studies of the shape of the distribution of consumption in the population are presented. Most studies indeed report highly skewed distributions. This skewness is often interpreted as an approximation of a lognormal shape, and even as a support for the Ledermann model. In most empirical studies, however, the observed proportion of the sample drinking at the highest levels is less than predicted by the Ledermann formula. Most authors do not make clear what they regard as a good approximation or a reasonable fit of the model. Furthermore, those studies that put the data to a statistical test, do not always come to the same conclusion (e.g., Guttorp & Song, 1979; Skog, 1971, 1979; Duffy, 1977a).

The review of literature on the distribution of alcohol is summarized in the following 6 points:

1.4.1 *Interpretation of cross-sectional analyses.*

For a test of the model usually cross-sectional data have been used. Longitudinal data are quite rare (except in Bruun et al., 1975; Cartwright et al., 1978; Kendell et al., 1983; Knibbe et al., 1985). Results of empirical analyses that reveal differences between countries (e.g. that are far apart in mean consumption) cannot, a-priori, be interpreted as analogues of longitudinal developments (Miller & Agnew, 1974, Room, 1973).

1.4.2 "What goes up, must come down".

Most studies deal with the question what will happen when mean consumption rises (except, Kendell *et al.*, 1983). However, neither cross-sectional, nor longitudinal studies of populations with increasing mean consumption secure that the observed developments will be reversed in the case of a downward trend. For example, Romelsjö & Ågren (see appendix 1.1) report that a downward trend in mean consumption has not been associated with a similarly large decline in excessive use in Sweden. Knibbe *et al.* (1985) have found that a decrease in mean consumption for young Dutch males between 1970 and 1981 has coincided, contrary to expectation, with a rise of those drinking 22 glasses or more per week.

1.4.3 Statistical tests of lognormality.

Few attempts have been made to put apparent lognormality to a statistical test. Often authors are satisfied with the observation that the distributions are highly skewed. Usually, authors compare, visually, a few frequency categories to the Ledermann distribution. Duffy and Cohen (1978) argue that these comparisons do not give information on the exact shape of the distribution.

Statistical testing as such is problematic. Finding a good fit of a distribution to a mathematical model does not imply that the correspondence is equally well at all levels. A good fit in the central part of the distribution is no guarantee for a good fit at the extremes. It is particularly the extreme part of the distribution in which one is interested (Duffy 1977b). Statistical uncertainty is therefore high, especially in small samples of populations with a low mean consumption.

Partly for this reason, Bruun *et al.* (1975), in their now famous book on control policies, have mitigated the strict distribution law, as presented by Ledermann, and propose a "weaker" version in which the statistical nature of the association between mean and excessive use is stressed. Skog has later referred to this as the "sometimes true" character of the predicted association (Skog, 1981, p. 320). Bruun *et al.* considered the Ledermann formula not suitable for estimation purposes, because of the observed, however small, variation in dispersion between populations with a similar per capita consumption. On the basis of a regression analysis of 15 survey data sets they conclude that: "...within the range which is of practical interest, the proportion of heavy consumers appears to be approximately proportional to the square of the mean consumption.".

1.4.4 Uni- or bimodality?

Some authors have questioned the unimodality of the distribution of consumption which is based on the observation of a smooth transition from low to high consumption levels (Miller & Agnew, 1974). Some authors have claimed that the distribution as it is observed may very well be bimodal,

composed of two distinct categories of drinkers each with their "own" distribution. In this view, the, by far, largest category consists of normal, social drinkers, the other consists of alcoholics, or heavy drinkers with a pathological drinking pattern who are not influenced by social norms (see, Tuck, 1980). The second peak (the mode of the distribution of heavy drinkers) and the relative invariance of the drinking pattern of alcoholics may very well go unobserved, because of the large differences in size of the two categories which causes large statistical variance at the higher levels. According to the advocates of a bimodal model, the Ledermann model might be able to predict incidence of excessive use when consumption rises but would fail to give an adequate description of the developments when the mean consumption drops.

1.4.5 *Poor validity of self-report data on consumption.*

A serious problem in surveys on alcohol consumption is the allegedly poor validity of the self-report data, often called the undercoverage problem (Pernanen, 1974, Midanik, 1982, 1988). Undercoverage refers to the poor coverage of sales and tax estimates of consumption (official figures of beverage alcohol that has been sold for consumption) with consumption data from surveys. On the whole, survey estimates cover only about 35% to 70% of the beverage alcohol that has been sold in a given period. In the Netherlands the reported coverage with the usual methods of measurement is about 50% (van Reek *et al.*, 1983). A poor validity of the empirical data is, of course, a weak basis for tests of a hypothetical model.

The fact that undercoverage is so large has lead researchers to hypothesize that underreporting of consumption by the subject interviewed is the main cause of the discrepancy. One of the reasons for this underreporting may be the threat imposed by questions about personal use of alcohol. It is often thought that underreporting of consumption increases with increasing consumption (e.g. Pernanen, 1974). However, others, such as Mulford & Fitzgerald (1981), have argued that there are many possible sources of error, not only respondent based but also researcher based. Examples are errors in estimates of consumption based on official sales statistics, errors that emanate from specific choices made in the survey design, such as selection of the sample frame, reference periods of questions on consumption, methods of measurement (options), choice of the unit of consumption, or calculation procedures. Related to survey design is the non-response that may bias estimates of total consumption based on survey self-reports. In section 7, an overview is given of sources of error and potential undercoverage. Several chapters in this thesis are devoted to an assessment of the effects several errors might have on the validity of estimates of consumption (chapter 2, 3, 4, 5, and 6), and, hence, on conclusions regarding the distribution of consumption.

1.4.6 *The genesis of lognormality.*

Ledermann has been very brief on theory as regards his distribution law. He does not offer substantial explanations for the alleged stability of the lognormal shape of the distribution. As noted in section 2, the reasons for claiming lognormality are mainly pragmatic, namely 1) that it is not realistic to assume normality, and 2) that a logarithmic transformation of a variable is not uncommon if the phenomenon develops according to a snowball or contagion effect (Ledermann, 1956, p.125). Ledermann explains the term contagion with the "*fact*" that the drinking behavior of an individual is not free, since it is determined in part by the pleasure alcohol may give, and for the other part, by the moral pressure toward moderation exercised by the social environment. Why the consumption should be lognormally distributed and not according to some other, e.g. exponential, model remains unclear.

Skog (1980a) notes that Ledermann's concept of contagion resembles the law of proportionate effects, formulated by Gibrat in 1931, which states that, for many socio-economic variables, change is proportional to the initial level. Aitchison and Brown (1969) show that this law is indeed a theoretical explanation of the genesis of lognormal distributions (see also chapter 7).

Other authors have criticized the directionality or causal link implicit in Ledermann's single distribution model, namely from mean to excessive consumption (see Duffy, 1980), thereby disregarding the possibly tautological reasoning. Duffy remarks that changes in consumption by individuals produce changes in mean consumption, rather than the other way around (ibid., p.150).

1.4.7 *Implications for prevention.*

In their review of the basic propositions of the single-distribution model of prevention, Parker and Harman (1978) conclude that the model is overly reductionistic. Apart from some of the points already mentioned above, they assert that specification of the total consumption variable is needed to be able to explain or predict differences between countries with regard to negative consequences. For example, total consumption should be specified into frequency of drinking, quantity per occasion, frequency of heavy drinking (e.g. intoxication). In support of their argument for an expansion of the model they report on empirical studies in which certain drinking patterns by certain groups (and not consumption, per se) are related to certain types of consequences (e.g. acute versus chronic).

Summarizing this section, one can conclude that the empirical support for the Ledermann model is rather weak, and that the model as such is not very useful. As the observed variation in dispersion between samples with similar mean consumption is obviously large, the model seems to be too rigid in its assumptions, and estimates of (changes in) excessive use can be expected to be biased. Similarly, the variation in prevalence of excessive use between

countries with similar mean consumption is larger than the model predicts. Since the empirical evidence, nevertheless, suggests that the distribution of consumption usually is skewed (and remains so when mean consumption rises), lognormality might be suitable as a first approximation. Similarly, the per capita consumption in a univariate prevention model should only be regarded as a crude indication of prevalence of alcohol related problems. There are many factors that intervene. Part of the controversy between authors seems to be caused by a different view on what is regarded a "*reasonable approximation*" of empirical data: some claim the bottle to be half full, while others stress the fact that it is half empty.

1.5 The theory of the collectivity of drinking cultures

In an attempt to provide a theoretical basis for the apparent regularity in the shape of alcohol distributions, O.-J. Skog, Ledermann's main critic, has proposed two hypothetical mechanisms by which a change in consumption is brought about (Skog, 1985). In the first hypothesis, Skog stressed the collective nature of drinking behavior. In his view, a change in mean consumption in a population is the reflection of a shift in consumption by all members of that population. The process responsible for this collective shift is social interaction (Skog, 1980a). Social interaction governs the diffusion of habits and norms about appropriate drinking occasions, for instance, via social pressure to drink (prescriptive norms are expected to be more pressing than proscriptive). Skog notes that the drinking practices in a person's "*local environment*" is an important factor of the person's own consumption. According to this hypothesis, the tendency to change one's drinking is, in part, determined by "*...the extent to which his consumption level deviates from that of his friends...*" (ibid, p.75). Other factors that cause people to change their drinking level, such as availability, price, and situational contagion are thought to be closely related to the process of social interaction. For instance, Sulkunen (1977) has proposed a process of situational contagion (in contrast to the "*contagion by persons*" of Skog) as an explanation for the relative invariance of alcohol consumption distributions. Since drinking is contingent upon situational cues -e.g. with meals, visitors-, new drinking habits will be added to already existing ones. So, diffusion of new drinking patterns occurs through additions to the old repertoire. In Skog's view, this process is essentially mediated by persons starting to drink in new situations. Thus, social interaction is seen as the mediating process by which changes occur. Similar to the single distribution model indicated earlier, Skog's dynamic model does not explain why people

change their drinking behavior, but rather how change will take place. Drinking is a highly social behavior (people drink mostly in the presence of others, in a group), and empirical evidence shows that people's drinking behavior is highly influenced by the drinking habits in the group. This mutual influence is reflected in a synchronization of drinking behavior of group members. In Skog's view, this finding can be extended from groups to larger networks, and he argues that a population can be seen as a large network of persons with social relations that differ in strength and structure (e.g. strong vs. weak; complex vs. simple). He hypothesizes that an impulse to behavioral change could be restricted to a small group of closely related persons, but he also shows that other impulses toward behavioral change, especially those in large networks with weak ties, may be very effective and reach nearly all persons in the network. In the latter case, a stimulus to behavioral change at one point in the network is transmitted to all elements and the resulting behavioral change will appear "*synchronized*". Skog concluded that empirical studies are needed that ascertain whether differences in degree of interdependence (or, in other words, heterogeneity) will indeed lead to different diffusion rates (for instance, one might expect differences across regions in interrelatedness of social substrata, such as men and women, and thus also in the way consumption patterns will change in the strata). Skog remarks that differences in interrelatedness between substrata, together with differences in strength of social control, could indeed produce variations in prevalence of heavy use, independent of per capita consumption (ibid., p.84).

In the second hypothesis, Skog has indicated in what way different drinkers will change their consumption. According to this hypothesis, when a factor causes drinkers to change their consumption, the change will be approximately proportional to their initial level. Skog mentions the Weber-Fechner law as a similar example of this proportional change. The law from early psychophysics established the relation between the (minimal) change in a perceptual stimulus and the perception of that change as a logarithmic one. More clearly, the law asserts that for a change in, for instance, the intensity of a sound to be noticed by the subject, the necessary *rise* in intensity of the sound increases as the sound becomes louder. Extended to alcohol consumption, Skog asserts that a change in consumption of 1 glass at a low level is not perceived equal to a similar change by drinkers at high levels. The process refers to the law of proportionate effects, mentioned already in connection to Ledermann's notion of contagion. Aitchison and Brown (1969) show that this proportionality of change generates lognormally distributed variates (see also ch7, footnote 2). The Ledermann model assumes lognormality. According to Skog, the law of proportionate effects implies that the factors responsible for a change in consumption combine multiplicatively rather than additively, similar to a situation of statistical interaction. Multiplicativity of effects is said to give an explanation of the skewness of distributions (see also ch7, fnt2). Specifically,

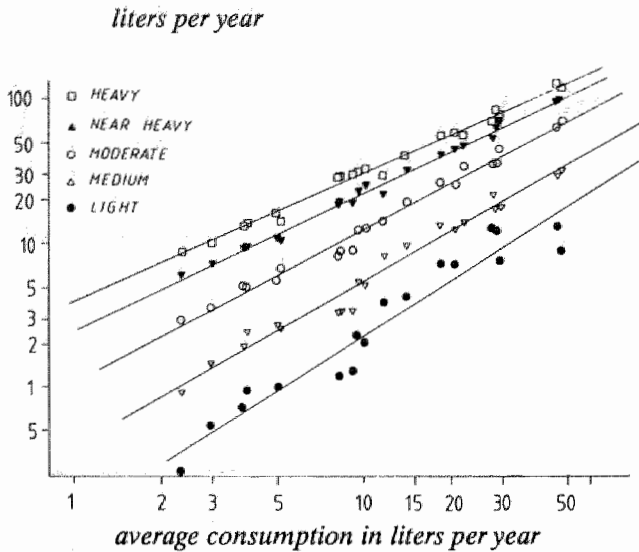


Figure 1.3. Relationship between average consumption and the consumption level of selected drinking groups (defined by percentiles), in 21 populations surveys. The straight lines are least square regressions. (copied from: Skog, 1985, p. 90).

Skog presented the result of empirical evidence from a Norwegian longitudinal study in which a change in consumption over time (from time 1 to time 2) was of the form $X_2 = \alpha X_1^\beta$. It is clear that a change in X from t_1 to t_2 is indeed proportional to X_1 (viz., a linear relation after logarithmic transformation of the data). In summary then, Skog assumes that a change in consumption is proportional to the initial level and that the change is collective, in the sense that “the population moves in concert upwards along the consumption scale.” (Skog, 1985, p.90).

As empirical evidence for his theory of collective change Skog presents data from 21 highly diverse populations. Five types of drinkers are operationalized in relative (ranking) terms defined by 5 percentiles. The assumption that drinkers at all levels of consumption will adjust their drinking relatively and proportionally to the mean of that population leads Skog to expect that all percentile values should vary as follows: $X_p = \alpha m^\beta$, where m denotes the mean, X_p the percentile value of the p -th percentile, and α and β constants. The analysis confirms the expectation. Additionally, it was observed that the slope of the regression (the β -parameter) decreases

Table 1.1 *Parameters of regression of 5 percentile points on mean consumption level (after logarithmic transformation of raw data); results from Skog (1985) and Lemmens et al. (1987) (R^2 = squared correlation coeff.).*

percentile	Skog, 1985			Lemmens, et al. 1987		
	b	a	R^2	b	a	R^2
25	1.278	.122	.930	1.20	.118	.985
50	1.146	.398	.976	1.41	.235	.984
75	1.065	1.102	.986	1.00	1.296	.997
90	.955	2.532	.992	.93	2.687	.999
95	.888	4.058	.988	.78	4.899	.994

with increasing percentile (this is not essential to multiplicativity). In other words, the relative change in percentile value drops as one moves up the consumption scale. Figure 1.3 depicts the regression of percentile from Skog study (1985, p.90). One can observe that the slope of regression (the b-parameter in table 1.1) decreases as the percentile considered increases. In his study, it is around the 75th percentile where the relative change in mean consumption is larger than the change in percentile value (b-parameter less than unity).

Lemmens *et al.* (1987a) have, more or less, reproduced these results using data from Dutch general population surveys, conducted in 1958, 1970, 1981. They do report some deviation at the lowest (25th) percentile value, which is probably due to peculiarities of the 1958 data (very low mean consumption, in connection with a weekly recall method of measurement, a consequently high number of drinkers with an observed zero consumption, viz., for which no exact data on weekly consumption are available, since their scores are rounded to zero). In table 1.1, the results of the regression of percentiles is given for Skog's and Lemmens' studies. The results of the regression in the two studies are surprisingly close.

Duffy (1986) has criticized the statement that the "population moves in concert..", since this suggests that all members change in the same direction. The above interaction model, according to Duffy, is not a model of (individual) change, but describes distribution around a given mean value. Duffy has objected (again) against Skog's interpretation of empirically observed deviations as "error". More or less in line with his own and Parker & Harman's earlier critic, he concludes that: "It could be that these deviations are related to factors causing important differences in distribution.." (ibid., p.738).

1.6 A distribution-free approach

The empirical evidence against and the fruitless debate over a mathematical, lognormal distribution model of consumption, and the observed regularity between the percentile points of distributions has lead to the conclusion that a distribution-free approach that concentrates on change in distribution (such as in Skog's dynamic theory) rather than (underlying) shape would probably be the most promising one (Lemmens, *et al.* 1987a). Of particular interest has been the mathematical and empirical background of the analysis and results reported in table 1.1.

A basic question one is confronted with when regarding the above regression analysis, a question which is also implicit in Duffy's comment above, is whether and how the observed relation between the percentiles of different distributions is related to distributional characteristics. In other words, how significant (as opposed to trivial) are the observed relationships and how specific are the observed relationships for the case of alcohol consumption. Would it be possible to specify conditions under which the observed relations would (not) occur? Translated into practical terms, two distinct questions have arisen that required further elaboration:

- 1) can the empirical relation between distributions of alcohol consumption indeed be described by a log-linear, multiplicative model as suggested by Skog; do different types of drinkers change their consumption proportional to their initial level?
- 2) is the β -parameter of regression of percentiles on mean consumption always positive and decreasing, as the empirical results in table 1.1 would suggest (a positive β -parameter means a positive change as the mean increases, a decreasing β -parameter means that the relative change decreases as the percentile considered increases)?

The first question will be further explored in chapters 7 and 8 of this study. In these chapters, the empirical relation between certain distributions of alcohol consumption is established, taking statistical uncertainty into account. In chapter 7 a procedure for comparing distributions is used that has been proposed by Doksum (1974), which relies heavily on the theory of empirical probability plots. The hypothesis is tested whether changes in distribution can be described by a multiplicative, that is, log-linear model. Empirical data from surveys conducted in the post-war Netherlands have been used. From a scientific point of view, the Netherlands are particularly relevant, because per capita (of D.A.P.) consumption has nearly quadruppled since the early fifties (from 3 liters 100% in 1955, to 11 liters pure alcohol in 1980). In chapter 8, the study is replicated with data sets from 4 western countries. The results with regard to the second question are not reported in this thesis, but in Tan *et al.* (1990). They will be briefly discussed in the last chapter. In its original form (Lemmens *et al.*, 1987b)

both questions were addressed simultaneously, however, several additions and modifications to this original report have been made. Part of these modifications concern the interpretation of Skog's concept of multiplicativity, which should be interpreted as a more general (regression) model (see: ch7, footnote 3). The consequences of the different interpretations are discussed in the last chapter.

1.7 Threats to validity of data on alcohol consumption

In section 4.5 several methodological problems have already been pointed out. It is evident that the impact of research on the distribution of consumption depends upon the validity of the basic data, viz., estimates of annual consumption. Looking for a good fitting model is useless if the data at hand are biased. The fact that with most surveys only about 35%-60% of the total sales are covered, indicates that the error in the data can be substantial. Therefore, assessment of this possible bias is necessary. Researchers in any field of alcohol research (either clinical or epidemiological) who need data on consumption of alcohol have been much concerned with problems of reliability and validity of their data. Because of the ambivalence with which most societies value drinking and the rather negative attitude towards heavy drinking in particular, one has always doubted the trustworthiness of self-reports of drinking, and, particularly, of heavy drinking. It goes without saying that a questionable validity of the basic data causes great uncertainty with regard to the results and conclusions (e.g. about distributional form).

In the first 5 chapters, explicit attention will be paid to the quality of self-reports of consumption obtained by survey methods. Several sources of a possible low validity (particularly, underestimation of consumption by surveys) are being presented in appendix 1.2. Research as well as respondent-based errors and their probable effect on the quality of consumption data are distinguished. The following 5 chapters (2 through 6) describe attempts to assess and quantify the effects of these, possibly contaminating, sources. In chapters 2, 4, 5 and 6 data from a Dutch survey on alcohol consumption will be used. The survey, conducted in 1985, has been specifically designed for this purpose (for a full report on this survey, see Lemmens, 1987). In the introduction of each chapter the main research questions are pointed out. The reader is referred to the single chapters for a literature review on these subjects. In chapter 9 an overview and a recapitulation is given. In the rest of this section, seven main errors listed in appendix 1.2 are briefly elucidated.

Non-coverage (Appendix 1.2.1) One of the reasons for an apparent underestimation of sales figures (the external criterium) with survey data is when mean consumption based on sales is an overestimate. Apart from the validity problems already noted in section 3 of this chapter concerning the assumption that all alcohol sold is also consumed in the population in that period, non-coverage of the target population (which means that the sampling frame does not coincide with the target population: not all inhabitants who have drunk the alcohol sold in that period are in the sample frame) may also produce a discrepancy between average consumption estimated with survey and sales data. Usually, total sales are averaged over all persons but, for practical reasons, sample frames do not cover the younger and oldest age categories. Still, a small part of the total sales will be consumed by youngsters and people older than 70. So, disregarding the consumption of these categories will inflate the numerator (sales) of the sales estimate of mean consumption. Other effects of non-coverage are discussed in chapter 2. In chapter 4, in the sample section, an estimate is given of the consumption of 12 to 15-year-olds, an age category that is usually excluded from the sample frame.

Non-reponse (Appendix 1.2.2) Rates of non-response by those selected in the survey sample have been steadily growing since the sixties. In the Netherlands it amounts to about 30% of the original sample (personal interviews; other methods show even larger rates). Survey researchers confronted with such large non-response rates should be concerned with especially selective non-response. In chapter 2, two methods are presented that may be helpful in detecting a possible bias.

Time Sampling (Appendix 1.2.3) Temporal variation in individual drinking is large. Because of this variation, consumption at the aggregate level will vary not only over days and weeks, but also over the months of the year. Chapter 5 discusses the problems that may arise because surveys are usually restricted to a limited periods of time (time-sampling). The analyses will use data from the Dutch 1985 alcohol survey, collection of which has been spread out over 10 months of 1985, including the Christmas holidays.

Basic Unit of Measurement (Appendix 1.2.4) Usually basic unit of measurement in surveys is the "standard glass", people are asked to report the "number of drinks" they have had. Bias is introduced because there is not an accepted standard for the amount of alcohol a glass or a drink contains. In licensed places (e.g. pubs), drinks are served in glasses more or less standardized for each type of beverage. What is then called a standard glass contains about 1.25 cl or 10 g of pure alcohol for any type of drink (in Holland, it still varies between 1.4 and .7 cl; see Produktschap voor Gedistilleerde Dranken, 1986, appendix 33). Since most alcohol is consumed in private settings where people are free in the way they consume their drinks, even more variation is probably introduced. For instance, Wilson (1981) has found that distilled drinks consumed in private settings actually

contain half as much alcohol as a standard type of glass. In a similar vein, the difference between standard and actual type of glass has been investigated in the Dutch 1985 survey for distilled spirits, wines and fortified wines. People who said to drink the particular type of drink were asked to pour water into the glass they typically use. The content was then measured in a measuring glass (see ch4, footnote 1).

Memory Effects (Appendix 1.2.5) Probably the main error in retrospective data collection is the error in recall, due to a limited memory capacity (see Moss & Goldstein, 1979). In the case of alcohol consumption, people will easily forget the exact number of drinks, especially if saliency of the event is low (for instance, when it is part of a daily routine). Apart from forgetting (omission of events), incorrect dating of past events may also occur when the reference period for the behavior is restricted. This may take the form of so-called forward telescoping, the respondent includes an event in the reference period when it actually occurred before that date, and backward telescoping, when an event is placed outside the reference period when it actually took place during the reference period. It is clear that backward telescoping will lead to underestimation of events. Saliency and frequency of events appear to be factors affecting telescoping: a low frequency and high saliency of events will increase the risk of forward telescoping (Sudman & Bradburn, 1982).

Memory effects will be covered in chapters 3 and 4. In chapter 3 recall of consumption in self-reports of recent drinking occasions (a 7-day recall period) will be compared with data from a "prospective" diary method, with a maximum lapse of only 1 day between event and report. Special attention is paid to the nature of the error (e.g. linear vs. non-linear relationship with consumption). The data are from a Dutch nationwide survey in 1981. Chapter 4 reports about aspects of the drinking pattern in a comparison of 5 different self-report methods. The emphasis is on accuracy of recall of quantity and frequency of drinking. The data come from the Dutch 1985 alcohol survey. Both chapters try to assess which method yields the most valid data.

Collateral Reports (Appendix 1.2.6) An important distinction in research on survey designs is between threatening and non-threatening questions (ibid., 1982). It has been established that questions about drinking are mildly threatening (Bradburn & Sudman, 1979). It is conceivable that perceived threat will increase the risk of distortion of self-reports on drinking to more (socially) acceptable levels. A way to circumvent this problem is by asking respondents to report about other's behavior (e.g. best friend). In chapter 6, a comparison is made between self-reports of subjects and reports on the subjects' drinking made by their partners. Not only consumption variables but also reports on problems and disapproval are compared. It was expected that underestimation of self-reported consumption (and drink related

problems and disapproval) would be higher at the higher consumption levels. A multivariate model is presented that might rule out rival explanations concerning disagreement between subject and partner.

Time-variation of drinking (Appendix I.2.7) It seems rather obvious that reports about highly complex behaviors is more prone to error than reports about simple ones. Temporal variation of drinking seems indeed rather large, and obtaining, or, for that matter, supplying accurate information about past drinking is a difficult problem. Over the years, several methods have been developed that have tried to deal with these problems. In chapter 4, temporal variation of drinking behavior in relation to method of self-report will be further explored.

References

- AITCHISON, J. & BROWN, J.A.C. (1969) *The Lognormal Distribution* (Cambridge, Cambridge University Press).
- BRADBURN, N. & SUDMAN, S. (1979) *Improving Interview Method and Questionnaire Design* (San Francisco, Jossey-Bass).
- BRUUN, K., EDWARDS, G., LUMIO, M., MÄKELÄ, K., LYNN, P., POPHAM, R., ROOM, R., SCHMIDT, W., SKOG, O.-J., SULKUNEN, P., ÖSTERBERG, E. (1975) *Alcohol Control Policies* (Helsinki, Finnish Foundation of Alcohol Studies).
- CARTWRIGHT, A., SHAW, S. & SPRATLEY, T. (1978) The relationships between per capita consumption, drinking patterns and alcohol related problems in a population sample, 1965-1974. Part II: Implications for alcohol control policy, *British Journal of Addiction*, 73, pp. 247-258.
- DE LINT, J. (1974) The epidemiology of alcoholism. In: N. Kessel, A. Hawker, H. Chalke (eds) *Alcoholism: a Medical Profile* (London, Edsall).
- DE LINT, J. & SCHMIDT, W. (1968) The distribution of alcohol consumption in Ontario, *Quarterly Journal of Studies on Alcohol*, 29, pp. 968-973.
- DE ZWART, W. (1983) *Het Alcoholgebruik en het Alcoholprobleem. Een Statistisch Rapport, 1960-1980* (Amsterdam, SWOAD).
- DOKSUM, K.A. (1974) Empirical probability plots and statistical inference for non-linear model in the two-sample case, *Annals of Statistics*, 2, pp. 267-277.
- DUFFY J.C. (1977a) Estimating the Proportion of Heavy Drinkers. In: *The Ledermann Curve. Report of a Symposium* (London, Alcohol Education Centre).
- DUFFY, J.C. (1977b) Alcohol consumption, alcoholism and excessive drinking. Errors in estimates from consumption figures, *International Journal of Epidemiology*, 6, pp. 375-379.
- DUFFY, J.C. (1980) The association between per capita consumption of alcohol and the proportion of excessive consumers - a reply to Skog, *British Journal of Addiction*, 75, pp. 147-151.

- DUFFY, J.C. (1986) The distribution of alcohol consumption - 30 years on, *British Journal of Addiction*, 81, pp. 735-741.
- DUFFY, J.C. & COHEN G.R. (1978) Total consumption and excessive drinking, *British Journal of Addiction*, 73, pp. 259-264.
- EKHOLM, A. (1972) The lognormal distribution of blood alcohol concentration in drivers, *Quarterly Journal of Alcohol Studies*, 33, pp. 508-512.
- FITZGERALD, J.L. & MULFORD, H.A. (1982) Alcohol consumption in Iowa; 1964-1979, *Journal of Studies on Alcohol*, 43, pp. 1171-1189.
- GREGSON, R.A.M. & STACEY, B.G. (1980) Distribution of self-reported alcohol consumption in New Zealand, 1978-1979, *Psychological Reports*, 47, pp. 159-170.
- GUTTORG P. & SONG H.H. (1979) A note on the distribution of alcohol consumption. A rejoinder to Skog, *Drinking and Drug Practices Surveyor*, 15, p. 6.
- JELLINEK, E.M. (1960) *The Disease Concept of Alcoholism* (Highland Park, Hillhouse).
- KENDELL, R.M., DE ROUMANIE M., RITSON, E. (1983) Effect of economic changes on Scottish drinking habits, 1978-1982, *British Journal of Addiction*, 78, pp. 365-379.
- KNIBBE, R.A., DROP, M.J., VAN REEK, J, SAENGER G. (1985) The development of alcohol consumption in the Netherlands: 1958-1981, *British Journal of Addiction*, 80, pp. 411-419.
- LEDERMANN, S. (1956) *Alcool, Alcoolisme, Alcoolisation*. Tome I (Paris, Presses Universitaires de France).
- LEDERMANN, S. (1964a) Can one reduce alcoholism without changing total alcohol consumption in a population? Paper presented at the 27th International Congress "Alcohol and Alcoholism", Frankfurt a.M., 1964.
- LEDERMANN, S. (1964b) *Alcool, Alcoolisme, Alcoolisation*. Tome II (Paris, Presses Universitaires de France).
- LEMMENS, P.H.H.M. (1987) *Het Ledermann-model Nader Bezien. Gevolgen van Onderschatting voor de Verdeling van Alcoholconsumptie* (Maastricht, University of Limburg).
- LEMMENS, P.H.H.M., KNIBBE, R.A., DROP, M.J. (1987a) De verdeling van alcoholconsumptie in de bevolking. Zin en onzin van het Ledermann-model, *Gezondheid en Samenleving*, 8, pp. 23-36.
- LEMMENS, P.H.H.M., KNIBBE, R.A., TAN, E.S. (1987b) The distribution of alcohol consumption: a non-parametric approach. Paper presented at the 13th ICAA/Alcohol Epidemiology Section meeting, Aix-en-Provence, 7-12 June 1987.
- MIDANIK, L. (1982) The validity of self-reported alcohol consumption and alcohol problems: a literature review, *British Journal of Addiction*, 77, pp. 357-382.
- MIDANIK, L. (1988) Validity of self-reported alcohol use: a literature review and assessment, *British Journal of Addiction*, 83, pp. 1019-1029.
- MILLER G.H. & AGNEW, N. (1974) The Ledermann model of alcohol consumption, *Quarterly Journal of Studies on Alcohol*, 35, pp. 877-898.
- MOSS, L. & GOLDSTEIN H. (Eds) *The Recall Method in Social Surveys* (London, Institute of Education).

- MULFORD, H.A. & FITZGERALD, J.L. (1981) "Words and deeds": responses to Popham and Schmidt, *Journal of Studies on Alcohol*, 42, pp. 362-367.
- NORDLUND, S. (1989) Estimating true alcohol consumption from false data. Paper presented at the 15th Alcohol Epidemiology Symposium, Maastricht, 11-17 June 1989.
- PARKER, D.A. & HARMAN, M.S. (1978) The distribution of consumption model of prevention of alcohol problems. A critical assessment, *Journal of Studies on Alcohol*, 39, pp. 377-399.
- PERNANEN, K. (1974) Validity of survey data on alcohol use. In: R. Gibbens, Y. Israel, H. Kalant, R. Popham, W. Schmidt, R. Smart (Eds) *Research Advances in Alcohol and Drug Problems*, Vol. 1, pp. 355-374 (New York, Wiley).
- PLANT, M. (1980) The Ledermann Curve: a Review of the Recent Debate. In: *Use of the Ledermann Curve in Determining Policy on Alcohol Related Problems* (Brewers' Society).
- PRODUKTSCHAP VOOR Gedistilleerde Dranken (1986) *Jaarverslag 1985* (Schiedam, PGD).
- PRODUKTSCHAP VOOR Gedistilleerde Dranken (1988) *How Many Alcoholic Beverages Are Being Consumed Throughout the World?* (Schiedam, PGD).
- ROMELSJÖ, A. & ÅGREN, G. (1985) Has mortality related to alcohol decreased in Sweden? *British Medical Journal*, 291, pp. 167-170.
- ROOM, R. (1973) Notes on the implications of the lognormal curve, *Drinking and Drug Practices Surveyor*, 7, pp. 18-20.
- SKOG, O.-J. (1971) Alkoholkonsumenters Fordeling i Befolkningen. Cited in: O.-J. Skog (1983) *The Distribution of Alcohol Consumption. Part II: a review of the first wave of empirical studies* (Oslo, National Institute for Alcohol Research).
- SKOG, O.-J. (1979) A note on the distribution of alcohol consumption: a reply and rejoinder, *Drinking and Drug Practices Surveyor*, 15, pp. 3-6.
- SKOG, O.-J. (1980a) Social interaction and the distribution of alcohol consumption, *Journal of Drug Issues*, 10, pp. 71-92.
- SKOG, O.-J. (1980b) Is alcohol consumption lognormally distributed? *British Journal of Addiction*, 75, pp. 169-173.
- SKOG, O.-J. (1980c) Total alcohol consumption and rates of excessive use: a rejoinder to Duffy and Cohen, *British Journal of Addiction*, 75, pp. 133-145.
- SKOG, O.-J. (1981) Alcoholism and social policy: are we on the right lines? *British Journal of Addiction*, 76, pp. 315-321.
- SKOG, O.-J. (1985) The collectivity of drinking cultures: a theory of the distribution of alcohol consumption, *British Journal of Addiction*, 80, pp. 83-99.
- SMART, R. & SCHMIDT, W. (1970) Blood alcohol levels in drivers not involved in accidents, *Quarterly Journal of Studies on Alcohol*, 31, pp. 968-971.
- STACEY, B. & ELVY, G. (1981) Is alcohol consumption lognormally distributed among fourteen- to seventeen-year-olds? *Psychological Reports*, 48, pp. 995-1005.
- SUDMAN, S. & BRADBRUN, N. (1982) *Asking Questions* (San Francisco, Jossey-Bass).

- SULKUNEN, P. (1977) Behind the Curves: on the Dynamics of Rising Consumption Level. In: *The Ledermann Curve. Report of a Symposium* (London, Alcohol Education Centre).
- TAN, E.S., LEMMENS, P.H.H.M. & KONING, A.J. (1990) Regularity in alcohol distributions: implications for the collective nature of drinking behaviour, *British Journal of Addiction*, 85, pp.745 - 750.
- TUCK, M. (1980) *Alcoholism and Social Policy: Are We on the Right Lines?* (London, HMSO).
- WATTS, R., PHIL, C., RABOW, J. (1981) The role of tourism in measures of alcohol consumption, alcohol availability and alcoholism, *Journal of Studies on Alcohol*, 42, pp. 797-801.
- VAN REEK, J., KNIBBE, R.A., DROP, M.J. (1983) Ontwikkelingen in alcoholgebruik in Nederland tussen 1958 en 1981, *Tijdschrift voor Sociale Gezondheidszorg*, 61, pp. 954-959.
- WEVER, O.R. (1976) Methoden ter schatting van het voorkomen van alcoholisme in een bevolking, *Tijdschrift voor Alcohol en Drugs*, 29, pp. 19-25.
- WEVER, O.R. & GIPS, C.H. (1977) Schatting van het aantal excessieve drinkers en alcoholisten in Nederland met de Ledermann-formule, 1875 tot 1977, *Tijdschrift voor Alcohol en Drugs*, 30, pp. 43-49.
- WILSON, P. (1981) Improving the methodology of drinking surveys, *The Statistician*, 30, pp. 159-167.

Bias due to Non-response in a Dutch Survey on Alcohol Consumption

ORIGINALLY PUBLISHED IN THE BRITISH JOURNAL OF ADDICTION, 83, 1069-1077, 1988. PAUL H.H.M. LEMMENS, FRANS E.S. TAN, RONALD A. KNIBBE.

Summary

Increasing non-response rates force survey researchers to focus on the danger of biased results. In this study two methods were used to determine whether a non-response rate of 31% in a survey of drinking habits in the Netherlands seriously affects estimates of alcohol consumption. Firstly, several indices of mean consumption were corrected for differences in response probabilities of the distinguished substrata. Secondly, data from a follow-up among non-respondents were used to study differences in consumption between the response and non-response stratum. Both methods led to the general conclusion that the effect of non-response on estimates of consumption for the entire population appears to be small. Among female non-respondents, a higher proportion of abstainers and a lower average consumption was found. Although occasional heavy alcohol use in the previous 6 months was more frequent among male non-respondents than among male respondents, the reverse was true for frequent, weekly heavy use.

2.1 Introduction

In recent decades survey researchers have been confronted with increasing rates of non-response. Even in surveys where overall response rates are still relatively high, refusal rates show a marked increase over the years (Madow *et al.*, 1983). In general, non-response leads to loss of precision of population estimators. The major concern, however, is the bias in estimates when a relationship exists between the variables of interest and the occurrence of non-response. As bias also depends on the rate of non-response, the danger of a serious bias is greater with high non-response rates. Non-response rates appear to depend on many factors, such as survey method (mail, telephone, face-to-face), type of questions (e.g. level of perceived threat), level of experience of the interviewers and type of sample frame (Madow *et al.*, 1983; Bradburn & Sudman, 1979; Kish, 1965). Though much cost and effort is usually devoted to keep non-response small, high response rates are not always feasible.

In surveys on alcohol consumption non-response has frequently been put forward as a factor causing discrepancies between official sales statistics and survey estimates of total consumption. Pernanen (1974) states that *"it seems probable that heavier drinkers are harder to locate for an interview ...[and]...are more likely to refuse an interview."*(p.360). Furthermore, he observed that in many surveys subpopulations with a higher proportion of heavy drinkers tend to show higher non-response rates. Wilson (1981) found that respondents who were most difficult to contact had a consumption which was 17% higher than the average. The effect on the total coverage of sales estimates, however, was marginal. Wilson nevertheless warned that because of the highly skewed distribution of alcohol consumption in the population, under-representation of heavy drinkers can have marked effects on estimates of mean consumption. Similar results from a Canadian survey were mentioned by De Lint (1981). In this survey, larger purchases of alcoholic drinks were reported by respondents requiring several house calls than by respondents who were at home at the first call. Similarly in line with Pernanen's argument, Nillson and Svensson (1971) reported that non-respondents in a Swedish survey were three times more likely to have a police record for drunkenness than were respondents. In contrast, Garretsen (1983) found that none of 200 randomly selected non-respondents in a Dutch survey were registered as clients of the alcohol clinics compared to 9 out of 453 randomly selected respondents who were registered at the time of the interview. In the same survey, 106 of the 850 initial non-respondents (who refused or were not available after 2 call-backs) later complied with a shortened questionnaire by telephone. The results showed that abstention rates, especially among male non-respondents, were much higher in this category (34% compared to 19% among respondents) and that there were no

indications of more frequent or heavier alcohol use. Knibbe (1984) also reported higher rates of abstention among a selection of non-respondents in another Dutch survey but heavy use (6 glasses or more per occasion) was more frequent among the non-abstaining non-respondents. Mulford and Miller (1959) also suggested a higher abstention rate among refusers in a survey in Iowa. In a recent follow-up study among non-participants (*"not available"* or *"never in"*) from a British survey on alcohol consumption, female non-participants were more likely to be employed and to be non-manual workers (Crawford, 1986). This difference was not reflected in higher scores on consumption indices for non-participants. Among men, the average amount of alcohol consumed on the last drinking day was even significantly lower for non-participants. Thus, the overall evidence for Pernanen's statement seems far from conclusive. That high response rates do not guarantee a high coverage of sales estimates (often used as an external criterium of validity) is clear from Finnish surveys with response rates between 97% and 94%, yet with coverages of sales data ranging between 34% and 38% (Simpura, 1987).

The main goal of this paper is to assess whether estimates of alcohol consumption from a Dutch general population survey on drinking habits are biased as a result of the rather high non-response rate of 31%. The analysis is restricted to the so-called unit-non-response, which occurs *"if units in the selected sample and eligible for the survey do not provide the requested information, or the provided information is unusable"* (Madow, et al., 1983, p.3). The definition of non-response includes selected persons who either refuse or are unable to cooperate, those who cannot be reached at the listed address, or those who provide inadequate information. Persons not belonging to the target population are excluded from the non-response category and are labeled as overcoverage. The effects of other sources of incomplete data, such as noncoverage (Pernanen, 1974), are thought to be relatively small in the present case [footnote 1].

