‘The missing drink’

Nonresponse and data quality in mail surveys on alcohol consumption

Viviënne Lahaut

'The missing drink'. Nonresponse and data quality in mail surveys on alcohol consumption.
Lahaut, Viviënne M.H.C.J.
Thesis ErasmusMc, University Medical Center Rotterdam
– with summary in English and Dutch.

Cover design by Suzanne Tuynman, S.C.T. Producties, info@sctproducties.nl

Printed by Grafisch Bedrijf Ponsen & Looijen

ISBN-10: 90-74234-61-5
ISBN-13: 978-90-74234-61-0

‘The missing drink’
Nonresponse and data quality in mail surveys on alcohol consumption

‘Het missende glas’
Nonrespons en datakwaliteit bij postenquêtes over alcoholconsumptie

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus

Prof.dr. S.W.J.Lamberts

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
22 november 2006 om 15.45 uur
door

Viviënne Maria Hubertina Catharina Josephina Lahaut
geboren te Maastricht

Promotiecommissie

Promotor:	Prof.dr. H.F.L. Garretsen
Overige leden:	Prof dr. J. Passchier
	Dr. P.H.H.M. Lemmens
	Prof.dr. H.G.M. Rigter
Copromotoren:	Dr. H. van de Mheen
	Dr. H.A.M. Jansen

CONTENTS Page

Chapter 1	General introduction	7
PART I Nonresponse error		21
Chapter 2	Nonresponse bias in a sample survey on alcohol consumption.	23
Chapter 3	Estimating non-response bias in a survey on alcoholconsumption: comparison of response waves.	37
Chapter 4	Effects of survey sponsorship, questionnaire length, topic label and design on response rate in mail alcohol surveys.	53
PART II Measurement error		65
Chapter 5	Comparison of two formats of the weekly recall and quantity-frequency alcohol measures.	67
Chapter 6	Measuring the year consumption of alcohol: the development of a questionnaire.	81
Chapter 7	General discussion	99
Summary		113
Samenvatting		121
Dankwoord		131
Curriculum Vitae		137

The chapters 1-7 are based on the following papers:

Viviënne M.H.C.J. Lahaut, Harrie A.M. Jansen, Dike van de Mheen, Henk F.L. Garretsen. Non-response bias in a sample survey on alcohol consumption. *Alcohol & Alcoholism* 2002, 37(3): 256-260.

Viviënne M.H.C.J. Lahaut, Harrie A.M. Jansen, Dike van de Mheen, Henk F.L. Garretsen, Jaccqueline E.E. Verdurmen, Ad van Dijk. Estimating non-response bias in a survey on alcohol consumption of response waves. *Alcohol & Alcoholism* 2003, 38(2): 128-134.

Viviënne M.H.C.J. Lahaut, Harrie A.M. Jansen, Dike van de Mheen, Henk F.L. Garretsen. Comparison of two formats of the weekly recall and quantity-frequency measures. *Journal of Substance Use* 2003, 8(3): 164-169

Viviënne M.H.C.J. Lahaut, Harrie A.M. Jansen,. Henk F.L. Garretsen, Dike van de Mheen. Effects of survey sponsorship, questionnaire length, topic label, and layout on response rate. *Submitted*.

Harrie A.M. Jansen, Viviënne M.H.C.J. Lahaut, Henk FL Garretsen, Dike van de Mheen. Measuring the year consumption of alcohol: the development of a questionnaire. *Submitted*.

Chapter 1

General introduction

Introduction

The level of alcohol consumption in a population is an important determinant of health and social well-being (WHO, 2005). In the Netherlands, since the 1980s the year consumption of alcohol has been around 8 litres of pure alcohol per person. In 2003, the year alcohol consumption in Europe ranged from 6.9 liters in Italy to 12.6 liters in Luxembourg (van Haren, van der Wilk, van Laar, 2005). In societies where drinking is part of social life drinking is often seen as positive. There is even some evidence that moderate drinking (consumption levels of less than 10 g alcohol a day) has a protective effect against the risk of coronary heart disease. Overall, however, there is a well-documented relationship between the level of alcohol consumption and health damage. Levels of alcohol consumption higher than 20-30 g a day can have harmful effects in the long term on the brain (e.g. Korsakov's syndrome), liver (e.g. liver cirrhosis), stomach (e.g. irritable stomach) and heart and blood vessels (e.g. high blood pressure, cardiac arrhythmias, cerebral haemorrhages) (Anderson, Cremona, Paton et al. 1993).

Alcohol consumption is moderately well accepted by the Dutch society. At the national level the Dutch Ministry of Health, Welfare and Sports aims to arrive at an effective alcohol policy that prevents health damage by drinking. The alcohol policy includes a coherent package of health, social and fiscal measures. Examples of these measures are excise on alcohol sales, age restrictions on the sale of alcoholic beverages, national prevention campaigns, and care programmes for those with alcohol addiction or alcohol problems. At the regional level by means of the Collective Prevention Act (in Dutch: Wet Collectieve Preventie Volksgezondheid) the municipalities are obliged to gain insight into public health and to carry out a four-yearly health policy. Part of this health policy is to develop and apply a local alcohol policy. Besides the reduction of alcohol-related traffic and nuisance problems, a local alcohol policy should focus on the reduction of health damage by means of preventive and curative measures.

To develop an effective policy, the policymakers need data on how much the local population drinks. Besides the total amount of alcohol consumption, drinking patterns need to be determined because the extent of health damage due to alcohol consumption is also dependent on drinking patterns. For example, binge drinking (drinking a lot of alcohol in a short period of time) in the short term is more harmful than drinking the same amount spread over a longer period of time. Therefore, it is also important to identify drinking patterns such as binge drinking.

Indirect and direct methods are used to get information on alcohol use. Indirect methods include analysis of secondary data sets that have been collected for other purposes, for example morbidity and mortality registration, or the registration of

emergency care in relation to alcohol use. Besides such registrations, alcohol sales figures can also be used to estimate alcohol consumption. The advantage of secondary data sets is that data are already available and costs are therefore relatively low. However, data collection, the registration procedure and the quality of data is then beyond the researcher's scope and, most importantly, these secondary data sets can not provide representative data on a population's drinking patterns. Therefore, municipalities and their local health authorities prefer to use survey research as a direct method of alcohol consumption measurement.

Alcohol consumption measurement can be part of a specific survey on drugs and alcohol use, or may be part of a general health survey. These general population surveys provide data on alcohol use, drinking patterns and risky drinking behavior. In large Dutch cities health surveys including alcohol questions are carried out periodically (i.e. about every four years).
However, besides having advantages, this type of survey research also has drawbacks. Four types of errors can threaten the survey data: coverage error, sampling error, nonresponse error, and measurement error (Groves, 1989). Coverage error occurs when members of the target population have unequal probabilities to be included in the survey sample. Sampling error occurs when only a subset of the population is surveyed. Nonresponse error occurs when survey respondents differ from nonrespondents in relevant characteristics related to the research topic. The fourth error, measurement error, occurs when the measurement instruments (such as questionnaire, interview) produce false outcomes.

Our institute the Addiction Research Institute has considerable experience in conducting alcohol and drug surveys. The methodological issues arising from the 1994 alcohol postal survey in Rotterdam (Bongers, 1998), which is a slightly adapted version of the postal survey as developed and conducted in 1981 by Garretsen (1983) formed the basis of this thesis. Of the four above-mentioned types of errors, coverage and sampling errors are not serious problems for the data quality of a postal alcohol survey because in the Netherlands we have municipal registers that contain basic information on every inhabitant. This enables to draw random samples from the general population. However, it should be noted that some subgroups of the population are difficult to reach via a postal survey, such as the homeless, prisoners and those with a long-term stay in health care institutions.

Conversely, nonresponse and measurement errors are serious threats to the quality of the alcohol survey data. These two errors are frequently propounded as explanatory factors for the discrepancy between survey-based alcohol estimates and alcohol estimates based on sales data; surveys show a significant

undercoverage of alcohol consumption (Lemmens et al, 1992; Leifman et al, 2000). In the Netherlands, response rates on mail surveys have been decreasing during the past decades (de Leeuw & de Heer, 2002); this is also the case in Rotterdam. While the 1981 alcohol survey had a response rate of 70%, the 1994 alcohol survey had only 44%. The question arises whether this relatively low response rate has consequences for the interpretation of the alcohol estimates. Besides decreasing response rates measurement problems also arise due to self-reports. For example, cognitive studies on alcohol survey have shown that many respondents have problems with answering questions on their alcohol consumption (Midanik, 1999; Strunin, 2001), which may explain at least part of the under coverage.

In summary, two main types of error may threaten the quality of alcohol surveys: nonresponse and measurement error. Using the 1994 Rotterdam alcohol survey as reference in this thesis, we address both nonresponse and measurement error in a postal survey on alcohol consumption. In the introductory chapter we first describe the nonresponse problems and subsequently the problems related to measurement of alcohol data by survey. We then present the research questions, and end this chapter with an outline of the thesis.