Though non-response is easy to define, its effects are difficult to assess. There are several methods with which selective non-response can be evaluated (Bethlehem & Kersten, 1986), two of which are used in this study. The first method compares respondents and non-respondents on background variables known for all units in the sample. By means of a logistic regression analysis, a model of differential response probabilities is constructed. Information about the relation between the background variables, ensued from the model, and the target variable (namely alcohol consumption) is then used to determine the effect differential response probabilities have on 5 indices of alcohol consumption.

The second approach consists of attempts to obtain at least part of the desired information from (a sample of) non-respondents (Hansen & Hurwitz, 1946), by means of a follow-up interview by telephone. By comparing consumption estimates from the main survey with those from the

follow-up, one can determine whether non-response has been selective and whether the resulting bias seriously affects estimates of consumption. As either method has its limitations the results of the two approaches are presented concurrently.

2.2 Methods

Between March 1985 and January 1986 a survey of drinking habits was conducted among a random sample of the Dutch population between 16 and 70 years of age. A two-stage sampling technique was used, common in the Dutch large-scale survey tradition. First, a random sample from all municipalities was drawn after stratification for region and urbanization (16 cells). Next, a random sample of persons between 16 and 70 years of age was obtained using municipal population registers (these also include registered foreign nationals). Contact was established by visiting the person's home address, up to a maximum of 4 call-backs. Participation involved a face-to-face interview and a subsequent diary to be completed in the week following the interview. Successful completion of both forms was required to be labeled response. On the basis of the questionnaire and self-administered diary several indices of alcohol use can be composed (to be discussed below).

To compute the response fraction, it is necessary to compensate for overcoverage of units in the sample, which do not belong to the target population (Wiseman and Billington, 1984). In our case overcoverage only consists of persons younger than 16 or older than 70 years of age at the time of the interview. Overcoverage appears to be relatively small (0.6%) which offers another indication of the accuracy of the municipal registers. Table 1 summarizes the outcome of the fieldwork.

The response rate of 69% is rather low, yet not uncommon in current large-scale surveys in the Netherlands (Bethlehem & Kersten, 1986). Main reasons for non-response were refusal (64%), unavailability (8%), not at home (9%) and non-response due to an incomplete diary (6%).

Because it was expected that the time-consuming diary would evoke more refusals than usual, it was decided to follow up the non-response with telephone interviews. The telephone questionnaire contained a selection of items on drinking which are consistent with the ones used in the main survey. Eligible for the follow-up were the not-at-homes and those persons who had refused for reasons other than personal or principal. No further attempts were made to contact persons whose addresses were unavailable. As can be seen in table 2.1, 295 non-respondents (56%) were selected for the follow-up

Table 2.1. *Sample size, overcoverage and nonresponse for the Dutch 1985 main survey and follow-up of nonresponse.*

main survey	N	Follow-up nonresponse	N
gross sample size	1807	follow-up sample	295
gross response	1244	response follow-up(c)	116
overcoverage response	4	nonresponse follow-up(d)	179
net response (a)	1240		
gross nonresponse	563	response fraction(c/(c + d))	39.3%
overcoverage nonresponse	6		
net nonresponse(b)	557		
response fraction (a/(a + b))	69.0%		

sample. Selection for the follow-up of non-respondents did not lead to differences with regard to sex. Persons married and persons between 26 and 35 or over 46 years of age tended to be over-represented in the subsample. Those from the southern region were less likely to be selected. No difference reached significance at the .01 level (chi-square tests). A maximum of 5 calls at different times were permitted to establish contact. The response fraction was low (39.3%), mainly due to refusals (41%), no or unlisted telephone numbers (34%) and not-at-homes (10%).

On the basis of questions on drinking habits in the main survey 5 indices of overall amount of drinking can be composed. Two indices are based on actual consumption in, respectively, the week preceding the personal interview (weekly recall; *WR*-index) and the diary week following the interview (*D*-index). Two indices are based on modal quantity and modal frequency of alcohol use (*QF*-index) and a-typical or variable heavy use (*V*) in the previous 6 months (*QFV*-index). The fifth index was uni-dimensional and was based on one question about typical daily number of glasses in the previous 6 months (*T*-index). In addition, a respondent was labeled abstainer if no alcohol had been consumed in the past half year. Based on these indices, the coverage of sales data ranges between 44% for the *QF*-index and 70% for the diary. The wide difference in average consumption, and, hence, coverage, seems in part to be caused by memory failure since the time interval between actual consumption and reporting that consumption is smallest in the case of the diary. Furthermore, it is conceivable that questions about modal or typical consumption are particularly prone to distortion, as they do not directly refer to actual behavior but require generalisations of drinking behavior over a long period (Redman *et al.*, 1987). In spite of the large differences in average consumption, the correlations between methods are still rather high (Kendall's *tau* between .58 and .66). This result suggests that the rank order of drinkers is fairly stable across methods

(Lemmens,1987). In order to avoid problems concerning the validity of the measurements, the effect of differences in response probability will be determined for all five indices.

In the telephone interview the same questions about abstaining, typical number of glasses per day (T) and variable heavy use (V) were asked. Respondents of the main survey and respondents of the telephone interview will be compared on these three variables. Even though the questions are identical in wording, type of survey technique (face-to-face and telephone) could, nevertheless, produce differences in response to these questions. Since distortion of response to threatening questions seems to be greater when the survey technique is more personal (Bradburn & Sudman,1979), one would expect the face-to-face interviews to have more errors, namely a bias towards lower self-reports of alcohol consumption, than the telephone interviews in the follow-up. At worst, this effect would be interpreted as evidence for a higher consumption among non-respondents. Usually, however, the bias seems relatively small.

2.3 Results

Response probabilities

As can be seen from the exploratory analysis in table 2.2, men and women do not clearly differ in response rate. Although women seem less likely to refuse, the effect of gender on non-response appears to be small. Age and region do lead to large differences in response rate. The young (up to 35 years) show a higher response rate than persons in the older age categories, because they seem less likely to refuse an interview. In the western part of the Netherlands non-response is highest, probably caused by a greater tendency to refuse. Marital status does not seem to be very important at first glance. Only widow(er)s clearly show an elevated response rate. Rural and small urban communities show a somewhat higher response rate than urban municipalities.

The above, univariate analysis does not take into account the relationship between the background variables, nor does it allow one to determine which (set of) variables are sufficient to predict response. Therefore, non-response is evaluated by means of a multivariate, logistic regression analysis with which a linear model of response probabilities is constructed. In the section to follow, the variables ensuing from the model are used to assess the possible effect differences in response probability have on estimates of mean alcohol consumption.

Table 2.2. *Shift of response and nonresponse for 5 background variables, relative to overall percentage¹.*

	Response	Unavallable	Refusal	Other reason
gender	69.0	2.4	20.0	8.6
men	.5	.0	.5	-1.0
women	-.4	.0	-.7	1.1
age	69.2	2.4	20.0	8.4
16-25	2.0	1.8	-2.4	-1.4
26-35	3.9	.9	-3.3	-1.6
36-45	-2.2	-1.4	3.1	.5
46-55	-5.2	-1.1	3.7	2.6
56-70	-.1	-.5	-.1	.7
marital status	69.3	2.4	20.0	8.3
married	.3	-1.0	.8	-.2
unmarried	-1.7	2.4	-1.2	.5
divorced	1.1	1.3	-2.7	.3
widow(er)	7.8	-2.4	-5.4	.0
urbanization	69.2	2.4	20.0	8.4
rural municipalities	2.1	-1.9	.4	-.5
urban countryside	-.5	.4	.2	.0
urban municipalities:				
small	3.7	-.2	-4.8	1.3
middle	-.5	-.3	1.0	-.2
large	-2.5	.7	2.1	-.3
region	69.0	2.4	20.0	8.6
north	.2	-.1	-.8	.7
east	6.8	-1.5	-2.3	-2.9
west	-4.7	.0	3.4	1.2
south	4.5	1.4	-5.3	-.6

¹ Total percentages slightly vary because of missing values for some variables

In linear modelling, response probability (p) is often expressed with a binomial distribution. To secure that predicted values will fall within the range of 0 and 1, the probability (p) is transformed into the so-called logit of p ($L = \log(p/1-p)$), where p denotes probability of response and $1-p$ probability of non-response (McCullagh & Nelder, 1983). Furthermore, it is assumed that the logit of p is linearly related to the independent variables.

Several models were tested which predicted the response variable as a function of the background variables. Goodness of fit was evaluated by inspecting the deviance between predicted and observed values.

The explanatory variables in the most satisfactory model are age, region and marital status. Gender and urbanization did not contribute in a way to warrant inclusion. The fitted model involved only the main effects of the categorical variables (deviance = 2158.4; df = 1773; p = .01). No significant interaction effects between the variables were found when a method for all subsets (up to first order interactions) was used. In table 2.3 the parameter estimates of the fitted equation are presented. Interpretation of table 2.3 is somewhat difficult, as the parameter values do not directly signify response probabilities. A positive value of a parameter denotes a positive contribution, a negative value denotes a negative contribution to the response probability, relative to the value of the reference group (aged 16-25 years, married and northern region). Bold figures in table 2.3 indicate significance at a .05 level.

It becomes clear that the youngest age category has a higher probability of response. All parameter estimates for older age categories are negative, with the 36-45 and 46-55 categories showing the significantly lowest response probability, probably due to the relatively high refusal rate, noted in table 2.2. Those who are married appear to have a significantly greater tendency to respond than those who never married (after correction for age and region). This result is probably caused by the greater mobility of the latter category (see table 2.2). The results for the variable region seem less conclusive. The response probability for persons in the western region is lower than for those

Table 2.3. *Parameter estimates for the logit-model1, including age (5 categories), marital status (4) and region (4).*

	parameter estimates	stand.error
grand mean	1.234	.227
age		
26-35	-.144	.187
36-45	-.510*	.200
46-55	-.644*	.214
56-70	-.435*	.215
marital status		
unmarried	-.394*	.159
divorced	.030	.254
widow(er)	.467	.359
region		
eastern	.349	.194
western	-.187	.165
southern	.252	.191

1 GLIM, statistical program (Baker & Nelder, 1978).

* significant at a .05 level.

in the other parts of the Netherlands (particularly the eastern region). However, differences between the northern and other regions do not reach significance.

For an overall comparison of the distinguished categories, the predicted response fractions, based on the model, are depicted in figure 2.1. Two categories, the divorced and widowed, are left out of the presentation for purpose of convenience. From the figure the differences in expected response probability for the discrete substrata become apparent. Those persons that are young, married and from the eastern part of the Netherlands show the highest response probability, those that are middle aged, unmarried and from the western part show the lowest probability. Note that the lines in the figure do not intersect.

Correction of Consumption Indices Based on Response Probability

Before quantifying the effect of differential response probabilities on estimates of mean consumption, it should be noted that no use is made of information about the population from which the sample is drawn as we are not primarily interested in population estimates and their variances. Furthermore, it should be kept in mind that, in applying the procedure described above, it is assumed that the relation between the background variables and alcohol consumption is identical in both the response and the non-response. In other words, units in the non-response substrata

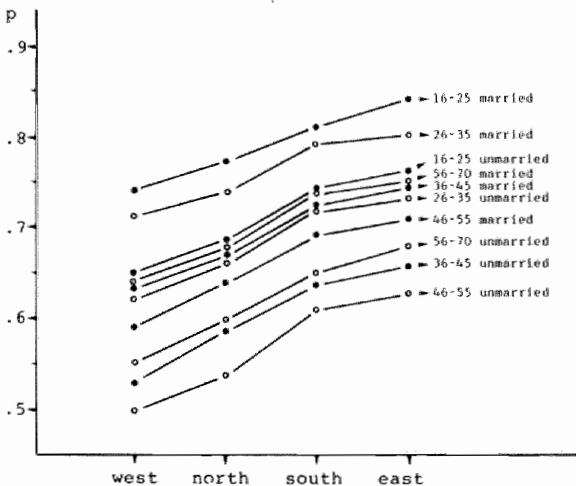


Figure 2.1. Predicted response fractions (p) of substrata, determined by the variables age, marital status and region.

(determined by age, marital status and region) are assumed not to differ in drinking habits from those in the respective response substrata. A check on the validity of this assumption is obtained in the following section, where additional information from the subsample of non-respondents is available. The procedure to quantify the effect of differences in response probability is fairly straightforward. First, one assesses mean consumption in all relevant response substrata determined by the model, resulting in 80 substrata. Next, these figures are weighted according to the observed response probabilities in the substrata (Bethlehem & Kersten, 1986). The weighting procedure produces corrected averages for the 5 indices of alcohol consumption. The results, presented in table 2.4, show that correction for differences in response probabilities of the specified substrata does not lead to adjustments in estimates of mean consumption for any index considered.

The outcome of the above correction procedure depends on two factors, namely, the explanatory power of the response model and the nature of the relationship in the initial sample between the variables in the model and the target variable (namely alcohol consumption). As the response model seems quite effective, the obvious conclusion would be that in a general population survey, contrary to Pernanen's observation mentioned earlier, bias in estimates of alcohol consumption due to differences in response probability is negligible.

Although correction for age, region and marital status decreases the risk of bias, there is no guarantee that the non-response stratum does not differ from the response on other factors, for which no information from both response and non-response strata was available. The second approach to the non-response problem, using additional information from a follow-up among the non-response, might clarify this issue.

Table 2.4. *Estimates of mean consumption (glasses per week) based on indices from main survey, before and after correction for differential response probabilities.*

Index	Before correction	After correction	Proportional change
typical consumption(T)	9.64	9.66	0.2%
quantity-frequency (QF)	7.57	7.59	0.3%
quantity-frequency- variability (QFV)	9.36	9.33	-0.3%
weekly recall (WR)	9.62	9.59	-0.3%
diary (D)	11.61	11.56	-0.4%

Follow-up of Non-response

When comparing follow-up data on alcohol consumption from the non-response stratum with data of respondents from the main survey, one assumes that neither selection for follow-up, nor subsequent non-response has been selective with regard to alcohol consumption in the non-response stratum (Bethlehem & Kersten, 1986).

Table 2.5. *Comparison of alcohol consumption indices from the main survey to those from the follow-up, before and after stratification of the latter for sex and region.*

	Main survey		Follow-up			
	Total sample	Drinkers only	Unweighted		Weighted	
			Total sample	Drinkers only	Total sample	Drinkers only
Typical consumption¹						
all	9.64	11.43	7.68	9.64	7.99	10.32
men	12.33	13.60	11.54	12.28	11.58	12.67
women	6.46	8.38	3.82	5.87	3.84	6.26
Fraction abstaining²						
all	0.160	-	0.244	-	0.226	-
men	0.095	-	0.086	-	0.086	-
women	0.236	-	0.362	-	0.387	-
Frequency of heavy use³						
a) at least weekly						
all	15%	17%	8%	10%	8%	10%
men	22	24	12	13	12	13
women	6	8	4	6	3	5
b) less than weekly						
all	22%	26%	29%	36%	30%	39%
men	29	32	51	55	51	55
women	13	17	6	9	6	10
(c) never						
all	64%	57%	63%	54%	62%	52%
men	49	44	37	32	38	33
women	81	75	91	86	91	85
Total(1%)	100%	100%	100%	100%	100%	100%

¹ Converted into glasses per week; bold figures: significant difference between response and non-response stratum of the main survey (Mann-Whitney U test, $p < 0.05$, one-tailed).

² Bold figures: significant difference between response and non-response of the main survey (chi-square test; $p < 0.05$).

³ Frequency of events in which 6 drinks or more were consumed in previous 6 months; significant difference between response and non-response of the main survey for all and male category (chi-square test; $p < 0.05$).

To limit the risk of invalidity of this assumption, post-stratification is often applied. As the efficacy of stratification depends on the quality of the relation between background variables and target variable, sex and region were chosen as background variables, since in previous research these showed the highest correlation with alcohol consumption (Knibbe, *et al.*, 1985). An alternative procedure, using "*reason for non-response*" for stratification had to be discarded, because of insufficient size of the substrata.

After stratification, estimated mean consumption of the non-response stratum rose only slightly from 7.68 to 7.99 glasses per week, according to self-reports of typical daily consumption (*T*). Adjustments in estimates of the fraction abstaining from alcohol use and frequency of occasional heavy use of alcohol are also marginal (table 2.5).

Table 2.5 also shows the results of the comparison between consumption indices of the response and non-response of the main survey. It appears that average consumption (*T*) in the non-response stratum is lower than in the response stratum. This is true for both men (though not significantly so) and women and regardless of the differences in rate of abstention between the two strata (only for women at a .05 level). Abstinence from alcohol use appears to be more prevalent only among female non-respondents, which is in clear contrast to the outcome reported by Garretsen (1983), mentioned earlier. Males in the two strata, on the other hand, differ in frequency of occasional heavy drinking. Whereas average consumption seems lower in the non-response stratum, occasional heavy use is reported more frequently. For instance, among male drinkers in the non-response stratum 33% report never, in the previous 6 months, to have exceeded 6 glasses on one occasion, compared to 44% of the respondents. Conversely, frequent heavy use (at least weekly) appears to be more common among males in the response stratum (24%) than in the non-response stratum (13%).

In all, the results from the analysis of the follow-up data do not indicate that non-respondents generally drink more, nor that alcohol abuse is more common in this category. To put it more strongly, there is evidence that female non-respondents generally drink less and show higher rates of abstention. Correction of the consumption indices from the main survey, based on the results of the follow-up would, in fact, result in lower population estimates.

2.4 Discussion

The results of the analyses in this study do not lend support to the hypothesis that among non-respondents from a general population survey alcohol consumption is higher than among respondents. The danger of underestimating alcohol use, if one relies solely on data from the response stratum, therefore seems limited.

Before generalizing this conclusion, however, one should consider the restrictions associated with the methods by which non-response is analyzed in this study as well as the type of sampling procedure.

The very nature of the phenomenon, namely the lack of information from a part of the sample, considerably limits the options for research on non-response bias. The information about the non-response usually available is, at best, sufficient to correct for differences in response probability. The flaw of this method lies in its assumption that response probability does not bear a relation to alcohol consumption. The outcome of the follow-up study among non-respondents suggests that this may be true for the male substratum, but the assumption could well be violated in the case of the female substratum.

The follow-up among initial non-respondents, in turn, offers no guarantee that estimates of alcohol consumption, based on data from the response to the follow-up, are unbiased. If one imagines, as does Pernanen (1974), that heavy drinkers are more difficult to contact or are more likely to refuse an interview as a consequence of their specific habits (e.g. if heavy drinkers are less likely to have fixed addresses or are more likely to perceive questions about their alcohol use as threatening), then neither an extensive call-back procedure nor a follow-up would greatly reduce the expected bias. From this point of view, one could argue that call-backs and follow-up would yield only data from those respondents whose reason for non-response is not related to their drinking habits. It is clear that no single method, adopted during or after the fieldwork stage, can give a decisive answer to the problem of non-response bias. What can be done, firstly, is to keep non-response rates as low as possible, which decreases the magnitude of any non-response error in estimates and, secondly, to collect as much information as possible about non-respondents, either from within the survey procedure or from auxiliary sources, with which estimates may be corrected.

A further comment has to be made about the sampling procedure. It was noted earlier in footnote 1 that the sample frame in this study (municipal population records) is fairly complete. Accordingly, undercoverage of the target population is easily avoided. With other frames, like area samples, non-coverage and its assessment becomes a problem (Kish, 1965). Undercoverage might lead to underestimation of alcohol use, especially when heavy drinkers are concentrated into certain areas (Pernanen, 1974). If

non-coverage is large, analyses of non-response will not cover the entire field of non-observational errors (Kish,1965) and additional procedures are needed to assess the effect of non-coverage.

Footnote

1. For example, undercoverage of the target population by the sample frame is probably small as the frame consisted of municipal population registers which are quite accurate. Only those persons labeled "without permanent residence" are not included in the sample frame (less than 0.1 o/oo. Any person registered and selected for the sample but not actually living in the municipality will be labeled unit-non-response. Thus, administrative incompleteness of the municipal records generates unit-non-response rather than undercoverage.

References

- BAKER, R. J. & NELDER, J. A. (1978) *The GLIM-manual*, Release 3 (Oxford, Numerical Algorithms Group).
- BETHLEHEM, J. & KERSTEN, H. (1986) *Working with non-response* (Amsterdam, University of Amsterdam).
- BRADBURN, N. M. & SUDMAN, S. (1979) *Improving interview method and questionnaire design* (California, Jossey-Bass)
- CRAWFORD, A. (1986) A comparison of participants and nonparticipants from a British general population survey of alcohol drinking practices, *Journal of the Market Research Society*, 28, pp. 291-297.
- DE LINT, J. (1981) "Words and deeds": Responses to Popham and Schmidt, *Journal of Studies on Alcohol*, 42, pp. 359-360.
- GARRETSEN, H. (1983) *Probleemdrinken* (Lisse, the Netherlands, Swets and Zeitlinger).
- HANSEN, M. & HURWITZ, W. (1946) The problem of nonresponse in sample surveys, *Journal of the American Statistical Association*, 41, pp. 517-529.
- KISH, L. (1965) *Survey sampling* (New York, Wiley).
- KNIBBE, R. A., DROP, M. J., VAN REEK, J. & SAENGER, G. (1985) The development of alcohol consumption in the Netherlands: 1958-1981. *British Journal of Addiction*, 80, pp. 411-419.
- KNIBBE, R. A. (1984) *Van gangbaar tot problematisch drankgebruik* (Maastricht, University of Limburg).

- LEMMENS, P. (1987) *Het Ledermann-model nader bezien. Gevolgen van onderschatting voor de verdeling van alcoholconsumptie* (Maastricht, University of Limburg).
- MADOW, W. G., NISSELSOON, H. & OLKIN, I. (1983) *Incomplete data in sample surveys*, Vol.I (New York, Academic Press).
- MCCULLAGH, P. & NELDER, J. A. (1983) *Generalized linear models* (London, Chapman and Hall).
- MULFORD, H. & MILLER, D. (1963) The prevalence and extent of drinking in Iowa, 1961; a replication and evaluation of methods, *Quarterly Journal of Studies on Alcohol*, 24, pp. 39-53.
- NILLSON, T. & SVENSSON, P.-G. (1971) The alcohol habits and alcohol attitudes of the Swedish people, in: *Svenska Folkets Alkoholvanor*, Stockholm: Statens offentliga utredningar. Finansdepartementet (cited in Pernanen, 1974).
- PERNANEN, K. (1974) Validity of survey data on alcohol use, in: Gibbins, R., Israël, Y., Kalant, H., Popham, R., Schmidt, W., Smart, R.G. (eds) *Research advances in alcohol and drug problems*, Vol.1, pp. 355-373 (New York, Wiley)
- REDMAN, S., SANSON-FISHER, R.W., WILKINSON, C., FAHEY, P.P., GIBBERD, R.W. (1987) Agreement between two measures of alcohol consumption, *Journal of Studies on Alcohol*, 48, pp. 104-107.
- SIMPURA, J. (1988) Comparison of indices of alcohol consumption in the Finnish 1984 drinking habits survey data, *The Drinking and Drug Practices Surveyor*, 22, pp. 3-10.
- WILSON, P. (1981) Improving the methodology of drinking surveys, *The Statistician*, 30, pp. 159-167.
- WISEMAN, F. & BILLINGTON, M. (1984) Comment on a standard definition of response rates, *Journal of Marketing Research*, 21, pp. 336-338.

Weekly Recall and Diary Estimates of Alcohol Consumption in a General Population Survey

ORIGINALLY PUBLISHED IN THE JOURNAL OF STUDIES ON ALCOHOL, 49, 2, 131-135, 1988. PAUL H.H.M. LEMMENS, RONALD A. KNIBBE, FRANS E.S. TAN.

Summary

Diary and weekly recall measures of alcohol consumption in a representative sample ($N=399$) of the Dutch population are compared. The weekly recall method consisted of a personal interview with questions about actual alcohol consumption on the previous 7 days. The diary consisted of 14 daily self-reports of consumption and followed the interview. The diary method yields estimates of consumption which are on average 22% higher than those based on weekly recall measures, reducing total undercoverage by about 11%. The difference between the two methods cannot be attributed to a variation of consumption over weeks but seems to stem from a difference in accuracy of recall. There was an increase in glasses underreported at the upper levels of consumption, but underreporting did not seem to be of a non-linear nature. Considering the large individual variation in consumption over weeks, the ranking of individuals according to their self-reports is relatively stable across method.

3.1 Introduction

The use of alcohol consumption data from general population surveys has invariably focused attention on methodological issues concerning the measurement of alcohol consumption. Estimates of total alcohol consumption in a population based on self-reports typically cover only 40 to 60% of the estimated amount of alcohol available for consumption in a particular period and this is usually seen as evidence of the poor validity of these data (Midanik, 1982).

Figures on alcohol available for the home market are based on annual taxes, import and production figures (Produktschap voor Gedistilleerde Dranken, 1986). Some authors (cf. Midanik, 1982) have criticized the use of these official statistics as an external criterion since they do not include untaxed alcohol (airports, shipping), or illegally produced and homemade alcohol. On the other hand, purchases by persons not within the sample frame (e.g. tourists), stockpiling and alcohol sold but not consumed in a certain period in the target population will lead to overestimation of consumption. Considering the degree of undercoverage it seems unlikely however that errors in sales estimates fully account for the differences between survey estimates of consumption and availability. For instance, in the Netherlands home and illegal production of alcohol is believed to be very low. Still, the coverage rate of sales estimates across surveys, ranging between 44 and 49% (Knibbe *et al.*, 1985), does not differ greatly from surveys in countries where illicit production is high (e.g. Norway).

Several other sources of error have been distinguished such as sampling bias, question wording, memory failure and conscious underreporting (Alanko, 1984; Pernanen, 1974; Poikolainen and Kärkkäinen, 1985; Alanko, 1985), which in part originate in the survey design (Mulford, Fitzgerald, 1981). More specifically, much attention has been given to the type of consumption measurement. In this study we will compare two methods of data collection, the weekly recall and diary method of which the former is an established questionnaire method in European surveys. The two methods will be compared with regard to their estimates of overall weekly consumption, their coverage of annual sales and the rank order of drinkers. Although both methods rely on self-reports of actual consumption, the diary method is thought to be more reliable since it is supposed to reduce the time interval between actual consumption and the reporting of that consumption to about 1 day. For the weekly recall this interval runs to 7 days, during which failure of recall might more easily occur.

Special attention will be given to an important issue in interpreting the consequences of measurement error namely *whether underestimation is uniform across all levels of consumption* (de Lint, 1981; Hyman, 1981; Mulford and Fitzgerald, 1981; Popham and Schmidt 1981; Wechsler, 1981).

If not, in which case the reporting error is nonlinear, a low coverage rate means that results based on questionnaire data are systematically distorted. Past studies comparing both methods are not conclusive in that diary self-reports do not invariably yield higher estimates of consumption (Gerstel *et al.*, 1980, Poikolainen and Kärkkäinen, 1983; Uchalik, 1979; Williams *et al.*, 1985). In the discussion section a possible explanation for the differences in results will be formulated.

3.2 Method

The data on alcohol consumption used in this study were collected April-June 1983 as part of a Dutch nationwide survey (20-64 year-olds) on life-style and subjective health assessment. The first part of the survey consisted of face-to-face interviews with items concerning health problems, medicine consumption, smoking and alcohol consumption. The overall response rate was 71%. Response rates were slightly lower for the young (20-24 year-olds) and the lowest socioeconomic class. In a previous report (Halfens, 1985) no significant differences were shown in alcohol consumption between participants and those nonparticipants about whom additional information on consumption was available.

Respondents were asked to recall their consumption of alcoholic drinks on each day of the previous week, with a detailed description of location. No specification as to type of beverage was applied. The amount of alcohol in each reported glass, the basic unit of measurement, is assumed to be equal for all beverage types (approximately 10 g). Those respondents who reported no alcohol consumption in the foregoing year were labeled abstainers.

At the end of the interview a randomly selected subsample of 599 respondents was asked to keep a diary for 14 consecutive days; 496 respondents (83%) completed the diary. The major reason for not completing the diary was refusal (57%), followed by reasons of ill-health (7%) and holidays (5%). For 32 persons (31%) the reason for nonparticipation was not recorded by the interviewer.

The diary consisted of questions regarding consumption of medicines and beverages including alcoholic drinks, in the morning, afternoon and evening. Each day in the diary period was specified by the interviewer. No specifications of type of alcoholic beverage or location were made. Respondents were instructed to record their drinks the morning of the following day. After 2 weeks the interviewer personally collected the diary questionnaire.

The diary subsample does not differ from the original sample with regard to sex, age and social class. Average consumption per week (9,8 glasses), as measured by the weekly recall method, of the 697 respondents for whom there are no diary reports is only slightly lower than for those who completed the diary (10.1 glasses). No difference in abstention was found between the two subsamples. In effect, there are no indications of systematic deviations on relevant variables due to sampling errors or nonparticipation in the diary subsample. The subsequent analysis of diary and weekly recall estimates of alcohol consumption will be confined to drinkers ($n=399$). Therefore, 89 abstainers are excluded as well as eight respondents for whom data on alcohol consumption are missing.

3.3 Results

Average reported alcohol consumption in either diary week is significantly higher than in the week preceding the interview ($p < .001$). The diary estimates of mean (\pm SD) weekly consumption (12.4 ± 12.9 and 12.3 ± 12.9 glasses, respectively) are 22% higher than the mean weekly recall estimate (10.1 ± 10.1 glasses). (A standard glass contains 10 g alcohol). The difference in mean consumption between the 2 diary weeks is negligible (0.1 glass).

According to sales estimates, per capita consumption in 1983 of the population aged 15 years or over was 17.3 glasses per week (Produktschap voor Gedistilleerde Dranken, 1984). After correction for abstainers, the coverage rate of the survey estimates of consumption increases from 48% for the weekly recall estimate to 59% for the diary estimate. Because of seasonal variation (sales are highest in December) per capita consumption in spring is likely to be somewhat lower. On the other hand, the limited age range in the sample (20-64 years) and the assumed higher mean consumption in this age category inflate the above coverage rates. No exact information is available which allows us to adjust for these two factors. It could be argued that the difference between diary and weekly recall estimates of mean consumption might be a result of systematic variation in consumption over different weeks. Although individual consumption will surely vary over weeks, causing large random error and consequently low reliability, there is reason to believe that with sufficient sample size and dispersion of survey weeks, the variation over weeks will not systematically affect aggregate consumption measures (Alanko, 1985).

The very small difference in mean consumption between the two diary estimates supports this supposition. Visual inspection of the distribution of the individual differences in consumption between these two diary estimates

Table 3.1. *Weekly recall and diary estimates of average alcohol consumption at different levels of consumption (N = 399).*

Consumption level level (glasses per week)	N	Weekly recall (A)	Diary week1 (B)	Difference (A-B)	Difference proportionate to diary week1
3 or less	93	2.45	1.23	1.22	+ .98
3 < c ≤ 10	124	5.83	6.35	-0.52	-.08
10 < c ≤ 17	87	10.62	13.75	-3.13	-.23
17 < c ≤ 28	53	17.30	21.17	-3.87	-.18
More than 28	42	29.67	41.20	-11.53	-.28

showed that the values are symmetrically distributed around the mean (0.1 glass). Consequently, the individual variation in drinking over weeks seems to be of a random nature and does not systematically affect aggregate measures. The higher diary estimates of alcohol consumption can therefore be attributed to a real difference between the two methods of data collection. A plausible explanation is the aforementioned reduction of the time interval between actual consumption and reporting that consumption in the diary method.

Table 3.1 shows the mean number of glasses in the first two weeks for 5 different consumption categories. Drinkers are classified according to their average consumption in the diary weeks. It is clear that underreporting increases with consumption. At the lowest level (up to 3 glasses), there is a tendency to overreport consumption. Room (1985) suggested that the so-called "forward telescoping" tendency among infrequent drinkers -a

Table 3.2. *Drinkers with an average weekly consumption over a given level, according to weekly recall and diary consumption measures, in percent (N = 399).*

Consumption level (glasses/week)	Weekly recall ^a	Diary week 1	Diary week 2
3 or over	73.9	77.4	78.7
7 or over	52.6	59.6	58.1
14 or over	27.8	34.8	35.8
21 or over	14.8	18.8	19.5
28 or over	6.8	11.0	11.3
35 or over	3.3	5.8	5.5
42 or over	1.8	4.5	3.8

^a Two-tailed Wilcoxon test for difference between weekly recall and both diary weeks, $p < 0.0001$.

tendency to report drinking occasions prior to the reference period- might lead to overreporting of consumption in this category. The Spearman rank order correlations between weekly recall and diary week 1 (.75) and week 2 (.78) estimates are high compared to the correlation of .84 between the two diary estimates, indicating that the relative position of drinkers on the consumption scale is fairly stable across methods.

Estimates of the prevalence of drinkers at the upper levels of consumption however increase considerably if diary measures are used (table 3.2). The proportion of consumption underreported with the weekly recall method is highest in the upper three consumption categories (table 1). This does however not, however, imply a nonlinear relation between consumption level and underreporting. To inspect this relation we have plotted the differences in consumption between the two methods as a function of average consumption (figure 3.1). Both a linear and a nonlinear model were fitted to

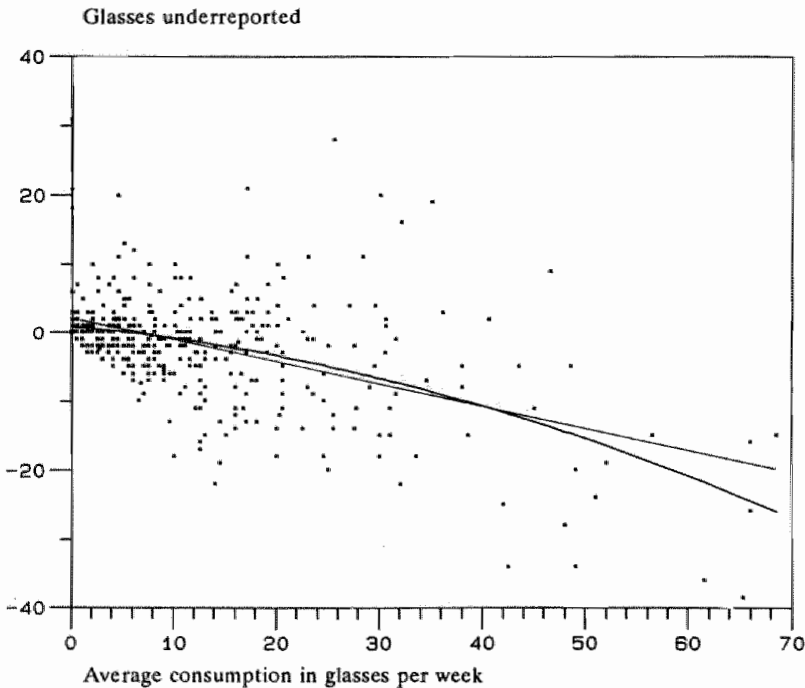


Figure 3.1 *Plot of differences between weekly recall and diary week 1 consumption measures as a function of average consumption and least square regressions (linear and nonlinear).*

the data, of which the regression lines are drawn in figure 3.1. Although the nonlinear model gives a somewhat better fit, the difference with the linear model is marginal. Moreover, inspection of the residuals of the linear model revealed no systematic variation. The assumption, therefore, of a linear relationship between consumption level and underreporting, due to the difference between the two methods, seems warranted. Linear correlation between average consumption (defined by consumption level in the 2 diary weeks) and degree of underreporting (difference between diary week 1 and weekly recall consumption) is relatively high (.49).

3.4 Discussion

The results show that diary estimates of average weekly alcohol consumption in the Dutch population are 22% higher than those based on a weekly recall questionnaire. With the use of diary self-reports the overall underreporting of consumption is reduced from 52% to 41%.

An obvious explanation for the reduction in underestimation is the assumed shorter time interval between actual consumption and reporting that consumption which minimizes recall failure. In support of this supposition, Philipsen *et al.* (1983) using similar weekly recall data from a Dutch survey held in 1981, found a decline in average number of reported glasses consumed with an increase of recall period. When the length of the recall period was limited to a maximum of 2 days, the reporting of average consumption rose by about 12%. Note, however that since it is not known whether respondents in fact recorded their consumption every day of the diary week, no definite statement about this alleged effect of length of recall period can be made.

Uchalik (1979) and Poikolainen and Kärkkäinen (1983), comparing diary and questionnaire methods, also report significantly higher diary estimates. It should be noted that the questionnaire measures in these studies were not based on actual consumption but on usual quantity and frequency of drinking, and that their subjects were sought by advertisements or personal contacts. In the latter study many subjects felt, while keeping a diary, that their drinking was considerably higher than expected. As they were explicitly instructed not to change drinking habits during the research period, many of them resolved to reduce their alcohol consumption in the future. This effect of self-monitoring might help explain the conflicting results reported by Gerstel *et al.* (1980) and Williams *et al.* (1985) who did not find the diary method to be superior. The diary in their surveys covers consumption over a

period of 30 days (14 days in the present study). Because of this longer period of self-monitoring and increased awareness of drinking habit, a gradual reduction of actual consumption may have occurred. In the present study there is no systematic decline in consumption in the second week.

Apart from the length of the recall period, another distinction between diary and questionnaire methods can be made. Diary self-reports are recorded in the absence of an interviewer. Because questions about drinking are to some extent threatening, self-monitoring of consumption may have resulted in more honest reports. In an extensive study on response effects, however, Bradburn and Sudman (1979) concluded that presence or absence of an interviewer made little difference in terms of response distortion. Although both effects cannot be separated in this study, it seems likely nevertheless that failure of recall is more important in explaining the difference in estimates of consumption.

The outcome, that part of the undercoverage in surveys is linearly related to consumption level, reduces to some degree the possibility that undercoverage will seriously affect the validity of estimates of consumption. This does not preclude that error from other sources will systematically bias consumption estimates. Research strategies that control various sources of error are therefore needed to assess the impact of these errors on validity. For example, most surveys do not cover consumption during special events like Christmas, carnival or holidays when drinking is supposed to be particularly high. Systematic exclusion of these events from the survey period may invalidate consumption measures. Another problem concerns the unit of measurement (usually 1 glass). As Wilson (1981) has pointed out, the average glass of spirits consumed at home contains more alcohol than the assumed 10 g. Furthermore, there are indications that, at least in some instances, selective nonresponse or incomplete sample frames may bias consumption data from general populations (Crawford, 1986; Wilson, 1981). Knowledge of the contribution of these sources of error make to the problem of undercoverage and of the conditions under which these errors are likely to occur, will enable a better assessment of the threats they constitute to validity.

References

- ALANKO, T. (1984) An overview of techniques and problems in the measurement of alcohol consumption. In: Smart, R.G., Capell, H.D., Glaser, F.B., Israël, Y., Kalant, H., Popham, R.E., Schmidt, W. and Sellers, E.M. (eds) *Research advances in Alcohol and Drug Problems*, Vol. 8, pp. 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/Alcohol Epidemiology Section meeting, Rome, 1985.
- BRADBURN, N.M. and SUDMAN, S. (1979) *Improving interview method and questionnaire design: Response effects to threatening questions in survey research* (San Francisco, Jossey-Bass).
- CRAWFORD, A. (1986) A comparison of participants and non-participants from a British general population survey of alcohol drinking practices, *Journal of the Market Research Society*, 28, pp. 291-297.
- DE LINT, J. (1981) "Words and deeds": Responses to Popham and Schmidt, *Journal of Studies on Alcohol*, 42, pp. 359-361.
- GERSTEL, E., HARFORD, T.C. and PAUBLER, C.C. (1980) The relative validity of a chronological drinking record, *Drug and Alcohol Dependence*, 6, pp. 359-364.
- HYMAN, M.M. (1981) "Words and Deeds": Responses tot Popham and Schmidt, *Journal of Studies on Alcohol*, 42, pp. 361-362.
- KNIBBE, R.A., DROP, M.J., Van REEK, J., SAENGER, G. (1985) The development of alcohol consumption in the Netherlands: 1958-1981, *British Journal of Addiction*, 80, pp. 411-419.
- MIDANIK, L. (1982) The validity of self reported alcohol consumption and alcohol problems: A literature review, *British Journal of Addiction*, 77, pp. 357-382.
- MULFORD, H.A. & FITZGERALD, J.L. (1981) "Words and Deeds": Responses to Popham and Schmidt, *Journal of Studies on Alcohol*, 42, pp. 362-367.
- PERNANEN, K. (1974) Validity of survey data on alcohol use. In: Gibbens, R.J, Israel, Y. Kalant, H., Popham, R.E., Schmidt, W. & Smart, R.G. (eds) *Research Advances in Alcohol and Drug Problems*, Vol. 1, pp. 335-373 (New York, Wiley).
- PHILIPSEN, H., KNIBBE, R.A. & Van REEK, J. (1983) Alcohol consumption in the Netherlands as a social phenomenon. In: Hermus, R.J. (ed) *Alcohol, Health and Society*, pp. 53-70 (Zeist, TNO).
- POIKOLAINEN, K. & KÄRKKÄINEN, P. (1983) Diary gives more accurate information about alcohol consumption than questionnaire, *Drug and Alcohol Dependence*, 11, pp. 209-216.
- POIKOLAINEN, K. & KÄRKKÄINEN, P. (1985) Nature of questionnaire options affects estimates of alcohol intake, *Journal of Studies on Alcohol*, 46, pp. 219-222.
- POPHAM, R.E. & SCHMIDT, W. (1981) Words and deeds: the validity of self-reported data on alcohol consumption, *Journal of Studies on Alcohol*, 42, pp. 355-358.
- PRODUKTSCHAP VOOR gedistilleerde dranken (1984) *Jaarverslag 1983* (Schiedam, PGD).

- PRODUKTSCHAP VOOR gedistilleerde dranken (1986) *How many alcoholic beverages are being consumed throughout the world?* 25th Edition (Schiedam, PGD).
- ROOM, R. (1985) Measuring alcohol consumption in the U.S.: Methods and rationales. Paper presented at the ICAA/Alcohol Epidemiology Section meeting, Rome, 1985.
- UCHALIK, D. (1979) A comparison of questionnaire and self-monitored reports of alcohol intake in a nonalcoholic population, *Addictive Behaviors*, 4, pp. 409-413.
- WECHSLER, H. (1981) 'Words and Deeds': Responses to Popham and Schmidt, *Journal of Studies on Alcohol*, 42, pp. 367-368.
- WILLIAMS, G.D., AITKEN, S.S. & MALIN, H. (1985) Reliability of self-reported alcohol consumption in a general population survey, *Journal of Studies on Alcohol*, 46, pp. 223-227.
- WILSON, P. (1981) Improving the methodology of drinking surveys, *The Statistician*, 30, pp. 159-167.

Measuring Quantity and Frequency of Drinking in a General Population Survey.

A Comparison of 5 Indices.

SUBMITTED TO THE JOURNAL OF STUDIES ON ALCOHOL, DECEMBER, 1989.
PAUL H.H.M. LEMMENS, FRANS E.S. TAN, RONALD A. KNIBBE.

Summary

The paper compares 5 indices of alcohol consumption in a general population survey conducted in 1985 in the Netherlands. Self-reports of consumption were obtained with a prospective diary, a retrospective 7-day recall method, and 3 summary measures, such as a quantity-frequency index. The coverage of sales data appeared highest for the diary (67%), which suggests a higher validity. Special attention was given to comparisons of quantity and frequency of drinking between the diary, on the one hand, and the weekly recall and "summary" measures, on the other. It was found that underreporting, relative to the diary reports, was generally higher in the frequency than in the quantity domain. This result, together with the finding from longitudinal studies that intra-individual variation is also higher for drinking frequency, leads to the conclusion that forgetting is a potent source of undercoverage in surveys and to the hypothesis that large differences in overall drinking pattern between populations (e.g. in regularity of drinking) may account for the large differences in coverage rates of sales data. Furthermore, the subjectively assessed probability of drinking by means of a "usual" frequency question, appeared a poor predictor of (diary) drinking

frequency for respondents reporting a low or moderate frequency. For subjects claiming a high "*usual*" drinking frequency, a reasonable correspondence between diary and summary measures was found. This mitigates the fear often expressed that heavy drinkers particularly underreport their consumption.

4.1 Introduction

The problem of the measurement of alcohol consumption, in particular the issue of the validity and reliability of self-reports, is a regularly recurring topic in scientific discussions. It is clear to everyone that self-reports on alcohol consumption cannot be trusted in all circumstances. No one would recommend traffic police to rely on drivers' self-reports of consumption. In population surveys the threat imposed by questions about drinking is probably much less and deliberate under-reporting is, therefore, not the only or most important factor responsible for the large discrepancy between survey and sales estimates of consumption, often referred to as the undercoverage problem.

Sampling errors, non-response bias, errors in procedures for calculating total consumption, in questionnaire design and in question wording are researcher-based and more or less avoidable (Mulford & Fitzgerald, 1981). Other errors are attributable to the respondents' failure to answer questions correctly, due to factors such as failure to recall drinking events, incorrect dating of these events or incorrect details of the events. These factors are less easy to manipulate. For instance, the degree of forgetting in surveys that collect retrospective data can be minimized by limiting the length of the recall period, however, only at the expense of losing information on individual behavior patterns, an enlarged required sample size and, consequently, higher costs (Kemsley, 1979).

In the past decades several types of retrospective self-report methods have been used in surveys on alcohol. They can be roughly divided into two categories, so-called summary measures and measures based on recent drinking occasions (Room, 1985). The summary measures require the respondent to evaluate (aspects of) his or her behavior over a certain period of time, usually ranging between one month and one year. They are based on questions about frequency and quantity of drinking and are used mostly in North American surveys. The recent occasion approach uses self-reports of actual consumption. The respondent is asked to list all the drinking events that have occurred in the near past. This type of method is often used in European surveys.

It is clear that annual consumption estimated by these two types of method is liable to different distortions (Redman et al., 1987). Those estimates based on a listing of recent occasions (e.g. during the previous 7 days) are not only sensitive to memory failure but also to the rather large time-variation of individual drinking behavior. The shorter the recall period and the larger the time-variation of drinking, the larger the risk that atypical periods are covered by the recent occasion methods. This latter risk is thought to be absent in the summary approach. Consumption in these questionnaires is summarized over a much longer period (from 30 days to 1 year) and the subject is asked to correct for episodes in which drinking was atypical. Main disadvantage of this type of method is that there is neither a uniform procedure nor knowledge of the mental processes by which the respondents (should) arrive at an estimate of their, often complex, drinking behavior. One can easily imagine that other factors besides recall ability affect the accuracy of an answer.

Apart from the more fundamental issues, such as parameters of the process of memory decay, one could raise the question of comparability of empirical data obtained with different methods. In an effort directed to a quantification of the effects different questionnaire methods have on empirical, self-reported consumption data, a Dutch nationwide population survey on alcohol consumption was designed that permitted the comparison of 5 indices of total alcohol consumption.

In this chapter special attention is given to two methods that record actual, recent drinking occasions but which vary in recall period (a retrospective, weekly recall questionnaire and a prospective diary) and to the relative accuracy of subjective, summary measures of typical and atypical drinking frequency and quantity per occasion over a longer period of time.

4.2 Methods

4.2.1 *Measures of average weekly consumption*

All methods in the survey on alcohol, conducted in 1985, have their particular rationale and serve specific purposes. The choice of a particular version of a type of method (out of many alternatives) has been made mainly on practical grounds, such as comparability with previous Dutch surveys. The survey consisted of a personal interview at home and a diary, to be completed in private. A respondent is labeled abstainer if he or she reported not to have drunk any alcoholic beverage during the past 6 months.

Weekly recall (WR)

The so-called weekly recall method asks respondents to report retrospectively about their drinking that took place on the 7 days prior to the day of the interview. In the 1985 survey, type of day, location (home, outdoors) and type of beverage (beer, wine, fortified wine and distilled spirits) were specified. The unit of home beer consumption was the bottle or can, to be chosen from 5 current types. This was converted into glasses of 25 cl accordingly. For the other beverages, unit of consumption was the standard glass (which contains about 10 g or 1.25 cl of 100% alcohol, for any alcoholic drink - *see footnote 1*).

Diary (D)

Respondents were asked to keep a diary for 7 consecutive days, in the week following the interview (N=918) or during the Christmas holidays (N=322). Among other items, such as type of activity, smoking, consumption of non-alcoholic drinks and non-illicit drugs, type of beverage (equal to the one above) and number of glasses drunk had to be specified by the respondent for 15 minute-periods. The diary was to be filled out at least once a day. After one week, the diaries were collected by the interviewer. In contrast to the methods described below, the diary and weekly recall methods record actual drinking occasions over a certain period of time and are, therefore, best suited for the assessment of temporal features of drinking patterns.

Quantity-frequency (QF)

A third index has been constructed, widely known as the quantity-frequency index. From a list of 8 options, ranging from "never" to "every day", the respondents were asked to select their modal or usual frequency of drinking during the past 6 months. The frequency categories are exhaustive and cover the entire 6 months period. A second, open-ended question enquired about the number of glasses usually consumed on a drinking occasion in that period. Neither type of beverage, nor location was specified. After conversion into an average weekly frequency (by taking the midpoints of the categories; e.g. 1-3 times per month becomes 0.5 times per week), the quantity and frequency scores were multiplied to obtain an estimate of total weekly consumption.

Quantity-frequency-variability (QFV)

Since questions about modal or usual frequency and quantity do not cover exceptional heavy consumption, respondents were also asked about their frequency of atypical or variable heavy drinking in the 6 month reference period. Heavy drinking was defined as 6 glasses or more on one occasion (the so-called variability score, after Cahalan, *et al.* 1969). The frequency options were equal to the ones used in the usual frequency question (F). Similarly, the variability score (V) was converted into a weekly frequency and multiplied by 6 (glasses) to obtain a volume estimate of unusual, heavy consumption per week. This product was added to the QF score of each

individual. In a few cases a different procedure was adopted to avoid duplication of reported drinks. If a person reported a usual quantity (Q) of 6 glasses or more per occasion, no consumption was added to the QF index, given $F \geq V$. If $F < V$, and $Q \geq 6$ then the product of Q and V was taken as the QFV score. If $F = V$, and $Q < 6$, then QFV consisted of the product of $F \cdot 6$. As in the case of the QF index, location or type of beverage cannot be specified. Initially, the QF and QFV indices have been developed for classification of drinkers into only a few, basic categories. In later years, many different versions have been introduced which aim at a finer discrimination between categories of drinkers. As in the present case, these indices often approximate a continuous scale (e.g. Fitzgerald & Mulford, 1984, 1987).

Typical consumption (T)

This uni-dimensional index is based on only one question about usual daily consumption of any alcoholic beverage. Daily consumption has been multiplied by 7 to obtain weekly consumption. Methodologically unsophisticated, though easy to administer and low in costs, the method has been included in this survey mainly for comparative purposes.

Since accuracy of recall of many, not only behavioral, events decreases rather rapidly with increasing reference period (e.g. in many studies an exponential relation has been found, see: Sudman & Bradburn, 1973), the diary method is thought to be superior to the weekly recall method because of the assumed shorter time interval between consumption and reporting that consumption. In a study using Dutch survey data from 1983, Lemmens *et al.* (1988a) indeed found that a prospective diary method yielded a 22% higher average weekly consumption than a method based on recall over 7 days preceding the interview. They conclude that recall failure on part of the respondent is probably an important source of undercoverage in surveys.

A possible drawback of the diary method is what has been called a sensitization or self-monitoring effect: while keeping a diary people become aware of their drinking habits. Often, people drink more than they themselves expect and this awareness may influence their drinking habits. For periods as short as one or two weeks, however, this problem seems trivial (see: Lemmens, *et al.*, 1988a). Advantage of both the diary and the recall method is that they refer directly to behavior and do not require respondents to make abstractions or a subjective evaluation of their drinking habits, as in the other three methods.

In the sections that follow, differences between measures in estimates of total consumption and in distribution are presented first. The analysis proceeds with an investigation of aspects of the drinking pattern, namely frequency, quantity and stability of drinking. The conclusion that measures empirically differ mainly in reported frequency of drinking, leads to the hypothesis that the coverage rate of survey data, is, at least in part, dependent upon prevailing drinking patterns in the target population.

4.2.2 *The sample*

The sample frame consisted of all 16-70 year-olds registered at municipal population offices. These registers are fairly complete and accurate. A response rate of 70% was obtained. Non-response analyses, consisting of a follow-up by telephone and a correction for differences in response probabilities, revealed no significant bias (Lemmens, *et al.*, 1988b).

Exclusion of persons younger than 16 years and older than 70 years from the sample frame inflate undercoverage rates. Since the per capita sales figures of beverage alcohol have not changed much between 1981 and 1985, it was thought that drinking habits would not have changed very much either and that adding a weighted sample of the Dutch population, aged 12-15 years, who kept a diary in 1981 to the 1985 sample would give a reliable estimate of the proportion of the total volume consumed in the youngest age category. It was found that consumption in this age category amounted to only 1.1% of the consumption of the total population (up to 70 years of age). Estimates of total sales in 1985 have been corrected accordingly. No estimates of the consumption of the population over 70 years of age are available and the sales estimates could not be corrected. Therefore, the sales estimates of average consumption in this study are still somewhat inflated.

Interviews were spread over 10 months in 1985 (March through December) in order to correct for seasonal variation. Those subjects who completed the diary during the Christmas period ($N=322$) were excluded from most analyses since these weeks are exceptionally "wet".

No scores on *WR*, *QF*, *QFV* are available for self-reported abstainers ($N=213$) and they are also excluded from further analyses. The consistency in reported abstinence (in the personal interview) was checked with scores on the diary, in the week following the interview. It appeared that 7 men and 17 women (ca.13%) reported to have drunk during the week following the interview. The possible error made by excluding them seems, however, small since their diary consumption was very low.

4.3 Results

4.3.1 *Coverage rates*

A comparison of the 5 measures of mean weekly consumption and respective coverage rates of sales data is shown in table 4.1. Consumption scores have been weighted according to sex, age and region (32 cells) to correct for deviations in representativeness. Mean per capita consumption according to sales estimates in 1985 was 16.2 glasses per week, corrected for

the consumption of 12-15 year-olds. Respondents who filled out the diary during the Christmas holidays (two, exceptionally "wet" weeks) were left out of the comparisons with the diary.

The coverage rate appears lowest for the *QF* index and highest for the prospective diary. Addition of variable heavy use (*V*) to the *QF* in the *QFV*-index, raises average consumption considerably by about 25%. Quite surprisingly, the simple measure *T* does not really deviate from the *WR* or *QFV* averages. The coverage rate of 58% of the *WR* is about 10% higher than the rate of similar methods in Dutch surveys in 1958 and 1981, 49% and 44%, respectively (Knibbe *et al.*, 1985).

Although average consumption is the only measure that can be compared directly to sales, it is not the only criterium of validity. It is also useful to consider the agreement between the 5 measures and to have a closer look at the distributions. The many methodological differences between the indices (e.g. reference period) are expected to result in correlations that are far from unity. Rank order correlations are presented in table 4.2. The table shows that the correlations between the measures are of about the same magnitude in all comparisons. The highest agreement (.90) is found between the *QF* and *QFV* indices. Obviously, addition of infrequent heavy use to the *QF* index does not result in a large shift in rank order.

The lowest correlations in table 4.2 are found for the typical daily consumption measure *T*, which seems to stand apart from the other measures. Closer examination of the distributions, presented in table 4.3, show that the lowest consumption levels (below 7 glasses per week) are underrepresented in *T*. Apparently, the question about daily use of any alcoholic beverage leads to an undesirable grouping of scores. The scores on the question about typical daily use have more often been rounded to 1 than to 0, if the person claims to drink at all. Multiplication with a factor 7 has lead to a over-representation of the "7 drinks per week" category, and,

Table 4.1. Comparison of estimates of average weekly alcohol consumption (in # glasses) based on 6 indices and their respective coverage rates in the 1985 survey (*N* = 918).

method	average weekly consumption	coverage rate
1. sales in 1985	16.2	100%
2. diary (D)	10.9	67%
3. weekly recall (WR)	9.4	58%
4. typical daily cons.(T)	9.4	58%
5. QFV index	9.2	57%
6. QF index	7.4	46%

Table 4.2. Rank order correlations (Kendall's tau-b) between 5 indices of alcohol consumption from the 1985 survey ($N \approx 750$ for diary; $N \approx 1014$ for other measures; abstainers excluded).

index	QF recall	QFV	Typical	Weekly
QFV	.90			
Typical	.61	.61		
Weekly recall	.66	.66	.60	
diary	.61	.61	.58	.65

consequently, an over-estimated mean consumption. The rounding of scores is also present at other values, but most probably bi-directional, and thus with less effect on estimates of mean consumption.

4.3.2 Comparison of frequency and quantity: actual consumption measures

Total consumption can be thought of as being composed of a frequency component and a quantity component. In order to specify the difference between the weekly recall and diary estimates of consumption, found in the previous section, total volume has been broken down into frequency of drinking and average quantity consumed on a drinking occasion.

Usually, frequency of drinking refers to occasions or events that are defined by situational aspects (e.g. people, places) rather than by temporal ones (e.g. time of day, duration). Drinking frequency, then, equals the number of times (per day, week or month) a person enters a situation in

Table 4.3. Cumulative distributions (percentages) of 5 measures of alcohol consumption in the 1985 survey ($N \approx 750$). Consumption level in glasses per week.

gl/wk	T %	QF %	QFV %	WR %	D %
<0.8	8	17	16	17	15
<2.0	8	28	25	22	20
<3.5	8	40	35	34	29
<7.0	8	52	47	49	41
<12	66	71	67	63	57
<18	84	86	79	79	71
<31	95	95	90	90	88
<53	99	99	98	97	97
<71	99	99	99	99	99

which at least one drink is taken. In the 1985 survey, unfortunately, neither the *WR* nor the diary method allow for a full specification of drinking occasions. The *WR* inquires only about drinks that are consumed at home or outdoors (a full specification would require a large number of questions). The diary defines a drinking occasion by one of 27 activities and the reported activity does not necessarily coincide with the situational definition. Because of these limitations, frequency of drinking in this paper is operationally defined as "drinking days per week" and quantity per occasion as "*number of drinks per day*".

Since it does not seem likely that respondents, recalling their consumption on the previous 7 days, as in the *WR* method, differentiate between drinks consumed before or just after midnight, "days" in the diary week have not been defined by the usual calendar, with the transition at midnight, but by "days" with the transition at 7.30 a.m. At this hour, all respondents have stopped drinking or have not yet begun. This adjustment of the time of transition in the diary index changes the diary frequency of drinking in 74 cases (11.5% of those who report drinks in the diary week). Average frequency drops slightly (0.7%) and average quantity per occasion rises by 1.5%.

The comparison of *WR* and diary estimates of frequency and quantity per occasion is shown in table 4.4. It appears that the difference in total consumption between the diary and the weekly recall method is primarily caused by a difference in average drinking frequency rather than by a difference in average quantity per occasion. Average frequency, measured by the *WR* method, is lower both in men and in women, whereas estimates of average quantity do not clearly differ.

Postulating decay of memory as the primary source of the discrepancy between the *WR* and diary method, one could also assume that respondents who try to remember their drinking behavior on the previous 7 days (*WR*) fail to recall the exact number of drinking occasions rather than the number of drinks on the recalled occasions. Figure 4.1 shows that the moderate frequencies (4,5 and 6 days) are most likely to be underreported when the *WR* method is used.

A possible explanation for the above result comes from a Finnish study using individual, longitudinal data (Alanko, 1985). It was found that intra-individual variation in weekly consumption over 27 weeks was primarily caused by a fluctuation in average frequency rather than by a variation in the average quantity consumed on an occasion. Alanko suggested that "*time-variability of drinking is dominated by a stochastic process governing the occurrence of drinking occasions and not by the variability in amounts consumed*" (p.8). As Alanko pointed out, the larger the time-variation in frequency, the larger the random error in estimates of average consumption, based on a single week. However, it is equally plausible that the larger the

Table 4.4. *Average drinking frequency (days per week) and average quantity per occasion (glasses per day) for non-abstaining men and women (s.d. in brackets); comparison of WR and QF with diary reports.*

	Diary	WR	QF index
average frequency			
men (n≈430)	4.12(2.4)	3.71(2.6) ¹	3.47(2.6) ¹
women (n≈320)	3.10(2.5)	2.65(2.5) ¹	2.35(2.4) ¹
average quantity ³			
men (n≈380)	3.94(2.8)	4.08(3.1) ²	3.72(2.9) ¹
women (n≈250)	2.53(1.4)	2.42(1.6) ²	2.42(1.7) ²

1 Wilcoxon matched-pairs signed-ranks test; p. < 01.
 2 id.; difference with D not significant at a 1% level.
 3 arithmetical average over those persons drinking at least on one occasion.

variability in the number of occurrences over weeks, the more mistakes will be made in recalling these occasions, which then results in a larger undercoverage.

It does not seem far-fetched to hypothesize that a greater stability in drinking frequency over weeks will result in fewer memory errors in

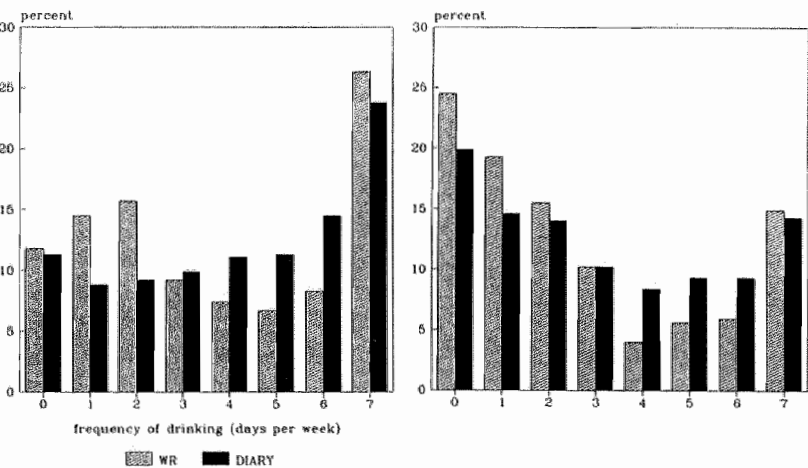


Figure 4.1. *Frequency of drinking (days per week) according to weekly recall method and diary, for men (left) and women (right); abstainers excluded.*

retrospective data. A possible reason, then, for the apparent under-representation of intermediate frequencies at the aggregate level is that drinking at these levels is less stable over weeks. In the present study, stability at the individual level is, however, hard to investigate empirically since all scores are with measurement error and the methods each cover only a single week. Nevertheless, and at the risk of inclusion of false positives and exclusion of false negatives, a stable frequency in this paper has been defined as an equal reported frequency in both *WR* and diary. Stability in reported drinking frequency over the diary and weekly recall weeks, presented in figure 4.2, is highest for the most infrequent drinkers (0 days in both weeks) and for the most frequent drinkers (7 days in both weeks), thus lending at least partial support to the hypothesis that under-reporting is greater at the less stable, intermediate frequency levels (4, 5 or 6 days).

If the recall problem is in fact smaller for respondents whose drinking rhythm is more regular, then one might speculate that differences in regularity of drinking at the aggregate level (e.g. between countries) could lead to differences in the proportion of drinking occasions recalled, and, consequently, in total reported consumption. Hence, this would account for the large differences in coverage rates of the *WR* method between countries that differ in prevailing drinking style. For instance, the coverage of sales data with the *WR* in the Netherlands (ca.50%) is consistently higher than the

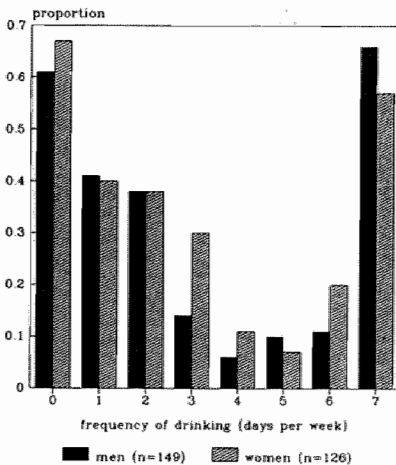


Figure 4.2. Proportion of drinkers in each frequency category with an equal drinking frequency in the diary and weekly recall reports (abstainers excluded).

coverage in Finland (ca. 30% ; Simpura, 1988). For the moment, a test of this hypothesis is not possible because of the lack of comparable longitudinal data on drinking rhythms.

If the frequency and quantity component of consumption are affected differently by failure of memory, then one would expect the relation between the two variables in the *WR* and diary to differ, as well. Figure 3 reveals a clearly positive correlation between diary frequency of drinking and diary quantity consumed per occasion in the male subsample (Pearson's $\rho = .39$, after adjusting for outliers). For women the correlation between diary frequency and diary average quantity seems less pronounced ($\rho = .31$, after adjusting for outliers) which might be a reflection of the more strict proscriptions for women in society as regards (maximum) drinking levels. The correlation is absent in the weekly recall data ($\rho = .07$ and $.05$ for men and women respectively, after adjusting for outliers). Retrospectively reported average (*WR*) consumption per drinking day is highest for men who drink 4 and women who drink 5 days per week. This inconsistency casts some doubt on the validity of empirically established relationships between drinking variables. The relationship in the present study is clearly dependent upon the type of method used.

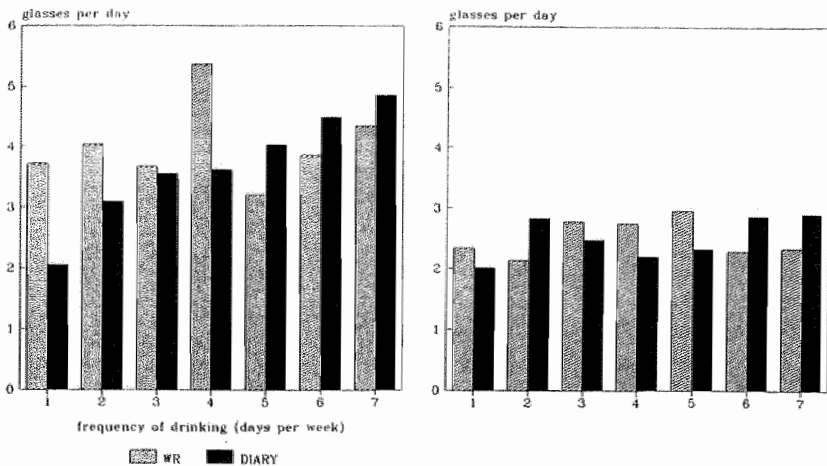


Figure 4.3. Relationship between average quantity consumed per day (# glasses) and reported frequency (days/week) according to the diary and the weekly recall method, for men (l) and women (r); drinkers with zero week consumption are excluded.

4.3.3 Comparison of frequency and quantity: actual versus usual consumption measures

The QF index of alcohol consumption requires respondents to abstract or generalize their drinking behavior, in particular, their usual frequency of drinking and the quantity usually consumed on an occasion. Apart from other disadvantages over actual consumption measures (Redman *et al.*, 1987), one can imagine that this subjective evaluation of one's drinking behavior is particularly prone to both deliberate and unconscious distortion. A comparison of QF reports with those from actual consumption measures, and the diary in particular, can give information about their validity.

A comparison of average drinking frequency and average quantity per drinking occasion, as measured by the diary and QF-method, is shown in table 4 and leads to a conclusion similar to the one in the previous section: the difference in total volume between diary and QF index is primarily caused by a difference in the frequency domain. Though average usual quantity in the male sample is also significantly lower than in the diary, this is only 0.2 glass.

Moreover, and contrary to the results for the diary reports, the positive relation between frequency of drinking and average quantity per occasion is absent in the case of the usual consumption measure for both sexes, as shown in figure 4. Among drinkers reporting a frequency of at least once a week, even a negative relation between frequency of use and average quantity can be observed.

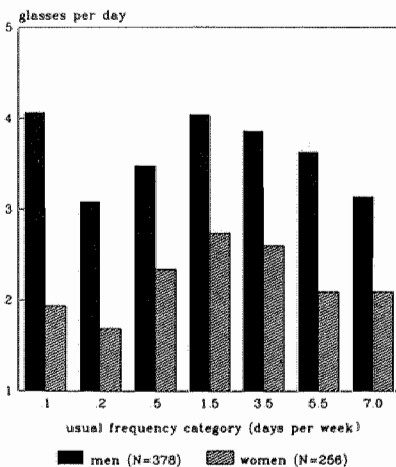


Figure 4.4. Average "usual" quantity consumed per occasion in the "usual" frequency categories, for men and women (abstainers excluded).

It is worth considering how the difference in average frequency between D and QF is distributed over the usual frequency (F) categories. Or, to put it differently, how well does reported usual frequency of people at different F -levels correspond with their actual (D) drinking behavior. In line with Alanko's finding that time-variability in drinking is dominated by a stochastic process, the reports on usual frequency have been reformulated as subjective estimates of the *probability of drinking* in the 6 months reference period. The individual probabilities have been quantified by taking the midpoints of the initial frequency classes and converting them into probabilities of drinking on a random day in the 182-day reference period. For example, the category 'once or twice a week' becomes 1.5 times per week, or 39 times per half year, and results in a probability of drinking of $39/182 = .21$. Those respondents reporting a usual frequency of 7 days per week were, arbitrarily, given a probability of .95. This procedure resulted in 7 probability classes from a very low ($p = .008$) to a very high ($p = .95$) probability.

Since the assumption that drinking events occur independently in time does not seem unrealistic, a binomial model of the distribution of the expected number of occurrences in a 7-day period for each probability class has been adopted. These expected frequencies were then contrasted with the observed frequencies in the diary week. The null hypothesis is that the (diary reports of) actual events can be regarded as realizations of the subjectively assessed probability, provided, of course, that the diary week following the interview does not systematically differ from the 26 weeks in the reference period, preceeding the interview.

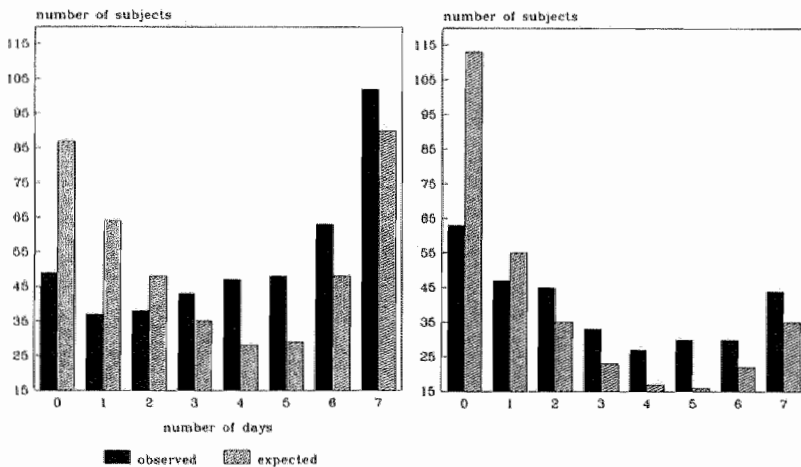


Figure 4.5. Comparison of observed (D) and expected number of drinking days in a 7-day period under a binominal model (F), for men (l) and women

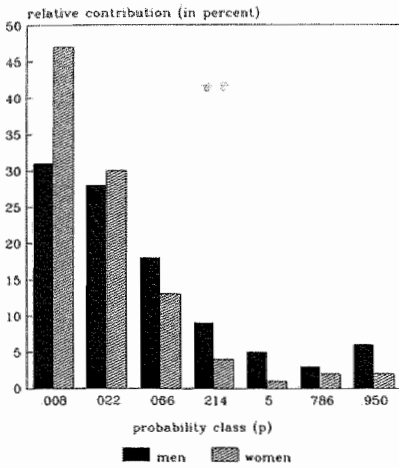


Figure 4.6. *Relative contribution (in percent) of each frequency class (F) to the total discrepancy between expected and observed number of occurrences in a 7-day period, for men and women.*

Figure 4.5 depicts the distributions of the number of drinking occurrences to be expected under the above binomial model and of the observed frequencies in the diary week. The figure shows that, in the aggregate, frequencies greater than 2 days for men and 1 day for women in a 7-day period are underreported by the usual frequency method. As a measure of the discrepancy or relative distance between observed (o) and expected (e), the following statistic was chosen: $\sum (No_{ij} - Ne_{ij})^2 / \text{var}(e_i)$; N_{ij} denotes the number of respondents in each cell, determined by frequency category (i) and number of days in a 7 day period (j); $\text{var}(e)$ stands for the variance in each category. The statistic approximates a chi-square distribution. The measure of the discrepancy between predicted and observed frequencies reveals that the difference is greatest in the lower and moderate probability classes (less than weekly). Figure 4.6 shows the relative contribution (in percent) of each usual frequency class to the total discrepancy between expected and observed frequencies.

The poor representation of actual frequency at the lower F -levels is also illustrated in graphical comparisons of cumulative distributions of expected and observed frequencies. In figure 4.7, examples of 3 such comparisons from the male sample are given. The relative distance between the two graphs decreases as F increases. The results for women are more or less equal.

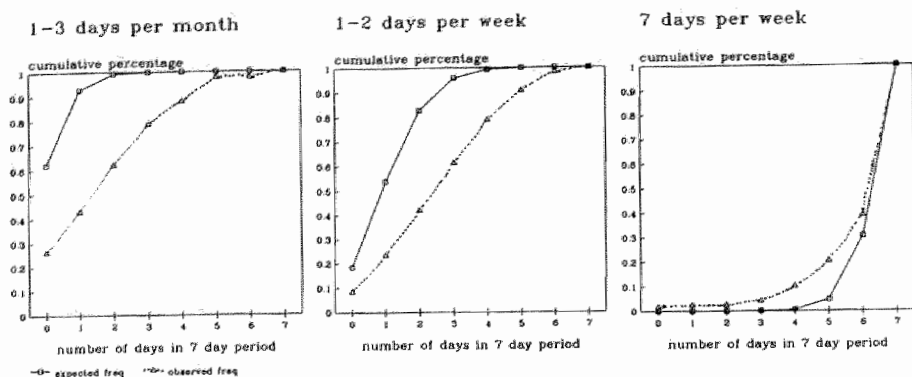


Figure 4.7. Examples of a comparison of cumulative distributions of expected (F) and observed (D) number of respondents in subsequent frequency categories for 3 classes of reported "usual frequency" (F); male drinkers only

4.4 Discussion

In this chapter several types of survey methods on alcohol consumption have been compared. A prospective diary method showed the highest average consumption. This result suggests that recall failure is an important source of undercoverage of consumption in retrospective survey methods. It has been found that underreporting of consumption, relative to the diary method, is greater in the frequency domain than in the quantity domain, for any type of retrospective method. In other words, respondents seem to have more trouble recalling the correct number of occasions that have occurred in the reference period than the correct number of glasses consumed on these occasions.

This conclusion is in line with results from a Finnish study in which frequency of drinking showed a larger intra-individual variation over weeks than quantity consumed per occasion. If this result can be generalized to the present population, one could also hypothesize that underreporting should be less for those with a more regular drinking pattern. This, in turn, could explain the higher underreporting at the, assumingly less stable, intermediate drinking frequencies (WR method). Similarly, one might speculate that large differences in overall drinking pattern (e.g. regularity of drinking) between populations could be responsible for the different coverage rates of sales data found in nation-wide surveys. Such an interaction of drinking pattern

and validity of self-report data has negative consequences for comparability of research data, not only across populations but also within populations over time.

It appears to be a difficult task most respondents are faced with when they are asked to summarize their drinking frequency over a long period of time (*QF* method). The large undercoverage of the *QF* index seems primarily caused by a lower reported frequency of drinking. It has been shown that under a binomial model the subjectively assessed probability of drinking (*F*) is a poor predictor of actual frequency of drinking, as measured by the 7-day diary method. The results showed that the relative accuracy of self-reported frequency (*F*) is poorest among those who claim to drink infrequently. The highest self-reported usual frequencies seem rather well in accordance with the corresponding diary reports. This finding mitigates the fear often expressed by researchers that undercoverage is caused by underreporting of the heavy drinkers.

The inconsistency that was found in the correlation between the frequency and quantity component of drinking behavior when measured by different methods, has also been reported by Alanko (1988). In 2 out of 4 Finnish surveys the relationship between the two variables (measured by 3 different methods) was not in the same direction for all methods. Alanko points to the temporal structure of the relationship and he suggests that a positive association is a sign of change in drinking pattern, whereas lack of association indicates a stable state of affairs. However, this does not explain the inconsistency between methods, found in his own study and in the present case. Furthermore, there are some technical difficulties when drawing conclusions about the relationship, such as the risk of spurious positive correlation due to (large) non-homogeneity of the sample. For instance, men have a higher drinking frequency as well as a higher average quantity per occasion than women. When both subsamples are mixed, one will find a positive correlation, even if this correlation is absent in the separate subsamples. In conclusion, these results show that one should be very careful in drawing conclusions about relationships between drinking variables on the basis of only a single method.

There is little disagreement over the complexity of the measurement problems in surveys on alcohol consumption. It is also clear that designs more rigorous than the one presented in this paper are necessary to test and study the many different aspects of self-reporting behavior, particularly in connection to drinking behavior. For example, to test hypotheses concerning temporal dynamics, a longitudinal design is appropriate. Other issues, such as memory decay and recall strategies, can be adequately studied only under carefully controlled laboratory conditions. Some promising experiments have been undertaken in the field of medical consumption (e.g. Fienberg, *et al.*, 1985). For alcohol methodology, such an approach has been suggested recently by Midanik (1989). She suggests qualitative analyses of "cognitive

protocols" in order to investigate the strategies respondents use in reconstructing their past drinking behavior. Such analyses could result in more meaningful questions, options for answering, time frames, etc. Equally, they could provide insight into how people solve the task set by the interviewer and provide possibilities for (differential) routes and paths along which respondents could be aided or guided in their recall of the past.

Footnote 1.

1. Basic unit of measurement (see Appendix 1.2): the typical glass (out of which people typically consume the particular drink) contained on the average more than the standard glass for all types of drink. For fortified wines the difference was greatest (+29%), for wine smallest (+2%). The content of distilled drinks was on the average 18% higher than the standard 3.5 cl. The resulting correction of survey estimates of total consumption (only weekly recall method) amounts to about 5% of the sales-based consumption.

References:

- ALANKO, T. (1985) Effects of time-variation of drinking on the validity of consumption surveys: an empirical investigation using Finnish drinking rhythm data. Paper at the ICAA/Alcohol epidemiology section meeting, Rome, June, 1985.
- ALANKO, T. (1988) Connection between drinking frequency and per occasion consumption in a series of three drinking surveys. Paper at the Annual Alcohol Epidemiology Symposium of the Kettil Bruun Society, Berkeley, June, 1988.
- CAHALAN, D., CISIN, I.H. & CROSSLEY, H.M. (1969) *American Drinking Practices; a National Study of Drinking Behavior and Attitudes* (New Brunswick, N.J., Rutgers Center of Alcohol Studies).
- FEINBERG, S.E., LOFTUS, E.F. & TANUR, J.M. (1985) Cognitive aspects of health survey methodology: an overview, *Milbank Memorial Fund Quarterly/Health and Society*, 63, pp. 547-564.
- FITZGERALD, J.L. & MULFORD, H.A. (1984) Seasonal change in alcohol consumption and related problems in Iowa, 1979-1980, *Journal of Studies on Alcohol*, 45, pp. 363-368.
- FITZGERALD, J.L. & MULFORD, H.A. (1987) Self-report validity issues, *Journal of Studies on Alcohol*, 48, pp. 207-211.
- KEMSLEY, W.F.F. (1979) Collecting Data on Economic Flow Variables Using Interviews and Record Keeping. In: Moss, L. & Goldstein, H. (eds) *The Recall Method in Social Surveys* (London, University of London Institute of Education).

- KNIBBE, R.A., DROP, M.J., VAN REEK, J. & SAENGER, G. (1985) The development of alcohol consumption in the Netherlands: 1958-1985, *British Journal of Addiction*, 80, pp. 411-419.
- LEMMENS P.H.H.M., KNIBBE, R.A. & TAN, F. (1988a) Weekly recall and diary estimates of alcohol consumption in a general population survey, *Journal of Studies on Alcohol*, 49, pp. 131-135.
- LEMMENS, P.H.H.M., TAN, E.S. & KNIBBE, R.A. (1988b) Bias due to non-response in a Dutch survey on alcohol, *British Journal of Addiction*, 83, pp. 1069-1077.
- MIDANIK, L. (1989) The use of protocol analysis in alcohol research. Paper presented at the 15th Alcohol Epidemiology Symposium, Kettil Bruun Society. Maastricht, June 1989.
- MULFORD, H.A. & FITZGERALD, J.L. (1981) "Words and deeds": responses to Popham and Schmidt, *Journal of Studies on Alcohol*, 42, pp. 362-376.
- PRODUKTSCHAP VOOR Gedistilleerde Dranken (1986) *Jaarverslag 1985* (Schiedam, PGD).
- REDMAN, S., SANSON-FISHER, R., WILKINSON, C., FAHEY, P. & GIBBERD, R. (1987) Agreement between two measures of alcohol consumption, *Journal of Studies on Alcohol*, 48, pp. 104-108.
- ROOM, R. (1985) Measuring alcohol consumption in the U.S.: methods and rationales. Paper presented at the 31st ICAA/Alcohol Epidemiology section meeting, June 1985, Rome.
- SIMPURA, J. (1988) Comparison of indices of alcohol consumption in the Finnish 1984 drinking habits survey data, *The Drinking and Drug Practices Surveyor*, 22, pp. 3-10.
- SUDMAN S. & BRADBURN N.M. (1973) Effects of time and memory factors on response in surveys, *Journal of the American Statistical Association*, 68, pp. 805-815.