Nonresponse

Survey nonresponse may cause two problems. First, nonresponse leads to a smaller sample size and therefore a loss of accuracy in population estimates. A solution for this problem is to recruit larger sample sizes or, in case of an anticipated high nonresponse among specific population subgroups, oversampling of these subgroups beforehand. A second and more serious problem is a possible selection bias due to nonresponse. Selection bias occurs when a significant number of people in the survey sample do not respond to the survey and have characteristics that differ from those who do respond, when these characteristics are related to the central survey variable (Dillman, 2000). For example, when people who do not drink refrain from participation in an alcohol survey. A low response rate does not necessarily mean that there is no nonresponse error; there also has to be a relationship between the response and the survey variable of interest in order to produce error.

Estimating and correcting nonresponse error

There are indirect and direct methods to deal with nonresponse error. One widely-used indirect method of correcting for nonresponse error is weighting, based on known characteristics of respondents and nonrespondents. This method generally uses socio-demographic characteristics of both respondents and nonrespondents (e.g. age, gender, marital status) to weight survey cases. However, this method is only valid when respondents and nonrespondents

within the same socio-demographic category are equal on the outcome variable; this assumption seems untenable in most cases (Bradburn, 1992; Presser & Traugott, 1992; Van Goor & Stuiver, 1995). Another indirect method to deal with nonresponse error is to use survey estimates from late respondents (i.e. subjects who required more reminders before they respond) as a proxy for nonrespondents. This method is based on the 'continuum of resistance' model, applied by Lin & Schaeffer (1995) and Voogt et al. (1998). However, neither of those studies provided strong support for this model: the patterns of the relationship between the response of different response waves and the variable of interest were irregular.

In the present study we analyze differences between response waves from an alcohol survey and two additional Dutch alcohol surveys, in order to replicate the test of the 'continuum of resistance' model (see Chapter 3).

Indirect methods, such as weighting on demographic variables or extrapolating estimates of late respondents, are only useful if the weighting variables and the response delay are strongly and directly related to the dependent survey variable of interest. In most surveys strong evidence for this relationship is not available. Therefore, there is a need to develop and apply direct methods to assess nonresponse error. The direct approach is to collect information on the nonrespondents themselves which can be used to provide data to correct for possible nonresponse error.

The two main methods applied to study nonrespondents directly use follow-up surveys among (a sample of) nonrespondents. The first method is the strategy of double sampling which implies approaching the complete sample via an 'inexpensive' survey mode and subsequently drawing a sample of nonrespondents and approaching these with an 'expensive' mode and achieving 100% response among this group (Hansen & Hurvitz, 1946). Bethlehem & Kersten (1986) developed a variant of the double sampling method, namely the 'Basic Question Approach". This implies that all nonrespondents are asked one central research question. A high response rate on these follow-up surveys is a condition for the validity of these direct methods; only with a high response rate in the follow-up of nonrespondents can one be confident that the follow-up is really representative for all nonrespondents. Until now, follow-up studies among nonrespondents of Dutch alcohol surveys had relatively low response rates. For example Garretsen (1983) approached 850 nonrespondents again with a short questionnaire of which 106 nonrespondents participated. Another nonresponse follow-up survey of Lemmens, Tan & Knibbe (1992) reached a response rate of 39% (116 of 295 nonrespondents).

In order to achieve a high response rate we opt for a small-scale intensive follow-up study to estimate nonresponse bias. In our study we follow-up a small

sample of nonrespondents intensively with a small number of questions about their alcohol consumption (see Chapter 2).

Enhancing response rates

Because, it is obviously better to prevent nonresponse error, it is important to try to enhance response rates based on theories on survey participation. Groves, Cialdini & Couper (1992) distinguished factors at four levels which can influence survey participation: societal level factors (e.g. survey taking climate or legitimacy of societal institutions), attributes of the survey design (e.g. questionnaire length or survey topic), interviewer attributes and respondent-interviewer interaction (e.g. persuasive strategies or interviewee expectations), and characteristics of the sample person. Dillman (2000) applies the social exchange theory to explain survey participation; in this theory three elements are important: rewards, costs and trust. Rewards are what one expects to gain from survey participation e.g. social validation, appreciation, 'liking to do' or tangible rewards, such as money. Costs are what one gives up or spends from survey participation, e.g. inconvenience, embarrassment, great physical or mental effort, or privacy invasion. Trust is the expectation that in the long run the rewards of participating will outweigh the costs, e.g. participants may see a legitimate authority or a token of appreciation in advance as a form of trust. In our survey we will manipulate four survey characteristics that may influence the reward-costs-trust-balance of participating: sponsorship, questionnaire length, topic label and questionnaire layout

Relationship of (non)response and survey topic: theory of the 'two-tailed nonresponse' pattern

In epidemiological survey research on alcohol consumption a general idea prevails that excessive drinkers may participate less in surveys than others, because of the social sensitivity to the topic (i.e. shame). However, since 1983 nonresponse research within this field has showed that nonresponse has been prevalent especially among abstainers, while no serious nonresponse bias has been found among excessive drinkers. Garretsen (1983) found an underreporting of both the frequency and the quantity of alcohol consumption. Lemmens, Tan & Knibbe (1988) found no evidence of more heavy drinkers among nonrespondents. Indeed, they found evidence for higher rates of abstention among female nonrespondents. However, the response rates of these two follow-up surveys among nonrespondents were respectively 12% and 39%.

Results from a nonresponse study in an area other than alcohol research (i.e. municipal policy on caravans) showed that nonresponse was overrepresented in the extremes of the survey variable of interest for different reasons (van Goor & Stuiver, 1995). In this latter survey on caravan policy among municipalities, nonresponse was overrepresented among municipalities that do not succeed in

their caravan policy (lowest extreme) and among municipalities that have a successful policy (highest extreme). This may be caused by topic saliency (municipalities that already have a successful policy are not interested anymore in this topic), or by prevailing societal norms on the topic (municipalities are threatened by the survey and feel ashamed by their lack of policy). This theory of 'two-tailed nonresponse' pattern may be useful to apply to alcohol surveys. In our alcohol survey we expect relatively more abstainers among the nonrespondents because of lack of interest in the survey topic, and we expect relatively more excessive drinkers among the nonrespondents because of shame. In the Dutch society drinking alchol is accepted, but only in moderation.

Measurement error

Comparison of alcohol data from surveys with tax-based sales data revealed a serious undercoverage: survey estimates generally cover 40-60% of the recorded sales data (Lemmens, Tan & Knibbe, 1992; Leifman, Österberg & Ramstedt, 2002). Besides nonresponse error as causal factor undercoverage may also be explained by measurement error due to self-reports. Subjects may deliberately of inadvertently underreport or overreport their alcohol consumption. Respondents may deliberately misreport their alcohol consumption because they feel threatened or ashamed by the survey topic. One method to stimulate a respondent to give a valid answer is by making the survey method as confidential as possible. This can be done by using a self-administered questionnaire instead of a personal interview, and/or reassuring the respondent that the information is confidential (de Leeuw, 2001). Respondents may inadvertently misreport their alcohol consumption because of recall failure or because the questions are unclear or too complicated to fill in. Questions and their wording, question comprehension, response categories may determine the cognitive challenge of the survey.

Self-reports can be enhanced by improving respondents motivation and facilitating cognitive processes (Tourangeau, Rips & Rasinski, 2000). For this, insight into the question-answering process is necessary. Cognitive processes of responding to the question can be revealed by 'think aloud' procedures (i.e. asking the respondent to think aloud while answering the questions), and cognitive interviews (i.e. probing respondents retrospectively on their interpretations of questions). Knowing the cognitive problems, there are two main cognitive psychological principles to enhance self reports: to specify substantive or quantitative elements of the behavior; and to specify the contexts of the behavior. These two principles may improve the retrieval of information from the memory.

From the field of alcohol survey methodology much research has aimed to enhance alcohol self-reports by applying these two cognitive-psychological principles in the development of alcohol questions (Del Boca & Noll, 2000). For example, a disadvantage of the widely-used quantity-frequency measure is that subjects may find it difficult to report an average of their alcohol consumption over time. To simplify this averaging task, anchors of quantitative elements of drinking behaviour are provided in the graduated frequency measure (Greenfield, 2000). This measure ask subjects to report frequencies for each of graduated series of consumption levels, starting with the highest consumption level. Recent drinking measures (i.e. weekly recall) are good examples of minimizing recall failure by shortening the time span of the question. In the timeline follow-back method another strategy of minimizing recall failure is used: situational contexts of drinking are specified by providing respondents with a calendar with special dates (i.e. Queen's Day, birthdays, holidays) to be memorized (Sobell & Sobell, 1992).

From a pilot study (Jansen & Hak, 2002), respondents' cognitive problems of some alcohol questions in the 1994 Rotterdam alcohol survey (Bongers, 1998) were identified. To solve these problems we applied the two cognitive psychological principles to develop some alternative question formats for two alcohol measures. In our study we test these alternative formats in an experimental survey (see Chapter 5). Based on our cognitive interviews and the specific-settings approach of Single & Wortley (1994) we developed a new year consumption measure aimed to overcome the problem of long-term memory of drinking events and the problem of context-free memorizing of frequencies and quantities. In our explorative study the newly-developed year consumption questionnaire is tested in an alcohol survey (see Chapter 6).

Research questions

As described before, the methodological issues remaining after the 1994 Rotterdam alcohol survey (Bongers, 1998) represent the starting point of this thesis. Two main types of error may threaten the quality of alcohol surveys: nonresponse and measurement error. In this thesis we address both of these errors in a postal alcohol survey in Rotterdam. The purpose of this thesis is twofold: 1) to gain more insight into estimating nonresponse error and enhancing response rate and 2) to gain more insight into possibilities to enhance alcohol self-reports.

The following research questions are addressed in this thesis:

- Is there a bias in the alcohol survey results due to nonresponse and is there evidence for the theory of the 'two-tailed nonresponse pattern'?

- Is there evidence for the 'continuum of resistance' model: Can alcohol estimates of late respondents be used as proxies for estimates of nonrespondents? Are there differences in alcohol estimates between the different response waves?
- What are the response effects of different sponsors, questionnaire length, topic label and layout?
- Could formats including more memory cues be better alternatives for the commonly-used formats of the weekly recall and quantity-frequency alcohol measure in enhancing self-reports?
  What problems do respondents experience in answering the different formats of the weekly recall and quantity-frequency alcohol measure?
- Could a newly-developed year consumption questionnaire be a better alternative to a typical week measure in enhancing year alcohol consumption?

Outline of this thesis

This thesis consists of two parts. In part I, which includes Chapters 2, 3 and 4, we address aspects of nonresponse error. Firstly, nonresponse on a postal alcohol survey among the general population is investigated by a follow up study of nonrespondents. Here, the theory of a 'two-tailed pattern of nonresponse' is applied and discussed (Chapter 2). In Chapter 3 we discuss the validity of the 'continuum of resistance' model in estimating nonresponse bias. Can alcohol estimates of late respondents be used as proxies for alcohol estimates of nonrespondents? Therefore early respondents, late respondents and nonrespondents are compared on alcohol consumption. The effects of manipulating four survey characteristics (sponsorship, questionnaire length, topic label and design) on the response rate are described in Chapter 4.
The second part of this thesis is concerned with aspects of measurement error. Part II includes Chapters 5 and 6. In Chapter 5 we aim to improve two commonly-used alcohol measures (the weekly recall, and the weekly quantity-frequency measures) by rephrasing and restructuring the questions. Therefore, two formats of each alcohol measure are compared with regard to alcohol estimates and item nonresponse. We also aim to gain insight into respondent's problems in completing the alcohol questions by means of cognitive interviewing. In the last chapter of part II the validity of a newly-developed alcohol questionnaire is tested in a pilot survey. Therefore, the year alcohol consumption measured by this new instrument is compared to the year consumption based on a typical week report.

The final chapter of this thesis contains a general discussion: in this chapter we present a summary of the main findings, address methodological limitations,

and finally discuss theoretical and practical implications of the work presented in this thesis.

References

Anderson P, Cremona A, Paton A et al (1993). The risk of alcohol. Addiction, 88(11): 1493-1508.

Bethlehem JG and Kersten HMP (1986). Werken met non-respons. Statistische Onderzoekingen M30, CBS publicaties,'s Gravenhage: Staatsuitgeverij.

Bongers IMB (1998). Problem drinking among the general population: a public health issue? PhD thesis. Erasmus University Rotterdam.

Bradburn NM (1992). A response to the nonresponse problem. Public Opinion Quarterly, 56: 391-397.

Del Boca FK and Noll JA (2000). Truth or consequences: the validity of self-report data in health services research on addictions. Addiction, 95(3):S347-S360.

Dillman DA (2000). Mail and Internet surveys: the Tailored Design Method. New York: Wiley.

Garretsen HFL (1983). Probleemdrinken: prevalentiebepaling, beïnvloedende factoren en preventiemogelijkheden. Theoretische overwegingen en onderzoek in Rotterdam. Proefschrift. Katholieke Universiteit Nijmegen.

Goor van H and Stuiver B (1995). A wave-analysis of distributional bias, substantive bias and data quality in a mail survey among Dutch municipalities. Acta Politica, 33: 179-196.

Greenfield TK (2000). Ways of measuring drinking patterns and the difference they make: experience with graduated frequencies. Journal of Substance Abuse, 12: 33-49.

Groves RM, Cialdini RB & Couper MP (1992). Understanding the decision to participate in a survey. Public Opinion Quarterly, 56: 475-495.

Groves RM (1989). Survey Errors and Survey Costs. New York: John Wiley & Sons Inc.

Hansen M and Hurvitz W (1946). The problem of nonresponse in sample surveys. Journal of the American Statistical Association, 41: 517-529.

Haren van N, Wilk van der EA and Laar van M (2005). Alcoholgebruik. Geografische verschillen: zijn er verschillen tussen Nederland en andere landen. Nationaal Kompas Volksgezondheid. Utrecht: Trimbos Instituut.

Jansen H and Hak T. Assessing the quality of quantity-frequency-variability questioning in a mail alcohol survey. Results from a field study with three-step-test interviews (TSTI) compared to desk expert analysis. Paper presented at the International Conference of Improving Surveys (ICIS, 2002). Copenhagen.

Leeuw de ED and Heer de W (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In: RM Groves, DA Dillman, JL Eltinge & RJA Little (eds.) Survey Nonresponse (pp. 41-54). New York: Wiley.

Leeuw de ED (2001). Reducing missing data in surveys: an overview of methods. Quality & Quantity, 35: 147-160.

Leifman H, Österberg E and Ramstedt M (2002). Alcohol in postwar. Europe: a discussion of indicators on consumption and alcohol-related harm. Finnish National Research and Development Centre for Welfare and Health and Swedish National Institute of Public Health: University of Stockholm.

Lemmens PHHML, Tan ES and Knibbe RA (1992). Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. Journal of Studies on Alcohol, 53 (5): 476-486.

Lin IF and Schaeffer NC (1995). Using survey participants to estimate the impact of nonparticipation. Public Opinion Quarterly, 59: 236-258.

Midanik LT (1999). Face-to-face versus telephone interviews: using cognitive methods to assess alcohol survey questions. Contemporary Drug Problems, 26: 673-693.

Presser S and Traugott M (1992). Little white lies and social science models: correlated response errors in a panel study of voting. Public Opinion Quarterly, 56: 77-86.

Single E and Wortley S (1994). A comparison of alternative measures of alcohol consumption in the Canadian National Survey of alcohol and drug use. Addiction, 89: 395-399.

Sobell LC and Sobell MB (1992). Timeline followback: a technique for assessing self-reported ethanol consumption. In: Allen J, Litten RZ (eds) Measuring alcohol consumption: Psychosocial and Biological Methods (pp. 41-72). Totowa, NJ: Humana Press.

Strunin L (2001). Assessing alcohol consumption: developments form qualitative research methods. Social Science & Medicine, 53:215-226.

Tourangeau R, Rips LJ and Rasinski K (2000). The psychology of survey response. Cambridge: Cambridge University Press.

Voogt RJJ, Saris WE and Niemöller B (1998). Non-response and the gulf between the public and politicians. Acta Politica, 33: 250-280.

World Health Organization (2005). Factsheet Alcohol policy in the WHO European Region current status and the way forward.

PART I

Nonresponse error

Chapter 2

Nonresponse bias in a sample survey on alcohol consumption

Viviënne M.H.C.J. Lahaut, Harrie A.M. Jansen, Dike van de Mheen, Henk F.L. Garretsen. Non-response bias in a sample survey on alchol consumption. *Alcohol & Alcoholism* 2002; 37 (3): 256-260.

Abstract

In a non-response follow-up study non-respondents of the original mailed questionnaire were approached again by house visits in order to compare their alcohol consumption with that of the respondents of the same mailed questionnaire. Differences in alcohol consumption between respondents and non-respondents were found. There is strong evidence for overrepresentation of non-response among abstainers, but weak evidence among frequent excessive drinkers.

Introduction

During recent decades, survey response rates in most Western countries have declined. In the Netherlands, response rates of official governmental surveys are very low, with rates dropping from 80% in the 1980s to about 60% at the end of the 1990s; moreover, this negative trend is continuing. The low response rates in The Netherlands may be due partly to the fact that participation in governmental surveys is not mandatory (De Heer, 1999).

Non-response leads to a smaller final sample size and, therefore, to a loss of accuracy in population estimates. However, if the non-response is not related to the research variable of interest, taking larger samples can compensate for this loss. Conversely, if non-response is directly related to the research topic, errors may occur, which can seriously distort the survey results. This non-response bias occurs when a significant number of people in the survey sample fail to respond and have relevant characteristics that differ from those who do respond (Dillman, 2000). In such cases, the non-response is selective.