Seasonal Variation in Survey and Sales Estimates of Alcohol Consumption

SUBMITTED TO THE JOURNAL OF STUDIES ON ALCOHOL, AUGUST 1990.
PAUL H.H.M. LEMMENS, R.A. KNIBBE.

Summary

Time-variation of drinking is large and has an effect even on aggregate estimates of consumption. In this chapter, it is shown that because of a considerable seasonal variation in consumption (20%) a serious bias in annual consumption estimates is expected to occur for surveys with a limited time-frame. The present study analyses drinking data collected in the general population of the Netherlands from March 1985 through December 1985 (including Christmas and New Year's Eve). Several indices of consumption have been used, since sensitivity to temporal fluctuations was expected to differ. Although the assessed seasonal effect varies indeed across types of measurement, across male and female sub-sample and across types of alcoholic beverage, the general tendency for consumption is to be highest in the spring season and lowest in the autumn. Sales figures fluctuate accordingly. It is evident that the risk of biased estimates is larger the shorter the time-frame of the survey. Contrary to reports for seasonal changes in Iowa, seasonal variation was highest in the frequency domain. Furthermore, exclusion from the time-frame of collective holidays, during which people drink more often and more per occasion (viz. Christmas), increases the risk of biased estimates. Even estimates of abstention, especially among women, appear to vary considerably over the 3 seasons in this study. The main

conclusion is that results of comparisons of survey data on drinking, particularly those over time, are more or less invalid if the respective time-frames of the surveys do not correspond.

5.1 Introduction

Drinking is a highly variable behavior. Temporal variation of drinking is large not only at the individual level but also, though apparently less, at the aggregate level. Evidence for a large variation at both levels has been reported by Alanko (1985) who, in a secondary analysis of longitudinal Finnish survey data on drinking, has found a considerable weekly variation in aggregate estimates of alcohol consumption. In his study, estimates of mean consumption based on consumption scores in a single week ($N = 41$) ranged between 186 and 47 percent of the mean consumption, calculated over a total of 54 weeks.

One could expect, then, that estimates of annual alcohol consumption derived from surveys that are conducted in a restricted time-period are prone to a seasonal bias. Similarly, results on trends in consumption based on surveys that do not cover equal periods of the year should be cautiously interpreted. For instance, Fitzgerald and Mulford (1978) have reported substantially different coverage rates of sales records for estimates of overall annual consumption as well as for estimates of annual consumption of 3 beverage types in 2 surveys conducted in Iowa in 1958 and 1961, in the summer and winter season, respectively. These differences could indeed be attributed in part to seasonal variation in consumption. The authors made clear that if not taking seasonal variation into account, one would have interpreted the differences as a downward trend.

The main, and legitimate, reason for survey interviews to be planned within a limited time period is to reduce variation between individuals that is attributable to temporal variability (e.g. Kish, 1965). The relevance of a restricted interview period is quite obvious in the case of, for instance, opinion polls. Yet another reason for limiting the fieldwork period is the higher cost often associated with longer time periods in geographically widespread samples. If, however, the aim of the inquiry is to obtain aggregate estimates of a particular behavior over a longer period of time (e.g. annual consumption) and the behavior is expected to vary considerably in this period, one should be less concerned with the larger inter-individual variation and spread the interviews over the total length of the period.

In this chapter an assessment is made of seasonal variation in aggregate estimates of consumption, based on sales figures and on survey measurements. Seasonal fluctuation of consumption is compared for 5 survey methods, for the male and female sub-samples. Since the reference periods for some methods differ (from 1 week, for the *WR* and diary, to 6 months, for *QF* and *QFV*) or are absent (*T*), and because some aspects of drinking behavior may vary more than other (such as, the variable heavy use *V* component in the *QFV* index) the methods will probably differ in their sensitivity to seasonal variation. A comparison of seasonal fluctuation of 4 types of alcoholic beverage is possible for the two methods dealing with actual consumption. The last two weeks of December are considered separately. Being weeks of celebration, they are thought to be exceptionally "wet" (Wilson, 1981).

5.2 Methods

In 1985, a national survey on alcohol consumption was conducted in the Netherlands which has been described in detail in chapters two and four. In the personal interviews drinking were assessed by means of 4 different methods. The interviews were spread over the months March through December. For the most part ($N=918$), subjects were asked to keep a diary on drinking in the week following the interview (an ordinary week, or diary 1). The other 322 subjects were instructed to complete the diary not immediately after the interview but in one of the last two weeks of 1985, the Christmas holiday and New Year's Eve (Christmas week, or diary 2).

The approach in the 1985 survey does not allow an assessment of monthly variation because of the variable size of the monthly sub-samples. Instead, subjects are classified into 3 seasonal categories, roughly covering spring, summer and autumn. The 3 seasons comprise diaries and interviews collected in, respectively, the months March to May, June to September and October to December. The last two weeks of December are analysed separately.

Monthly sales figures have been calculated from excise revenues for distilled spirits (Central Bureau of Statistics, 1985/86a), sales statistics for beer (ibid. 1985/86a) and export and import statistics for wine (ibid. 1985/86b). The variation in monthly sales figures of alcoholic beverages in the Netherlands in 1985 is shown in figure 5.1. In the case of a correlation between sales and consumption, a considerable variation could be expected in monthly Dutch alcohol consumption data, as well.

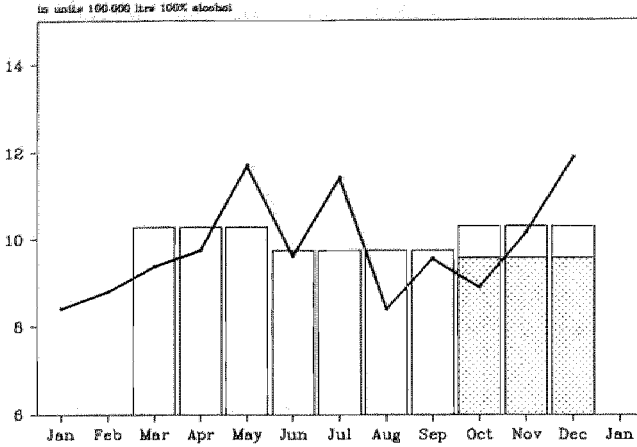


Figure 5.1. *Monthly sales (solid line) and seasonal average per month (bars) of total alcohol consumption in 1985, in units of 100.000 liters 100% alcohol. Shaded area is average corrected for exceptional consumption in the last two weeks of December (source: CBS, 1985/86).*

Monthly sales vary between 121 and 86 percent of the monthly average in 1985. Because sales or purchases of alcoholic beverages do not necessarily coincide with the consumption of these beverages, the sales figures in a given month should be regarded as approximations of actual consumption in that month. One can imagine that there will be different time-lags between purchase and registration, on the one hand, and between purchase and consumption, on the other (see, for instance, Fitzgerald & Mulford, 1987). Still, it may be expected that, at least, part of the seasonal fluctuation in sales is a reflection of the variation in consumption. It is worth noting that no major changes in price have occurred in 1984, 1985 or in the first months of 1986, so it is not likely that excessive stockpiling has occurred. Furthermore, per capita consumption in the eighties has been fairly stable in the Netherlands (Produktschap voor Gedistilleerde Dranken, 1988).

Table 5.1. *Estimates of average weekly alcohol consumption (gl/wk) and respective coverage rates in the 1985 survey, for subjects with a diary during Christmas week (N = 322) and for those in normal weeks (N = 918).*

index	Diary in Christmas week		Diary in normal week	
	average weekly consumption	coverage rate	average weekly consumption	coverage rate
1. Diary	18.4	114%	10.9	67%
2. WR	7.8	48%	9.4	58%
3. QF	6.3	39%	7.4	46%
4. QFV	7.5	46%	9.2	57%
5. T	8.9	55%	9.4	58%

5.3 Results

5.3.1 Consumption during the Christmas holidays

Table 5.1 shows survey estimates of consumption according to the 5 indices available for two sub-samples, one with diaries completed during normal weeks of the year (*diary1*) and the other one with diaries during one of the last two weeks of December (*diary2*). Estimates have been weighted for sex, age and region.

The table (table 5.1, see next page) clearly reveals that consumption in the last two weeks of December is about 70% higher than in "normal" weeks. Average consumption according to the other 4 indices is slightly lower in the "Christmas" sub-sample than in the sample with diaries in a "normal" week. This suggests that the difference found for the diary is even a slight underestimation. When average *diary1* consumption over the normal weeks was to be corrected for the allegedly higher consumption in the two exceptional weeks, average weekly diary consumption would rise from 10.9 to 11.2 glasses, which means a coverage rate of 69% (see also table 4.1). Since surveys usually do not cover other collective holidays (such as carnival and Easter) either, one could expect survey measures of consumption to miss even more than only these 2 "wet" December weeks. As could be expected, the consumption of, especially, wine and fortified wine is higher than usual during these two weeks of celebration (more than twice the usual volume). The consumption of distilled drinks rises with 70% during these weeks. The increase of consumption in these two weeks is lowest for beer (+40%). The higher average consumption is caused by a higher prevalence of drinkers who actually drink during the diary week, by a higher frequency of drinking (number of days), and by a higher average quantity per occasion.

Table 5.2. Rank-order correlations (Kendall's tau-b) between diary and four other indices of alcohol consumption from the 1985 survey (*diary1* = normal weeks; *diary2* = Christmas weeks). Abstainers are excluded.

index	QF	QFV	WR	T
<i>diary1</i>	.61	.61	.65	.58
<i>diary2</i>	.55	.56	.55	.51

In the 1985 survey, consistency in reports on abstinence in the 6 months reference period, given at the time of the interview, can be cross-checked with reports on consumption during the diary week, following the interview. About 17% of the abstainers did report one or more drinks in the diary week following the interview (average weekly consumption: 2.8 glasses). Inconsistency in reports of abstinence between interview and diary was greatest among female "abstainers" with a diary during the Christmas holiday (28%). Average weekly diary consumption was also higher in this category (5.3 glasses).

Correlations between the exceptional *diary2* scores and the four other consumption indices are expected to be lower than the corresponding correlations for *diary1* scores, obtained during normal weeks of 1985.

Table 5.2 shows that the rank-order correlations are indeed lower for those with a diary during the last two weeks of December than for *diary1*, which indicates that drinking behavior during the Christmas holidays differs somewhat from the usual drinking practice in the "normal" weeks (see also table 2, chapter 4).

5.3.2 Seasonal variation in aggregate consumption estimates.

Abstention rates, especially for women, vary quite considerably over the 3 seasons of 1985. For men, the rate varies between 7% in the interviews conducted in spring 1985 and 11% in the summer interviews (autumn, 9%). In the case of women, the rate varies between 16% in the spring and 30% in the summer (autumn, 23%).

Table 5.3 shows average consumption levels in 3 seasons of 1985 for the male and female sub-population, according to the 5 different methods of measurement. In general, average consumption is highest in the spring and lowest in the autumn of 1985, for both men ($p < .01$) and women (*n.s.*). This rank-order equals the rank-order of sales figures, when these are corrected for the exceptionally high consumption in the last two weeks of December, estimated to be twice the consumption in "normal" weeks (corrected autumn sales: shaded area in figure 5.1). The variation in sales figures across seasons, however, is smaller than the variation in survey estimates.

Table 5.3. *Average consumption (gl/wk), average frequency of drinking (days/wk) and average quantity per occasion (gl/drinking day) over 3 successive seasons of 1985, according to 5 survey estimates, corrected for differences in age and region (abstainers and diaries in last two weeks of December are excluded).*

	men spring	summer	autumn	women spring	summer	autumn
1.a. T	15.8	15.8	12.6 ³	8.5	8.3	8.1
2.a. QF index	14.3	12.6	11.2 ³	6.3	5.9	5.1
b. F	3.8	3.5	3.2 ³	2.5	2.2	2.4
c. Q	3.6	3.5	3.5	2.4	2.4 ²	2.2
3.a. QFV index	19.2 ¹	15.1	13.6 ³	8.2	7.6	6.5
4.a. WR index	19.4 ¹	18.8 ²	13.3 ³	9.1	7.6	5.8 ³
b. frequency	4.1	3.9 ²	3.4 ³	3.1	2.7	2.6
c. quantity	3.5	4.0	3.4	2.4 ¹	2.0	1.6 ³
5.a. DIARY	20.9 ¹	19.1	16.6 ³	9.4	8.4	7.8
b. frequency	4.7 ¹	4.2	4.0 ³	3.3	3.1	3.1
c. quantity	3.5	3.6	3.4	2.2	2.0	1.9

Mann-Whitney test: significance ($p < .01$) between spring and summer¹, summer and autumn², and spring and autumn³.

In the male sub-population average consumption in the autumn, according to the 5 measures, ranges between 69 and 80 percent of the average consumption in the spring season. For women, average consumption in the autumn ranges between 64 and 95 percent of the consumption in the spring. Although not consistently so, the fluctuation in average consumption between seasons can be attributed to differences in average frequency of drinking rather than to differences in average quantity per occasion. The relatively largest variation between seasonal estimates, in either sub-population, is found in the case of the weekly recall method, and the smallest variation in the case of the unidimensional "typical consumption" measure.

The fluctuation over seasons seems to be primarily the result of variation in beer consumption which is predominantly a male affair (table 5.4). The variation as well as consumption among women is highest for wine, with beer in second place. Table 5.4 also shows that the apparent fluctuation in each type of beverage differs somewhat across method (beverage type was specified only for diary and weekly recall) but this does not affect the main conclusions. The seasonal survey estimates of consumption for the 4 types of drinks follow the aggregate sales figures reasonably well. The sales figures in

Table 5.4. Average consumption (gl/wk) of 4 types of alcoholic beverages in 3 successive seasons in 1985, according to weekly recall, diary measures and sales. Abstainers and those with a diary in the last two weeks of December are excluded; survey figures weighted for sex, age and region.

	men			women		
	spring	summer	autumn	spring	summer	autumn
DIARY						
beer	15.1 ¹	13.6 ²	11.3 ³	1.6 ¹	3.2 ²	1.7
wine	2.1 ¹	1.3 ²	1.7 ³	4.9 ¹	2.8	3.1 ³
fortified wine	0.8 ¹	0.5	0.6 ³	1.5	1.0	1.5
distilled spirits	2.8	3.7	2.9 ³	1.4	1.3	1.5 ³
WEEKLY RECALL						
beer	14.9	14.3 ²	9.0 ³	1.9 ¹	3.0 ²	1.0 ³
wine	1.3	1.2	1.4	4.2	3.0	2.5 ³
fortified wine	0.6 ¹	0.7 ²	0.8	1.6	0.6 ²	1.7
distilled spirits	3.5	3.6	2.7	2.3 ¹	1.2	1.2 ³
SALES*						
beer	10.7	10.7	8.5			
wines	4.8	4.5	3.9			
distilled spirits	5.8	5.1	3.8			

Mann-Whitney test: significance ($p < .05$) between spring and summer¹, summer and autumn², and spring and autumn³.

*mean # glasses per week over 83% of drinking age population in January 1985; autumn figures corrected for extra consumption during Christmas weeks.

table 5.4 are corrected for consumption during the last two weeks of December, in which consumption is estimated to be twice as high as in normal weeks. No distinction could be made between sales statistics of wine and fortified wines.

Table 5.5. Fraction of the consumption (*100) of 4 types of alcoholic beverages that is consumed outside the own home in 3 successive seasons in 1985 (weekly recall measure; abstainers and Christmas diaries excluded; weighted for sex, age and region).

	men			women		
	spring	summer	autumn	spring	summer	autumn
beer	36 ¹	29	33 ³	15	14 ²	6 ³
wine	6	7	8	19	21 ²	14
fortified wine	5 ¹	0 ²	3	12 ¹	2	5 ³
distilled spirits	8 ¹	8	10	21 ¹	8	8 ³

Mann-Whitney test: significance ($p < .05$) between spring and summer¹, summer and autumn², and spring and autumn³.

Table 5.6. *Percentage of drinkers reporting a heavy drinking occasion (≥ 6 glasses) at least once a week; in percentage (abstainers and those with a Christmas diary are excluded).*

	spring	summer	autumn
men	28	25	24
women*	16	8	8

*chi-square test: $p < .05$

A factor that is known to cause consumption to vary is the weather. Part of the variation in (especially the thirst-quenching beer) consumption, may be attributed to climatic conditions. Consumption on hot and sunny days will not only be higher, but drinks will more often be consumed outside the own home, as well. It is expected that the extra consumption will be added to the existing pattern. Table 5.5 shows the fraction of the consumption of the 4 types of drinks taken outside the own home. Outside the own home refers to all drinks that are taken while being away from the own home, so it refers not only to outdoor drinks. Male beer consumption outside the own home appears highest in the spring.

Contrary to the above expectation, the fraction of male beer consumption consumed outside the own home is lowest during the hottest summer season. The fraction of male consumption for the 3 other beverages is low throughout the entire year. The fraction of the consumption drunk outside the home other than beer is higher for women than for men and the expected sharp decline in the fraction consumed outside after the summer can indeed be observed for female wine and beer consumption.

Another seasonal difference between male and female drinking behavior is presented in table 5.6. The number of women reporting at least one weekly heavy drinking session (drinking 6 or more glasses) is twice as large in the spring than in the summer and autumn. For men, this percentage is fairly stable over the seasons.

5.4 Discussion

The results of the study indicate that survey estimates of annual mean consumption which are based on data collected in a particular season will most probably be biased. Systematic exclusion of collective holidays in time-frames of surveys increases this bias. Compared with the average over 10 months of 1985, estimates of annual consumption based on data collected in a particular season vary in this sample between 120% (spring, *QFV* index) and 77% (autumn, *WR* index) for men, and, similarly for women, between 121% (spring, *WR*) and 77% (autumn, *WR*). Although the degree of under- and overestimation differs between men and women and across method, the general tendency is that estimates based on interviews are lowest in the autumn (after correction for the expected peak in consumption during the last two weeks of December). This tendency is paralleled by sales estimates over the seasons.

The finding that the weekly recall measure of consumption shows a larger variation than the other measures could be the result of a greater sensitivity of the former to temporal fluctuations. The *WR* measure is based on actual events within a single week, whereas the *T*, *QF* and *QFV* indices are based on subjective estimates of drinking behavior over a much longer period of time. It is conceivable that these last measures allow the subject to correct for seasonal and/or atypical variation in drinking behavior. However, the results for the diary method, which is also based upon actual drinking occasions in a specific week, only with a shorter recall period, contradict this explanation. The low variation in estimates based on the typical consumption measure *T* could be the result of the rounding effect at lowest consumption levels, already discussed in chapter 4.

The higher average consumption during the last two weeks of December confirm the expectation of quite a large change in total volume. The slightly lower rank-order correlations between *diary2* and the other indices suggest this is more than a general, overall change in total volume only: some people change their drinking behavior more than others. The finding that consumption of the four beverage types is raised differently during these weeks supports this suggestion. However, even though the reference periods of the separate indices for those with a Christmas diary are quite far apart and total consumption is about twice as high during these weeks, the rank-ordering of subjects seems only mildly affected.

Because Christmas and New Year's Eve are days of celebration, which means that more people drink more often more drinks during these weeks, and because the interval between interview and diary was longer for subjects who completed a diary during these weeks, one would have expected reports on abstinence to be more often inconsistent with the *diary2* than with the *diary1* reports. Only for women did this effect occur. Since the diary week is

not part of the retrospective reference period of the interview, an answer to the question on abstinence that is inconsistent with the subsequent diary report does not necessarily imply that the person has been lying or that the answer has been false in any other way. The finding does suggest, however, that the number of, particularly female, abstainers will probably be lower for surveys conducted just after Christmas and in the spring season than in the summer. It should be kept in mind that the reference period for abstinence was 6 months. The differences might disappear when the question addresses drinking in the past year. Still, the seasonal variation could be the reason for the variation found in abstinence rates in Dutch alcohol surveys, held in the past 30 years (Knibbe, *et al.* 1985).

The finding that the variation is caused by fluctuations in the frequency domain rather than in the quantity domain is contrary to the results found by Fitzgerald & Mulford (1978), already mentioned. The inconsistent outcomes could be the result of different cultural norms. Traditionally, Iowa is a low consumption state (high abstinence rates), whereas the Netherlands now rank 18th place on the world per capita consumption scale and can be regarded a "medium" consumption country (low abstinence rates). Maybe the "dry" Iowa tradition proscribes new, additional drinking situations, which would result in a higher frequency of occasions in the summer, but is less strict on the number of drinks on appropriate occasions. In other words, it may be more acceptable in Iowa to increase one's consumption on particular summer nights, than to drink more often and, hence, in unusual situations. Circumstantial evidence for this explanation can be found in another study by the same authors (Fitzgerald & Mulford, 1984) who report a 36% increase in the number of persons consuming more than 4 glasses per day from winter 1979 to summer and, again, a similar decrease in the following winter. Estimated sales went up and down, accordingly (17%). Equally, the number of persons who reported at least one atypical heavy drinking day (5 glasses) went up during the summer and declined again during the following winter. This trend suggests indeed an increased quantity per occasion in the summer rather than an increase in the frequency of drinking. Contrarily, in the Dutch male sample prevalence of atypical heavy drinking behavior (6 glasses) did not vary over seasons (see table 5.6).

In summary, the main conclusions in this chapter are (1) that when survey data on drinking are compared, the results should be interpreted with due respect to the survey time-frames and the populations under study, and, (2) that measurement of consumption based upon recall of actual consumption in a short time-period is probably more sensitive to seasonal fluctuations than so-called summary measures.

References

- 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/Alcohol Epidemiology Section Meeting, Rome, June, 1985.
- Central Bureau of Statistics (1985/86a) *Maandstatistiek voor de industrie*, 27-28, 85/5-86/12 (Staatsuitgeverij, The Hague).
- Central Bureau of Statistics (1985/86b) *Maandstatistiek voor de landbouw*, 33-34, 85/5-86/12 (Staatsuitgeverij, The Hague).
- FITZGERALD, J.L. & MULFORD, H.A. (1978) Distribution of alcohol consumption and problem drinking. *Journal of Studies on Alcohol*, 39, pp. 879-893.
- FITZGERALD, J.L. & MULFORD, H.A. (1984) Seasonal changes in alcohol consumption and related problems in Iowa, 1979-1980. *Journal of Studies on Alcohol*, 45, pp. 363-368.
- FITZGERALD, J.L. & MULFORD, H.A. (1987) Self-report validity issues. *Journal of Studies on Alcohol*, 48, pp. 207-211.
- KISH, L. (1965) *Survey Sampling* (New York, Wiley).
- KNIBBE, R.A., DROP, M.J., VAN REEK, J. & SEANGER, G. (1985) The development of alcohol consumption in the Netherlands: 1958-1981. *British Journal of Addiction*, 80, pp.411-419.
- PRODUKTSCHAP VOOR Gedistilleerde Dranken (1988) *How Many Alcoholic Beverages are being Consumed throughout the World?* (Schiedam, the Netherlands).
- WILSON, P. (1981) Improving the methodology of drinking surveys. *The Statistician*, 30, pp. 159-167.

Agreement between Respondent and Partner on Reports of Drinking and Drinking Problems in a General Population Survey.

Summary

In this chapter concurrent validity of self-reports has been assessed in a design that uses so-called collateral reports as criterion. Comparisons have been made of reports by the subject and his or her partner present at the time of the interview on several indices of alcohol consumption, on disapproval of drinking by persons in the subject's environment, and on 8 drinking problem indicators have been compared, both for the male and female sub-sample of a general population alcohol survey held in the Netherlands in 1985. Results for reports on consumption are not univocal. Female self- and collateral reports do not differ significantly. Male self-reports are higher in the case of a weekly recall (*WR*) measure (+16%) but lower for a quantity-frequency (*QF*) type of measurement (-8%). Reports on infrequent heavy use (*V*) is, again, higher for male self-reports. Concordance between self- and collateral is quite high for reports on "usual" frequency and quantity (60-70%). For composite indices of drinking concordance is much less. Still, correlations between reports on indices of total consumption are quite high ($\pm .75$). Pairs at higher or lower levels tend to mitigate each other's response. Collateral reports on 8 drinking problem indicators and on disapproval did not yield higher estimates than would self-reports. A slight tendency among men could be observed to report a higher frequency of heavy drinking and disapproval, whether reporting about their own or about their partner. A multivariate test of a model with factors (psychological, cultural, physical) that are thought to determine level of correspondence between both sources revealed that discordance between

both sources is likely to increase with consumption level and when there are indications of a drinking problem. This result supports the hypothesis that a higher perceived threat in the interview situation enhances denial or repression of behavior, and, hence, a lower correspondence. Other factors that emerge are level of daily routine, which enhances predictability of behavior, and the fraction of the total consumption that is taken in the own home, which is thought to increase the chance that the partner witnesses the behavior.

6.1 Introduction

In the past two decades, methodological research in the alcohol field has focussed on validity and reliability of, particularly, self-reports of drinking and drinking related problems. Several criteria of validity have been used in a variety of populations. Examples are the coverage rate of official sales data by self-reports in general population surveys, breath and blood analysis and direct observations in clinical samples (Babor *et al.*, 1987).

Another, potentially strong, indicator of concurrent validity is the agreement between self-reports and reports from so-called collateral informants. These collateral reports are obtained from significant others in the subject's environment, such as, parent, spouse, or peer. In this paper, the reports by the partner about the drinking behavior and drinking related problems of the originally selected subject are compared to the subject's self-reports.

Interest in collateral reports arises from the fact that questions about alcohol use and alcohol problems are more or less threatening which may lead to denial of reported drinks and repression of drinking related problems. It is thought that perceived threat is less if the questions concern drinking behavior of others (Bradburn, *et al.*, 1979). The desire to present oneself favorably is less apparent in such a situation and the reports are expected to be more accurate.

The effectiveness of collateral reports as a criterion of validity depends on the plausibility of a number of assumptions, however. Firstly, the person who is reporting about the subject's behavior should have accurate knowledge of the (drinking) behavior and thus have a reasonable amount of contact with the subject. For this reason it is usually someone close to the subject. However, this requirement of intimacy entails an inbuilt disadvantage. The person will be inclined to psychologically protect the subject with which he or she is emotionally involved from criticism (Guze, *et al.*, 1963). The exact opposite could also occur, the person reporting

exaggerates the subject's problems or use of alcohol out of dissatisfaction with the subject's drinking behavior. Furthermore, the person should have an accurate memory, at least, not worse than the subject. It is clear that these conditions will not be easily fulfilled.

Most studies on this subject in the past have been directed toward an assessment of the accuracy and reliability of alcoholic's self-reports (Babor, *et al.* 1987). According to Midanik (1988), the results in these studies "*range from having no confidence in the accuracy of self-report measures to concluding that self-reports are extremely valid*" (p. 1020). However, little research on agreement over reports of consumption and problems in the general, non-alcoholic populations has been done. In her reviews, Midanik (1982, 1988) has reported only 5 general population studies, 2 of which have compared aspects of the drinking behavior and 3 have compared reports on alcohol problems. Cahalan *et al.* (1967), comparing distributions of "*usual*" frequency of consumption, concluded that reports from the spouse do not enhance validity. Female partners even reported somewhat lower consumption levels for their husbands. Bradburn *et al.* (1979) found that asking about the subject's own and friends' intoxication in the past year yielded higher reports for the "*friends*", irrespective of the level of perceived threat. Bailey *et al.* (1966) found a high agreement between the partners on family problems due to respondent's drinking, while Knupfer (1967) found a much lower agreement, and the discrepancy being in the direction of underreporting by original subjects. In Mulford & Wilson's study (1966), conducted in households which included "*known*" alcoholics, collateral informants were found to report more alcohol problems than subjects reporting about their own behavior.

In a Dutch survey, Garretsen (1983) interviewed partners of subjects drinking more than 4 glasses per day and found that the differences between reports of regular, daily drinking (versus not daily) were small. Differences were noted in reports on frequency of heavy use, the collateral reports yielding slightly higher estimates, especially in the case of female subjects. On questions about problems, no systematic differences between the two sources could be established.

In her review of the literature, Midanik (1982, 1988) has mentioned several problems with the use of collateral reports as a criterium for validity. Firstly, reports from other sources need not be more accurate than self-reports. Because of the basic idea that respondents (especially the heavy drinkers among them) underreport their drinking, there is a tendency to interpret differences along the "*more is better*" principle. So, every time the collateral report is higher than the self-report, the collateral is regarded more valid (as in, for instance, Garretsen, 1983). If the discrepancy changes sign, "*the collateral loses status as a validity measure*" (Midanik, 1982, p.359) and the subjects are thought to be overreporting in order to present themselves favorably (or unfavorably, such as alcoholics who exaggerate their drinking at

the intake for treatment; Armor, *et al.*, 1976, p.165). This “*more is better*”-approach may be appropriate in a clinical situation for a conservative assessment of the result of or need for treatment, it can hardly be defended in a design that uses samples from a general population. Secondly, and related to the above argument is Midanik’s notion that a high correlation between measures from the two sources does not necessarily imply a high validity. Both sources could be “*contaminated*” in a similar way.

The inconsistency in results from studies that compare both sources could in part be attributed to differences in research design, in target population, or, more specifically, in the plausibility of the assumptions mentioned in the introduction. To examine this, the potential bias in self-reports in this chapter is studied *a*) by varying some of the features of the research design (e.g. by comparing more than one measure or index of drinking, and using several items which relate to drinking problems and disapproval of drinking), *b*) by not selecting subjects, and, *c*) by testing or ruling out rival hypotheses. Comparisons are made of self-reports of alcohol consumption and alcohol related problems with reports on the same topics by the partners of the subjects in a general population survey. Several indices of alcohol consumption were available, based upon recall of actual, recent occasions of drinking (weekly recall, *WR*) as well as upon summary, quantity-frequency (*QF*) and variability (*QFV*) type of questions. Scores on 8 items on symptomatic drinking behavior and on overt disapproval of the drinking by others have been obtained from both sources. A special feature of this study is that also information about the partner’s own consumption is available. Because interpretation of results is often hampered by alternative hypotheses about the cause of the discrepancy between the two sources, a multivariate analysis is subsequently performed in which two alternatives to the “*perceived threat*” hypothesis are tested, namely the level of daily routine and the partner’s ignorance of subject’s drinking behavior. The model examines which factors are best predictors of correspondence.

6.2 Methods

The study is part of a random, nation-wide, representative survey on drinking habits conducted in the Netherlands in 1985. The survey consisted of a personal interview, an interview with a partner (if present) and a subsequent 7 day diary. A total of 1240 persons agreed to the personal questionnaire and diary. The response rate was 70%. Non-response did not lead to a detectable bias (Lemmens, *et al.*, 1988).

No formal criterion has been used in the selection of partners for an interview. Interviewers were instructed to invite the partner who happened to be present at the time of the interview until a certain quatum was reached. The aim at the onset was a total sample of about 600 partners. In all, 544 partner interviews were succesfully completed. For the most part, partners were of the opposite sex (only 13 were of equal sex). Similar to the original interviews, the partner questionnaires were spread over the entire survey period (March 1985 to December 1985). No information about reasons for non-response of partners were recorded.

A comparison between self-reports and partner reports can be made for three indices of alcohol consumption. Two indices are based on a subjective evaluation or a summary of components of the subject's drinking behavior. A third index is based on recall of actual, recent drinks. The so-called quantity-frequency index (*QF*) is the product of reported usual frequency (*UF*) and usual quantity per occasion (*UQ*) in a 6 months reference period. *UF* was one of 8 options, ranging from "never" to "every day" in the reference period. Since questions about usual, modal frequency of drinking do not cover irregular heavy drinking, a question on the frequency of variable heavy use is asked (how often do you drink 6 glasses or more per occasion). After conversion into a weekly amount, this is added to the *QF* score, resulting in a *QFV* index. The third method consists of a recall of drinking occasions on the previous 7 days (*WR*). Type of day, location (outdoors versus home), and type of beverage (beer, wine, fortified wines, distilled spirits) were specified. All questions recorded number of glasses, which contain about 1.25 cl or 10 g 100% alcohol. Following the personal interview, the originally selected subjects were asked to complete a diary, which recorded activities, consumption of all beverages and licit drugs, for 15 minute periods, on 7 consecutive days. The diary was to be filled out at least once a day.

Subjects' drinking related problems were assessed by means of 7 on questions symptomatic drinking and 1 on alcohol related accidents. The items resemble the ones developed by Cahalan (1976) and Knibbe (1984). A positive score on the items may indicate dependance and loss of control over the drinking (e.g. black-outs, drinking in the morning or drinking secretly). The questions were imbedded in a list of 14 Yes/No questions which also inquired about negative effects of coffee use. The scores on the items are combined in an additive "problem index", every affirmative answer counting as 1 (and thus a maximum score of 8). Disapproval of drinking (too much in the past year) has been assessed for the partner, friends, family and physician.

The format of the questions in the partner questionnaire was equal to the formats in the original one. For reasons of efficiency, there are 4 exceptions: in the partner questionnaire abstention has not been asked in a separate question as in the original sample (but in the *UF* question), the *WR* index has been simplified (in that in the partner questionnaire no specification has

been made of types of drink or of location), the 6 questions on negative effects of coffee use have been omitted from the partner questionnaire, and the reference period in the partner questionnaire on disapproval of drinking is year, instead of 1 year.

Partners who agreed were interviewed on the same day as the subjects. Most questions were asked orally, except for the questions on actual consumption in the previous week and the 8 "problem" items. These questions had to be filled out on a separate sheet by the interviewee.

Although the interviewers were instructed to interview subject and partner apart, one cannot preclude that interviewing (or the mere presence) of the partner has had an effect on the responses of the subjects, or vice versa. A comparison between answers of subjects whose partner has been interviewed and of subjects without a partner (or whose partner was not present or whose partner did not agree to an interview) did not reveal significant differences in the percentage abstaining, or in average consumption according to the three indices (table 6.1). The two groups only differ in mean age, subjects from the partner sub-sample being an average 6.6 years older.

Due to mistakes made by some interviewers, 31 cases are systematically missing in the comparisons. All 31 cases have abstaining partners and are very light drinkers according to their own reports, and it is expected it will only marginally affect results. Apart from these 31 cases, there is a varying number of cases missing in single comparisons.

Table 6.1 *Percentage abstainers and average consumption according to 3 measures in the sub-samples of subjects of which the partner has and has not been interviewed.*

	partner interview	no partner interview
Abstention		
men	9%	10%
women	26%	22%
Consumption		
Weekly recall index	12.0	11.0
Quantity-frequency index	8.9	9.2
QF-variability index	11.1	11.3

6.3 Results

6.3.1 Consumption indices

In only 7 out of 495 cases (1.4%), concerning 5 female and 2 male subjects, did the partner and the subjects differ in their report on the subject's abstinence of alcohol in the past 6 months. In 6 of these cases the partner claimed that the subject did *not* have a drink in the previous 6 months. The consumption of these 6, self-reported drinkers is very low (*QF* index less than 1.5 glass per week), and the reason for the discrepancy is obvious. Thus, it seems that use of collateral reports would not lead to a downward adjustment of abstinence rate. Subjects are accurate, and seemingly even more so than their partners, in their self-reports on this topic. Those subjects for which either the partner or the subject claimed to abstain from alcohol during the past 6 months have been omitted from the following analyses.

For the sample as a whole, average consumption level as measured by the two evaluative, summary measures, the *QF* index and the *QFV* index, is higher for the collateral reports (table 6.2). In the case of the *WR* index, which is the sum of recent, actual drinking occasions, collateral reports lead to lower estimates of mean consumption (table 6.2). In percentages, the collateral *QF* index is 8% higher than the self-reports, while the collateral *WR* index is 16% lower than the self-reports. The observed differences between collateral and self-reports are caused mainly by differences between the male subject and his female partner. Statistical significance is reached only for the *QF* index and the *WR* index of male subjects. The differences in the case of female subjects do not reach statistical significance at a .05 level. Furthermore, it is interesting to note that on the average self-reported frequency of excessive drinking (*V*) of men is higher than the corresponding report of their, mostly, female partners.

Table 6.2. *Estimated mean values of 3 drinking variables and 3 composite indices of total volume, according to three indices from the two sources, the subjects (SR) and their partners (CR).*

	Men			Women		
	SR	CR	p*	SR	CR	p*
Usual frequency (UF) ²	3.8	3.9	-	2.5	2.6	-
Usual quantity (UQ) ³	3.1	3.3	-	2.1	2.1	-
Variability (V) ²	.9	.6	< .01	.3	.3	-
<i>QF</i> index ¹	11.6	12.6	< .05	5.6	5.7	-
<i>QFV</i> index ¹	14.3	14.7	-	6.8	7.2	-
<i>WR</i> index ¹	15.4	12.4	< .001	6.7	6.5	-

1) glasses per week; 2) occurrences per week; 3) glasses per occasion

*) Wilcoxon matched-pairs signed-rank test, at a .05 level.

Table 6.3 shows the agreement between subject and partner at the individual level. The results seem to be well in line with those presented in table 6.2. For instance, the positive mean difference between *SR* and *CR* on the *WR* index is associated with a higher percentage of discordant pairs of which the male self-report (*SR*) is higher than the, usually, female collateral (*CR*), a surplus of pairs $SR > CR$ of $53\% - 22\% = 31\%$. This is also true for the variable heavy use (11%) and, though *SR* and *CR* in reverse order, for the *QF* index (a surplus of pairs of which $CR > SR$ of 12%). The results for the variable (*V*) furthermore suggest that men tend to report a higher frequency of irregular, heavy use than women, whether reporting about their own or about their, usually, female partner (surplus of $SR > CR$ in the male sample of $26 - 15 = 11\%$, in the female sample a surplus of $CR > SR$ of $15 - 10 = 5\%$). On the whole, however, absolute agreement between the two sources is quite high. Between 60-70% of all paired observations are concordant on variables concerning usual drinking behavior in the past 6 months. Agreement in the case of the composite, 'last week', *WR* measure is lowest. This is not very surprising since questions with a fixed format (the *UF* and *V*) are expected to lead to a higher absolute agreement between the two sources (restricted variation).

The above results leads to the question whether disagreement between the two sources is dependent upon consumption level. Or, stated differently, do heavy drinkers underestimate their consumption more than moderate or light drinkers? In line with the assumption of more threat experienced at higher levels of consumption, one would expect denial of consumption to increase with consumption level, and so would underreporting, viz. the difference between collateral and self-report, the operational terms in this study. A preliminary analysis of the data indicate that the discrepancy is often in the direction of a mitigation of responses, as shown in figure 6.1a. Average consumption for self- and collateral reports are presented for 5 consumption levels, which, in figure 6.1a are based on self-reports. Each of the 5 categories represent about one-fifth of the total sample of drinkers.

Table 6.3. *Percentage (absolute) agreement and disagreement between subject's own (self-report = SR) and the partner's report about subject's drinking (collateral report = CR).*

	Men			Women		
	agree	CR > SR	CR < SR	agree	CR > SR	CR < SR
Usual frequency(UF)	69	19	12	70	17	13
Modal quantity(UQ)	62	20	18	69	14	17
Variable heavy use(V)	59	15	26	75	15	10
QF index	45	34	22	52	26	22
QFV index	32	38	30	44	31	26
WR index	25	22	53	50	23	27

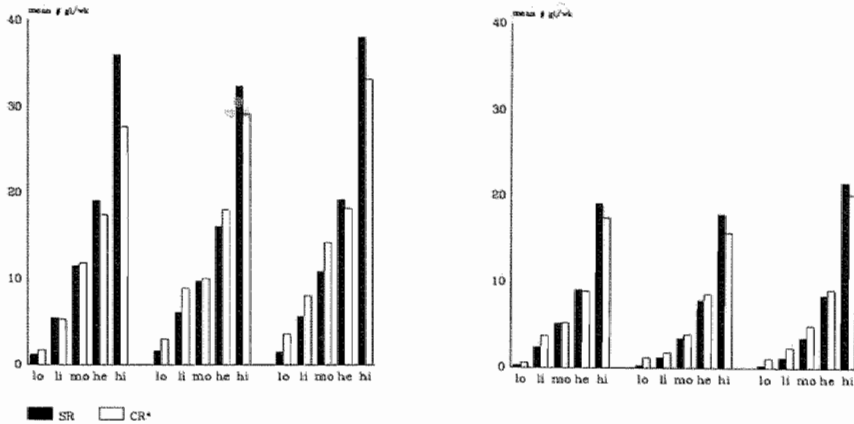


Figure 6.1a. Average consumption of male (left) and female (right) subjects according to self-(SR) and collateral report (CR) at 5 consumption levels defined by the SR. Three indices of consumption: WR, QF, QFV.

It can be seen that collateral reports in the two highest consumption categories tend to be lower than the self-reports ($CR < SR$), and vice versa: partners of low to moderate drinkers tend to give higher reports than the subjects themselves ($CR > SR$).