A widely used method to correct for a non-response bias is corrective weighting of the survey data by use of socio-demographic variables (Lemmens et al., 1988; Bongers et al., 1997). However, it has been shown that this method does not correct the bias sufficiently, because the inherent assumption that respondents and non-respondents within the same socio-demographic category are also equal on the outcome variable seems untenable (Bradburn, 1992; Presser & Traugott, 1992; Van Goor & Stuiver, 1995).

In the field of alcohol research, non-response is frequently considered as a causal factor of the discrepancy between the survey-based estimates and the official (taxes-) based estimates of alcohol consumption. Pernanen (1974) has reported that many survey subpopulations with a higher proportion of heavy drinkers tend to show higher non-response rates; this can cause underreporting of the survey data on alcohol consumption compared to the official sales data on alcohol consumption. Garretsen (1983) showed an underreporting of both the frequency and the quantity of alcohol consumption. The underestimation varied according to the subpopulations; for example, underreporting of alcohol consumption was higher among women than men, but there was no support for a higher non-response among the problem drinkers. Knibbe (1982) found some evidence for a higher percentage of abstainers and 'excessive' drinkers among the non-respondents whereas Lemmens et al. (1988) found no support for the hypothesis of higher alcohol consumption among non-respondents, compared with respondents. In the same study, Lemmens and colleagues did show that female non-respondents generally drank less and had higher abstention rates. Furthermore, 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 heavy alcohol use. A follow-up of non-respondents of the (American) National Household Survey of Drug Abuse showed no differences between non-respondents and the respondents in the prevalence of alcohol consumers (Caspar, 1992). Gmel (2000) found no differences between non-respondents and respondents in higher alcohol consumption; although the mean consumption and the percentage of heavy drinkers were higher in non-respondents, these differences were not significant. Wild et al. (2001) showed that nonresponders to a follow-up questionnaire on alcohol use and beliefs about drinking reported consuming five or more standard drinks once per week or more at a slightly greater rate (19,1%) than did respondents (14,3%). A limitation of most follow-up studies among non-respondents is the relatively low response rates of these secondary surveys. In the follow-up of non-respondents, generally there are many initial non-respondents who cannot be reached or who refuse again (Sosdian and Sharp, 1980; Jansen and Hak, 1999). It is often assumed that excessive drinkers and/or problem drinkers are difficult to reach because of their lifestyle characteristics (e.g. homelessness, seldom at home, not answering their mail). It is also assumed that excessive drinkers and/or problem drinkers, if reached, are more likely to refuse to participate in a survey than light and moderate drinkers.

Various models and theories have been developed to identify and explain non-response. Dillman (2000) applied the social exchange theory as a basis for survey design methods aiming at maximising the response rate. According to this theory, actions of individuals, in this case responding or not responding, can be predicted on the basis of three elements: rewards, costs and trust. Rewards are what one expects to gain from a particular activity, e.g. social validation, appreciation, 'liking to do' or tangible rewards such as money. In conducting a survey, a phrase such as 'we very much appreciate your help' has a reward value for many people. 'Liking to do' is also a powerful determinant of behaviour; most people enjoy participating in a survey if the topic is of interest to them. Costs are what one gives up or spends to obtain the rewards, e.g. inconvenience, embarrassment, feeling subordinated, anxiety, great physical or mental effort, or invasion of privacy. A simple act such as inclusion of a (paid) envelope for a mailed questionnaire increases the response rate, whereas questions that may cause embarrassment or anxiety lower the response rate. The third element in the social exchange theory is trust. This is the expectation that in the long run the rewards of doing something will outweigh the costs e.g. participants may see a legitimate authority or a token of appreciation in advance as a form of trust.
On a more concrete level, Groves et al. (1992) reported various (reward/cost) factors which can influence survey participation. They distinguished factors at four levels: i.e. societal-level factors; attributes of the survey design; attributes of the interviewer and respondent-interviewer interaction; and, characteristics of the

sample person. Examples of societal level factors are the survey-taking climate, the legitimacy of societal institutions or social cohesion. Attributes of the survey design can be the mode of the survey, the length of the questionnaire or the survey topic. Relevant personal characteristics are age, gender, income and health status of the sample person. These latter characteristics are also important for the interviewer. Factors related to respondent-interviewer interaction can include strategies of the interviewer to persuade the respondents or expectations of the interviewer and the interviewee.

These theories are tools for understanding the decision to participate (or not) in a survey, but do not specifically address the question of non-response bias. Most factors mentioned by Groves et al. (1992) may cause bias only in an indirect way, whereas non-response bias is strongest when respondents select themselves in relation to the perceived topic of the survey. Therefore, here the focus is on 'topic' as the main explanatory factor of non-response in our earlier survey on alcohol consumption. The two main aspects of the topic with regard to responsiveness are reported to be salience and social desirability. Heberlein and Baumgartner (1978) showed that salience (interested or not interested in the topic) has a strong influence on the response rate. When the topic of the survey is salient to the respondent, the costs of responding may be reduced. Martin (1994) verified that people's interest in the survey topic can have considerable impact on response rates: persons were almost twice as likely to participate if the topic was of high interest. Similarly, Dillman and Carley-Baxter (2000) showed that salience is a significant determinant of response rate. Others have shown that people who feel threatened by a topic behave less socially desirable or feel embarrassed/ashamed with respect to the survey topic and thus respond less (Gannon et al., 1971; Green, 1991). For example, socially acceptable behaviour, such as exercise and good nutrition may be frequently overreported, whereas undesirable behaviour such as smoking and drinking may be underreported (Warnecke et al., 1997; Bongers, 1998). Van Goor and Stuiver (1995) showed a pattern of overrepresentation of non-response in both 'extreme' categories of his outcome variable, i.e. the effectiveness of governmental organisations in policy making with regard to trailer camps. In his explanation about this pattern, the factor 'topic' plays an important role in that the very effective organisations lost interest in the topic of effectiveness and the less effective organisations did not respond because they felt threatened by the survey because of their poor performance. This latter study gave rise to a two-tailed pattern of non-response bias. We hypothesize that this pattern can also be applied to surveys on alcohol consumption. Two hypotheses are formulated: 1) non-response is high at the lowest end of the research variable, in our case the abstainers; and 2) non-response is also high at the upper end of the research variable, in our case the frequent excessive drinkers.

These hypotheses rely on the same reasoning as used by Van Goor and Stuiver (1995). Abstainers do not respond because they have no experience with alcohol and therefore may not be interested in it and/or may not see the relevance of their response. Frequent excessive drinkers do not respond because this group may perceive their amount of consumption as socially undesirable.

In the present study, the two hypotheses were tested by conducting a follow-up investigation of non-respondents to a mailed questionnaire on alcohol consumption, in order to compare them with respondents.

Subjects and Methods

Procedure and sample

The non-response follow-up study presented here is part of a main methodological study conducted in 1999 on alcohol surveys. For the main study, a random sample of 1000 persons aged 16-69 years was drawn from the municipal registry in Rotterdam, The Netherlands. These people received a mailed questionnaire on alcohol consumption. After two reminders (one of which included the questionnaire again) the overall response rate was 44%. In order to facilitate and intensify our follow-up study on non-response, we selected 25 postal areas in the centre of Rotterdam. This sample (the 'secondary' sample) consisted of 310 persons: 133 respondents of the primary survey (i.e. the mailed questionnaire) and 177 non-respondents. These persons were designated as primary respondents and primary non-respondents, respectively. The persons who did or did not respond to our non-response survey were called secondary respondents and secondary non-respondents, respectively.

Information about the primary respondents' alcohol use was derived from their answers to two alcohol questions in the mailed questionnaire, which were completed by all of them. All primary non-respondents were approached by means of house visits without prior notice. A maximum of five attempts was made to reach the right person at home. During the fieldwork, 22 of the 177 primary non-respondents appeared to have moved house or were absent for a longer period; these persons were deemed ineligible. Moreover, it proved impossible to contact another 26 primary non-respondents (they were not at home or did not open the door). Finally, having lost 48/177 persons to follow-up, 129 primary non-respondents were contacted during the house visits; this yielded a contact rate of 83% [129/(177-22)]. Of the several questions asked of the primary non-respondents, two were about alcohol use. Of the 129 primary non-respondents, 102 answered at least one alcohol question and 80 of these 102 secondary respondents answered both alcohol questions; this yielded a net response rate of 52% [80/(177-22)].

Both the primary respondents and secondary respondents were asked the same two alcohol questions.

Measures

Alcohol consumption was measured by the following two items: 'Do you ever drink alcoholic beverages?' and 'Do you ever drink six or more units of alcoholic beverages in one day?' The frequency of the alcohol consumption was assessed in the first instance by a nine point-scale (every day, more than 3 times a week, 2 or 3 times a week, once a week, 2 or 3 times a month, once a month, less than once a month, never drinking ≥ 6 units in one day and never drinking alcohol at all). Table 1 presents the classification used for alcohol consumption. This classification is based on a combination of those used by Garretsen (1983) and Wild (2001). The socio-demographic measures used for the analyses included: gender, age and nationality. Age was classified as a categorical variable (16-25, 26-35, 36-69 years). Nationality was divided into two categories: Dutch only and non-Dutch only.

Table 1. Classification of alcohol consumption

Classification of alcohol consumption	frequencies (based on 'drinking 6+ units in one day')
Frequent excessive drinker	every day more than 3 times/week 2 or 3 times/week
Occasional excessive drinker	once a week 2 or 3 times/month
Moderate drinker	once a month less than once a month never drinking 6+ units in one day
Abstainer	never drinking alcohol at all

Statistical analyses

The differences in self-reported alcohol consumption between the primary respondents and secondary respondents (i.e. primary non-respondents who responded to our non-response survey) were first analysed by bivariate cross tabulation. Next, the cross tabulation was stratified by gender, age and nationality. Statistical significance was estimated by the chi-square test. Each category of drinking frequency was tested separately. $P \leq 0.05$ was regarded as statistically significant.

Results

Table 2 gives information on the differences between primary and secondary respondents in their self-reported alcohol consumption. There were higher proportions of frequent excessive drinkers and of abstainers among secondary respondents than among primary respondents. The proportions of occasional excessive and moderate drinkers were greater among primary respondents than among secondary respondents. Thus, non-response is highest among the frequent excessive drinkers and abstainers. When stratified by gender (Table 3), a trend to overrepresentation of non-response of both abstainers and frequent excessive drinkers was seen amongst males and amongst females, but the differences for males were not significant.

Table 2. Distribution of alcohol consumption of primary respondents and secondary respondents (%)

Alcohol consumption	Primary respondents (n=133)	Secondary respondents (n=80)	P-value
Frequent excessive drinker	3.8	6.3	0.405
Occasional excessive drinker	18.8	16.3	0.638
Moderate drinker	50.4	25.0	0.000
Abstainer	27.1	52.5	0.000

n is number of subjects

Table 3. Distribution of alcohol consumption of primary respondents and secondary respondents by gender (%).

Gender	Male			Female		
Alcohol consumption	Primary respondents (n=67)	Secondary respondents (n=41)	P-value	Primary respondents (n=66)	Secondary respondents (n=39)	P-value
Frequent excessive drinker	6.0	7.3	0.783	1.5	5.1	0.283
Occasional excessive drinker	26.9	24.4	0.776	10.6	7.7	0.623
Moderate drinker	41.8	31.7	0.295	59.1	17.9	0.000
Abstainer	25.4	36.6	0.216	28.8	69.2	0.000

n is number of subjects

Table 4 shows that in the age group of 16-25 years there were relatively more abstainers among the secondary respondents, whereas there were relatively more frequent excessive drinkers among the primary respondents. An overrepresentation of non-response on both 'extreme' sides of the variable 'alcohol consumption' was seen in the age group 26-35 years. Relatively more secondary respondents aged 36-69 years were frequent excessive drinkers, occasional excessive drinkers and abstainers. Significant differences in alcohol consumption between primary and secondary respondents were found for moderate drinkers aged 26-35 years and 36-69 years and for abstainers aged 16-25 years and 26-35years. Table 5 gives data on the differences between primary and secondary respondents in their self-reported alcohol consumption after stratification by nationality. Dutch secondary respondents tended to be abstainers more often than the Dutch primary respondents, and were less likely to be moderate drinkers. Among the non-Dutch secondary respondents the abstention rate was also higher than among the non-Dutch primary respondents. (No frequent excessive drinkers were found among either the non-Dutch primary respondents or the non-Dutch secondary respondents.)

To summarize: this study showed significant underrepresentation of several categories in the primary survey: 1) abstainers were under-represented, whereas frequent excessive drinkers were not significantly underrepresented; 2) the underrepresentation of abstainers is greater for females than for males, greater for those aged <35 years, and is greater for the non-Dutch than for the Dutch group.

Table 4. Distribution of alcohol consumption of primary respondents and secondary respondents by age (%).

Age	16-25 years			26-35 years			36-69 years		
Alcohol consumption	Primary resp (n=34)	Secondary resp (n=23)	P-value	Primary resp (n=37)	Secondary resp (n=27)	P-value	Primary resp (n=62)	Secondary resp (n=30)	P-value
Frequent excessive drinker	8.8	0.0	0.143	0.0	3.7	0.238	3.2	13.3	0.066
Occasional excessive drinker	26.5	17.4	0.423	27.0	18.5	0.427	9.7	13.3	0.597
Moderate drinker	44.1	26.1	0.166	51.4	22.2	0.018	53.2	26.7	0.016
Abstainer	20.6	56.6	0.005	21.6	55.6	0.005	33.9	46.7	0.236

n is number of subjects

Table 5. Distribution of alcohol consumption of primary respondents and secondary respondents by nationality (%)

Nationality	Dutch-only			non-Dutch-only		
Alcohol consumption	Primary respondents (n=103)	Secondary respondents (n=29)	P-value	Primary respondents (n=50)	Secondary respondents (n=29)	P-value
Frequent excessive drinker	4.9	10.0	0.227	0.0	0.0	-
Occasional excessive drinker	18.4	22.0	0.604	17.2	6.9	0.227
Moderate drinker	58.3	32.0	0.002	24.1	10.3	0.164
Abstainer	18.4	36.0	0.017	58.6	82.8	0.043

n is number of subjects

Discussion

This study first explored whether there were any differences in alcohol consumption between primary respondents (i.e. respondents in our primary survey, the mailed questionnaire) and secondary respondents (i.e. primary non-respondents who responded only to our secondary survey, the non-response follow-up study). Then, the hypothesis was tested that non-response is over-represented in the extreme categories of our alcohol consumption variable. There was strong evidence for overrepresentation of non-response abstainers. The differences in the abstention rates between the primary and secondary respondents were also present when data were stratified by gender, age and nationality. However, the evidence for differences in the other extreme, i.e. the frequent excessive drinkers, was weak and inconsistent. This latter result may be due to the low prevalence of frequent excessive drinking in the study population, which led to very small numbers in the relevant categories.

Some other limitations of the study should be addressed. First, because the secondary non-response rate was about 52%, our conclusions rely on the assumption that the secondary non-response has the same effects on the estimates of alcohol consumption as the primary non-response. Another limitation concerns the representativeness of the sample. Only 25 postal areas in the city centre were included to obtain the sample for the follow-up study and the age group 36-69 years were under-represented in this second sample. Nevertheless, there were no significant differences in gender and nationality between the primary survey sample and the secondary survey sample. The study methods also differed. In the primary survey a mailed questionnaire was used, whereas the follow-up of non-respondents consisted of face-to-face interviews. It has been suggested that, due to greater anonymity, mail surveys encourage fuller report of sensitive topics than face-to-face interviews (De Leeuw, 1992; Gmel, 2000; Kraus & Augustin, 2001), whereas others found higher reported alcohol consumption in personal interviews (Cutler, 1988; Rehm&Arminger, 1996). Because of the conflicting reports, it is difficult to assess the impact of the different methods used in our alcohol surveys.

Despite these limitations, the results of this study reveal a serious non-response bias in our primary survey, which is directly related to the topic of the survey, namely alcohol consumption. This bias cannot be corrected by weighting data on the basis of socio-demographic variables because, within our subgroups, the same bias exists. Therefore, the aim should be to minimize non-response in a survey by, for example, developing more appealing survey materials. Furthermore, this study confirms the need for a thorough non-response follow-up study to evaluate non-response biases, as also emphasised by Hill and Roberts (1997) in the field of lifestyle surveys. Because of the time-consuming effort

required to contact the primary non-respondents it is advisable to restrict the follow-up study to a small subsample thus allowing more attention to be given to the individual subject.

References

Bongers IMB, Lemmens PHHM, Oers HAM and Tan van der FES (1997). Methode ter correctie voor vertekening van surveyresultaten ten gevolge van indirecte selectieve non-respons. Tijdschrift voor Sociale Gezondheidszorg 75: 122-128.

Bongers IBM (1998). Problem drinking among the general population: a public health issue? PhD thesis. Erasmus University Rotterdam.

Bradburn NM (1992). A response to the nonresponse problem. Public Opinion Quarterly 56: 391-397.

Caspar R (1992). Follow-up of non-respondents in 1990. In Survey measurement of drug use.Turner CF, Lessler JT and Gfroerer JC (eds). pp. 155-173. Rockville:National Institute on Drug Abuse,.

Cutler SF, Wallace PG. and Haines AP (1988). Assessing alcohol consumption in general practice patients - a comparison between questionnaire and interview. Alcohol 23: 441-450.

Dillman DA (2000). Introduction to tailored design. In Mail and internet surveys: the tailored design method. Dillman DA (eds). pp. 3-31. New York: John Wiley & Sons, Inc.