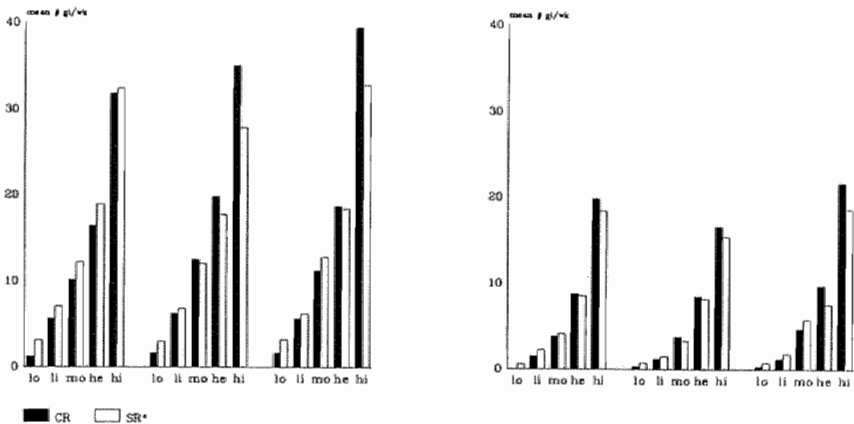


Figure 6.1b. Average consumption of male (left) and female (right) subjects according to self-(SR) and collateral report (CR) at 5 consumption levels defined by the CR. Three indices of consumption: WR, QF, QFV.

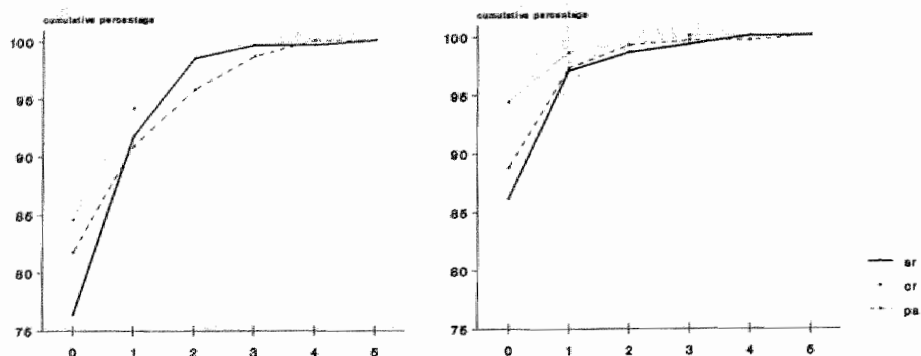


Figure 6.2. Cumulative distribution of scores on the additive problem index for male (left) and female (right) subjects, according to SR and CR. In the figure the partner's own score on the problem index is given (*pa*).

The results for the case in which the partner scores were taken as the reference for categorization (figure 6.1b) are similar. These results again confirm that collateral reports would not lead to an upward adjustment of consumption scores. They also signify the difficulty one will have deciding which report, the *SR* or *CR*, is closest to the subject's true score, since the analysis does not give a definite answer as to who should be taken as the reference. Whether the mitigation of responses is a real effect or a methodological artifact cannot be answered at this point.

6.3.2 Reported problems and disapproval.

Most respondents do not report any of the 8 "problems" of the additive index (items on symptomatic drinking and accidents). About 75% of the male subjects and 86% of the female subjects have a zero self-reported (*SR*) score (figure 6.2). Reports by collaterals (*CR*) are even more modest. In only 15% and 6% of all cases, concerning male and female subjects, respectively, do partners report one or more problems. As can be seen in figure 6.2, the tendency of partners to report fewer problems than the subjects themselves is present at nearly all levels of the additive problem index. The partners' reports about their own symptomatic drinking behavior (*PA*) have also been depicted in figure 6.2, and they seem halfway between the subjects' *SR* and the collateral *CR*.

In terms of agreement, thus at the individual level, 80% of problem reports by and about male subjects and 86% of the problem reports in the female subsample are in perfect agreement. Collaterals reports are higher in

6% and 2% of all cases, respectively, while self-reports of problems are higher in 15% and 12% of all cases (difference is significant at a .05 level, Wilcoxon matched-pairs, signed-ranks test). It can be concluded then that collateral reports do not lead to higher estimates of problems in a general population study, nor do they seem to be more reliable.

A comparison of reports from both sources about disapproval of drinking behavior is shown in table 6.4. It concerns disapproval of, especially, the amount drunk as expressed by the partner, relatives, the G.P., and the subjects' friends in the past year (subject's questionnaire) and past 6 months (partner's questionnaire).

Even though the reference periods differ somewhat, agreement is very high. Prevalence and disagreement of disapproval is highest between male subjects and their usually, female partners. Disapproval of drinking is reported (*SR* and/or *CR*) for about one third (29%) of the non-abstaining men. It is mostly the partner (viz., the wife) who is said to disapprove of the drinking (25%). Disagreement is also higher in this "*partner's disapproval*" category, namely 12%, but equal in both directions ($\pm 6\%$). Besides, it is obvious that disagreement increases with prevalence. For the other 3 categories, concerning disapproval by "*third parties*", the collateral source reports slightly more often disapproval than the male subjects themselves. Since disapproval of drinking is hardly reported for women, the disagreement between female subjects and their, mostly male, partners is, likewise, very low. The husband answers slightly more often affirmatively when asked about having disapproved of his wife's drinking than do their female partners.

On the whole, there does not seem to be a systematic difference between male *SR* and *CR* since in about half of the disagreeing pairs the collateral is higher than the self-report (see combined score on bottom line table 6.4; additive, each category counts as 1). Male subjects and their wives agree on the combined "*disapproval*" score in 81% of all cases. The male combined "*disapproval*" *SR* is higher in 9%, and lower in 10% of all cases. In the female sample there is agreement in 94% of all cases, in 4% does the husband report a higher score.

6.3.3 *A multivariate model of correspondence.*

The results thus far do not suggest that collateral reports would invariably lead to higher estimates of drinking, abstention rates or to higher rates of disapproval or problems, nor do they suggest that at the aggregate level underreporting increases with consumption level (viz. figure 6.1).

As already pointed out in the introduction, there are probably many factors determining level of agreement, and, consequently, estimated reliability and validity of self-reports. In this section, indicators of four factors that are thought to influence level of correspondence between reports of subject and partner are described in detail. Subsequently, these

Table 6.4. Agreement between self- (SR) and collateral (CR) reports of disapproval of drinking expressed in the past year (CR: past year) by the partner, relatives, general practitioner and by friends (in percentages).

MEN	disapproval*	agree		disagree	
		SR = CR	SR + /CR +	SR - /CR +	SR + /CR -
disapproval by:					
Partner	25	87	13	6	6
Relatives	11	92	2	5	4
G.P.	2	98	0	1	11
Friends	5	95	0	4	1
Combined score	29	81	10	10	9

WOMEN	disapproval*	agree		disagree	
		SR = CR	SR + /CR +	SR - /CR +	SR + /CR -
disapproval by:					
Partner	5	97	2	3	0
Relatives	3	98	1	1	1
G.P.	1	99	0	0	1
Friends	3	98	1	2	0
Combined score	8	94	2	4	2

+) CR + or SR + means an affirmative answer by collateral or self-report, respectively.

*) an affirmative answer on disapproval by either source or both

factors will be entered in a multivariate model in which some of the assumptions, mentioned in the introduction, are tested. In explaining the variation in agreement in reports on drinking between subject and partner one could think of four hypothetical factors or mechanisms. These mechanisms are not necessarily incompatible, but they can be operative at the same time. The factors concern:

1. subject's level of *perceived threat* of questions about drinking;
2. the partner's *desire to protect* the subject from criticism, which is reformulated in threat perceived by the partner;
3. the partner's knowledge, or better, *ignorance* of the drinking behavior of the subject, which is probably related to the amount of time partners spend together;
4. the predictability or level of *routine* of the subject's activities.

Since most factors have not been measured directly in the 1985 survey, several indicators have been used. The choice of the explaining variables has been made on theoretical and/or empirical grounds. The rationale for the choice is described below (variables are in italics). In figure 6.3 the variables are presented schematically.

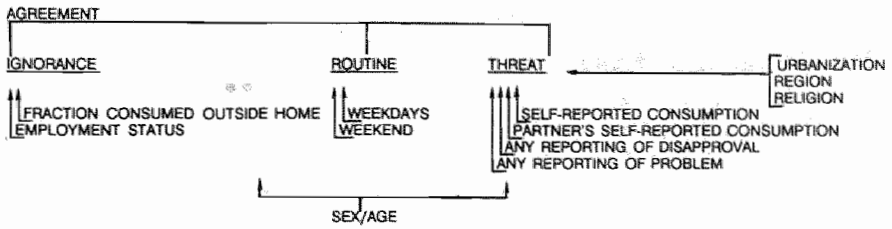


Figure 6.3. Schematic model of correspondence between self- and collateral reports of the subject's drinking

1 & 2 threat: The basic idea in a research design that uses reports from subjects' partners, relatives, collaterals, or significant others is that for some behaviors it is less threatening to report about someone else's behavior than about one's own. These collateral reports are thought to be less distorted. The general expectation is that a higher *alcohol consumption of the subject* (SR) will increase feelings of threat (e.g. Bradburn & Sudman, 1979) and, hence, will induce a lower agreement with the corresponding collateral report (CR).

Matters are more complicated, however, since it is known that partners protect each other from a revelation of embarrassing details. The desire to protect the subject, also referred to as spousal courtesy (R. Room, 1989) can be reformulated into threat experienced by the partner, and it may explain why partners do not freely report about the subjects' behavior, but instead "*stay on the safe side*". It is conceivable that the threat experienced by the partner is greater not only when the subject's consumption but also when the *partner's own consumption* is higher. On the other hand, one might argue that it is not as much the partner's own consumption but its relation to the subject's consumption that will determine threat. For example, threat is aroused when subject's and partner's consumption patterns are dissimilar. Therefore, an interaction term has been included in the model. Furthermore, one might speculate that the (interview) situation is more anxiety provoking in the case when either party reports *problems* connected to the drinking or reports about *disapproval* of drinking.

There are probably also cultural factors that co-determine the experience of threat in an (alcohol survey interview) situation. It seems plausible that threat is more easily aroused (or, people are more sensitive to questions about alcohol) in *dry regions* than in the traditionally *wet regions*. Attitudes toward drinking are less liberal in dry regions than in wet regions, and, as a consequence, one would expect disagreement or underreporting to be higher in relatively dry areas. For instance, Room (1971) has found, comparing

aggregate sales and survey data on consumption from US-states, that in states with a history of temperance (which, nowadays, are still characterized by a lower mean consumption and a pattern of irregular heavy drinking) survey coverage of sales data was lowest. A low coverage was seen as an indication of higher underreporting. At another occasion, Room has reported that wives' reports in dry regions about themselves as well as about their husbands are more conservative (women report more often abstinence and a lower drinking frequency than their husbands) than in the more liberal, wetter regions (ibid., 1989). In the present study, the 3 Northern provinces are lowest in average consumption.

Room has also stressed that other socio-cultural factors are important, especially when relations between married men and women differ greatly. For this reason, a distinction is made between *rural and urban* households, since in the Netherlands, the more traditional marital values are to be found in rural rather than in urban areas (Kooy, G.A., 1977). Similarly, reporting a *religious affiliation* is also seen as an indicator of a more traditional orientation (e.g. toward marital harmony, parental authority, and a greater sex-role differentiation). Like threat experienced by the subject, the "*desire to protect*"-factor refers to a subjective perception of the (interview) situation and to the level of cohesion between partners. This is in contrast to the more objective factors such as ignorance (the amount of time spend in each other's presence) or the level of routine.

3. Ignorance: It is obvious that partners will differ in the knowledge they have of the behavior of their spouse. One might expect ignorance of the subject's drinking behavior to be positively related to disagreement, the disagreement in the direction of a lower reported consumption (and less reported problems) by the partner. The contradictory results for the evaluative QF index and the WR index, which is based on observations of actual behavior, reported in table 6.2 lend support to this assumption (see discussion section). Ignorance is probably negatively related to the amount of time partners spend in each other's presence. Unfortunately, in the 1985 survey no data on time spend together are available. Analyses of four data sets from a large international time-budget study (Szalai, 1972) revealed that the time partners spend together was less for those who were employed but did not increase with other variables, such as age (except in Hungary. By the way, time spend alone did increase with age in 3 out of 4 countries). Because *employment* is strongly (& inversely) related to time shared together, it is used in the model as an indicator for the ignorance factor. In the present case it was found that for male subjects employment was, indeed, negatively correlated with agreement. For female subjects a relation between employment and agreement was only reached for the QF index. Employment status is defined by an outdoor job of either the subject or the partner.

One might, successfully, argue that the time spend together in drinking situations is the crucial factor rather than shared time in general. Therefore, another indicator of ignorance has also been included: *the fraction of the reported consumption which is drunk outside the own home*. It is expected that a large fraction will lead to a lower agreement, the assumption being that drinks taken at home are more likely to be witnessed by the partner than drinks taken outside the own house. Only data obtained by means of the weekly recall method can be used for this variable. Univariate analyses indicate that subjects who disagree with their partner drink relatively more often outside the own house, but the effect is not very large (32% of all drinks are drunk out of the own home; among disagreeers 40%). At least for male subjects, this factor does seem to account for some variation in agreement. The percentage of the outdoor-drinks in total weekly recall appears to decrease (monotonically) with age (male subjects only: F-test, $p < .05$). The difference between lowest and highest age categories is quite large: for the 16-25 year group 44% and for the 56-70 year-olds 26% of the total consumption during one week.

4. Routine: An alternative to the ignorance hypothesis is that, rather than differences in actually witnessing the subject drinking, it is the degree of predictability of the subject's behavior, operationalized in the level of daily routine, which causes discrepancies in reports. In the model, similarity of activities over successive days is used as an index of routine. Main activity (out of 27 categories) has been registered for every quarter of an hour in a 7 day diary (see Lemmens, 1987). Using Dutch time-budget data from 1980 (see Knulst & Schoonderwoerd, 1983), Van Gelooven et al. (1988) have constructed an index for daily routine in which they use all possible combinations of days in a week. Mean values of this routine index varied between 9 quarters/day (between Sunday and Thursday) and 27 (Monday and Tuesday), or 9% and 28% respectively. In the present study, two indices of routine have been used: an index of similarity between *weekdays* (averaged over 10 pairs) and an index of similarity between *weekenddays* (viz., Saturday and Sunday) are made. Both variables can vary between 0 and 100 (= total equivalence).

Two additional factors, *sex* and *age*, are included in the model because they seem to have a strong impact on some of the variables in the model. One does not, in essence, assume sex of the partner to affect level of agreement per se. But it is quite conceivable that women drink more often in the presence of their spouse than men, in which case men more often than women witness their spouse's drinking. To test for this possible effect, an interaction term of sex and the fraction of consumption out of the house has been included. In the present study (see table 6.2), agreement is indeed better for self- and partner reports in the case of female subjects. Since the association between agreement and employment was only found for men, the

interaction between sex and employment is also tested. Univariate analyses, furthermore, show that level of agreement is gradually increasing with age for all 3 indices of total volume and the large age differences in fraction consumed outside the own home. Therefore, the first order interaction between age and fraction consumed outside the own home is tested.

6.3.4 *Test of the model of correspondence.*

The model is tested in a logistic regression analysis (GLIM 3.77, 1985). The response variable is a dichotomous agreement versus disagreement of self- and collateral reports on the subject's consumption, regardless of direction. A pair is said to agree when the reports did not differ more than 10%, with the self-reports taken as reference. Because a 10% difference for $SR < 10$ glasses per week is not exactly measureable (the minimum difference is 1 glass), a pair ($1 \leq SR \leq 9$) is said to disagree when the difference with the collateral report is at least 1 glass.

Two models were contrasted, model1 with SR 's and CR 's based on the weekly recall method (actual observations during the previous week), and model2 with reports based on the more subjective, summary QF index (usual quantity and frequency in past 6 months). The corresponding response variable (viz., agree versus disagree) is also based upon the respective SR 's or CR 's, and, consequently, not the same in the two models. Apart from a similar effect of threat on WR and QF measures of consumption, the expectation is that correspondence between weekly recall reports is more affected by the ignorance factor (since the partner may not be present in all drinking situations), whereas the agreement between QF reports is enhanced by a larger routine (actual witnessing the drinking is less important than predictability). The results of the two models are presented in table 6.5.

Model1 (weekly recall data):

In model1 four variables emerge as significant contributors to the response variable. None of the interaction terms appear significant. The fit of the basic model was only moderate, and could be improved slightly by transformation of some of the continuous variables (deviance = 303; d.f. = 271; $p > .01$). The effects of 3 variables, self-reported consumption (\sqrt{SR}), weekday routine ($WEEKDAY^3$) and (reported) alcohol problems, are in the expected direction, in that an increase in SR and problems (indicators of threat), and a decrease in routine enhance the probability that a pair disagrees. The contribution of the socio-cultural variable region is contrary to expectation: those from a wetter region are more likely to disagree. Neither the partner's self-reported consumption as such, nor in interaction with SR (not included in the table), seems to affect level of agreement.

Model2 (summary QF data):

In model2, four estimates appear significantly larger (.05 level) than 0. The fit of the model is, even after transformation of the SR term, worse than model1 (deviance 343; d.f. = 269; $p < .01$). As in model1, a higher

Table 6.5. *Parameter estimates of logistic regression (GLIM 3.77, 1987); model1 with weekly recall measures, and model2 with usual frequency and quantity measures. Response variable is agreement (0) versus disagreement (1).*

	model1 estimate	s.e.	model2 estimate	s.e.
1	.406	1.73	-3.16	1.06
SR consumption				
√SR	.510*	.108	.	
ln(SR)	.		2.320*	.550
(ln(SR)) ²	.		-.523*	.142
Partner's SR cons	.017	.016	.002	.011
SR/CR of problem	1.017*	.381	.772*	.352
SR/CR of disapproval	-.373	.409	.094	.378
FRACTION away from home	.005	.004	.034*	.013
FRAC x Age	.		-.0006*	.0003
Employment	-.198	.381	-.363	.357
Routine weekend	-.022	.015	.0008	.014
Routine weekday	.		-.363	.357
Weekday ²	.013	.007	.	
Weekday ³	-.0002*	.0001	.	
Urban	-.884	.483	-.304	.444
Region	1.119*	.526	.228	.446
Religion	-.077	.290	-.125	.267
Sex	-.431	.410	.257	.374
Age	-.0004	.013	.017	.015

(* significant at p. < .05)

self-reported consumption (lnSR) and (reported) alcohol problems increase the probability a pair will disagree. Instead of the routine variable in model1, the variable "fraction of consumption consumed outside the own house (FRAC)" appears significant. The effect is in the expected direction, in that an increase in FRAC increases the chance partners will disagree. The interaction between FRAC and age also gives a significant contribution: the effect of the FRAC variable differs with age. The other interaction terms did not enhance the fit of the model. Neither the partner's self-reported consumption as such, nor in interaction with SR, seems to affect level of agreement.

6.4 Discussion

Main conclusion that can be drawn from this study is that there is no consistent evidence for a systematic bias in self-reports when reports by the partner are taken as the criterion. Although estimated mean consumption is significantly larger for collateral reports for the male *QF* index, the opposite result is obtained for the male *WR* index. For the female subsample, no significant differences were found. Furthermore, partners do very well agree with the subjects on topics such as abstinence (98%), drinking problems (male subjects, 80% and female, 86%) and disapproval (men, 87% and women, 99%). Even in about 60% to 70% of the cases do partner and subject exactly agree on the usual frequency (*UF*), usual quantity (*UQ*) and frequency of variable heavy use (*V*). For the composite indices agreement is, of course, less. Still, rankorder correlations (τ -b ranging between .72 and .75) for the three indices are relatively high compared to, for instance, the correlation of .84 found between consumption scores in two successive diary weeks (see ch3).

The question whether underreporting by subjects increases with consumption level cannot be answered affirmatively. It appeared (see figures 6.1a & 6.1b) that, with either self- or collateral reports taken as the reference (or "true") score, both sources mitigate the other's response: the higher scores are deflated by the concurrent response, the lower scores are inflated. From these results, it is impossible to decide which source is more trustworthy, the subject or the collateral. The finding may well be interpreted as a methodological artifact, similar to a regression to the mean effect that can be expected in a situation of repeated, but imperfect measurements, with correlations less than unity. Translated to this design, the two reports (*SR* and *CR*) may be regarded as imperfect measurements of a single "true" score. Assuming the error in both measurements to be normally distributed, with an expected 0 mean and equal variance, mitigation of responses at the highest and lowest deciles could indeed be expected and the results could be explained as a methodological artifact (e.g. Crano & Brewer, 1973, pp. 278-299; Davis, 1976).

A salient result in this study is that the type of method with which consumption is assessed does make a difference. Two explanations for the contradictory results concerning the subjective, summary *QF* index on the one hand, and the more objective *WR* measure on the other can be given. The first, and most obvious one concerns the fact that the partner is not present in all drinking situations of the respondent and does not have full knowledge of all the drinking. This of course affects both measures, but one might speculate that in the case of the usual consumption questions (*UF*, *UQ*), the partner can very well compensate for "missing observations", an opportunity that is absent when asked to recall actual drinking of the subject.

A second reason for a lower collateral report on actual consumption may be the less differentiated format of the *WR* questions in the partner-questionnaire. Studies in which questionnaire-format has been varied (Poikolainen & Kärkkäinen, 1985; Bradburn & Sudman, 1979) have found that differentiation indeed affects magnitude of the reported behavior. Whatever the cause, the fact that collateral *QF* reports are somewhat higher seems to point to a slight tendency of the, particularly male, respondent to mitigate his self-reported usual or modal quantity (*UQ*) and frequency (*UF*).

Tables 6.2 and 6.3 suggest that, except for the *QF* index, reports given by men are more often higher than reports by women, whether about their own or about their (female) partner. A similar result has also been reported by Room (1989). This could point toward by a general tendency among women to be more conservative in their reports than men. In the light of a total coverage of 40-50%, the impact of a possible gender effect on total undercoverage is probably small.

The logistic regression analysis, with which plausibility of certain assumptions implicite in a subject/collateral design has been explored, indicates that discordance between partners over the subject's level of consumption is not a chance phenomenon but increases with reported consumption level. When either party reports an alcohol related problem, disagreement is even more likely to occur. These findings are interpreted as evidence of a higher threat experienced in the interview situation. This finding also implies that reliability of self-reported consumption data decreases with consumption (see also the heteroscedasticity in figure 3.1, ch3). Together with the result that systematic underreporting by the subject does not seem to increase with consumption, one might conclude that the error is present in both sources. With regard to the response variable, agreement vs. disagreement, the following word of caution: the choice of *SR* ($\pm 10\%$) as basis for the dependent variable introduces a bias in the response variable. Since the margin of 10% is quite large, the effect of this error is probably small, but it cannot be precluded that the above results would have been different if the *CR* would have been taken as reference. The problem of choice of criterion, however, cannot be solved here (see Crano & Brewer, 1973, pp. 263-270).

Neither level of the partner's (self-reported) own consumption as such, nor in interaction with subject's consumption seems to have an impact on agreement. Also contrary to expectation, pairs from the dryer regions (viz., the three Northern provinces) are less likely to disagree. The effect of other two factors that emerge is as expected. The fraction of consumption that is taken while being away from home (indicator of ignorance) appeared, however, not to be related to correspondence between pairs of *WR* reports but to agreement between *QF* reports. Similarly, weekday routine emerges as a factor in the analysis of correspondence between *WR* reports, instead of between *QF* reports. This result contradicts the first, "*compensation*"

explanation given in the third paragraph of this section. This "cross-over" is not easy to interpret. Given these rather disappointing results, one may doubt either the effectiveness of the variables (as indicators of the factors) or the theoretical model itself. As already pointed out, the choice of indicators is made ad hoc, and there are many ways to improve the operationalization of the factors. The results should thus be seen as exploratory and be regarded with due caution.

The fact that the sample is representative for a general adult population is an advantage but it also has drawbacks. The advantage is the generalizability of the findings. One is not confronted with an atypical group of people who assign a special meaning to the questions about drinking and related problems, such as alcoholics who are recovering or who are entering treatment. Another advantage is the large size of survey samples, which facilitates multivariate approaches. A drawback of a general population approach, however, is the low prevalence of heavy alcohol use, and, similarly, low prevalence of alcohol related problems and disapproval. It is apparent from, for instance, table 6.4, that correspondence between reports is inversely related to prevalence of the behavior in question. For instance, among women, there are only a few persons for which questions on problems are relevant. Statements about the reporting of alcohol related problems are thus based upon a relatively small number of women and the advantage over a design which uses clinical samples becomes less pronounced.

The research design in this study resembles the psychometric design which estimates the reliability of measurement of certain individual characteristics by means of a parallel test. One of the conditions in such a design is uniformity of the test-environment. Although it has been attempted to secure uniformity by means of written and oral instructions to the interviewers (esp. regarding the presence of subject and partner during each other's interview) the survey researcher does not have actual control over the test situation. Furthermore, the relation between a particular behavior and the reporting of that behavior is complex, and many more variables than presently available may (co-)determine the eventual measurement. For instance, one may criticize most of the operationalizations (indicators), or argue for an alternative to the concept of threat as explanation of the observed, inverse relationship between correspondence and consumption. For instance, an alternative to the threat factor may be complexity or difficulty of the task both subject and partner are confronted with. This complexity might very well increase with consumption, regardless of feelings of threat. One might argue then for a model that would also specify a cognitive factor besides the above personal, socio-cultural and physical factors (see Midanik, 1989). This study shows that results of a simple subject-collateral design are limited in scope, when it does not specify the factors responsible for the observed discrepancies. The study has attempted to clarify the conditions or mechanisms that cause partners to disagree in their reports on drinking. The

study should not be seen as a rigorous test of the assumptions underlying a subject-collateral design. It is clear that a more experimental approach, in which factors are better operationalized and conditions can be better controlled than in a survey could be helpful in answering the fundamental questions initially formulated. It is clear that different methods of measurement referring to the same behavior but stressing different features of that behavior, and two "observers" who judge and rate the behavior (and make mistakes) lead to a (measurement) model that is much more complex than commonly presented.

References

- ARMOR, D.J., POLICH, J.M., STAMBUHL, H.B. (1976) *Alcoholism and treatment*. Report R-1739-NIAAA (St. Monica, Rand Corporation).
- BAILEY, M., HABERMAN, P., SHEINBERG, J. (1966) Identifying alcoholics in population surveys, *Quarterly Journal of Studies on Alcohol*, 27, pp. 300-315.
- BRADBURN, N., SUDMAN, S. AND ASSOCIATES (1979) *Improving interview method and questionnaire design* (San Francisco, Jossey-Bass).
- CRANO, W.D. & BREWER, M.B. (1973) *Principles of Research in Social Psychology* (New York, McGraw Hill).
- DAVIES, C.E. (1976) The effect of regression to the mean in epidemiologic and clinical studies, *American Journal of Epidemiology*, 104, pp. 493-498.
- GARRETSEN, H. (1983) *Probleemdrinken* (Lisse, Swets & Zeitlinger).
- GLIM 3.77 (1987) Numerical Algorithms Group (Oxford, NAG).
- GUZE, S.B., VICENTE B.T., STEWARD, M.A., PICKEN B. (1963) The drinking history: a comparison of reports by subjects and their relatives, *Quarterly Journal of Studies on Alcohol*, 24, pp. 249-260.
- KNIBBE, R.A. (1984) *Van gangbaar tot problematies drinkgedrag* (Maastricht, University of Limburg).
- KNULST, W. & SCHOONDERWOERD, L. (1983) *Waar Blijft De Tijd* (The Hague, Staatsuitgeverij).
- KNUPFER, G. (1967) *The Validity of Survey Data on Drinking Problems: A Comparison between Respondents' Self-reports and Outside Sources of Information* (Berkeley, University of California, Social Research group).
- KOOY, G.A. (1977) *Het Modern-Westerns Gezin* (Deventer/NL, Van Loghum Slaterus).
- LEMMENS P.H.H.M., TAN, E.S., KNIBBE, R.A. (1988a) Bias due to non-response in a Dutch survey on alcohol, *British Journal of Addiction*, 88, pp. 1069-1077.
- LEMMENS, P.H.H.M., TAN, E.S., KNIBBE, R.A. (1988b) Comparison of indices of alcohol consumption. Issues of validity of self-reports. Paper presented at the Alcohol Epidemiology Symposium, Berkeley, Cal.

- MIDANIK, L. (1982) The validity of self-reported alcohol consumption and alcohol problems: a literature review, *British Journal of Addiction*, 77, pp. 357-382.
- MIDANIK, L. (1988) Validity of self-reported alcohol use: a literature review and assessment, *British Journal of Addiction*, 83, pp. 1019-1029.
- MIDANIK, L. (1989) The use of protocol analysis in alcohol research. Paper presented at the 15th Alcohol Epidemiology Symposium, Maastricht, June 11-16, 1989.
- MULFORD, H. & WILSON, R. (1966) *Identifying Problem Drinkers in a Household Survey*. (Washington D.C., National Center for Health Statistics).
- POIKOLAINEN, K. & KÄRKKÄINEN, P. (1985) Nature of questionnaire options affects estimates of alcohol consumption, *Journal of Studies on Alcohol*, 46, pp. 219-222.
- ROOM, R. (1971) Survey versus sales data for the U.S., *Drinking & Drug Practices Surveyor*, 71, pp. 15-16.
- ROOM, R. (1989) Spouse Reports Versus Self-Reports of Drinking in General Population Surveys. Paper presented at the 15th Alcohol Epidemiology Symposium, Maastricht, June 11-16, 1989.
- SZALAI, A. (1972) *The Use of Time* (The Hague, Mouton).
- VAN GELOOVEN, R., DIEDERIKS, J., DROP, M.J., KNIBBE, R. (1988) Health and Everyday Life: How to Measure Structure of Everyday Life. Paper presented at the 2nd Congress of the European Society of Medical Sociology, Zagreb, 28-30 September.

Comparing Distributions of Alcohol Consumption : Empirical Probability Plots.

ORIGINALLY PUBLISHED IN THE BRITISH JOURNAL OF ADDICTION, 85, 751-758, 1990. LEMMENS P.H.H.M. , TAN E.S., KNIBBE R.A.

Summary

Parametric approaches to the problem of the distribution of alcohol consumption have not been very successful. In this article, it is shown that regularity in distribution can be studied without making assumptions about a distribution model underlying the data. For this purpose, a method is used with which distributions are compared graphically in so-called probability plots. It appears that, up to a proper linear transformation on a logarithmic scale, a surprisingly large regularity over time can be observed between distributions taken from Dutch samples in 1970, 1981 and 1985. Equally, distributions from male and female sub-samples do not appear to differ up to a linear shift. The finding of a relative equality in distributional form is in accordance with the Ledermann model. However, the difference with the Ledermann's model is that no assumptions about the exact shape of the distributions are being made.

7.1 Introduction

Since Ledermann (1956, 1964) formulated the notion of a fixed distribution of alcohol consumption in homogeneous populations, many studies have been conducted evaluating its usefulness, both as a descriptive and as an explanatory model. Despite the many scientific reservations, the model is still a dominant feature in various alcohol control programs, not in the least because of its apparent simplicity.

Ledermann's model of the frequency distribution of alcohol consumption assumes that the distribution in a population in a given time period can be described adequately by a special variant of the lognormal curve (*footnote 1*). Specifically, the model hypothesizes a fixed relationship between the mean and the dispersion parameter of the lognormal distribution, thereby reducing possible variation in distribution.

Several authors have extensively criticized the Ledermann model on empirical as well as on theoretical grounds (Parker & Harman, 1978; Skog, 1980; Duffy, 1986). Regarding empirical validity, various studies have shown that the two basic propositions of the Ledermann-model, viz. lognormality and invariance of the relation between the two parameters, are not always tenable. Attempts to discriminate between applications of theoretical distributions, like gamma, Weibull or lognormal, have produced divergent and sometimes conflicting results (Guttorp & Song, 1977; Skog, Guttorp & Song, 1979). Others have shown that even small deviations of the dispersion parameter in the model would lead to large differences in estimates of prevalence of heavy drinking (Duffy, 1977; Duffy and Cohen, 1978). Often, the Ledermann model was found to over-estimate the number of drinkers at the higher levels. In a Dutch general population survey, for instance, the Ledermann formula was found to over-estimate the proportion of drinkers with a mean consumption level over 7 glasses (ca. 8.7 cl 100% alc.) per day in the male sample and 4 glasses (ca. 5.0 cl 100% alc.) per day in the female sample (Lemmens, 1987). Most analyses of empirical data are in support of the argument that the Ledermann model is too rigid in its assumptions. On grounds such as these, the usefulness of the Ledermann model as an indirect method for estimating prevalence of (heavy) alcohol use in a population can be seriously questioned. Furthermore, theoretical interpretation of the parameters of the distribution proves to be difficult. Ledermann himself has not given a clear explanation of the relative invariance of the dispersion parameter, nor of the lognormal shape of the distribution. Nevertheless, his remarks about a possible snowball-mechanism (1956) seem to point toward a principle of proportionate change (see: Skog, 1980).

Controversy over the choice of mathematical models is by no means restricted to the field of alcohol consumption. Examples of problems of parametric choice can be found in Eaton and Whitmore (1977) who studied

length of hospitalization (inverse Gaussian is preferred over a lognormal approximation because of ease of interpretation), in Lawrence (1984) on duration of strikes in the UK (inverse Gaussian, Weibull or lognormal), and in Eppink (1988) for geographical data, such as the lognormal versus the Pareto model of the distribution of city size. Equally, neither a Pareto nor a lognormal model seems to give a good approximation of the entire distribution of personal income, comparable to the situation regarding alcohol consumption (Central Bureau of Statistics, 1972). Obviously, problems of parametric choice are not restricted to the alcohol field. In all, parametric approaches to the study of alcohol consumption distributions have not been very fruitful up till now and they seem of limited value for estimation purposes.

An alternative approach to the distribution problem has been presented by O.-J. Skog (1980, 1985), who proposes a theory of drinking behavior, in terms of (1) a process of social interaction or diffusion and, (2) the multiplicative character of the process underlying change in consumption. In particular, the latter hypothesis states that the many factors responsible for a change in consumption do not combine additively but multiplicatively. Multiplicativity is in effect when the factors do not operate independently (as in statistical interaction). At the individual level it means that people will tend to change their consumption proportional to their initial level. Furthermore, the hypothesis of multiplicativity is also said to give an explanation of the apparent skewness of empirical distributions (*footnote 2*). Point of departure for the present paper was the question whether regularity in the distribution of alcohol consumption over time could be studied without making assumptions about the underlying distributional form. For this purpose, a graphical method known as probability plotting is adopted with which relationships between distributions of alcohol consumption can be investigated.

How well can the expected relationship over time be specified? From the Ledermann model but also from Skog's notion of multiplicativity, a linear relation on a logarithmic scale could be expected between consumption levels over time. In the present paper, such a relationship is also hypothesized between consumption distributions. So the distributions are thought not to change substantially. With this hypothesis, an intermediate position is taken between Ledermann, who is very explicit about the one-parametric lognormal form of distributions, and Skog, who does not presume anything regarding the relationship between distributions over time. Our model hypothesizes equality of distributions over time, like Ledermann did in his model, but the test is distribution-free, like in Skog's approach (*footnote 3*).

The hypothesis of equality in distribution after a transformation of the form $\log Y = \alpha + b \log X$ is put to a statistical test. The procedure is illustrated with Dutch longitudinal data. The results indicate that the distribution of

alcohol consumption can indeed be thought of as relatively fixed over time (viz. belonging to one family of distributions). Moreover, the change in distribution in the Netherlands seems indeed to be linear (on a logarithmic scale) in nature, as was suggested in the hypothesis.

7.2 Empirical Probability Plots

Previous research has suggested a more or less mathematical similarity between distributions of alcohol consumption at different points in time, even when mean alcohol consumption changes dramatically (e.g. Bruun *et al.*, 1975). Yet, there is debate as to how strong this relationship between distributions is. As was explained above, results from parametric approaches, when models are postulated from which the data originate, have not proven very satisfactory. Two factors could be held responsible for this disappointing outcome: (1) the consumption data may not be specific or accurate enough to infer a model and/or (2) the phenomenon cannot be captured within a single, mathematical model. With regard to the former factor, many researchers have, over the years, expressed doubts as to the quality of alcohol consumption data (e.g. Midanik, 1988; Lemmens, 1987). The small coverage of sales of alcoholic beverages by survey data is seen as evidence of a lack of validity. Particularly, it is thought that under-representation and under-reporting is strongest for the heavier drinkers, with consequences for the right tail of the consumption distribution (Pernanen, 1974). Another problem is that consumption data are often grouped or that they cluster around specific values. Since consumption distributions are usually positively skewed, small deviances in the consumption scores in the right tail of the distribution have a dramatic effect on the estimates of the parameters of the fitted distribution. Although many aspects of the measurement problem have received attention in recent decades, a definite statement cannot be made. In general, the poor quality of the data minimizes the chances for a successful test of a theoretical model underlying alcohol consumption data.

With regard to this latter topic, it is argued in this paper that without making model choices (e.g. lognormality) it is still possible and relevant to study changes in the distributional pattern of alcohol consumption. For this purpose, empirical probability plots for a graphical comparison of two distributions are used. The technique of probability plotting for analysing the differences between or contrasting two empirical distributions, has been developed by Doksum (1974, 1977) and Doksum and Sievers (1976). No preliminary assumptions about an underlying theoretical distribution model

are necessary. With the procedure one can compare independent samples from two distinct populations or from one population at two different points in time.

In short, Doksum has shown that in a comparison of the distributions of two random variates X and Y (not necessarily independent), there is essentially only one, unique transformation function T for which holds: $Y = T(X)$. In other words: Y is distributed as $X + \Delta(X)$, and " $\Delta(\cdot)$ is the amount of shift needed to bring the X 's up to the Y 's in distribution and it is also referred to as the shift function" (Doksum and Sievers, 1976, p.421). Doksum (1977) clarifies that this shift function is a generalization of the usual location (or linear, or shift) model, where $X + \Delta$ has the same distribution as Y and Δ is a constant. The shift function is also a generalization of the scale model, where Y has the same distribution as ΘX . If neither a location model nor a scale model applies, which appears to be the case for the data in this paper, Doksum's method of "plotting with confidence" can give a good indication how two distributions are related. From the hypothesis of linear change on a logarithmic scale it is expected that two distributions X and Y are equal after a transformation of the form $\log Y = \alpha + b \log X$ where α and b are a location and scale parameter, respectively, and constants to be estimated. Note that Ledermann's model also implies this linear change, but additionally assumes that X and Y are lognormally distributed variates.

A graphical expression of the transformation function T can be obtained by a so-called empirical probability plot, where corresponding percentiles of the empirical distributions are plotted, in a way suggested by Doksum (1974). The hypothesis of linear change implies that a graphical plot of the empirical shift function $\Delta(x)$ approximates a straight line. Systematic deviations from linearity can give clues about possible alternative hypotheses. The corresponding percentiles are estimates derived from the X and Y samples (Doksum, 1974, p.271).

Statistically, the hypothesis of linear change can be tested by computing the simultaneous $(1-\alpha)$ weighted Kolmogorov-Smirnov confidence bands, suggested by Doksum and Sievers (1976), between which the transformation function $\Delta(x)$ can vary for a given value of x . Interpretation of the plots is quite simple. If a straight line can be drawn within the confidence bands, one cannot conclude that the two distributions are statistically different, up to a linear shift.

Although originally designed to test treatment effects, namely the distribution Y in the experimental sample is compared with a reference distribution X in the control condition, this application is not essential to the technique. One can choose any distribution as the reference (or control) distribution, provided that one keeps this choice in mind while drawing conclusions.

7.3 Data

The plotting procedure, outlined above, will be applied to alcohol consumption data from three Dutch, general population surveys held in 1970, 1981 and 1985 (cf. Jessen, 1974; Adriaanse *et al.*, 1981; Lemmens, 1987, respectively). The first two surveys dealt with medical consumption and medical risk factors, the third one was exclusively on drinking habits. The samples are independent. In table 7.1 a description of the surveys is given.

All surveys relied on self-reported alcohol consumption but differed in the method of data collection. In 1970, respondents were asked for their typical or usual weekly consumption (range from 0 to 100 glasses). This clearly differs from the recall approach in 1981, when the interviewer asked for actual consumption on the previous 3 weekend-days (Friday included) and on 2, most recent, weekdays. An estimate of weekly consumption was obtained by multiplying the consumption on weekdays by 2 and adding it to the reported consumption during the weekend. The survey in 1985 specifically focussed on alcohol consumption and related issues, and its design permits the calculation of several indices of alcohol consumption. From these indices, the one based on a diary has been chosen, which resulted in the lowest undercoverage of sales data (only 29%), the assumption being that this would yield the smallest error and, hence, the highest validity. Support for this assumption is given by, for example, Poikolainen & Kärkkäinen (1983) who found a test-retest reliability for diary data of about 90%.