Dillman DA and Carley-Baxter LR. (2000). Structural determinants of mail survey response rates over a 12 year period, 1988-1999. Proceedings of the section on survey method. The American Statistical Association, Va. Alexandria. Published on web page http://survey.sesrc.wsu.edu/dillman.

Gannon MJ, Northern JC and Carrol Jr. SJ (1971) Characteristics of nonrespondents among workers . Journal of Applied Psychology 55: 586-588.

Garretsen HFL (1983). Probleemdrinken: prevalentiebepaling, beïnvloedende factoren en preventiemogelijkheden. PhD thesis. Katholieke Universiteit Nijmegen.

Gmel G. (2000). The effect of mode of data collection and of non-response on reported alcohol consumption: a split-sample study in Switzerland. Addiction 95: 123-134.

Goor van H and Stuiver B (1995). Succes en falen van beleid en non-respons: een empirisch onderzoek naar het terugzenden van schriftelijke enquêtes door Nederlandse gemeenten. Sociologische Gids 42: 388-406.

Green KE (1991). Reluctant respondents: differences between early, late and nonresponders to a mail survey. Journal of Experimental Education 56: 268-276.

Groves RM, Cialdini RB and Couper MP (1992). Understanding the decision to participate in a survey. Public Opinion Quarterly 56: 475-495.

Heberlein TA and Baumgartner RM (1978). Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. American Sociological Review 43: 447-462.

Heer de W (1999). International response trends: results of an international survey. Journal of Official Statistics 15: 129-142.

Hill A and Roberts J (1997) Non-response bias in a lifestyle survey. Journal of Public Health in Medicine 19: 203-207.

Jansen H. and Hak T (1999). Nonresponse to mail surveys in a lower-class urban area - a two stage exploration of access failure and refusal. Bulletin de Méthodologie Sociologique 62: 5-27.

Knibbe RA (1982). Probleemdrinken in Limburg. Rijkuniversiteit Limburg.

Kraus L and Augustin R (2001). Measuring alcohol consumption and alcohol-related problems: comparison of responses from self-administered and telephone interviews. Addiction 96: 459-471.

Leeuw de ED (1992). Data quality in mail, telephone, and face-to-face surveys. PhD thesis. Vrije Universiteit Amsterdam.

Lemmens PHHM, Tan ES and Knibbe RA (1988). Bias due to non-response in a Dutch survey on alcohol consumption. British Journal of Addiction 83: 1069-1077.

Martin CL (1994). The impact of topic interest on mail survey response behaviour. Journal of Market Research Society 36: 327-338.

Pernanen K (1974). Validity of survey data on alcohol use. In Research Advances in alcohol and drug problems. Gibbens R, Israël Y, Kalant H, Popham R, Schmidt W and Smart RG (eds). pp. 355-374. New York: John Wiley & Sons, Inc.

Presser S and Traugott M (1992). Little white lies and social science models: correlated response errors in a panel study of voting. Public Opinion Quarterly 56: 77-86.

Rehm J and Arminger G (1996). Alcohol consumption in Switzerland 1987-93: adjusting for differential effects of assessment techniques on the analysis of trends. Addiction 91: 1335-1344.

Sosdian CP and Sharp LM (1980). Non-response in mail surveys: access failure or respondent resistance. Public Opinion Quarterly 44: *396-402.*

Warnecke RB, Johnson TP, Chávez N, Sudman S, O'Rourke DP, Lacey L and Horm J (1997). Improving question wording in surveys of culturally diverse populations. Annals Epidemiology 7: 334-342.

Wild TC, Cunningham J and Adlaf E (2001). Nonresponse in a follow-up to a representative telephone survey of adult drinkers. Journal of Studies on Alcohol 62: 257-261.

Wild TC (2001). Effects of personal drinking on perceived norms for alcohol consumption, approval of others' drinking, and criteria for defining alcohol problems. Paper presented at the 27th Annual Alcohol Epidemiology Symposium of the Kettil Bruun Society for Social and Epidemiological Research on Alcohol, May 28 - June 1, 2001, Toronto, Canada.

Chapter 3

Estimating non-response bias in a survey on alcohol consumption: comparison of response waves

Viviënne M.H.C.J. Lahaut, Harrie A.M. Jansen, Dike van de Mheen, Henk F.L. Garretsen, Jacqueline E.E. Verdurmen, Ad van Dijk. Estimating non-response bias in a survey on alcohol consumption of response waves. Alcohol & Alcoholism 2003; 38 (2): 128-134.

Abstract

Aims
According to 'the continuum of resistance model' late respondents can be used as a proxy for non-respondents in estimating non-response bias. In the present study the validity of this model will be explored and tested in three surveys on alcohol consumption.

Methods
The three studies collected their data by means of mailed questionnaires on alcohol consumption whereby two studies also performed a non-response follow-up.

Results
Comparisons of early respondents, late respondents and non-respondents in one study showed some support for 'the continuum of resistance model' although another study could not confirm this result. Comparison of alcohol consumption between three time response groups showed no significant linear pattern of differences between response waves.

Conclusions
The hypothesis that late respondents are more similar to non-respondents than early respondents could not be confirmed or rejected. Repeated mailings are effective in obtaining a greater sample size, but seem not effective in improving the representativeness of alcohol consumption surveys.

Introduction

Survey researchers are constantly confronted with the problem of non-response. Moreover, with a low response rate the validity of survey results may well be questioned. However, a low response rate does not necessarily mean that the results are biased. Non-response bias occurs when respondents and non-respondents differ in outcome variable(s), in which case the population parameters of these variable(s) can be over- or under-estimated.

Methods are available for survey researchers to deal with the problem of non-response. One way is to build-in strategies during survey development and data collection in order to positively influence the response rate. Such strategies include financial incentives, repeated mailings, and an appealing survey design. Dillman (2000) describes these approaches in his so-called Tailored Design Method: a method to maximise both quantity and quality of responses. These approaches are sometimes successful: but biased estimates due to non-response may still remain. Non-response bias can be estimated and/or corrected for in various ways. An indirect approach is to weight cases, whereby weights are allocated to various substrata, which are mostly defined by background variables; this approach is justified if the background variables are strongly related to outcome variable(s). Direct approaches include collecting valid information from objective sources, or conducting a non-response follow-up to collect data on outcome variable(s) of non-respondents to get insight into differences between respondents and non-respondents. However, follow-up studies tend to be costly and time-consuming; moreover, it is often difficult to contact initial non-respondents and get their participation again.

An alternative widely-used approach is to estimate the non-response bias by comparing early and late respondents; a late respondent is then used as a proxy for a non-respondent. The underlying assumption behind this approach is that every subject in the study population has a position on the response continuum that ranges from 'will never respond' to 'will always respond'. Non-respondents will be concentrated on the side of 'will never respond'. Subjects who require more reminders before they participate would have been non-respondents if the data collection had stopped earlier. Therefore, late respondents most resemble non-respondents. This assumption has been called 'the continuum of resistance model' (Lin and Schaeffer 1995; Voogt, Saris et al. 1998). Questions arise, however, about the validity of this 'continuum of resistance model'. If it appears valid, then it is justified to use late respondents as proxies for non-respondents and then repeated mailings will help lower the degree of possible non-response bias. However, Table 1 shows that a recent literature review does not provide a consistent answer to these questions.

Some researchers found no support for the assumption that late respondents can be used as a proxy for non-respondents (Hébert, Bravo et al. 1996; Larroque, Kaminski et al. 1999), whereas others found that non-response bias can be estimated by analysing late respondents (Dalecki, Whitehead et al. 1993; Lin and Schaeffer 1995; Etter and Perneger 1997; Ullman and Newcomb 1998). In several surveys refusal conversions or repeated mailings appeared to be useful in lowering the degree of biased estimates. Some researchers found no significant differences between different time groups of respondents which means that the number of mailings or refusal conversions could be reduced without affecting the accuracy of survey estimates (Green 1991; Dalecki, Whitehead et al. 1993; Hébert, Bravo et al. 1996; Etter and Perneger 1997; Trinkoff and Storr 1997; Ullman and Newcomb 1998; Voogt, Saris et al. 1998; Larroque, Kaminski et al. 1999; van Goor and Stuiver 1998; Woodruff, Conway et al. 2000; Schmidt and Fletcher 2001).

Table 1 Literature review

Study	*Goal of study*	*Method*	*Survey Topic*	*Analysis: comparing groups*	*Conclusions*
Kellerman and Herold (2001)	To understand role of non-response bias	Five mailed surveys	Physician-specific topic	First- versus second-mailing respondents	Non-respondents are similar to respondents.
Schmidt and Fletcher (2001)	To demonstrate the importance of persistence in refusal conversions	Four random digit dial telephone surveys	Telecommunication, Beach use, Storm water quality and Driver satisfaction	Early versus late respondents	Refusal conversions are necessary to lower the degree of non-response bias
Woodruff et al (2000)	To examine non-response bias; To examine results of response-enhancing strategies	Mailed survey	Smoking	Initial respondents versus reluctant respondents versus non-respondents	Under-representation of hard-core smokers due to non-response; Response-enhancing strategies lower non-response bias
Larroque et al (1999)	To investigate non-response bias; To investigate the results of reminders/repeated mailings	Mailed survey	Child's temperament	Groups of respondents; Respondents versus non-respondents	Partly support for early-late comparison as proxy for non-response bias; Repeated mailings diminish bias
Ullman and Newcomb (1998)	To develop models that differentiate eager-, reluctant-, and non-responding participants.	Mailed survey	Unknown	Eager respondents versus reluctant respondents versus non-respondents	Preliminary evidence for differentiating participants based on personality characteristics.
van Goor and Stuiver (1998)	To examine distributional and substantive bias	Mailed survey	National caravan sites policy	Respondents (three waves) versus non-respondents	Higher response levels led to better results. At final response rate distributions and relationships are still biased.
Voogt et al (1998)	To investigate differences between respondents and non-respondents	Telephone survey	Election/Political interest	Groups of respondents	A systematic distortion due to non-response
Etter and Perneger (1997)	To investigate non-response bias	Mailed survey	Insurance plans	Early versus late respondents; Respondents versus non-respondents	Groups of respondents can differ. Late respondents are not similar to non-respondents.

Continuation of Table 1 Literature review

Study	*Goal of study*	*Method*	*Survey Topic*	*Analysis: comparing groups*	*Conclusions*
Trinkoff and Storr (1997)	To analyse non-response patterns	Mailed survey	Nurses worklife and health	Early versus late respondents; Groups of respondents	No large amount of non-response bias
Hébert et al (1996)	To study refusal bias	Mailed survey; Face-to-face interviews	Health risk factors	Early respondents versus late respondents versus non-respondents	Important bias due to postal nonresponse; Late espondents can be used as a proxy for non-respondents.
Lin and Schaeffer (1995)	To evaluate two models estimating non-response bias: 'continuum of resistance model' and 'classes model'	Telephone interviews; Mailed survey	Child support	Groups of respondents; Late respondents versus non-respondents	No support for both models.
Dalecki et al (1993)	To investigate non-response bias	Mailed survey	Wetland preservation	Early versus late respondents; Late respondents versus non-respondents	No support for early-late comparison as proxy for non-response bias.
Green (1991)	To investigate non-response bias	Mailed survey; Telephone interviews	Applications of research methods and findings to classroom teaching	Respondents versus non-respondents; Groups of respondents	Non-response bias due to delayed response and non-response.

Only three studies in our literature review had main survey topics that addressed substance use (Trinkoff and Storr 1997; Ullman and Newcomb 1998,Woodruff, Conway et al, 2000). Trinkoff and Storr (1997) did not question the justification of using late respondents as non-respondents, but they did investigate differences in substance use rates by mailing. Ullman and Newcomb (1998) compared reluctant respondents with non-respondents, early respondents with reluctant respondents and with respondents who participated at different time intervals; to estimate substance use of non-respondents they used data of earlier mailings in which these non-respondents had participated. Woodruff and colleagues (2000) also compared initial-, reluctant- and non-respondents, but focused more on the effects of incentives (such as financial incentives). All three studies drew their sample from specific subpopulations.

In a previous study, by means of a non-response follow-up (Lahaut, Jansen et al, in press) we investigated whether respondents and non-respondents differed in their alcohol consumption; the results showed a significant higher abstention rate and also a higher proportion of frequent excessive drinkers among non-respondents.

In the present study we first investigate whether 'the continuum of resistance model' fits the data of this follow-up study; this model is also tested in a data set from a larger Dutch survey on alcohol consumption. Then, we investigate whether repeated mailings are worthwhile to collect more representative data on our outcome variable, alcohol consumption. For this, two larger samples from the general population were used. These studies were conducted in the same time period and used similar questions about alcohol consumption. Specifically, we aimed to answer the following questions: 1) Are late respondents more similar to non-respondents than early respondents? and 2) Are there differences in alcohol consumption between response waves?

Methods

Procedure and samples

For this study three data sets were used. The data of all three studies were collected by means of mailed questionnaires on alcohol consumption. There were three mailing periods: Wave 1) the first mailing of the questionnaire accompanied by an explanatory letter; Wave 2) two weeks later a reminder was sent to those who had not yet responded; Wave 3) subjects who had not responded in the second wave, received another reminder with another copy of the same questionnaire. The questionnaire in each study used similar questions on alcohol consumption. Subjects, aged 16-69 years, were approached by mail in April/May 1999.

The sample of the first data set (called the small-scale Rotterdam survey) consisted of 310 persons, who were living in 25 postal areas in the centre of Rotterdam. This study performed a non-response follow-up study (wave 4) on a sample of 177 subjects who were approached mainly by means of house visits without prior notice. The random sample of the second study (called the Utrecht survey) consisted of 5229 persons drawn from the municipal registry. The study in Utrecht also included a non-response follow-up study (wave 4). A random sample, stratified for age, was drawn from all non-respondents and consisted of 662 persons who were approached by telephone. The random sample of the third study (called the large-scale Rotterdam survey) was drawn from the municipal registry in Rotterdam and consisted of 3226 persons. The large-scale Rotterdam survey did not perform a non-response follow-up study.

Measures

The outcome variable in the three studies was alcohol consumption. Alcohol consumption was measured by four questions: 1) Have you drunk any alcoholic beverages in the past year? 2) How many units of alcoholic beverages do you drink on average in a typical week? 3) Please indicate for each day in the previous week how many units of alcoholic beverages you have drunk and; 4) Have you ever drunk six or more units of alcoholic beverages on one day in the past six months?

Non-respondents of the small-scale Rotterdam survey and the Utrecht survey were asked several questions about alcohol consumption. In the two non-response follow-up surveys the interviewer asked non-respondents whether they had drunk any alcoholic beverages in the past year. If the answer was yes, these participants were asked an additional question on alcohol consumption. In the follow-up of the small-scale Rotterdam survey this question was: Have you ever drunk six or more units of alcoholic beverages on one day in the past six months? The follow-up of the Utrecht survey used the question: Please indicate for each day in the previous week how many units of alcoholic beverages you consumed on each day.

We constructed several variables that provided information on the amount of alcohol consumed. The variable 'drinking alcohol' made a distinction between abstainers and drinkers. According to the total alcohol intake in a typical week the frequencies of drinkers were categorised as: 1-14 units/week, 15-28 units/week or ≥ 29 units/week. This variable was called 'Total alcohol consumption in a typical week'. Based on the weekly recall question the frequencies of the total alcohol intake consumed in the previous week were calculated (variable 'Total alcohol consumption in previous week'). The question whether the subjects had ever consumed six or more units of alcoholic beverages on one occasion and with

what frequency (never; 1-5 times/half year; 1-3 times/month; 1-2 times/week; ≥ 3 times/week) was used for constructing the variable 'frequency of excessive drinking'.

Analysis
Differences in the distribution of alcohol consumption between response waves were analysed by cross-tabulation. Statistical significance was estimated by Chi-square tests. A p-value ≤ 0.05 was considered significant.

Results

Response rates
The total gross response rate of the small-scale Rotterdam survey was 42.9% (n=133). The first response wave yielded a response rate of 19.4%, the second response wave an additional 7.4% and the third response wave yielded an additional response of 16.1%. Of the 177 non-respondents in the small-scale Rotterdam's non-response follow-up study, 22 persons had moved or were absent for a longer period. Finally, we reached 129 non-respondents of which 102 persons answered at least one question on alcohol consumption. The contact rate of the follow-up is 83.2% (129/155).

The survey in Utrecht had a total gross response rate of 55.5% (n=2902). The response percentages of the three waves were 32.9%, 10.7% and 11.6%, respectively. The response time of 19 persons was unknown (response percentage of 0.3%). In the non-response follow-up study of Utrecht 370 correct telephone numbers of 662 selected non-respondents were found. The researchers contacted 254 non-respondents. During these contact attempts 133 persons answered at least one question on alcohol. The contact rate of the follow-up is 38.4% (254/662).

The total gross response rate of the large-scale Rotterdam survey was 50.5% (n=1630). The first wave yielded an additional 29.0%, the second wave an additional 9.5%, and the third response wave yielded an additional response of 12.1%.

Differences between early respondents, late respondents and non-respondents
Table 2 gives data on differences in alcohol consumption between early (first wave) respondents, late (third wave) respondents and non-respondents (fourth wave respondents) from the small scale Rotterdam survey. A significantly higher proportion of abstainers was found among non-respondents than among early respondents. There were no significant differences between early respondents and non-respondents in frequencies of excessive drinking. The comparison of

alcohol consumption between late respondents and non-respondents showed no significant differences, except a significant difference in excessive drinking with frequency 1-5 times/half year. There was a higher proportion of excessive drinkers with frequency 1-5 times/half year among late respondents than among non-respondents. The proportion of abstainers was also higher among non-respondents than among late respondents (third wave respondents) although the difference between these two groups was not significant.