Table 7.1. *Characteristics of three general population surveys held in the Netherlands in 1970, 1981 and 1985.*

	1970	1981	1985
Age-range original sample	≥ 16	21-65	16-70
Original sample size	1622	1301	1240
Non-response rate	15%	27%	31%
Average consumption(glasses/week)	6.2	7.5	11.5
Coverage of sales data	48%	44%	71%
Proportion abstainers	22%	22%	16%
Drinkers who report zero consumption in reference period:			
males	13%	10%	11%
females	25%	19%	20%
Number of respondents in comparisons:			
men	541	484	405
women	332	368	301

Similarly, in Williams *et al.* (1985) the diary is found to be superior in this respect to other measures (test-retest of 99%). Reported correlation between diary and recall methods ranges between .71 (Poikolainen & Kärkkäinen, 1983) and .78 (Lemmens, *et al.*, 1988).

Because of the allegedly superior qualities of the diary data, the 1985 distribution has also been taken as the reference distribution. So both the 1970 and 1981 distributions are contrasted with the 1985 distribution. In interpreting the results, however, one has to assume a linear relation (on a logarithmic scale) between the distributions obtained by the different data collection methods. Lemmens *et al.* (1988) suggest that one can indeed expect a linear relationship between a recall method, as the one used in 1981, and a diary method, like the one used in 1985.

Another, quite striking difference between the surveys is the abstention rate. This difference in abstention is probably due to the omission of a reference period in the question on abstention in the surveys in 1970 and 1981, rather than a result of a real change between 1981 and 1985. Drinkers at the lowest frequency levels will probably be most sensitive to the presence or absence of a reference period (Lemmens, 1987). In all three surveys, there are about twice as many abstainers among women than among men. Incidentally, this ratio is stable across surveys, which gives some support to the above assumption. Average consumption is about twice as large among men than among women in all three surveys. The same ratio is found for the fraction of drinkers with zero consumption in the survey period which is twice as large among women than among men. These drinkers are left out of the comparison because of insufficient differentiation at this level of measurement. Because of the expectedly high unreliability at the highest consumption levels (outliers?), the upper 2.5% of the drinkers has also been omitted from the analyses. The size of the samples permits a separate analysis of the male and female substrata over the respective years which enhances the homogeneity of the samples. For the same reason, abstainers and persons under 21 and over 70 years of age have been excluded.

7.4 Results

The results of the procedure are depicted in figure 7.1.

In the four plots, the horizontal axes represent the logarithmically transformed percentile values of the male and female, 1985 distributions. The vertical axes are log-transformed percentile values of the corresponding 1970 and 1981 samples. The solid line shows the relation between the successive percentiles of two distributions. The area between the dashed

lines denotes the .95-confidence band. In all four figures, the relations between successive percentiles do not seem to deviate systematically from linearity. Because a straight line can be drawn within the respective bands, the (null) hypothesis of a linear relation (on a logarithmic scale) cannot be rejected. Thus, the probability plots of longitudinal data confirm findings from previous studies which report a large degree of regularity of consumption distributions across populations. With regard to the restrictions imposed, the model confirmed here is in between the parametric Ledermann model and Skog's regression model.

The data at hand give the opportunity to address the question concerning alleged differences in the underlying distributional model of alcohol consumption between male and female subpopulations (Skog, 1977). Figure 7.2 shows a comparison of distributions of the three female subpopulations in 1970, 1981 and 1985 with the distribution of male consumption from the 1985 survey. Although the concavity of the plot of the women70 and men85 distributions indicates that the corresponding percentiles of the distribution

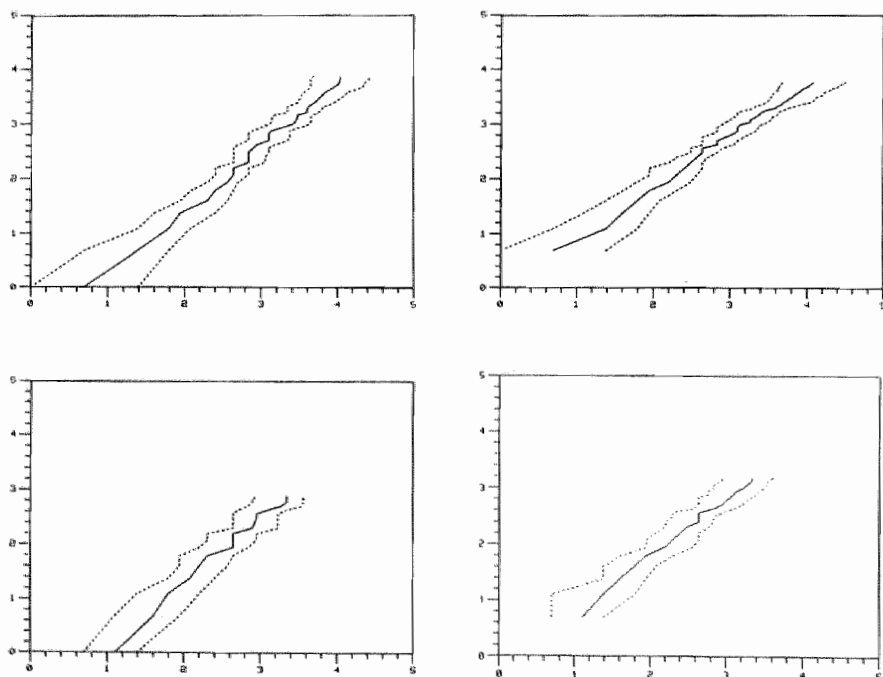


Figure 7.1. Empirical probability plots and the .95 confidence band of the male (top) and female (bottom) sample distributions in the years 1970 (left) and 1981 (right) with the respective 1985 distributions as reference (logarithmic scale: number of glasses per week)

of male consumption (after transformation) are somewhat higher than of female consumption, it is still clear, and quite convincingly so, that the linearity assumption cannot be rejected. It does seem plausible then that the same transformation function can be used to describe differences between distributions across male and female subpopulation and that one cannot discriminate male from female populations as regards underlying distribution model. The stability in the relation over a period of 15 years is surprisingly high.

An obvious next step in the procedure was to try and estimate the parameters indicative of the differences between distributions (viz. the parameters of location and scale, a and b , respectively). This proved not to be that easy, however. Bjerve and Doksum (1978) have explored several options which are useful as a rough impression of the magnitude of the parameters. It appeared difficult to assess the exact quality of the

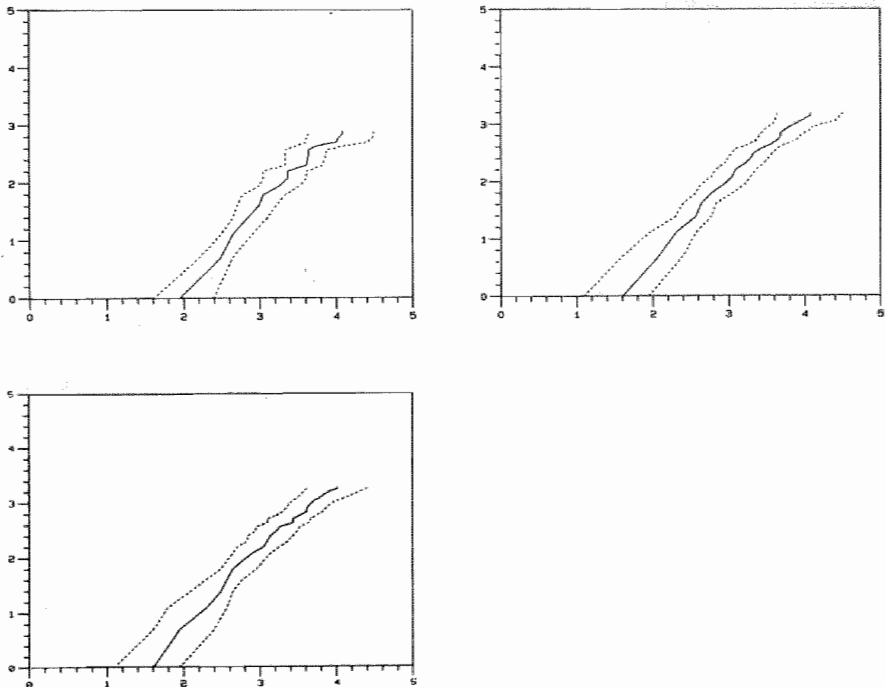


Figure 7.2. Empirical probability plots and the .95 confidence band of the female sample distributions in 1970, 1981 and 1985 with the male 1985 distribution as the reference (logarithmic scale: number of glasses per week)

point-estimates. In a recently published paper a statistical method is presented that relates mean consumption to the parameters of the transformation function by means of region estimates (Tan *et al.*, 1990).

7.5 Discussion

In this paper a generalized procedure has been applied to test the hypothesis about the linear relationship on a logarithmic scale between distributions of consumption. The graphical displays of the corresponding percentiles of the distributions, the so-called probability plots, yield information about the relationships between distributions of alcohol consumption. It was concluded that, up to a linear transformation, the distributions show a remarkable similarity over the years. This similarity is found in the comparisons of the distributions of male consumption as well as for the distributions from the female sample, though a little less pronounced in the latter case. The comparisons of the distributions across the male and female subsamples show that, again up to a proper linear transformation, the empirical distributions of consumption from the male and female sub-population are also closely related. It should be noted that, in general, the parameters of the linear transformation do not need to be equal over all comparisons.

The results reported in this paper seem to indicate that aggregated data-sets of alcohol consumption are indeed characterized by a certain regularity in distributional pattern. So, it is quite possible that the same underlying distributional model is applicable to male and female subpopulations, over time and even across sub-populations. One is tempted then to infer that this regularity in distribution is brought about by processes consistently operating, and equal for men and women. Processes whereby change in individual consumption is proportional to initial level, similar to Skog's concept of multiplicativity. It is, however, not a logical necessity that all factors that influence consumption level at one time should be present at all times. The distributions considered are aggregates, the final result of many simultaneous operations.

While interpreting the results in this manner one should, however, keep the following three technical and theoretical considerations in mind.

Firstly, the fact that the null hypothesis cannot be rejected does not imply that a lognormal or any other model actually underlies the data. It only signifies that with the given data-sets and the given test one cannot discriminate between the various models. Even though the power of the statistical test is not very high, one should not stress this fact too much, since

visual inspection clearly suggests that the data can be described rather well by one class of distributions that only differ from each other by a scale and location shift.

Secondly, it is open to debate whether an exact determination of the underlying distribution model, should it be possible to specify one, is crucial for an explanation of the dynamics of drinking behavior in large populations. It has pragmatic relevance in that it facilitates estimation of excessive use, in the same way the Ledermann model has been used. Knowledge of the parameters governing the distribution process does, however, not automatically contribute to a better understanding of the mechanisms and factors that determine consumption level. These are, as Skog and others have noted, probably manifold. The distribution could turn out to be very complex and difficult to specify. The method used in this paper does not address this issue of parametric choice. Instead, the method reveals whether one will be able to distinguish up to a scale and location shift between empirical distributions of consumption, irrespective of the particular form of the distribution(s). Given the large statistical uncertainty in the estimation of the parameters of the underlying distributions, it does not seem proper to make more explicit parametric statements about the relations between distributions than the one given in this paper. Nevertheless, the method has confirmed Ledermann's notion of a more or less fixed distribution and in this respect it is more restrictive than Skog's regression model.

Thirdly, part of the statistical uncertainty obviously stems from the limited (not to say, poor) quality of the available data on total volume of drinking. Research on the distribution of consumption should pay attention to this problem. Many times, also in the present case, comparisons are made without sufficient attention being paid to differences in the data being compared that are a result of differences in, for instance, data collection methods. Without knowledge of the effects of methodology, statements about relationships between "actual" distributions will be of a provisional nature. Clearly, if one aspires to elaborate on this subject more research should be directed to the study of the effects of methodology. Studies with simulated data on drinking patterns and reporting behavior could be of help in this area.

footnote 1.

The time period in the Ledermann model is one year, leading to a distribution of annual alcohol consumption or, actually, mean daily consumption in a certain year, much like income distributions. However, much unlike income data, data collection methods of alcohol consumption data do not cover an entire year, usually. The assumption in calculating annual individual consumption on the basis of a smaller period (for instance,

one week) is that time-variation of consumption can be neglected at the aggregate level. Reliability of individual consumption estimates is expected to be limited, but, with sufficient sample size and survey-period length, validity of the data with regard to the aggregated distribution is presumably unaffected by variation over time (Alanko, T. 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/Alcohol Epidemiology Section meeting in Rome, June 2-7, 1985). Thus, there is an expected loss of precision but the data do provide an unbiased distribution.

footnote 2.

Although this paper will not discuss the matter in detail, some elaboration seems appropriate here. As Skog (1985) explains, the idea of multiplicative change as an explanation of skewness comes from the law of proportionate effect, formulated by Gibrat in 1930, which applies to the situation where a change in a non-negative variate is proportional to the previous value of the variate. For large numbers a variate subject to this law approximates a 2-parameter lognormal distribution (Aitchison & Brown, 1969, p.23). However, Skog's illustration of multiplicative change of alcohol consumption (1985, p.87) does not, altogether, seem appropriate as an explanation of skewness, since it assumes an initial symmetry in distribution. Consequently, it implies an ever-increasing skewness since the process is supposed to operate continually, and the larger the number of steps in the process, the larger the variance. This sequential aspect in the genesis of the lognormal (or otherwise skewed) distributions is not essential, however. Aitchison and Brown propose an explanation in terms of a simultaneous operation of a large number of causes (in fact, they seem to suppose non-homogeneity of the drinking population), a solution also mentioned by Skog. The genesis of skewness or lognormality is not pursued any further here. The reader is referred to Aitchison and Brown (1969, ch.3). This paper will confine itself to questions regarding the specified hypothesis concerning change over time, without trying to explain the form of the distribution.

footnote 3

Actually, Skog concept of multiplicative change differs from the one proposed in this paper in that Skog's regression model assumes equality in expectation (viz. $E(\log Y/X) = \alpha + \log X$) whereas the present model assumes equality in distribution ($\log Y \stackrel{d}{=} \alpha + \log X$), after a proper logarithmic transformation of the variate.

References

- ADRIAANSE, H., DROP, M.J., HALFENS, R. & PHILIPSEN H. (1981) *Leeft Nederland Oké?* (Maastricht, University of Limburg).
- AITCHISON, J. & BROWN, J.A.C. (1969) *The lognormal distribution* (Cambridge, University Press).
- BJERVE, S. & DOKSUM, K.A. (1978) Estimation in some non-linear models, *Journal of the American Statistical Association*, 73, pp. 366-370.
- BRUUN, K., EDWARDS, G., LUMIO, M., MÄKELÄ, K., PAN, L., POPHAM, R., SCHMIDT, W., SKOG, O.-J., SULKUNEN, P. & OSTERBERG, E. (1975) *Alcohol Control Policies in Public Health Perspective* (Helsinki, The Finnish Foundation for Alcohol Studies).
- CENTRAL BUREAU OF STATISTICS (1972) Een Analyse van de Ongelijkheid van de Persoonlijke Inkomens in 1967, *Statistische en Econometrische Onderzoekingen* No. 13 (The Hague, Central Bureau of Statistics).
- DOKSUM, K.A. (1974) Empirical probability plots and statistical inference for non-linear models in the two-sample case, *Annals of Statistics*, 2, pp. 267-277.
- DOKSUM, K.A. (1977) Some graphical methods in statistics: a review and some extensions, *Statistica Neerlandica*, 31, pp. 53-68.
- DOKSUM, K.A. & SIEVERS, G. (1976) Plotting with confidence: graphical comparisons of two populations, *Biometrika*, 63, pp. 421-434.
- DUFFY, J.C. (1977) Estimating the Proportion of Heavy Drinkers, in: *The Ledermann Curve: Report of a Symposium*, pp. 11-24 (London, Alcohol Education Centre).
- DUFFY, J.C. (1986) The distribution of alcohol consumption - 30 years on, *British Journal of Addiction*, 81, pp. 735-741.
- DUFFY, J.C. & COHEN, G. (1978) Total alcohol consumption and excessive drinking, *British Journal of Addiction*, 73, pp. 259-264.
- EATON, W.W. & WHITMORE, G.A. (1977) Length of stay as a stochastic process: a general approach and application to hospitalization for schizophrenia, *Journal of Mathematical Sociology*, 5, pp. 273-292.
- EPPINK, Th.W.A. (1988) *Choice of Mathematical Models in Geographic Research Considering Alternatives* (Nijmegen, The Netherlands, University of Nijmegen).
- GUTTORG, P. & SONG, H. (1977) A note on the distribution of alcohol consumption, *Drinking & Drug Practices Surveyor*, 13, 7-8.
- JESSEN, J.L. (1974) *Medische Consumptie* (Groningen, the Netherlands, University of Groningen).
- LAWRENCE, R.J. (1984) The lognormal distribution of the duration of strikes, *Journal of the Royal Statistical Society, A*, 147, part 3, pp. 464-483.
- LEDERMANN, S. (1956) *Alcool, Alcoolisme, Alcoolisation*, Vol.1 (Paris, Presses Universitaires de France).
- LEDERMANN, S. (1964) *Alcool, Alcoolisme, Alcoolisation*, Vol.2 (Paris, Presses Universitaires de France).
- LEMMENS, P. (1987) *Het Ledermann-model Nader Bezien* (Maastricht, University of Limburg).

- LEMMENS, P., KNIBBE, R. & TAN, F. (1988) Weekly recall and diary estimates of alcohol consumption in a general population survey, *Journal of Studies on Alcohol*, 49, pp. 131-135.
- MIDANIK, L.T. (1988) Validity of self-reported alcohol use: a literature review and assessment, *British Journal of Addiction*, 77, pp. 357-382.
- PARKER, D.A. & HARMAN, M.S. (1978) The distribution of consumption model of prevention of alcohol problems: a critical assessment, *Journal of Studies on Alcohol*, 39, pp. 377-399.
- PERNANEN, K. (1974) Validity of survey data on alcohol use, in: Gibbins, R.J., Israel, Y., Kalant, H., Popham, R.E., Schmidt, W. & Smart, R.G. (Eds) *Research Advances in Alcohol and Drug Problems*, Vol.1., pp. 355-373 (New York, John Wiley & Sons).
- POIKOLAINEN, K. & KÄRKKÄINEN, P. (1983) Diary gives more accurate information about alcohol consumption than questionnaire, *Drug and Alcohol Dependence*, 11, 209-216.
- SKOG, O.-J. (1977) Does the Same Distributional Model for Alcohol Consumption Apply to Both Male and Female Populations? *Sifa-Mimeograph Series* No. 10 (Oslo, National Institute for Alcohol Research).
- SKOG, O.-J. (1980) Social interaction and the distribution of alcohol consumption, *Journal of Drug Issues*, 10, pp. 71-92.
- SKOG, O.-J. (1985) The collectivity of drinking cultures: a theory of the distribution of alcohol consumption, *British Journal of Addiction*, 80, pp. 83-99.
- SKOG, O.-J., GUTTORP, P. & SONG, H. (1979) A note on the distribution of alcohol consumption: a reply and rejoinder, *Drinking and Drug Practices Surveyor*, 15, pp. 3-6.
- TAN, E.S., LEMMENS, P.H.H.M., KONING, A.J. (1990) Regularity in alcohol distributions: implications for the collective nature of drinking behaviour, *British Journal of Addiction*, 85, pp. 745-750.
- WILLIAMS, G.D., AITKEN, S.S. & MALIN, H. (1985) Reliability of self-reported alcohol consumption in a general population survey, *Journal of Studies on Alcohol*, 46, pp. 223-227.

Comparisons of Distributions of Male and Female Alcohol Consumption from Surveys in 5 Western Countries.

Summary

Distributions from four Western countries have been compared with the Dutch distributions from the 1985 survey in order to assess whether regularities that have been found in chapter 7 can be generalized to other populations, or should be seen as a local phenomenon. Since the surveys differ with regard to the method with which the data have been collected, an assessment has been made first of the relation between two commonly used self-report methods, a quantity-frequency (QF) and a recent occasion (WR) method. It appeared that the null-hypothesis of a linear relation between methods could not be rejected. However, the quality of the resulting plots was better for data methods that register actual consumption behavior (WR-type). This methodological effect was also apparent in the plots of cross-sectional comparisons. After a proper linear transformation, the logarithmic consumption data of male and female sub-samples, from the USA in 1974, 1976 and 1977, from the UK in 1978 and 1980, from West Germany (1978) and Switzerland (1981), have been contrasted with the respective Dutch distributions. Though deviations from linearity are sometimes large, they do not seem to be systematic in most of the plots (an exception: female consumption distributions from the USA). These results indicate that with some methodological reservations regularity in distribution, one of the substantive hypotheses in the Ledermann model, is not only present in the Dutch alcohol situation, but can be observed in other Western countries as well. This conclusion suggests that change in distribution of alcohol consumption can be predicted, given the statistical uncertainty. The precision of the plotting procedure, and hence, of

statements about change in distribution, seems dependent not only upon sample size, but also upon type of measurement method.

8.1 Introduction

Main question in this chapter is whether the regularities in distribution that have been found in chapter 7 for Dutch, longitudinal survey data are unique to the changes in the alcohol situation in the Netherlands, or whether this relative 'stability' can also be observed between countries that presumably differ in drinking practices. For this purpose, 14 empirical consumption distributions from 7 male and 7 female samples from general population surveys in 4 countries have been compared with the respective sub-samples of the 1985 Dutch survey, already described in the previous chapter. All surveys are from general population samples, consisting of subjects older than 15 years of age. Sample sizes vary considerably, but none is less than 500 (female sub-sample in the Netherlands). In tables 8.1 and 8.2 a description of the samples is given.

At another occasion three of these surveys, from the Netherlands, Germany and Switzerland, have been used in a comparison of social correlates of drinking (Knibbe *et al.*, 1987). It has been found that though the countries (and sub-populations) differ in mean annual consumption, they do not differ very much with regard to social determinants of consumption level. This homogeneity may be reflected in a regularity of consumption distributions as outlined in the previous chapter.

On the basis of these findings, together with the hypothesis of multiplicativity and the results of the previous chapter, it is expected that the distributions do not differ from each other up to a linear transformation of the logarithmic consumption data. Similar to the procedure in chapter 7 and for the same reason (lowest undercoverage, and, hence, presumably highest validity), the male and female distributions of the Dutch 1985 diary reports of consumption are taken as the reference.

In the case of an observed deviation from linearity, two interpretations are possible. Either, the observed differences between distributions are real or they are caused by error. Error can be the result of the variation in the methodology of the surveys, among which differences in method with which the data are gathered. To assess the effect of this possible source of error, the same graphical comparison is presented of distributions obtained with two, widely used methods, a "quantity-frequency" (QF) and "recent occasion" (WR) type of method. The Dutch 1985 survey which simultaneously has gathered QF-, WR-, and *diary* data on consumption offers a unique

possibility for a test of linearity of the error that may be introduced by the type of measurement that has been used (see chapter 4 for a complete description of the 3 methods of measurement). The results should be regarded in relative terms (error is relative to the presumably more valid diary), as in chapter 3. The advantage over the regression approach in chapter 3 (see, figure 3.1), is that with the probability plots a direct visual inspection of the relation is obtained. One can easily observe (large) deviations from linearity at the different levels of consumption. It should be noted that the samples used in the procedure need not be independent (Doksum, 1974, pp. 270).

An overview of the methods with which consumption variables have been assessed in the various surveys and relevant descriptive statistics is presented in table 8.2. No comparison on overall coverage of sales data can be presented here. It has not been possible to obtain population estimates of consumption, since the necessary information on weights for the several sub-strata is often lacking and some sub-strata may be oversampled.

Another interesting detail of the data at hand is that 3 surveys have been held in the USA (1974, 1976, 1977) and 2 in the UK (*G.H.S.* in 1978, 1980) and thus a longitudinal comparison of distributions can be made, analogous to the one with Dutch data in the previous chapter.

8.2 Data

In table 8.1 a description of relevant information on the surveys is given. The column "*0-consumption*" depicts the proportion of drinkers, non-abstaining but who do not drink in the particular survey reference period, or who drink so little that they have a zero reported consumption in the data set. Because variation in this group is lacking, they have been omitted from the plots (actually, percentile values equal to 0 have been omitted pairwise; the group of zero-drinkers is, however, included in the computation of percentile values). The table shows that the percentage abstainers and zero-reporting drinkers varies considerably across the samples, but the figures in samples from one particular country (USA, UK) are quite similar. Again, the variation in these percentages may indicate real differences between countries but it may also be a result of national differences in survey methodology or response behavior. As the Dutch 1985 survey has learned, for instance, recent occasion methods that are necessarily restricted to consumption in 1 week, yield higher estimates of prevalence of non-abstaining zero-drinkers than do the summary measures. For reasons

Table 8.1. *Description of 16 data sets from 8 general population surveys held in 5 countries.*

	year	grossN	abstain	0- consumption	NetN
Netherlands	1985				
male		611	9%	11%	405
female		522	23%	20%	301
USA	1974				
male		1322	18%	8%	960
female		1565	37%	11%	804
USA	1976				
male		998	19%	7%	730
female		1486	32%	11%	846
USA	1977				
male		1423	17%	8%	1046
female		1805	31%	13%	1007
UK	1978				
male		9173	4%	8%	8090
female		10163	9%	24%	6797
UK	1980				
male		9088	4%	9%	7921
female		10079	10%	21%	6993
Germany (FRG)	1978				
male		817	3%	4%	756
female		860	7%	11%	708
Switzerland	1981				
male		584	4%	5%	530
female		835	14%	15%	591

discussed in the previous chapter, the upper 2.5% of the distributions are likewise deleted.

As can be seen in table 8.2, methods of data collection do not vary over surveys within one country (except for the special Dutch 1985 survey). Subjects older than 70 years of age and younger than 20, have been excluded from all samples. In the reference section, the principal research institutes that have carried out the survey are listed.

Basically, and already indicated earlier, two general types of assessment can be distinguished, a "quantity-frequency" (QF)-type and a "recent occasion", weekly recall (WR)-type. Even though some methods are of the same type, they still differ on many points (e.g. specification of type of drink and location). The basic self-report questions with which abstention and consumption have been assessed are presented in table 8.2.

Table 8.2. *Characteristics of the surveys and data sets that have been used in the comparisons.*

survey	survey period	age	abstention period	type of method	consump period	type of drink*	mean consumption (glases/week)
NL 85	Mar-Dec	16-70	½ year	diary	1 week	BWFS	men: 15.3 women: 6.8
NL 85	Mar-Dec	16-70	½ year	QF	6 month	any	men: 11.7 women: 5.6
NL 85	Mar-Dec	16-70	½ year	WR	1 week	BWFS	men: 15.0 women: 6.4
USA 74	-	>17	1 year	QF	1 month	any	men: 6.2 women: 2.6
USA 76	-	>17	1 year	QF	1 month	any	men: 6.4 women: 2.5
USA 77	-	18-90	1 year	QF	1 month	any	men: 6.6 women: 2.6
UK 78	Jan-Dec	0-99	1 year	QF	1 year	BWFS	men: 20.4 women: 5.4
UK 80	Jan-Dec	0-99	1 year	QF	1 year	BWFS	men: 19.1 women: 5.8
FRG	Dec	>15	none	WR	1 week	8 types	men: 22.1 women: 10.9
Switz	-	15-75	1 year	QF/WR	1 mo/1 wk	BWMS	men: 25.1 women: 7.8

* B = beer; W = wine(s); F = fortified wine; S = distilled; M = "most"

8.3 Assessment of the relation between different types of methods

An investigation of the possible difference in distribution resulting from *QF*- and *WR*-type of methods is given in figure 8.1. Distributions of consumption obtained with a *QF* and *WR* method from the Dutch 1985 survey have been compared with the *diary* (*D*). The basic question is whether the relation between “quantity-frequency” and “recent occasions” on the one hand, and the reference diary distribution on the other is linear, in which case the error-interpretation, mentioned in the introduction section, can be ruled out, or, at least, made less plausible.

The plot of the weekly recall and diary method is a fairly smooth and straight line for the male subsample (figure 8.1). In the female sub-sample, the *WR/D* plot seems a little concave, mainly in the lower part of the distribution. For the *QF* method the plots show a less smooth line, but the deviation from linearity does not seem to be systematic. The irregular

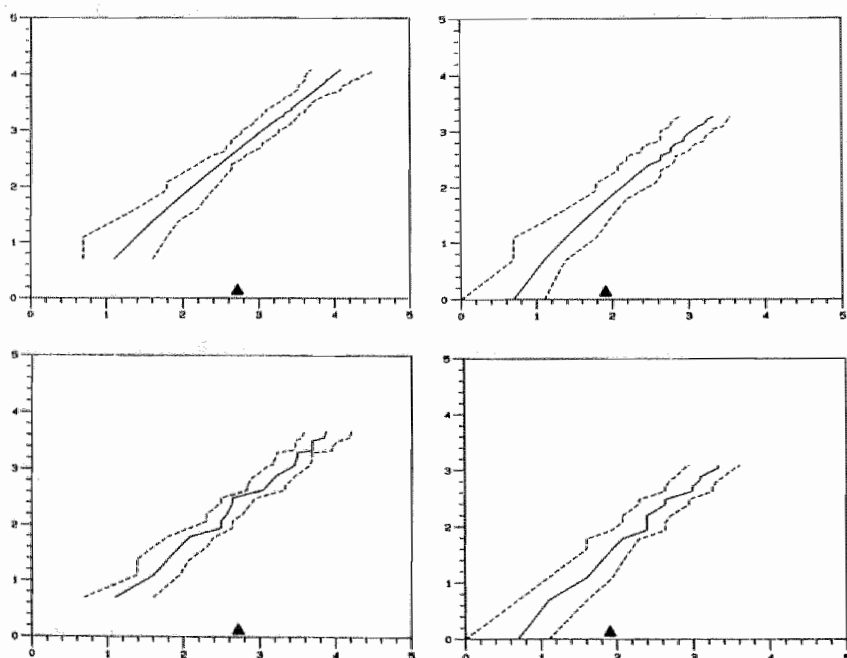


Figure 8.1. Plots of the distribution of Dutch 1985 male (top) and female (bottom) diary consumption data (x-axis) and linearly transformed distributions obtained by *WR* (left) and *QF* (right) method from the same sample (y-axis). All data are log-transformed. \uparrow = 50th percentile point

pattern may very well be the result of the larger grouping of observations in the QF data, caused by the restricted categories of its component variables, the *UF* and *UQ* (*usual frequency and quantity* - see chapter 4). The *UF* consists of only 8 categories, ranging between "every day" and "never". The *UQ* is an open ended questions, but the answers are usually round figures. The requirement in the plotting procedure of measurement on a continuous scale may be more or less violated. Because both *WR* and diary distributions are based on self-reports of actual drinking, their plots are less distorted.

8.4 Comparisons of distributions from different countries

The plots of the percentile values of distributions from the sub-samples, specified in table 8.1, after a proper linear shift of the (log-transformed) data are given in figure 8.2a and 8.2b. The X-axis depicts the reference, diary distributions of the relevant Dutch 1985 data. The Y-axis denotes the scores of the transformed distribution. Comments on the plots are presented pointwise below:

1) It can be seen that the null-hypothesis, namely linearity of the relation, cannot be rejected for any comparison. In all plots a straight line can be drawn in the area between the dotted confidence limits.

2) All three plots of male NL85 dairy distribution with male US74, US76 and US77 samples (figure 8.1a) are remarkably similar. Even the relatively large, though seemingly non-systematic deviations in the upper half of the distributions are quite constant over the years. Visually, the pattern resembles the one in figure 8.1 (*QF/Diary* plot).

3) The plots with the female samples US74, US76 and US77 (figure 8.1a) suggest a deviation from linearity (though the null-hypothesis cannot be rejected), since the line is slightly convex, particularly in the upper part of the distribution. Convexity indicates that the increment in p-value is higher in the US than in the Netherlands.

4) On the whole, respective comparisons of distributions of male and female UK78 and UK80 samples (figure 8.2b) show a clearly linear relation between the Dutch and British distributions. The line is relatively smooth, compared to the previous plots. The pattern is similar to the one in figure 8.1 (*WR/Diary* plot). The relation at the lower end of the male distributions in the UK78/NL85 plot is a little problematic, and may point toward a slight convexity.

5) As in the UK situation, deviations from linearity in the plots of male and female samples of the FRG (figure 8.2b) are relatively small and

non-systematic. At the lower end of the female distributions, the confidence band becomes very large and the line seems to “flatten”.

6) Similarly, deviations from linearity, especially at the higher levels, are small in the plots of NL85 diary distributions with male and female samples from Switzerland (8.2b). Contrary to the situation in the above female German sample, the line in the female Swiss plot seems to “drop” dramatically at the lowest levels.

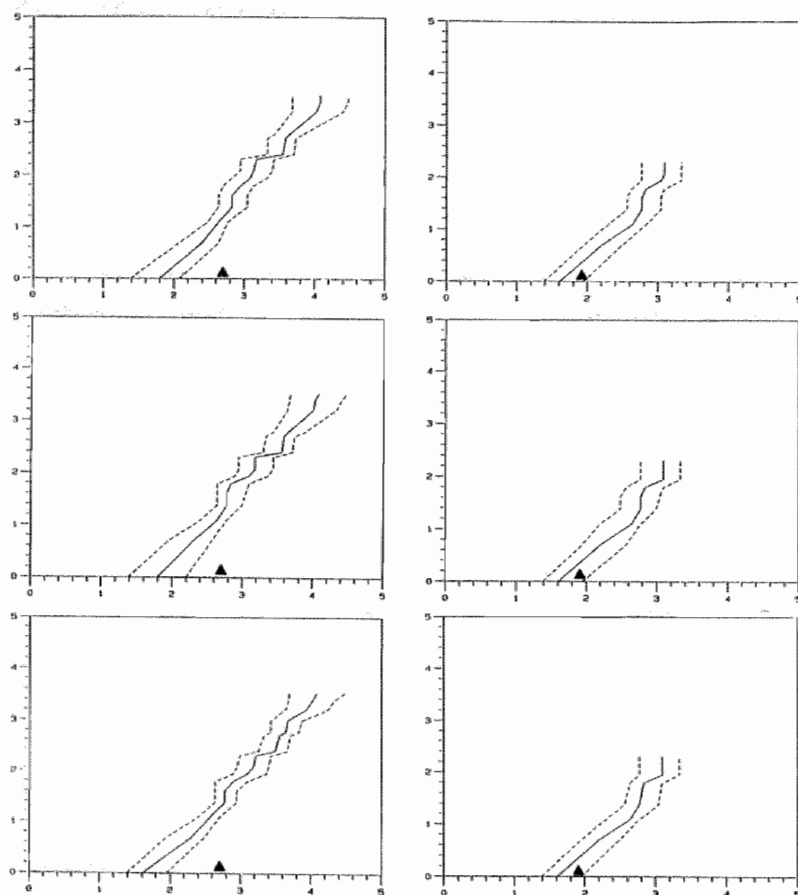


Figure 8.2 a. Plots of the distribution of the Dutch 1985 male (left) and female (right) diary consumption data (x-axis) and linearly transformed distributions from the USA surveys held in 1974, 1976 and 1977 (y-axis). All data are log-transformed.

↑ = 50th percentile point

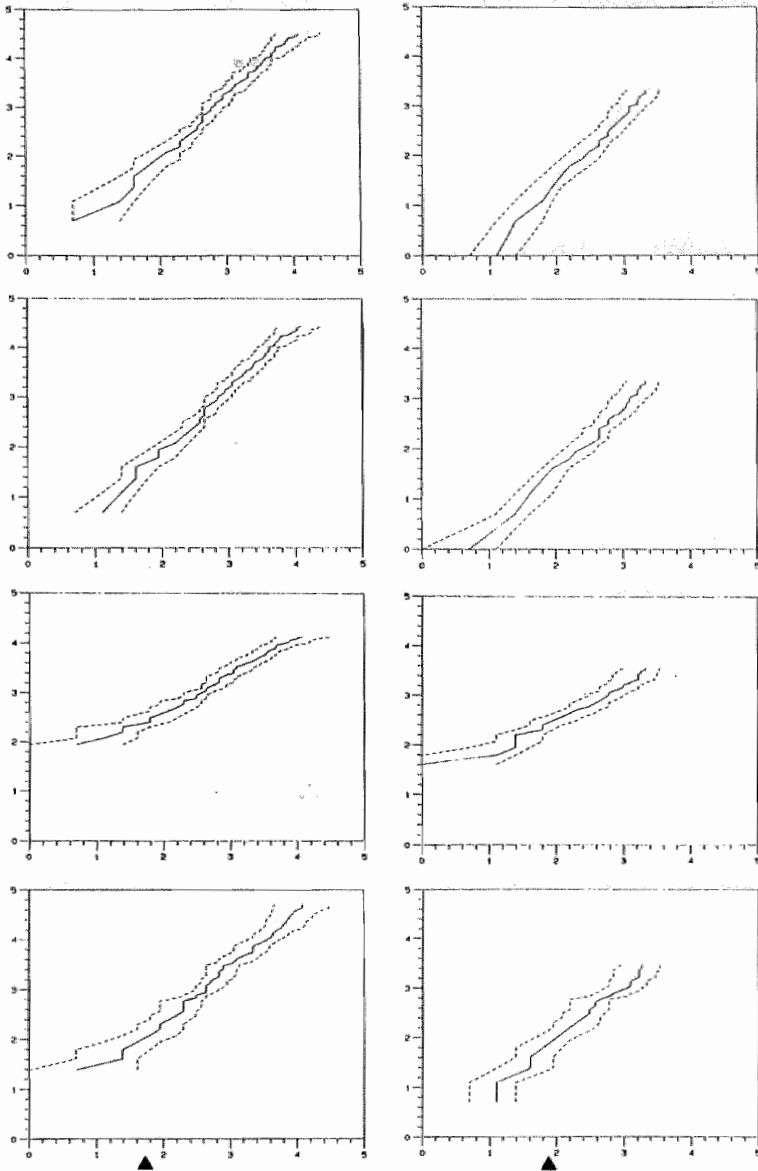


Figure 8.2 b. *Plots of the distribution of the Dutch 1985 male (left) and female (right) diary consumption data (x-axis) and lineraly transformed distributions from surveys held in the UK (1978 and 1980), the FRG (1978) and Switzerland (1981) (y-axis).. All data are log-transformed. ↑ = 50th percentile point.*

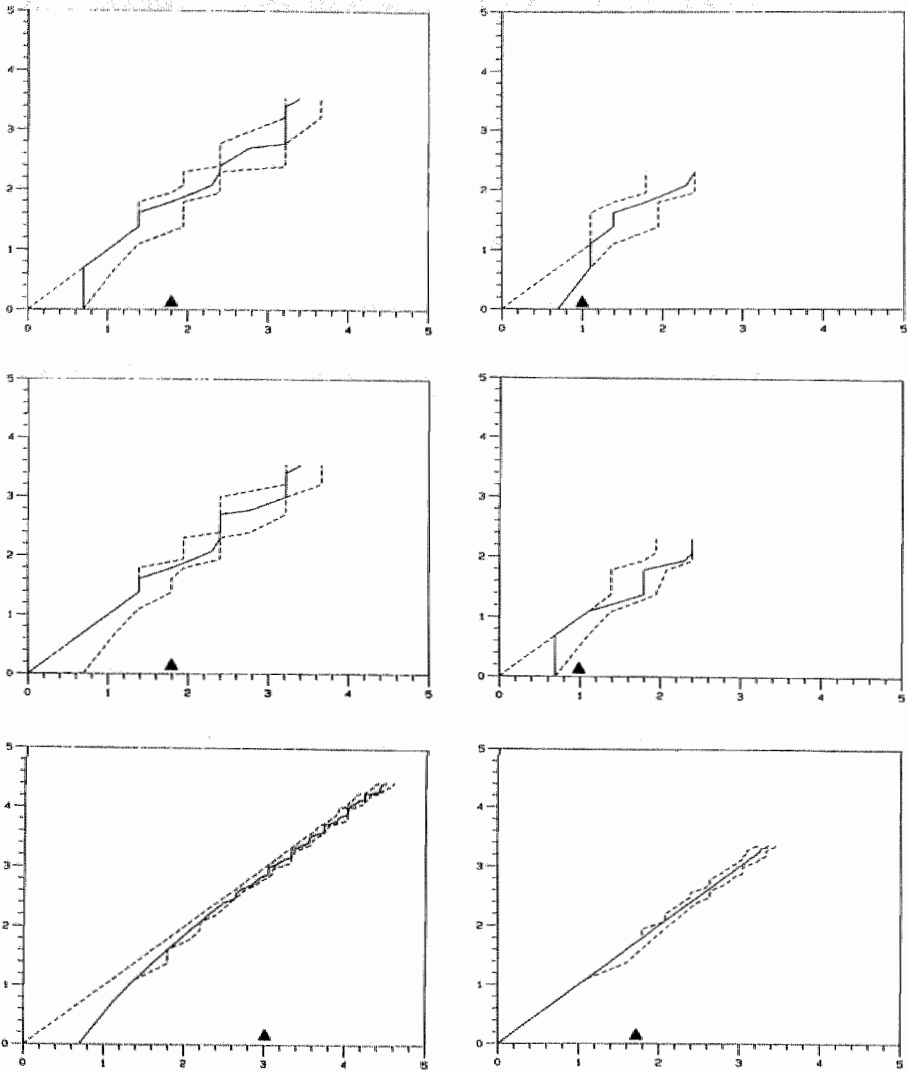


Figure 8.3. Plots of the distribution of the male (left) and female (right) consumption data from the US 74 sample (top 2) and UK 78 (bottom) on the one hand (x-axis) and respective linearly transformed distributions from the US 76, US 77 and the UK 80 sample (x-axis). All data are log-transformed.
↑ = 50th percentile point.

8.5 Longitudinal comparisons between distributions from the USA and the UK

Analogous to the analyses in the previous chapter, the plots of male and female distributions from surveys held in the USA and the UK in several years have been calculated (figure 8.3). The comparisons are made with the "oldest" survey as the reference distribution (for the USA, 1974; for the UK, 1978). The USA samples show large deviations from linearity. It is clear that it will be difficult to sustain the idea of linearity at all levels. This is all the more remarkable because the change in mean consumption over the years is only marginal. The plots of the UK samples, on the contrary, show a rather smooth line. The confidence band is small due, in part, to the large sample size.

The plot of the female 1978 and 1980 distributions is a clearly a straight line. The plot of male 1978/1980 distributions is somewhat concave. This concavity is caused mainly by the lower percentiles, and signifies that the (lower) percentile values of the 1978 sample are relatively higher than those from the corresponding percentiles of the (transformed) distribution of 1980.

8.6 Discussion

The main results of the previous sections are summarized below in the following 4 paragraphs.

On the whole, the relation between distributions of (log-transformed data form) *WR* and *diary* is linear, but the hypothesized linearity may not be tenable over the entire female distribution, particularly in the lower part. Linearity of the relation between distributions obtained by *QF* and *diary* method has not been firmly established. The deviations from a straight line are much larger than in the *WR/Diary* plot, even though they do not seem to be systematic in nature.

In the light of the rather high correlations between scores of *QF* and *WR* on the one hand and *diary* on the other (see chapter 4), different types of data collection method do not altogether seem to systematically distort the distributional characteristics. However, a salient detail that can be observed in all figures (8.1 - 8.3) is that comparisons between distributions of consumption obtained by means of methods that are based upon self-reports of actual drinking occasions yield a smoother "fit", a less wide confidence band and a less erratic pattern. As mentioned before, a strong concentration of scores on single values may be the cause of this phenomenon. This may

also explain the rather large deviations from a straight line in the plots of USA distributions in figure 8.3. One might infer, then, that the *QF*-type of method is less suited for the procedure used in this chapter.

As regards the international comparisons the results are clearly in favor of the linearity hypothesis. It is remarkable that even with such large differences between samples in mean consumption the distributions do not differ that much, that is after a proper linear shift. Of course, it has been noted that in some cases there seem to be systematic deviations, which may in part be attributed to methodological factors. The answer to the question posed in the introduction is that similarity in distribution is by no means a national, Dutch phenomenon but that it can also be observed in other populations, for instance, in the British general population (78/80).

Linearity of the relation between distributions as specified above (linear shift of log-data) implies a large predictability of distributional shape. This predictability is not present in all cases and at all levels of distribution. Still, it seems justified to conclude, after having made some methodological reservations, that in Western countries the transition from one distribution into the other can be predicted quite easily. With these results the idea of Ledermann of a certain invariance in the way alcohol consumption is distributed is affirmed and can be generalized to other populations than the Dutch. Equally, the results confirm Skog's concept of proportional change, and give substance to the empirical observation of differential change by different types of drinkers (as defined by percentiles, see figure 1.3). Practical use of this plotting procedure seems somewhat hampered by methodological factors. The UK comparisons in figure 8.3 show that with uniformity in type of measurement method and sufficient sample size, the procedure may be of practical value.

References

- DOKSUM, K.A. (1974) Empirical probability plots and statistical inference for non-linear models in the two-sample case, *Annals of Statistics*, 2, pp. 267-277.
- KNIBBE, R.A. & LEMMENS, P.H.H.M. (1987) Korrelate des Alkoholkonsums in der Schweiz, Deutschland, und den Niederlanden, *Drogalkohol*, 11, pp. 27-42.
- Principal research institutes:
- United States of America*: SOCIAL RESEARCH GROUP, Washington. National Survey on Drug Abuse. Source: Deadat, TCU, Fort Worth.
- West-Germany*: INSTITUTE FUER DEMOSKOPIE, Allensbach. Gefahr des Alkoholismus. Source: Zentralarchiv fuer empirische Sozialforschung. Köln.
- Switzerland*: SCHWEIZERISCHE GESELLSCHAFT FUER PRAKTISCHE SOZIALFORSCHUNG, Zuerich. Alkohol 81/301. Source: SFA/ISPA, Lausanne.
- United Kingdom*: OFFICE OF POPULATION CENSUSES AND SURVEYS. General Household Survey. Source: SSRC data archive, Colchester.

A recapitulation and extension.

Summary

In this closing chapter, results of the methodological studies have been reviewed and discussed, with special reference to the study of the distribution of alcohol consumption. For example, evidence has been accumulated that the relationship between the true and error component of consumption scores is a linear one. This means that conclusions about changes in distribution (as formulated in ch7 & ch8) are unaffected by this type of error. Furthermore, a linear type of error does not bias coefficients of correlation. Several alternative designs and techniques have been discussed which may prove useful for an assessment of the effects methodology has on the quality of data. In this respect, computer simulations of drinking, together with models of memory and telescoping effects in self-reports have been suggested. It is argued that future research should develop at least some theory of measurement, on how people respond to self-report questions on alcohol. This will enhance comparability of research across time and populations, and researchers will be better equipped to tune the methodology to their specific research questions.

In the second part of the chapter, research conducted within the same "*Ledermann-project*" has been presented on topics such as problems connected to estimates of the shift function, incompatibility of the hypotheses on regularity and collectivity, and stability of individual drinking

behavior over large periods of time. The chapter ends with a general discussion on the impact a “single-distribution” model has on the “degrees of freedom” of a primary prevention policy.

9.1 Introduction

In the previous chapters several issues concerning the measurement of alcohol consumption in the general population have been discussed. In the introductory chapter 1 it was claimed that an assessment of the validity of consumption data is a necessary requisite for the study of the distribution of alcohol consumption. A bias in the basic consumption data could lead to erroneous conclusions about characteristics of the shape of the distribution or about relations between distributions, for instance those presented in chapters 7 and 8. The link between both topics central in this thesis has already been made in chapters 2, 3 and 8. In section 9.2, the main results of the methodological studies will be resumed, with special attention being paid to their implications for the study of the distribution of alcohol consumption. In the other sections, practical and theoretical implications of the results of the three chapters on the distribution of consumption (1, 7, 8) will be considered, based partly on analyses performed in collaboration with other researchers working on the so-called “Ledermann-project” of the University of Limburg (see Tan *et al.*, 1990).

9.2 Recapitulation

In this section, the implications of the methodological studies are discussed.

9.2.1 *Linear or non-linear error.*

“Survey self-report data on drinking underestimate total consumption based on sales of alcoholic beverages”. To this general conclusion one could add that certain survey methods show a larger undercoverage of sales than others. In estimating alcohol consumption in the population, this underreporting certainly has to be taken into account. Important in this respect is whether it is possible to specify the nature of the relation between the “true” and observed consumption. In other words, do the results of the present study offer an indication how serious the apparent low validity of

empirical consumption data affects results and conclusions based on these data and how the downward bias could be corrected?

Results of the empirical studies in this thesis give support to the linearity assumption which states that the functional relationship between the true and error component in empirical consumption data is a linear one (for a discussion on the impact of non-linearity see also chapter 3). In chapter 3, the conclusion with regard to this assumption has been that, even though underreporting in *QF* and *WR* type of data, relative to the diary measurements, seemed to increase with consumption, the deviations from linearity were non-systematic (figure 3.1). From figure 8.1 it was concluded that the relation between distributions of data obtained with different type of methods (treated as two independent sets) was a linear one, that is, after log-transformation of the raw data. The finding of a linear type of error (or more precisely, a linear relationship between distributions of data from different methods) means that conclusions about the changes in *distribution* over time (and across different populations) are not affected by type of method (at least not dramatically). Comparisons similar to those in figure 8.1, but of distributions of the untransformed consumption data (since log-transformation may obscure non-linearity), confirm the earlier finding of figure 3.1, as can be seen in figure 9.1 (for reasons of efficiency, only the male comparisons are given). Furthermore, in chapter 4 it has been found that the correlations between different types of methods are quite high, especially in view of the large differences in the mean.

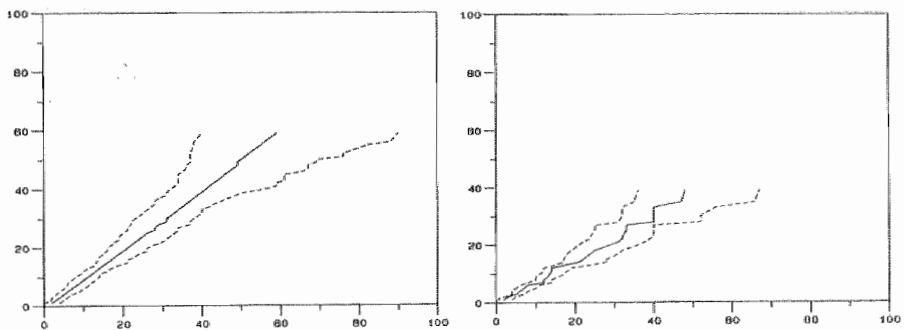


Figure 9.1. Empirical probability plots and the .95 confidence band of the male 1985 sample distributions of consumption data obtained with weekly recall (y-axis, left) and *QF* (right) and diary data (x-axis); raw untransformed data (abstainers and upper 2.5% excluded).

Altogether then, it seems justified to conclude that the alleged underreporting of "*recent occasion*" and *QF* type of method does not seriously affect results based on this type of data. For instance, correlations with and between data of this kind will, in the case of a linear type of error, not be seriously affected. It should be kept in mind that this conclusion is drawn relative to the allegedly superior diary data, which still show a 30% undercoverage of sales.

9.2.2 *Controlling the effects of recall failure.*

Estimates of alcohol consumption based on data obtained with a diary method are much higher than with the other, more traditional type of measurement methods, such as the weekly recall and quantity-frequency types. In chapters 3 and 4, diary data from two different general population surveys in the Netherlands, conducted 2 years apart, and each diary of a different format, have been compared to weekly recall data. In both cases the diary yielded a higher mean consumption. The main difference between the weekly recall and the diary type of method seems the smaller time interval in the diary between drinking and recording of details about the drinking situation. This suggests that recall failure is the main cause of the observed difference between these two methods.

One could argue, however, that the presented comparison of *diary* and *WR* method is merely circumstantial evidence for the above conclusion. Apart from the critic that the researcher has little control over the conditions under which the diary is completed (- do the subjects actually record their consumption each day, and is the maximum recall period really one day? -), one may add that a direct examination or modelling of memory decay (or, recall failure) would yield stronger evidence for a memory effect. In the present study, however, a modelling of decay cannot be further specified because the data needed for an operationalization of decay is lacking. Unfortunately, the exact date of the interview (which is the day the recall took place) has not been recorded in the 1985 survey.

In modelling a memory process, there may be mechanisms other than memory decay (a gradual loss of information as a function of time, resulting in omissions) that distort the self-report. Forward and backward telescoping, a tendency to date events as having occurred within or outside the reference period, are among these, resulting not only in underreporting but also in overreporting (see chapter 3). The large differences in memory capacity, telescoping, individual response tendencies (e.g. experienced threat) and large differences in parameters of the individual drinking behavior (e.g. frequency, quantity, regularity over time, location) would result in a very complex model of memory effects, which, besides its complexity, is probably not generally applicable across behaviors, populations or over time (see Sikkel, 1985).

As already referred to in chapter 3, Philipsen *et al.* (1983) have found a rapid decline in average number of glasses as the reference period increases. The simplest solution to the problem of recall failure would thus be a limitation of the reference period, as in the chosen diary design. However, other solutions are possible aimed at maximizing the contribution of retrospective data and based on assumptions about the nature of the processes governing the target behavior as well as its recall. For instance, Sikkel (1985), modelling behavior and forgetting, concluded that "*it does make sense to ask retrospective questions about events that are contaminated with memory effect, given a good model for the process that generated these events and the process of forgetting. With a plausible model events that lie relatively far in the past, or that are followed by many other (reported) events contribute to the efficiency of the relevant estimators*". As noted above, the results of his analyses do not appear to be generally applicable, however.

The starting point for a procedure as suggested above, is a model for the behavior that is being surveyed. An attempt to model drinking behavior has been reported by Alanko (1981). He distinguished separate statistical models for describing frequency and quantity aspects of drinking. Frequency of drinking could be modelled rather well with a Poisson process (see also the unique drinking rhythm study conducted by Ekholm, 1968) and a gamma model was chosen as underlying the quantity component. In a later unpublished manuscript (1982), Alanko suggested an inverse Gaussian distribution model for the quantity component. In both cases, however, he did not simultaneously take memory or other response effects into account and the empirical data came from only one population (Finnish). In his own words, the generalizability of his findings is probably limited.

At another instance, using this above model of drinking behavior, Alanko & Duffy (1989) performed computer simulations of the measurement process with three methods, the *QF*, *WR* and a so-called *survey period* method, in essence a recent occasions method with variable reference period. Simulation of *QF* data is, however, problematic since it was and remains rather unclear what subjects actually recall in a *QF*-type of method: is it their mean, modal or median frequency; do they actually summarize their drinking in the entire reference period, or do they emphasize recent drinking behavior? It is therefore not entirely clear how subjects assign themselves to the "usual frequency and quantity" categories, a problem that cannot be solved by simulation, of course. Some arbitrary solution has been adopted by the authors, and one of the results of the analysis was that "usual frequency" scale in the *QF* measure in simulated British data was the main source of a downward bias. This downward bias was due to the upper limit of "every day" in the highest optional frequency category, which is probably not high enough for the British population (there are many who drink more often than once per day). The authors concluded that a higher upper limit (more than once per day) would give less distortion (another possibility would be to

adjust the format of the "*usual quantity*" question). It is interesting to note that in the Dutch 1985 survey, discussed in chapter 4, the highest male *UF* category (every day of the week) showed the highest number of subjects (24% of all drinkers; see observed frequencies in figure 4.5). This finding could point toward a similar "*reservoir*" or "*ceiling*" effect of the highest category as in the simulated British data. Given a population as the British, it would be advisable to enlarge the *UF* question with an extra, heavy option of, for instance, "*twice daily*".

For the simulated Finnish data, Alanko and Duffy found that especially light drinkers would be underreported in a survey with a *QF* type of method. This result seems a corroboration of the finding reported in chapter 4 (figure 4.5 and 4.7), that expected weekly frequencies of drinking of subjects with a low score on *UF* ("*usual frequency*") show a larger deviation from "*observed*" diary frequencies than those with a high *UF* score. Thus the conclusion from both studies is that the *QF* method offers a reasonable classification of heavy drinkers. The above results from the simulation study, concerning the British as well as the Finnish data, seem inconsistent as to the main reason of the downward bias. It should be kept in mind that both populations differ considerably with regard to drinking (e.g. total volume) and that, contrary to the Dutch data in chapter 4, no response effects (memory, telescoping, concealing, etc.) have been taken into account. Further study is needed for more definite answers about the combined effect of the possible sources of variation. Another result of the simulation experiments worth mentioning, was that in the case of a large intra-individual variation (e.g. over time) the *WR* type of method (actual consumption) performed less well. This is similar to the conclusion regarding regularity of drinking and underreporting drawn in chapter 4 (fig. 4.2). It is also in line with the conclusion of a higher susceptibility of the *WR* method to seasonal fluctuations, done in chapter 5 (fig. 5.3). So it may be added to that conclusion that even without response effects, regularity of drinking enhances accuracy of consumption data. A simulation of drinking behavior (=input) and measurement (=output), such as the Alanko and Duffy trials, could very well incorporate models for memory effects, a suggestion also made by the authors themselves. Perhaps it would be a more comprehensive and parsimonious way to reach conclusions about optimum length of reference periods, questionnaire length (for instance, in the case of aided recall procedures) and, generally, choice of particular method than the elaborate design of the 1985 survey.

That aided recall procedures add to more valid measurements has been reported by Sudman and Bradburn (1973). In a meta-analysis of characteristics of some 500 surveys, they have demonstrated that surveys that use aided recall procedures show a higher response than surveys that use simple recall questions. They have also found that the use of available records about behavior decreases response. Their results suggest that aided recall procedures decrease omissions (forgetting), while the use of records

decreases error due to telescoping (overreporting). It is difficult to imagine (at least for the Dutch situation) which records could be used for a corroboration of self-reports of alcohol consumption. In the 1985 Dutch survey the *WR* method has been of an aided recall type, requiring specification of type of beverage and of location on each day of the week. This may well explain the fact that the coverage of sales in the 1985 survey (58%) is higher than in surveys held in 1958 (49%) and 1981 (44%), in which only a 5 day recall of drinks taken in a certain situation (1981) or type of drink (1958) has been specified.

9.2.3 *Time-variation of drinking behavior.*

"Drinking is a highly variable behavior, both within subjects over time as well as between subjects". The consequences of this general statement for the measurement of consumption have been worked out in chapters 3 and 4. In chapter 3, it has been found that variation over weeks indeed varies quite considerably (an indicator of stability is the correlation between diary scores in two consecutive weeks of .84). The distribution of the difference scores was, nevertheless, symmetrical around 0. This suggests that the error due to time-variation is random, at least within a time-interval as short as 2 weeks. This last comment refers to chapter 5 where it became clear that time of the year in which the survey is conducted influences the outcomes of measurement of drinking variables, and that it is advisable for surveys of which the goal is estimates of mean annual consumption to spread the interview period over the entire year.

The results reported in the methodological studies suggest that aspects of the drinking pattern, at the individual as well as at the aggregate level, interact with the error of measurement. The word *"suggest"* is used here, since all empirical data are with (self-reporting) error, and quantification and empirical verification of a possible interaction is difficult. The example presented in chapter 4, that a larger regularity of drinking (given a certain frequency) would reduce recall failure, implied that a difference in overall regularity of drinking between populations would affect the accuracy (validity) of self-reports on drinking accordingly. From Alanko & Duffy's simulation experiments, it can be deduced that even without a postulated memory loss, the accuracy and precision of estimates (reliability) would also be affected by differences in regularity of drinking.

Highly variable behavior is difficult to capture in a few survey questions and little is still known of the temporal dynamics of drinking behavior. The only empirical study which has gathered data on actual drinking occasions over a long period of time dates back to 1968 (Ekholm). More than anything else, the absence of empirical data on this subject signifies that the problems connected to such a design are manifold, not only with regard to the high costs and the strain put on the cooperativeness of subjects, but also in controlling the many possible attenuations, such as the effects of

self-monitoring on consumption. Moreover, populations are heterogeneous not only with regard to drinking but also with regard to (self-)reporting behavior. Consequently, it will be difficult to get a sample that is large enough to investigate these factors, but which is not too large so as to become too expensive.

A retrospective alternative to cover a larger time period than, say, 1 week are *QF* types of method. However, little is known about the process by which respondents come to an answer on what their "usual" frequency and quantity is. What people usually drink, is of course what (most) alcohol researchers want to know. It is, however, wishful thinking on the part of the researchers that by simply asking about usual drinking behavior, people would give a correct answer (within a few seconds). The measurement problem is in fact transferred to the subjects, since they themselves have to decide what is meant by "*usual*", or "*unusual, atypical*", which is even less well circumscribed. One of the problems is that for some (non-) drinkers the question is indeed easily and accurately answered in a few seconds. For others (and most), it will be quite a difficult task, which cannot be completed successfully (= accurately) without additional definitions of the concepts.

According to some scientists, "*the expansion of theories to include a theory of the measuring instruments is commonplace in physics...*" and research that records verbal responses of any kind needs "*at least a rudimentary theory of how subjects produce such responses*" (Ericsson & Simon, 1980, p. 216). Though their approach is directed toward verbalizations in experimental settings for the study of human information processes, it may be worthwhile to take a glance at their methodology. As already referred to in chapter 4, Midanik (1989) has proposed to tackle the above rudimentary question by means of a verbalization technique, called protocol analysis, in which the subject is asked to think aloud. Introspective techniques have been discredited particularly by behavioristic psychology, and many artefacts are lurking (see Linschoten, 1964, for some excellent examples). Still, it may be of help as a practical tool to the alcohol survey researcher who wants to optimize instructions or "*routes*" by which respondents construct their answers to evaluative "*usual ..*" questions. Midanik presents a few questions that may be subject of study: what is the reference period respondents are using in their evaluation, how do respondents interpret "*drinking occasion*", what are optimum response categories, how much time do they need for an answer, do respondents differ in their own assessment of "*accuracy*"? One could add several others, such as, in which direction do people search (forward or backward), what is the optimum reference period (how far do people go back in time), does specification help and which aspects of drinking behavior are meaningful to ask (e.g. types of drink, location, company you are with, type of activity, time of day, mood, and in what order), what cues or probes can be used to elicit memory of drinking events (e.g. anniversaries, parties, feasts), how do respondents value the interview

situation (do they feel embarrassed), what do they think about the interviewer during the interview, is it possible to detect response styles, or response "*personalities*", etc. The procedure could be extended to acquire information about positive and negative experiences, attitudes, and opinions about alcohol use. In contrast to the approach in psycho-physics, the survey researcher is less interested in ongoing mental processes but more in something that has already happened, namely past drinking. The information is, however, not readily available to the respondents, and the process to arrive at an answer (or, answers) can be seen as a problem-solving task. Perhaps a larger uniformity in response function (the relation between the target information and the individual response), a larger homogeneity in the way the task set by the interviewer is solved, and, hence, an increased comparability of scores could be established by an increase in points of reference, standards by which people judge their behavior (e.g. Saris, *et al.* 1987).

Whether alcohol surveys can actually benefit from such an approach and arrive at "a theory of how subjects come to their answers" is not yet clear. It is, however, evident that simple questions elicit simple answers and that there is a price to be paid for accuracy and precision of data. To quote one of the contributors to Moss & Goldstein (1978) "*A poorly discriminating measure tends to have a higher reliability than a more discriminating one*" (p. 47). In other words, the more detail you want, the more errors you elicit in responses. It is be advisable, then, to tune methodology to the specific research questions. Applying this advice to studies of the shape of the distribution of alcohol, one may reach a general conclusion that consumption measures based on information on actual, "*recent occasions*" are a better choice than the more easily contaminated "*summary*" measures. For more general purposes (e.g. screening or classification of heavy drinkers) the latter seem satisfactory.

9.2.4 *Selective response and denial.*

In the chapters 2 and 6, no evidence has been found that justifies the fear often expressed that especially heavy drinkers are more likely to refuse or more difficult to contact for an interview, or that they are inclined to deny a larger proportion of their consumption than subjects at more moderate levels. On the contrary, there are indications that (female) abstainers are over-represented (+14%) in the non-response category. Equally, the estimated mean consumption of non-abstaining females in the non-response follow-up category seems also lower than in the response sample. Similarly, the frequency of irregular heavy use is also lower among initial non-respondents (table 2.5). At first sight, this result suggests that cooperativeness (since refusal is the main reason for non-response) is not related (for men), or even positively related (in the case of women) to consumption. As already noted in the discussion of chapter 2, the follow-up

sample may, however, be a selection from the complete non-response category, consisting of only those persons whose reason for non-response is not related to their drinking habits (note, for instance, that older people are more likely to refuse). To shed some light in this darkness, it is advisable for future research to investigate the relation between target variable and the number of call-backs needed to establish contact (Hansen & Hurwitz, 1946). This could be combined with a further specification of the reason for non-response or other characteristics of non-respondents (e.g. attitudes), which could then be related to consumption in a follow-up (extended call-back) sample. Panel studies, in which many data of initial respondents are available, offer opportunities for further study on this topic (see also section 9.5).

Though total non-response is large (30%) and the difference between predicted response fractions is as large as 35% (which is the difference between persons "*married 16-25 year olds, living in the Eastern provinces*" and "*unmarried 46-55 year olds, in the Western provinces*"; see figure 2.1) no correction of consumption estimates results from this (table 4.2). It should be remembered that in calculating the effect of non-response, using the variables in the model of response probabilities, an equal relationship between target variable (consumption) and the variables in the model is presumed. A violation of this presumption results in different correction factors. Future research might explore the margins within which the effects of non-response are still tolerable, or how the structure of non-response might influence estimates of consumption. For instance, it can be inferred that the larger the skewness of the consumption distribution, the larger the effect of non-response of a single heavy consumer will be. Again with the relatively cheap simulation experiments one could model, not only the structure of non-response but also the relation between non-response and consumption, and estimate the effect on accuracy and precision of consumption estimates.

The effect of the other threat to validity of self-reports, concealment or denial of consumption and related issues, seems also limited. No evidence for a systematic underreporting of consumption of self-reports relative to collateral reports has been found. There are indications that there is a substantial error in both sources, but that the error does not lead to a systematic bias in self-report data. From the results in chapter 6, researchers could be confident about their respondents as the prime source for information on drinking and drinking related issues. No large differences in results and conclusions with regard to the distribution of consumption are to be expected when collateral instead of self-report measures are used.

Again, type of method is a factor that causes differences in aggregate measures of consumption. Those based on memory of actual drinking occasions seem to be better recalled by the drinkers themselves than by their partners. This also holds true for reported symptoms of abuse. Partners

report slightly higher consumption on the summary QF measure. Though the results of the model of correspondence in chapter 6 are disappointing, the attempt has made clear that without a model controlling for possible rival explanations of the difference in reports of subject and partner, the results of a comparison are only limited in value. This controlling can take either the form of a model such as the one presented in chapter 6 (quasi-experimental; e.g. survey), or the conditions under which the behavior (drinking) and reporting take place could be carefully controlled (experimental, laboratory).

Choosing the former, one should be critical over the choice of variables as indicators of the main concepts. There are certainly better indicators of, for instance, ignorance than the ones used in the model (figure 6.3). An interesting finding of a time-budget study on agreement between respondent and spouse over each other's presence in a particular situation has been reported by Juster (1985 - a time budget study, similar to the Dutch 1985 diary format, data were gathered for 15 minute periods). Both sources corresponded in about 85% of all 15-minute periods. The variation during one day ranged between 74% and 87% (lowest scores in the morning -less accurate memory?), which was regarded a good validity score by the authors. In the weekends correspondence over each other's presence was lower than during the week. During the weekends, in about 50% of all cases one of both sources did report the partner present, whereas during the week this was only 32%. This result suggests that in a design using collateral sources as the criterion, one should take these large margins into account. Couples differ quite much as regards the time they spend together, which on the average seems quite low, and it is not surprising then that partners' recordings of actual drinking are lower than self-reports. The same author also reports about a validation study (criterion: company records) on retrospective self-reports of unemployment (30 months) in which the error due to memory failure was very small compared to indicators of task difficulty and saliency (Duncan & Mathiowetz, 1985). A higher consumption level may possibly imply a more complicated drinking pattern and thus the task to reconstruct such a pattern may be more difficult, and more errors in reports are to be expected, and, hence, a lower agreement. Other indicators, such as those for more traditional spousal relations or values, marital cohesion, etc., can also be improved, as well as indicators of threat in the interview situation (e.g. Bradburn & Sudman, 1979).

9.2.5 *Ledermann revisited.*

The results of the methodological studies do not indicate that consumption data are biased, in the sense that heavy drinkers are underrepresented in the sample. On the contrary, respondents low on summary usual frequency tend to underreport, relative to the frequency reported during their diary week. On the other hand, the relationship

between *WR* and *diary* data, though expectedly linear, shows that at the higher levels, underreporting of consumption is also higher (fig. 3.1), and there is even some overreporting at the lower levels. It is difficult to decide whether diary data are more trustworthy than the other data. Nevertheless, when diary data are chosen for estimation purposes, it is likely that, compared to data that would have been obtained with other methods, (1) higher levels are not underrepresented, and (2) underreporting is not larger at higher observed levels. It seems justified then to state that for estimation purposes the use of diary is the best possible, "*conservative*" choice.

In table 9.1, comparisons of estimates of prevalence of consumers over a certain threshold are presented with empirical diary data and with data derived from the Ledermann formula. As has been pointed out in chapter 1, the only parameter in the formula is "*mean annual consumption*", which equals the mean of the diary scores in the respective samples. The mean in both columns is the same, only the shape parameter(s) will cause the variation in estimated proportion of drinkers. It can be concluded that at the lower levels the Ledermann formula clearly underestimates the proportion of drinkers, while it overestimates prevalence of high consumers. The "*turning-point*", where underestimation changes into overestimation, is different for men and women, 49 glasses per week and 28 glasses per week. The empirical distribution is clearly less skew than the Ledermann log-normal one.

Table 9.1. *Estimates of the percentage of drinkers with a consumption over successive levels of consumption, based on survey diary data and on the Ledermann formula. Equal mean consumption, abstainers excluded. Male and female samples considered separately (respective means: 20 and 9 glasses per week).*

consumption glasses/week	male		female	
	diary	formula	diary	formula
21 or more	38.7	26.5	11.1	9.3
28 or more	25.8	19.1	6.5	6.6
35 or more	18.6	14.4	1.5	4.9
42 or more	12.2	11.1	0.3	3.9
49 or more	8.8	8.8	-	3.1
56 or more	6.3	7.2	-	2.5
63 or more	4.1	5.9	-	2.1
70 or more	3.2	4.9	-	1.8

Table 9.2. *Distribution of total consumption over deciles of drinkers (abstainers excluded), according to diary estimates (1985 survey) and Ledermann-formula, for male and females.*

deciles	male		female	
	diary	formula	diary	formula
1-5th	11%	11%	12%	5%
6-7th	20%	13%	23%	8%
8-9th	35%	28%	33%	23%
10th	35%	48%	32%	64%

In table 9.2, the fraction of the total consumption for all fractiles are given. The table shows that the Ledermann formula apparently exaggerates the share of total consumption of the highest decile. For the empirical data about one-third of all consumption is consumed by the upper 10% of the drinkers, whereas the estimate for the highest decile with the formula ranges between 48% (men) and 64% (women). This confirms the above results. A word of caution thus to all those who use figures based on the Ledermann model.

9.3 Region estimates of the α -and β -parameter of the linear shift function

"Is the β -parameter of regression of percentiles on mean consumption positive and decreasing for all percentiles?" (see section 1.6). Tan *et al.* (1990) have shown that the relation between the slope of regression of successive percentiles on mean consumption depends upon the values of parameters in the shift function. In chapter 7 it has been noted that the point-estimates of the relevant parameters of the shift function, was not easily obtained. Tan *et al.* (1990) have tackled the problem by simultaneous estimation which produces so-called region estimates. In their article, they have shown analytically, that given the regularity in distribution as assessed in chapters 7 and 8 (linear shift of log-data; scale and location) there must be a turning point over which the alcohol consumption decreases with an increase in mean consumption. In other words, they have shown that, under certain circumstances (depending on the estimated values of location and scale), a percentile point exists over which the relation between the change in mean

and percentile changes sign (called point of attraction). This result means that the supposition of a regular, stable distribution as, for instance, in the Ledermann model is not totally compatible with the collectivity hypothesis in Skog's theory. This incompatibility means that either linearity cannot be sustained over the entire distribution (remember that in chapter 7 and 8, the upper 2.5% was not included in the comparison) or one has to doubt the collectivity of consumption change, which intuitively is difficult to grasp. By means of the above procedure (region estimates of location and scale parameter) Tan et al. have found a positive, decreasing relation (significant at a 95% level) up to the 90th percentile in the case of male Dutch data of 1970 and 1985 (see fig. 7.1). If the linear relation could be retained for the upper tail of the distribution as well, there must be a point of attraction which the authors estimated for these data at 157 glasses per week, which is by no means an unrealistically high level.

Skog's view on this matter is that such a point of attraction does not exist (personal communication; see also chapter 7, footnote 3). In the case it does not exist, and one might ignore it also for practical purposes, Tan *et al.* have shown that this is in contradiction with the regularity supposition, not only implicit in Ledermann's rigid mathematical model, but also in empirical analysis in chapters 7 and 8. If there would be strong evidence that even at very high levels of consumption, people adjust their consumption according to changes in the mean (the collectivity hypothesis), this would thus contradict the notion of a linear shift and it would make the regularity hypothesis unfit for extrapolation purposes in the tail of the distribution.

9.4 Impact and limitations of the notion of a "single"-distribution of alcohol consumption for a primary prevention model

The debate over the "control-of-supply" proposition that "*a major (perhaps the crucial) determinant of alcohol-related problems in a society is the overall amount of alcohol consumed*" (Peele, 1987, p. 66) is closely connected with the status of the single distribution theory of consumption. In this section the intention is not to summarize this debate nor to add entirely new arguments either in favor or against alcohol controls. It has not been the primary goal of the research project which has been reported here in this thesis, and there are many, sometimes excellent, reviews from which also this section has heavily borrowed (e.g. Peele, 1987; Moskowitz, 1989; Ravn, 1987; Rush *et al.*, 1986). The limited aim in this section is to place the findings of this study into a broader framework.

Point of departure for the “*Ledermann-project*” at the University of Limburg has been the evaluation of the usefulness of the Ledermann-model. The term “*usefulness*” refers to a practical goal, and, indeed, in the above mentioned debate, advocates of control measures have used the Ledermann-model in favor of their practical goals. These goals can be summarized as the prevention of (a rise in) alcohol related problems in society. In the first chapter the implications of a rise in consumption for the incidence and prevalence of “*heavy*” drinkers according to the hypothetical Ledermann model have already been discussed (e.g. figure 1.2). A rise in mean consumption would entail a growing number of people drinking at heavy consumption levels. In a sense, the distribution model can be regarded as an intervening variable between per capita consumption, on the one hand, and alcohol problems, on the other. However, there are many more factors than distributional regularity alone that have to be considered before per capita consumption can be equated to harm. In figure 9.2 several optional factors are depicted that might play a role in the primary prevention model, directed at a general decline of alcohol consumption.

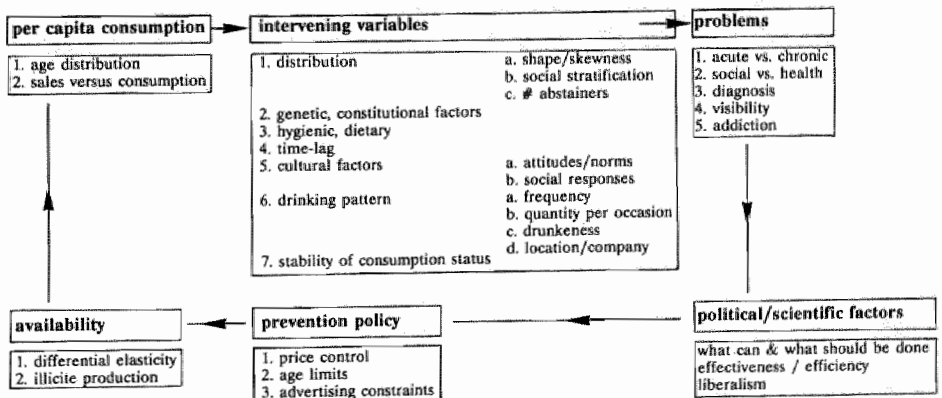


Figure 9.2. An extension of the single distribution model of prevention.

Probably the best documented health problem that has been found to be related to chronic heavy use is liver cirrhosis, at the individual level dependent not only upon heavy use per se, but also upon its duration (e.g. Leibach, 1974; Pequignot, *et al.* 1978; Skog, 1984). Though there is a clearly positive relation between per capita consumption and mortality of cirrhosis, large variations in mortality of liver cirrhosis can be detected.

In figure 9.3 the relation between cirrhosis in 1974 and average per capita consumption in 1972-74 is clearly positive but there are also large variations in mortality rates between countries with similar per capita consumption (e.g. *1,2,3; *4,5) or no variations in mortality between countries with divergent per capita consumption (*3,4). In the following paragraphs, liver cirrhosis mortality serves as an illustration of some of the mechanisms depicted in figure 9.2.

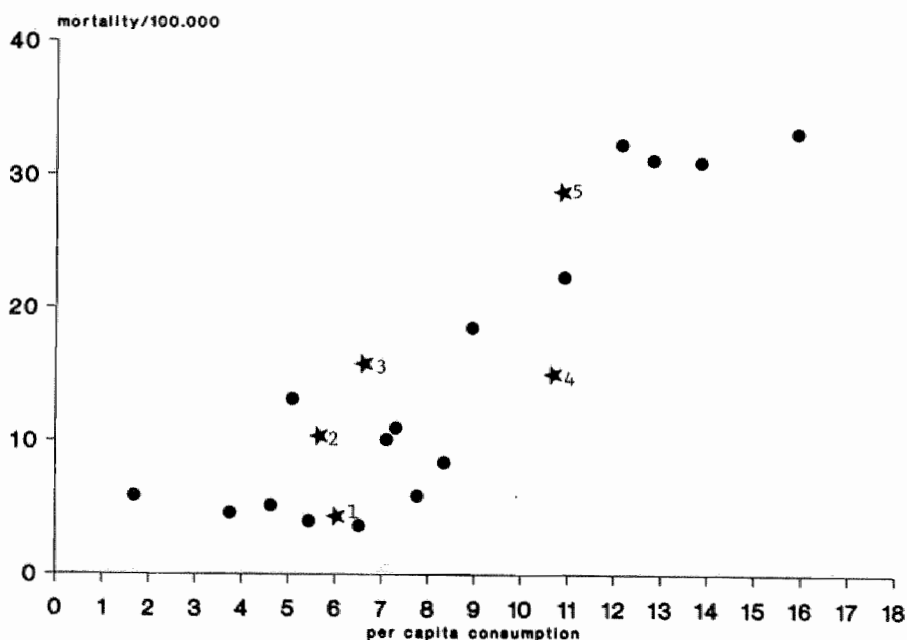


Figure 9.3. Liver cirrhosis mortality rates in 1974 (deaths/100.000) and per capita consumption (average 1970-72 in ltrs. 100% alcohol) in 21 Western countries (source: Hoogendoorn, 1978)

1 = Netherlands; 2 = Sweden; 3 = USA; 4 = Switzerland; 5 = West-Germany.

Before turning to the possible effects of "*per capita consumption*" it should be remembered (see chapter 1) that, mostly, sales are regarded equivalent to consumption. This assumption may be more or less true, which causes attenuation of the relation between "*true*" consumption and possible consequences. Similarly, "*per capita*" often means that the age distribution is disregarded and attenuation may occur because the total instead of the drinking age population is taken in the denominator.

Given a certain mean level, shape of the distribution between countries determines the prevalence of drinkers at the various drinking levels. When the risk function for liver cirrhosis is the same in all depicted countries, one can infer that differences in distributional form will cause variations in the number of people at risk of cirrhosis, and hence, will cause a variation in mortality. The above arguments concerning cross-sectional differences can also be made for longitudinal changes as in cases of "*redistribution*" of consumption. For instance, Norström (1987) has suggested that the exceptional rise in cirrhosis mortality in Sweden in the period 1950-80 could be attributable in part to a "*redistribution*" effect as a result of the abolition of the Bratt rationing system. This abolition, which was accompanied by a sharp increase in price of alcoholic beverage, has caused a larger increase at the initially higher consumption levels than at lower levels of consumption. Redistribution thus means that the risk of cirrhosis for heavy drinkers has increased. As an explanation of the redistribution, Norström suggests that the nominal price increase was not effective for those who, during the rationing, were used to purchase expensive drinks on the black market.

Another variable that may be responsible for variation is the social stratification of consumption. For instance, the age distribution of the drinking population may cause different mortality rates. Even if vulnerability for cirrhosis does not change with age (which may not be the case, see Skog, 1984), one might expect that average age of the population is positively related to cirrhosis mortality, since the probability for an older person to be exposed to high levels of consumption for a long period of time is higher than for a younger one.

Yet another distribution effect that can possibly affect mortality rate (given a certain per capita consumption) is the number of abstainers. Total sales might be consumed by 83% of all people older than 16 years of age (as in the Netherlands), or by 41% (as in Iowa, 1961- Mulford & Fitzgerald, 1988). Similarly, dramatic changes in number of consumers within a population (again, given a certain per capita consumption) might have large effects on the relation between per capita consumption and mortality rates (for example, the number of consumers in Iowa between 1961 and 1979 changed from 59% to 74% of those over 21 years of age).

Moreover, similarity between populations in the way total consumption is distributed over the drinkers does not necessarily imply that the population at risk of cirrhosis is the same. Overall stability of the position of drinkers on

the consumption scale (are the heavy drinkers in one year the same as the heavy drinkers the next year?) is a factor that may cause considerable variation in mortality, since not only volume but also duration of exposure to high alcohol consumption determines the eventual outcome. As an example, consider a country in which the consumption status of drinkers is highly stable (heavy drinkers remain heavy drinkers) for say, 10 years, and another country in which the highest category changes every year (for instance, certain age groups occupy the highest position). Depending in part upon the actual level and risk function, the former population will show a higher mortality rate than the latter. Longitudinal studies are needed to investigate the underlying dynamics of the resulting distribution. In the next section (9.5), one such study, conducted in the town of Zutphen in the Netherlands, will be discussed.

In considering the relation between mortality rates and consumption, one has to take a time-lag variable into account. This time-lag is connected to the above stability of high consumption habits, and refers to the fact that for a liver condition to become fatal, consumption has to be high for a relatively long period. The time-lag variable depends on the "*reservoir of already affected livers*". When in a low consumer society consumption is suddenly raised, the effect of this change in cirrhosis mortality will take several years to become noticeable. On the other hand, when in a high consumer society, consumption suddenly drops, this will have an instantaneous effect (since the cirrhotic process stops with abstention). An example of the latter is the immediate drop in cirrhosis mortality in Paris after the rationing of wine was introduced during World War II (see Ledermann, 1964, pp. 353). Attenuation of the relationship between per capita consumption and cirrhosis is likely to occur if the time lag is not included as a parameter, as in Skog's example of the negative correlation for England in the period 1931-58 (Skog, 1980). Other factors that remain undiscussed here are dietary (e.g. amount of fat), smoking, hygienic (non-alcohol hepatitis) and genetic (vulnerability) factors. Other authors have shown that medical and judicial practices (diagnostic issues) or developments in other concurrent causes of death (e.g. cancers, cardio-vascular diseases) may have a differential effect on morbidity and mortality rates (e.g. de Lint, 1981).

Aspects of the drinking pattern, such as frequency and quantity, are also factors that may cause variation in the relation between per capita consumption and negative consequences. For instance, in a large study in four Nordic countries, Hauge & Irgens-Jensen (1986) have reported that rate of intoxication and not per capita consumption could explain the differences in negative effects reported in surveys in the four countries. They note that, even though rate of intoxication was closely related with level of consumption and negative consequences in each country, rate of intoxication was lowest in Sweden and highest in Iceland (even when consumption was held constant), whereas per capita consumption in both countries was 7.1

and 4.5 (ltrs. 100%), respectively. The authors also report that social responses (arrests, disapproval, etc.) towards drinking differ (rate of intoxication held constant), being lowest in Sweden and highest in Finland. From this, the authors conclude that, given a certain per capita consumption, cultural factors determine not only whether drinking leads to intoxication (drinking pattern), but also whether this drinking will be regarded problematic (definition). This is closely related to another aspect of the drinking pattern, namely the extent to which drinking is regarded a social act, integrated in meaningful social settings (e.g. Douglas, 1987). For instance, many ethnic differences have been noted in the relation between consumption and rates of alcoholism (e.g. Vaillant, 1983).

Furthermore, it seems important to distinguish between patterns leading to chronic and those leading to acute effects of heavy consumption, the former more related to health consequences (cirrhosis, cancers), the latter to direct effects of intoxication (accidents, acts of violence).

Drinking pattern, cultural definitions of what is regarded problematic drinking, and social responses to drinking affect the extent to which a society "experiences" alcohol as a problem area. When alcohol is regarded a social problem of considerable proportion, it will attract attention of policy-makers and may become a political issue. Whether and how the perception of alcohol problems will result in primary prevention measures is in part dependent on ideology (e.g. liberalism, which rejects governmental regulations in many fields) and pragmatic relevance (e.g. Lemmens, 1989). In this process, results of scientific research are important arguments, and governmental funds are often spend on a scientific justification of political decisions (what should be done, tested against what can be done, what is effective).

The traditional approach to prevention is a restriction of supply, or a reduction of availability of alcoholic beverages. In its most extreme form restriction of availability has been installed by a total alcohol prohibition in several countries before World War II. Contemporary alcohol policies concentrate on price controls, governmental monopoly, licensing, opening hours, minimum age limits and constraints on alcohol advertisements. Of these, price controls draw most attention. At this point in the cycle, economic actors and factors enter the scene. To name but a few:

- not every country, social stratum, age groups, etc. will react in the same way to price increase;
 - there is a clear difference in the price and income elasticity of types of drinks. In most studies beer is found to be most inelastic;
 - other goals than a reduction in alcohol related problems may be served.
- For instance, Ornstein and Hanssens (1985) have noted that after prohibition was repealed both temperance groups and retail sellers had similar interests in a restrictive policy, the latter group for reasons of reducing competition.

- effects of a control policy might wear off. For example, Ornstein and Hanssens (1985) remarked that in monopoly states, a restrictive measure intended to keep prices high, nominal retail prices are lower than in neighbouring states with a more liberal policy.

- an increase in price may revitalize illicit markets, opening the alcohol arena for criminal activities.

Ornstein and Hanssens (1985), analysing demand for alcoholic beverages in an extensive model in which several economic, sociodemographic factors and alcohol control measures were included, found that factors which appeared to be significantly related to consumption were hard to influence by means of regulations (tourism, age-distribution and urbanisation). Of the control measures price (effect on distilled drinks) and minimum legal age (effect on beer) showed the strongest relationship with consumption.

The authors have criticized their own demand model as being inferior to a longitudinal one that uses time series. In the conventional correlational analyses one runs the risk of a potential endogeneity of the regulations, which means that regulations themselves are results of (changing) attitudes in the particular population, and that they do not have an effect on consumption. This potential artefact has also been noted by other authors, mentioned in Peele (1987, p.68, 69), who point toward the paradox present in the evaluation of a primary prevention model, namely that (per capita) *"consumption can be both an independent (or predictor) and a dependent (or criterion) variable"* (p.66). In the same journal issue, Room argues that *"the growth of addiction or loss-of-control is precisely one of the major methods by which American society has responded when levels of alcohol consumption come to seem intolerable."* (1987, p. 81). Awareness of the detrimental consequences of long-term use by a vast majority of the French has also been put forward as a reason for the decline in overall consumption in France in the past 3 decades, rather than legal controls (A. Tuyns, personal communication). This cyclical pattern between a rise consumption, growing awareness and concern over its consequences, described in this section, may be responsible for the large changes in consumption that have occurred in the past centuries (Skog, 1986). By just looking at one aspect of the process (e.g. legal alcohol controls and per capita consumption) one might miss the contingencies between availability, consumption, detrimental consequences and temperance, and interpret the cyclical pattern in consumption as a random process.

9.5 Longitudinal stability of drinking status

One of the intervening variables suggested in the previous section is the relative stability of individual drinking behavior. As has been pointed out, regularity in distribution may occur independent of the particular individual changes. What is a reality at the aggregate level (e.g. collective change of consumption), may not be observed (may not be observable?) by individual members of society. In the same vain, populations that show similar consumption changes at the aggregate level may differ in the way this change has been brought about by individual drinkers. Stability of consumption status (not to say, drinking pattern) can influence the transition of per capita consumption to negative consequences (to name but the most obvious, alcohol addiction), especially if the latter are dependent upon long-term heavy use. The "turnover rate" of drinkers has large consequences for the interpretation of the results from population surveys. For clinical populations assessment of drinking careers is important in the discussion on controlled drinking and chronicity of alcohol abuse. Methodological problems often make results of longitudinal studies incomparable or uninterpretable (Skog & Duckert, in press)

This section discusses the preliminary results from a study conducted at the University of Limburg in which consumption change in a cohort of men in a small Dutch town, Zutphen, was investigated (Knibbe, *et al.*, in preparation). The aim was twofold: firstly, whether change in consumption at the aggregate level was mimicked by change at the individual level; secondly, a solution to methodological problems was sought (Tan, 1989). The cohort consisted initially of 871 men aged 40-60 in 1960. Subsequent follow-ups were conducted in 1965, 1970 and 1985. Attrition in these years was not selective with regard to alcohol consumption. Average consumption rose sharply in the survey period, from 3.6 in 1960 to 13.0 glasses per week in 1985. It appeared that consumption in the panel study in Zutphen followed the national trend in these years. About half of the drinkers showed an increase in consumption between successive measurements. With the use of the Lisrel technique (error of measurement taken into account), a model of change (1960-65-70) has been tested in which the main assumption was a first-order Markov assumption (*viz.*, consumption in 1960 and in 1970 uncorrelated). Correlations between consumption level in successive years appeared to be as high as .80 (controlled for effects of marital status, age, employment status, and retirement). This means that variation in consumption between successive years can be attributed for a large part to initial consumption status. No test of regularity in distribution, as defined in chapters 1 and 7, is available yet, so no definite statements about the relation between change at the aggregate and individual level can be made. If, however, the results for middle-aged men in Zutphen can be generalized to

the total Dutch population, change in consumption is not as unpredictable as has often been assumed.

9.6 Final comment

The studies that have been collected in this thesis have in common that they all deal with complex behavior. The topics were rather technical in nature, and dealt with methodological problems of measurement of consumption and with the assessment of the frequency distribution of consumption. The connection between both is obvious: without an assessment of the validity of consumption data, no definite conclusion as regards the shape of the distribution of consumption can be made. Results of the methodological studies are not in support of the idea of a large bias due to selective non-response and conscious underreporting by the heavier drinkers. That diary data have been found superior with regard to coverage of sales, indicate that recollection of drinking is a difficult task for respondents to a survey. The effect of methodology on the empirical distribution of consumption is clearly shown, and recent occasion methods seem of better quality. However, with any method used, comparisons of empirical distributions reveal that the way in which consumption is distributed shows a large regularity over the years, as well as between different populations. They confirm one of Ledermann's basic ideas, namely a relative invariance of distribution.

References

- ALANKO, T. (1981) A statistical framework for describing the measurement of alcohol consumption from survey data. Paper presented at the 27th International Institute on the Prevention and Treatment of Alcoholism, Vienna, June, 1981.
- ALANKO, T. (1982) *A Statistical Model for Describing Drinking Behaviour and Distribution of Alcohol Consumption in Sample Surveys*. Unpublished manuscript (Helsinki, Finnish Foundation for Alcohol Studies)
- ALANKO, T. & DUFFY, J.C. (1989) Comparison of self-report alcohol consumption measures: preliminary results from simulation studies. Paper presented at the 15th Alcohol Epidemiology Symposium, Maastricht, 11-16 June, 1989.
- BRADBURN, N.M. & SUDMAN, S. (1979) *Improving interview method and questionnaire design: Response effects to threatening questions in survey research* (San Francisco, Jossey-Bass).
- DE LINT, J. (1981) Alcohol consumption and liver cirrhosis mortality, the Netherlands 1950-78, *Journal of Studies on Alcohol*, 42, pp. 48-56.
- DOUGLAS, M., (ed) (1987) *Constructive Drinking* (Cambridge, Cambridge University Press).
- DUNCAN, G.J. & MATHIOWETZ, N.A. (1985) *A Validation Study of Economic Survey Data* (University of Michigan, Ann Arbor).
- ERICSSON, K.A. & SIMON, H.A. (1980) Verbal reports as data, *Psychological Review*, 87, pp. 215-251.
- EKHOLM, A. (1968) A study of the drinking rhythm of Finnish males. Paper presented at the 28th International Congress on Alcohol and Alcoholism, Washington, September 1968.
- HANSEN, M. & HURWITZ, W. (1946) The problem of nonresponse in sample surveys, *Journal of the American Statistical Association*, 41, pp. 517-529.
- HAUGE, R. & IRGENS-JENSEN, O. (1986) The relationship between alcohol consumption, alcohol intoxication and negative consequences of drinking in four Scandinavian countries, *British Journal of Addiction*, 81, pp. 513-524.
- HOOGENDOORN, D. (1978) Het toenemende gebruik van alcohol en de stijgende frequentie van enkele (mede) door alcohol veroorzaakte ziekten, *Nederlands Tijdschrift voor Geneeskunde*, 122, pp. 1275-1280.
- JUSTER, F. TH. (1985) The validity and quality of time use estimates obtained from recall diaries, in: F.Th. Juster & F.P. Stafford (eds) *Time, Goods, and Well-Being* (University of Michigan, Ann Arbor).
- KNIBBE, R.A., TAN, E.S., LEMMENS, P., KROMHOUT, D. (in preparation) Changes in alcohol consumption in a cohort of men: 1960-1985 (working title) (University of Limburg).
- LEDERMANN, S. (1964) *Alcool, Alcoolisme, Alcoolisation*, Vol II (Paris, Presses Universitaires de France)
- LELBACH, W.K. (1974) Organic pathology related to volume and pattern of alcohol use, in: R. Gibbens, Y. Israel, H. Kalnt, R.E Popham, W. Schmidt, R. Smart (eds) *Research Advances in Alcohol and Drug Problems*, Vol 1 (Wiley, Toronto).

- LEMMENS, P. (1989) Communisme worstelt met alcoholprobleem, *NRC-Handelsblad*, 5 augustus.
- LINSCHOTEN, J. (1964) *Idolen van de Psycholoog* (Bijleveld, Utrecht).
- MIDANIK, L. (1989) The use of protocol analysis in alcohol research, Paper presented at the 15th Alcohol Epidemiology Symposium, Maastricht, 11-16 June, 1989.
- MOSKOWITZ, J.M. (1989) The primary prevention of alcohol problems; a critical review of the research literature, *Journal of Studies on Alcohol*, 50, pp. 54-88.
- MOSS, L. & GOLDSTEIN, H. (eds), (1978) *The Recall Method in Social Surveys* (London, Institute of Education).
- MULFORD, H.A. & FITZGERALD, J.L. (1981) "Words and deeds": Responses to Popham and Schmidt. *Journal of Studies on Alcohol*, 42, pp. 362-367.
- MULFORD, H.A. & FITZGERALD, J.L. (1988) Per capita consumption, heavy drinker prevalence and alcohol problems in Iowa for 1958-1985, *British Journal of Addiction*, 83, pp. 265-268.
- NORSTRÖM, T. (1987) The abolition of the Swedish rationing system: effects on consumption distribution and cirrhosis mortality, *British Journal of Addiction*, 82, pp. 633-641.
- ORNSTEIN S.I. & HANSSENS, D.M. (1985) Alcohol control laws and the consumption of distilled spirits and beer, *Journal of Consumer Research*, 12, pp. 200-213.
- PEELE, S. (1987) The limitations of control-of-supply models for explaining and preventing alcoholism and drug addiction, *Journal of Studies on Alcohol*, 48, pp. 61-77.
- PEQUIGNOT, G., TUYNS, A., BERTA, J.L. (1978) Ascitic cirrhosis in relation to alcohol consumption, *International Journal of Epidemiology*, 7, pp. 113-120.
- PHILIPSEN H. KNIBBE, R.A., VAN REEK, J. (1983) Alcohol consumption in the Netherlands as a social phenomenon, in: Hermus R.J. (Ed.) *Alcohol, Health and Society* (CIVO/TNO, Zeist).
- RAVN, I. (1987) The control-of-consumption approach to alcohol abuse prevention. II. A review of empirical studies, *International Journal of the Addictions*, 22, pp. 957-979.
- ROOM, R. (1987) Alcohol control, addiction and processes of change: comment on "The limitations of control-of-supply models for explaining and preventing alcoholism and drug addiction, *Journal of Studies on Alcohol*, 48, pp. 78-83.
- RUSH, B.R., GLIKSMAN, L. & BROOK, R. (1986) Alcohol availability, alcohol consumption and alcohol related damage.I. The distribution of consumption model, *Journal of Studies on Alcohol*, 47, pp. 1-10.
- SARIS, W.E., VAN DEN PUTTE, B., MAAS, K., SEIP, H. (1987) Vergelijkbaarheid van antwoorden: de noodzaak van meerdere referentiepunten, in: *De Vragenlijst in het Sociaal Onderzoek*, J. de Jong-Gierveld & J. van der Zouwen (eds) (Deventer, Van Loghum Slaterus).
- SIKKEL, D. (1985) Models for memory effects, *Journal of the American Statistical Association*, 80, pp. 835-841.
- SKOG, O.-J. (1980) Liver cirrhosis epidemiology: some methodological problems, *British Journal of Addiction*, 75, pp. 227-243.

- SKOG, O.-J. (1984) The risk function for liver cirrhosis from lifetime alcohol consumption, *Journal of Studies on Alcohol*, 45, pp. 199-208.
- SKOG, O.-J. (1986) The long waves of alcohol consumption: a social network perspective on cultural change, *Social Networks*, 8, pp. 1-32.
- SKOG, O.-J. & DUCKERT, F. (in press) The development of alcoholics' and heavy drinkers' consumption: a longitudinal study.
- SUDMAN, S. & BRADBURN, N.M. (1973) Effects of time and memory factors on response in surveys, *Journal of the American Statistical Association*, 68, pp. 805-815.
- TAN, E.S. (1989) Lisrel, with special attention for a longitudinal study of alcohol consumption. Paper accompanying didactic seminar Alcohol Epidemiology Symposium, Maastricht, 11-16 June.
- TAN, E.S., LEMMENS, P.H.H.M. & KONING A.J. (1990) Regularity in alcohol distributions: implications for the collective nature of drinking behaviour, *British Journal of Addiction*, 85, pp. 745-750.
- VAILLANT, G.E. (1983) *The Natural History of Alcoholism* (Cambridge, Harvard University Press).

Samenvatting

Volgens de Franse epidemioloog S. Ledermann kan de empirische frequentie-verdeling van alcoholconsumptie in een homogene populatie het best worden beschreven met een één-parametrische log-normale verdeling. Een dergelijk model impliceert dat de in een bevolking beschikbare hoeveelheid alcoholische drank op een voorspelbare wijze over de drinkers is verdeeld. In het bijzonder suggereerde Ledermann dat er een "vaste" relatie is tussen de gemiddelde alcoholconsumptie in een populatie en de prevalentie van drinkers op de verschillende consumptieniveau's. De regelmaat in de verdeling van consumptie heeft implicaties voor een beleid van primaire preventie. Het houdt bijvoorbeeld in dat een bepaalde toename van het gemiddelde met een nog grotere toename van het aantal drinkers op risicovolle consumptieniveau's gepaard gaat.

In hoofdstuk 1 van dit proefschrift wordt het Ledermann model aan een kritische evaluatie onderworpen. De theoretische en empirische grondslagen van het model worden besproken, alsook de praktische toepassingen en implicaties voor een preventiebeleid. De problemen die worden gesignaleerd zijn in een viertal punten onder te brengen, namelijk, 1) onvoldoende statistische bevestiging van het model, 2) twijfelachtige kwaliteit van empirische gegevens over consumptie in de bevolking, hetgeen onder andere tot uitdrukking komt in een lage dekking van verkoopcijfers, 3) schaarste aan longitudinale studies, er worden vaak uitspraken over ontwikkelingen in de tijd gedaan op basis van cross-sectionele gegevens, 4) ontbreken van een theoretische basis voor de veronderstelde vorm van en regelmatigheden in verdeling.

De conclusie is dat de empirische ondersteuning van het model tamelijk zwak is en dat aan populatieschattingen van met name overmatig drankgebruik verkregen met de Ledermann formule een grote onzekerheidsmarge moet worden toegekend.

Als alternatief voor het strikte Ledermann-model heeft O.-J. Skog een alternatieve theorie gepresenteerd, waarmee de regelmatigheden in verdeling zouden kunnen worden verklaard. Twee hypothesen staan hierin centraal, die over wederzijdse beïnvloeding van drankgebruik (collectiviteit)

en over de wijze waarop veranderingen in consumptie zich voltrekken (proportionaliteit). Deze theorie is uitgangspunt geweest voor de non-parametrische benadering van het verdelingsvraagstuk in de hoofdstukken 7 en 8. In de tussenliggende hoofdstukken wordt nader ingegaan op de kwaliteit van consumptiegegevens verkregen met een survey onder de algemene bevolking.

Hoofdstuk 2 behandelt de mogelijke vertekening die optreedt als gevolg van selectieve non-response. Vaststelling van deze vertekening geschiedt met behulp van een model van responsekansen en met gegevens verkregen uit een telefonische follow-up onder non-respondenten. Het responsekans model geeft aan dat non-response samenhangt met leeftijd, regio en burgerlijke staat. Het resultaat geeft echter geen bevestiging van eerdere vermoedens van een duidelijke onderrepresentatie van zware drinkers in de steekproef. Integendeel, onder de vrouwelijke respondenten van de telefonische follow-up bevonden zich significant meer geheelonthouders. Bovendien was de gemiddelde consumptie onder drinkende vrouwen in de follow-up lager dan in de initiele response categorie.

In hoofdstuk 3 worden twee surveymethoden van zelf-rapportage van alcoholgebruik met elkaar vergeleken, een prospectief (2 weken) en een retrospectief dagboek (herinnering van gebruik in afgelopen 7 dagen), uit een Nederlandse, landelijke survey (Nederland OK, 1983). Het prospectieve dagboek levert een gemiddeld 22% hogere schatting van consumptie op, hetgeen wordt toegeschreven aan de kortere periode tussen de feitelijke consumptie en het moment van rapportage. Onderschatting van consumptie met de retrospective methode lijkt toe te nemen naarmate de consumptie hoger is, doch de relatie lijkt niet non-lineair van karakter. Ten opzichte van de grote intra-individuele variatie over weken, is de invloed van de gebruikte methode op de ingeschatte positie van drinkers relatief gering.

In hoofdstuk 4 wordt verder ingegaan op aspecten van validiteit en betrouwbaarheid van meet-methoden, gebruik makend van gegevens verkregen uit een landelijke, speciaal voor dit doel ontworpen survey, gehouden in 1985. In totaal 5 methoden worden vergeleken op verschillende aspecten van het drinkpatroon. Het prospectieve dagboek resulteert in de hoogste dekking van verkoopcijfers, de methode gebaseerd op "gewoonlijke" frequentie en quantiteit (QF) de laagste. Relatieve onderrapportage van methoden (ten opzichte van het prospectieve dagboek) lijkt meer te worden veroorzaakt door verschillen in het frequentie- dan in het quantiteitsdomein. Dit resultaat, gevoegd bij de uitkomst dat intra-individuele variatie voornamelijk wordt veroorzaakt door schommelingen in de drinkfrequentie, geeft aan dat geheugen een grote rol speelt in de gesignaleerde onderschatting in alcohol surveys. Vooral respondenten met een relatief laag gebruik blijken een te lage inschatting te geven van hun "gewoonlijke" frequentie. Verschillen in gangbare drinkpatronen tussen landen, met name

in drinkfrequentie en regelmaat, zouden een verklaring kunnen zijn voor de grote verschillen in dekkingspercentages van verkoopcijfers door surveys.

Hoofdstuk 5 toont aan dat, vanwege het feit dat surveys vaak binnen een kort tijdsbestek worden gepland, onderdekking van verkoopcijfers kan worden veroorzaakt door schommelingen in consumptie in de loop van een jaar. De consumptie is het hoogst in het voorjaar en het laagst in de herfst, met een duidelijke uitschieter tijdens de laatste twee weken van december. Behalve het totale gebruik, méér mensen drinken vaker tijdens deze feestdagen, lijkt in de Kerst-tijd ook drankvoorkeur te veranderen. Bovendien zijn schattingen van met name het percentage geheelonthouders naar verwachting hoger in surveys die worden uitgevoerd vlak vóór de jaarwisseling dan ind de periode erna. Verschillen in consumptie tussen de seizoenen worden voornamelijk bepaald door de drinkfrequentie en vooral met de retrospectieve dagboekmethode, waarbij de respondent zich de consumptie op de voorbije 7 dagen moet herinneren. Dit laatste geeft aan dat deze methode gevoeliger is voor tijdsfluctuaties.

Hoofdstuk 6 doet verslag van een onderzoek naar de overeenkomst in rapportages van respondent en zijn of haar partner over het drankgebruik van de respondent, symptomen van problematisch drankgebruik en afkeuring. Centrale gedachte bij een dergelijk design is dat de rapportages over eigen drinkgedrag onbetrouwbaarder zijn dan die over andermans gedrag. Het resultaat geeft echter aan dat zelf-rapportages niet zonder meer leiden tot een grotere onderschatting van consumptie dan partner-rapportages. Voor de methode gebaseerd op feitelijk gedrag zijn zelf-rapportages hoger, voor de methode gebaseerd op een inschatting van "gewoonlijke" consumptie, leveren partner-rapportages enigszins hogere opgaven. Voor symptomen van problematisch gebruik en afkeuring leveren zelf-rapportages gemiddeld hogere schattingen. Overeenstemming tussen respondent en partner nemen af naarmate het consumptieniveau van de respondent hoger is. Van invloed op de overeenstemming zijn ook andere factoren als de mate van dagelijkse routine en de kans dat het drinkgedrag door de partner wordt waargenomen. De conclusie is dat bewust verzwijgen van consumptie, ontkenning van problemen, e.d. geen al te grote bedreiging vormt voor zelf-rapportages in algemene bevolkingssurveys.

Na de 5 methodologische studies, wordt in hoofdstuk 7,8 en 9 teruggekeerd naar het distributievraagstuk. Omdat een parametrische benadering (bijv. lognormaliteit) weinig zin heeft is in hoofdstuk 7 getracht een verdelingsvrije methode te vinden om de dynamiek in alcoholconsumptie te beschrijven. Deze is gevonden in zogenaamde empirische "probability plots", een grafische methode om veronderstellingen omtrent de verschillen in verdeling van een variabele tussen twee steekproeven te toetsen. De theorie van de empirische probability plots wordt uitgelegd en toegespitst op de aannames over de wijze waarop veranderingen in consumptie in een populatie plaatsvinden. Naar Skog's verwachting veranderen mensen hun

drinkgedrag proportioneel aan het intieële niveau. Deze zogenoemde "multiplicativiteit" ligt niet alleen ten grondslag ligt aan Skog's theorie, maar ook aan het Ledermann model. Dit betekent dat overeenkomstige consumptieniveau's (dat wil zeggen, de percentielpunten van de verdeling) proportioneel veranderen, en dus dat twee verdelingen van consumptie (na logaritmische transformatie van de ruwe data) aan elkaar gelijk zijn op een lineaire transformatie na (schaal- en locatie-parameter). De toetsing van de hypothese met longitudinale data series uit surveys onder de algemene Nederlandse bevolking, gehouden in 1970, 1981, en 1985, geeft een bevestiging van de veronderstelde gelijkvormigheid van verdelingen. Ook de verdelingen van mannelijke en vrouwelijke consumptie voldoen aan de bovenstaande aannames. Algemene conclusie is dat een van de assumpties van Ledermann, namelijk een relative stabiliteit in verdeling, dus wel kan worden bevestigd. Bovendien is er een duidelijke statistische onderbouwing aan gegeven.

In hoofdstuk 8 wordt de empirische ondersteuning van de conclusie uit het vorige hoofdstuk verder uitgebreid. Steekproeven uit de algemene mannelijke en vrouwelijke bevolking van een viertal Westerse landen, USA, UK, West-Duitsland en Zwitserland, worden vergeleken met die uit de data-sets van de Nederlandse 1985 survey. Omdat in de betreffende surveys consumptie met verschillende methoden (QF en retrospectief dagboek) is vastgesteld, wordt allereerst nagegaan of dit verschil in methode effect heeft op de ingeschatte relatie tussen verdelingen. Het blijkt dat de evaluatieve QF-methode een slechtere kwaliteit van de plot geeft dan de op feitelijke consumptie gebaseerde retrospectieve methode, hetgeen in latere analyses nog wordt bevestigd. De nul-hypothese van lineariteit kan echter niet worden verworpen. Uit de resultaten van de analyses kan worden geconcludeerd dat, hoewel de deviaties van lineariteit soms groot zijn (onder meer als gevolg van het gebruik van QF-methode), er geen sprake is van systematische afwijkingen (behalve USA vrouwen). Over het geheel genomen is de relatie tussen (log-getransformeerde) consumptie verdelingen lineair, de ontwikkeling in verdeling dus redelijk voorspelbaar, hetgeen een bevestiging van Ledermann's basale aanname is.

Hoofdstuk 9 bespreekt de resultaten van de methodologische studies in hun onderlinge samenhang en specificeert de implicaties voor (het onderzoek naar) de verdeling van consumptie. Zo lijkt de hypothese van een lineaire relatie tussen ware score en fout door de resultaten te worden bevestigd. Dit houdt in dat conclusies over de regelmatigheden in verdeling, maar ook, bijvoorbeeld, correlatie-coëfficiënten niet worden vertekend. Daarnaast worden er suggesties gedaan voor alternatieve designs en technieken waarmee de kwaliteit van surveygegevens kan worden onderzocht. In dit licht worden computersimulaties genoemd, waarin naast het drinkgedrag ook geheugen- en response-effecten worden gesimuleerd. Er wordt betoogd dat voor een betere vergelijkbaarheid van

onderzoeksresultaten tussen landen en over tijd, er een theorie dient te worden ontwikkeld waarin wordt aangegeven hoe mensen reageren en antwoorden op vragen naar alcoholgebruik en gerelateerde problemen.

In het tweede gedeelte worden de resultaten gepresenteerd van ander onderzoek dat binnen het "Ledermann-project" aan de Rijksuniversiteit Limburg heeft plaatsgevonden. Het betreft een onderzoek dat aantoonst dat de aanname van gelijkheid van verdeling (Ledermann, hoofdstuk 7 en 8) incompatibel is met Skog's hypothese van collectiviteit van drinkculturen. Voorts wordt een onderzoek besproken dat, als een logisch vervolg op de benadering in dit proefschrift, de stabiliteit van individueel drinkgedrag over langere periode tot onderwerp heeft. Dit aspect kan invloed hebben op de relatie tussen consumptie en negatieve gevolgen. De relatie tussen beide wordt nader uitgewerkt in een paragraaf waar de gevonden resultaten, i.c. de regelmatigheden in verdeling, in een breder kader van een preventiebeleid worden geplaatst. Met de alcoholische levercirrose als voorbeeld wordt verhelderd welke plaats de verdeling van alcohol inneemt in de totale cyclus van een restrictief alcoholbeleid. Hoewel de exacte vorm van de verdeling en zijn relatieve stabiliteit bepalend zijn voor het aantal excessieve drinkers, zijn er vele andere variabelen te onderscheiden die de relatie tussen per capita consumptie en alcohol gerelateerde problematiek zowel in positieve als in negatieve zin beïnvloeden.

Appendix 1.1 Empirical studies on the shape of the distribution of alcohol consumption.

SOURCE	POPULATION CHARACTERISTICS	MAIN CONCLUSIONS OF THE AUTHORS	REMARKS
Ledermann, 1956	5 Clinical and non-representative populations in France, US, Italy, and Canada	Ledermann has used these samples for estimation of the proportion drinking more than an average 100cl pure alcohol alcohol per day ($\approx 0.03\%$).	
Ledermann, 1964b	27 populations from France, of which 13 clinical ones (alcohol related health problems); average consumption over 20 liters pure alcohol per year; self-reports	Approximately lognormal distribution; in samples with a lower mean, greater discrepancies between expected and observed are found.	No statistical test; test for goodness of fit (reported by Smith, 1976) no confirmation of lognormality
De Lint & Schmidt, 1968	Data on individual purchases of wine and liqueur, in Ontario (Can); official registration of retailers.	Similarity between model estimates and observed frequencies at various levels. Ledermann model over-estimates prevalence of heavy drinking. Observed distribution is less skewed. No test of lognormality.	Duffy (1978) and Skog (1973) have criticized the implicit assumption that equals purchase and consumption behavior.
Smart & Schmidt, 1970	5 Samples of BAC's of car drivers, not involved in car accidents.	Approximately lognormal distribution. No test of Ledermann formula. Proportion of drivers with BAC's over a certain cut-off point, is strongly related to BAC sample.	Critical comment by O'Neill & Wells (1971) who report strong deviation from lognormality for 1 sample. However, Ekholm (1972) showed that a change in mean BAC implied a similar change in proportion of BAC over a certain cut-off point.
Skog, 1971	Survey data from 6 populations from different countries.	Distributions approximately lognormal. In 5 out of 6 samples an over-estimation of heavy drinkers by the Ledermann-model. In 3 out of 6 cases, a significant deviance from lognormality.	
Ekholm, 1972	Finnish longitudinal study in 1968 and 1969. General population. An increase in mean consumption between '68 and '69 of 49%.	Distributions in both samples are approximately lognormal. A decrease in dispersion of only 5%.	
Brun-Gulbrandsen, in Brun et al. (1975)	Longitudinal survey among 15-19 year old Norwegians, 2 successive years.	No details about shape of the distributional. Small decrease of dispersion as the mean increases. The increase of the prevalence of heavy drinking is larger than the increase in mean consumption.	
Duffy, 1977a	Secondary analysis of data from a Scottish survey	Large difference in magnitude of dispersion parameter between the male and female samples. The magnitude of the dispersion for female distributions differs from the ones presented in Bruun et al. (1975). Such differences would have a large effect on estimates of excessive use.	

cont

Appendix 1.1 (cont.1)

SOURCE	POPULATION CHARACTERISTICS	MAIN CONCLUSIONS OF THE AUTHORS	REMARKS
Buffy, 1977b	Re-analysis of data from the study by De Lint & Schmidt's Ontario study (1968).	Deviance between observed frequencies at various consumption levels and values predicted by the Ledermann model.	According to Skog (1980b), wrong choice of 'constants' in the Ledermann formula. After correction, distributions appear approximately lognormal.
Cartwright et al., 1978	Longitudinal survey in London suburb, '58 and '74	Rise in mean consumption between '58 and '64 has been accompanied by a rise in the prevalence of heavy use. No test of Ledermann formula.	Differences in problem rates between sub populations
Skog, 1979	Re-analysis of 6 surveys also used in 1971.	A two-parameter gamma distribution seems a better choice than a lognormal model. No test of Ledermann formula.	Statistical problems because of grouped data.
Guttorp & Song, 1979	Re-analysis of 6 surveys used by Skog, 1971.	Lognormal, two-parameter distribution model gives the best fit to the data. No test of Ledermann formula.	see Skog, 1979; Controversy as a result of statistical method applied.
Gresson & Stacey, 1980	General survey New Zealand. Male and female sample considered separately.	Skewed, approximately lognormal distribution. Deviance from lognormality in right tail. Under-estimation by Ledermann model of the proportion at the highest levels.	Skog (1981) has shown that estimates of consumption with last-occasion method are biased.
Skog, 1980b	Re-analysis of the Ledermann data (1964b).	Homogeneity of sample affects shape of distribution. Heterogeneity of a sample w.r.t average consumption results in deviance from lognormality.	
Plant, 1980	Survey data from 4 Scottish towns, 1976. Different mean consumption levels.	Large variability in consumption patterns. Survey data are ill-suited for estimation of frequencies at higher consumption levels. Differences in drinking pattern affect proportion of heavy drinkers. Ledermann model is inadequate for prediction and description of excessive use.	Plant does not take heterogeneity of population into account (e.g. # abstainers). Even though complex, a relation is found between mean consumption and prevalence of alcohol related problems.
Skog, 1980c	Secondary analysis of 1 Scottish and 2 London surveys (see: Dight, 1976; Cartwright et al. 1978)	Statistical test confirms (2 parameter) lognormal distribution. Ledermann model over-estimates prevalence of heavy consumers. The observed distribution is less skewed.	Skog rejects the mathematical assumptions of the Ledermann model as being too strict (see also Bruun, et al., 1975).
Stacey & Elvy, 1981	Survey among 14-17 year-olds in New Zealand in 1978 and 1979. Male and female samples are considered separately.	Distributions are approximately lognormal with deviances in the tail. Under-estimation by Ledermann model of proportion drinking at the highest levels. Equal shape of distributions for male and female sample.	Skog (1981) has shown that method of measurement of consumption (last-occasion) is biased. No statistical test.

cont

SOURCE	POPULATION CHARACTERISTICS	MAIN CONCLUSIONS OF THE AUTHORS	REMARKS
Fitzgerald & Mulford, 1982	Survey among general population in Iowa, USA. Self-reports in "dry", low mean consumption state.	Distribution is approximately lognormal. Over-estimation by Ledermann model of proportion at the highest consumption levels. Equal estimates of proportion drinking over 8 glasses per day. Observed distribution is less skewed.	No statistical test. Validity of consumption data can be questioned.
Kendall et al., 1983	Scottish longitudinal surveys conducted in 1978 and 1982.	Decrease in mean consumption of 18% (corrected for age) is the result of a decrease in consumption at all levels. Heaviest drinkers show a somewhat larger reduction.	Because subjects with zero consumption in 1978 are excluded in 1982, a possible bias may have been introduced: under-representation of light drinkers.
Knibbe, et al., 1985	Analysis of 3 general surveys in the Netherlands in 1958, 1970, 1981.	A rise in mean consumption has been accompanied by a even larger increase of the proportion of drinking more than 3 glasses per day.	No analysis of the shape of the distribution.
Romelsjö & Ågren, 1985	Swedish surveys in 1968 and 1979.	A rise in mean consumption of 8%, has been accompanied by an increase of 29% in the proportion of heavy drinkers. The result is interpreted as evidence for the single distribution model.	The large decline in mean consumption after 1979 has not resulted in a decrease of the prevalence of heavy use as large as could be expected on the basis of the results in the previous period.
Skog, 1985	Data sets from 21, mostly, general population surveys in different countries	All distributions are clearly skewed, with the mean/median ratio between 2.5 and 1.5. Variation in ratio is large especially in samples with means below 10 liters 100% alcohol.	Skog rejects Ledermann's rigid distribution law but presents two hypothetical mechanisms with which to explain dynamics in distribution of consumption.
Lemmens, 1987	Dutch general population survey in 1985. Male and female samples considered	Ledermann model over-estimates proportion of drinkers at the higher levels, for both sub-samples. Over-estimation is more apparent in the female sample. The model over-estimates the share of the total consumption that is consumed by the highest decile.	No statistical test of lognormality.

Appendix 1.2 Possible sources of error, type of bias, and methods that have been used to assess bias in chapters 2 through 6.

SOURCE OF ERROR	TYPE OF BIAS	ASSESSMENT OF BIAS
1. Non-coverage of target population by sampling frame (e.g., persons younger than 16 years of age or older than 70).	Overestimation of mean, per drinker consumption, based on official records; higher estimated under-coverage	Assessment of consumption in sub-samples that are not covered by the sampling frame (see ch4, sample section).
2. Non-response.	Selective non-response, due to a lower response probability of heavy drinkers; underestimation of consumption.	Follow-up of non-response by telephone model of response probabilities (see ch2).
3a. Incomplete time sampling. Reference period of survey does not cover entire year.	Depending on the temporal variation of aggregate consumption and the data collection period, under- or overestimation of annual mean consumption.	Spreading of interviews over the entire reference period (see ch5).
3b. Incomplete time sampling. Systematic exclusion of collective holidays from the time frame.	Because of the expected higher consumption during these periods mean consumption will be under-estimated.	Recording of consumption in a sub-sample during Christmas holidays (see ch5).
4. Basic unit of measurement; in self-reports usually one "standard glass of beverage alcohol".	a. Most beer is packed in bottles of varying content; b. Wines and spirits, drunk outside pubs, are of varying content. Mistakes are made in conversion to standard glasses; bias will expectedly be toward under-estimation of total volume.	a. Exact recording of number and type of bottles, cans, glass, and subsequent conversion into standard glass. b. Assessment of mean deviation from standard size of glasses from which drinks at home are consumed (see Lemmens, 1987, p. 40/49).
5. Decay of memory, forgetting, incorrect recall.	Forgetting leads to underestimation, relative to length of recall period. Incorrect dating of events can lead to overestimation (forward telescoping) or underestimation (backward telescoping).	Effect of length of recall period assessed by means of diary method (see ch3 and ch4).
6. Questions about alcohol consumption may induce anxiety; threat experienced while reporting may increase the risk of distortion.	Conscious or subconscious suppression of consumption, which will lead to underestimation, which is probably positively related to consumption.	Comparisons with reports from collateral informants, viz., partners of the originally selected subject (see ch6).
7. Temporal variation of individual drinking behavior.	Questions are about usual or "modal" consumption, disregarding temporal variation (incidental heavy drinking) which will underestimate mean consumption.	Question about incidental, infrequent heavy (6 glasses or more) use of alcoholic beverages (see ch4).

CURRICULUM VITAE

The author was born on 19th April 1955 in Beek, the Netherlands, graduating from Atheneum-B at the Sint Michielslyceum in Geertruidenberg. He has studied psychology at the University of Amsterdam, where he graduated in 1985 with a degree in clinical psychology.

Between 1982 and 1985, he has been intermittently employed by a rehabilitation research institute in Hoensbroek, as interviewer and assistant.

From July 1985 until November 1990 he has been a researcher at the Ledermann project, an initiative of the Department of Medical Psychology at the University of Limburg. The first 3 years he has been employed by the Netherlands organization for scientific research. From August 1988 several temporary assignments as a researcher by the Faculty of Health Sciences have enabled him to complete this book.