Table 2. Comparison of alcohol consumption between early respondents, late respondents and non-respondents in the small-scale Rotterdam survey (%).

Alcohol consumption	Early respondents (wave 1), late respondents (wave 3) and non-respondents (wave 4)					
	Wave 1 (n=60)	Wave 4 (n=89)	p-value	Wave 3 (n=50)	Wave 4 (n=89)	p-value
Drinking alcohol: abstainers	21.7	47.2	0.002	36.0	47.2	0.201
	Wave 1 (n=45)	Wave 4 (n=38)	p-value	Wave 3 (n=30)	Wave 4 (n=38)	p-value
Frequencies of excessive drinking:						
Never	44.5	39.5	0.668	36.6	39.5	0.813
1-5 times/half year	17.8	10.5	0.349	30.0	10.5	0.043
1-3 times/month	13.3	15.8	0.751	6.7	15.8	0.246
1-2 times/week	20.0	23.7	0.685	20.0	23.7	0.716
≥ 3 times/week	4.4	10.5	0.286	6.7	10.5	0.577

n= number of subjects

Table 3 gives data on alcohol consumption of early and late respondents from the Utrecht survey compared with non-respondents. These two comparisons showed no significant differences. The comparison of abstention rate between early respondents and non-respondents showed an almost significant difference (p-value 0.057). The abstention rate in the Utrecht survey was higher among non-respondents than among early respondents.

Table 3. Comparison of alcohol consumption between early respondents versus non-respondents and late respondents versus non-respondents in the Utrecht survey (%).

Alcohol consumption	Early respondents (wave 1) versus non-respondents (wave 4)			Late respondents (wave 3) versus non-respondents (wave 4)		
	Wave 1 (n=1650)	Wave 4 (n=133)	p-value	Wave 3 (n=562)	Wave 4 (n=133)	p-value
Drinking alcohol: Abstainers	15.5	21.8	0.057	18.1	21.8	0.332
	Wave 1 (n=1180)	Wave 4 (n=83)	p-value	Wave 3 (n=389)	Wave 4 (n=83)	p-value
Total alcohol consumption in previous week:						
0 units	11.9	17.0	0.137	12.2	17.0	0.197
1-14 units	58.6	57.0	0.757	62.3	57.0	0.325
15-28 units	20.9	19.0	0.652	17.8	19.0	0.784
≥ 29 units	8.6	7.0	0.584	7.7	7.0	0.818

n= number of subjects

Differences in alcohol consumption between the three response waves

Table 4 gives data on the comparison of alcohol consumption between the first, second and third wave of respondents in the large-scale survey in Rotterdam and in the Utrecht survey.

In both surveys there were no significant differences between the three response waves, except for the proportion of abstainers. The proportion of abstainers in the Utrecht survey was the highest among the second wave respondents and the lowest proportion was found in the first response wave. The pattern of differences between waves in alcohol consumption was not consistent across both surveys.

Table 4. Differences in alcohol consumption by response waves in the large-scale Rotterdam survey and in the Utrecht survey (%)

Alcohol consumption	Large-scale Rotterdam survey			p-value	Utrecht survey			p-value
	Wave 1 (n=924)	Wave 2 (n=301)	Wave 3 (n=382)		Wave 1 (n=1662)	Wave 2 (n=544)	Wave 3 (n=588)	
Drinking alcohol: Abstainers	26.2	24.6	24.9	0.803	17.6	25.0	20.6	0.001
	Wave 1 (n=595)	Wave 2 (n=194)	Wave 3 (n=252)		Wave 1 (n=1288)	Wave 2 (n=387)	Wave 3 (n=445)	
Total alcohol consumption in a typical week:								
1-14 units	75.6	72.2	74.6	0.627	71.3	70.0	73.0	0.436
15-28 units	16.5	17.5	15.5	0.845	18.8	22.5	17.3	0.165
≥ 29 units	7.9	10.3	9.9	0.463	9.9	7.5	9.7	0.354
	Wave 1 (n=666)	Wave 2 (n=220)	Wave 3 (n=281)	p-value	Wave 1 (n=1288)	Wave 2 (n=387)	Wave 3 (n=445)	p-value
Frequency of excessive drinking:								
Never	46.7	45.0	48.0	0.795	36.3	34.4	37.0	0.684
1-5 times/half year	28.5	25.5	24.6	0.385	28.9	26.7	31.0	0.391
1-3 times/month	12.3	15.9	12.8	0.385	17.5	18.6	15.8	0.529
1-2 times/week	8.3	9.1	9.6	0.781	11.7	15.2	11.3	0.129
≥ 3 times/week	4.2	4.5	5.0	0.867	5.6	5.1	4.9	0.844

n = number of subjects

Conclusions

In this study we aimed to answer the questions: 1) whether late respondents are more similar to non-respondents than early respondents; and 2) whether there are differences in alcohol consumption between response waves. For the first question, we can neither confirm nor reject the hypothesis that late respondents are more similar to non-respondents than early respondents. In the small-scale Rotterdam survey the abstention rate of late respondents was significantly more similar to non-respondents than early respondents. Thus, this finding supports 'the continuum of resistance model'. However, in the Utrecht survey this result was not confirmed. For the second research question neither of the surveys (i.e. the large-scale Rotterdam survey and Utrecht survey) showed a significant linear pattern of differences in alcohol consumption between response waves. We could not confirm the 'continuum of resistance model'. The results suggest that repeated mailings are not necessary in order to collect more representative alcohol consumption estimates. The results do indicate, however, that repeated mailings are effective in obtaining a greater sample size. The two reminders in the small-scale Rotterdam survey are responsible for 23.5% of the total response rate (42.9%), compared with 22.3% of the total response rate (55.5%) in Utrecht. The reminders in the large-scale Rotterdam survey together contribute an additional 21.6% to the total response rate of 50.5%. In both surveys, especially the mailing in the third wave yielded a relatively high additional percentage to the total response rate.

Discussion

Some limitations in our study have to be acknowledged. One limitation is the mode of data collection. The mode of the initial surveys (i.e. the mailed questionnaire) differed from that used in the follow-up of non-respondents (i.e. face-to-face interview/telephone interviews). Additionally, both non-response follow-ups (the small-scale Rotterdam survey and the Utrecht survey) used different modes: i.e. mainly face-to-face interviews versus telephone interviews. Because of the conflicting reports in literature, it is difficult to assess the impact of the different methods used on self-reported alcohol consumption, and on responsiveness of non-respondents in the follow-up (Rehm and Arminger 1996; Kraus and Augustin 2001). There was a difference in contact rates between both our follow-up studies (83.2% vs 38.4%). Possible explanations for this difference may be the different modes used for each non-response follow-up and/or the number of contact attempts.

Besides the problem of using different modes, another limitation is the analysis itself. Most non-response follow-up studies suffer from low response

rates. This is also true in our case; especially with regard to the Utrecht non-response follow-up. The response rates for questions on alcohol in the non-response follow-up study in Rotterdam and Utrecht were 79.1 % (102/129) and 52.4 % (133/254), respectively. This means that conclusions as to whether late respondents are more similar to non-respondents than early respondents rely on the assumption that the remaining non-respondents have the same pattern of alcohol consumption as fourth wave respondents. Due to a low response rate the results of the Utrecht non-response follow-up strongly relies on this assumption; however, the group of remaining non-respondents may become an increasingly deviant category. Although not in the field of substance use, the study by Van Goor and Stuiver (1998) showed that the group of hard-core non-respondents differed more and more from the other response groups.

We are able to compare our results with three other studies in the field of alcohol and drugs. Our results showed no significant differences in alcohol consumption between first, second, and third wave respondents. Similarly, Trinkoff and Storr (1997) found no significant differences in substance use rates by mailing. In the study of Ullman and Newcomb (1998) the pattern of results was complex and indicated few differences in behavioural low social conformity variables (i.e. frequency of alcohol use, cigarettes and marijuana) between early and reluctant respondents and between different groups of reluctant respondents. The results of Woodruff and colleagues (2000) showed significant differences in baseline smoking between 'on time' and 'reluctant' respondents. Smoking tobacco and marijuana use are similar to alcohol-sensitive survey topics. Ullman and Newcomb (1998) and Woodruff, Conway et al (2000) included financial incentives in their reminders, which meant that they did not measure the relationship between response and substance use directly. Our results also suggest that there are no significant differences in alcohol consumption between late respondents and non-respondents, although the group differences in the Rotterdam survey were not small. Reluctant respondents and non-respondents in the study of Woodruff, Conway et al (2000) showed considerable similarity in their baseline smoking characteristics.

The answer as to whether the alcohol consumption distribution of late respondents can be used as a proxy for the distribution of non-respondents remains ambiguous; our results are mixed. Thus, more studies are needed to establish whether late respondents can be used as proxies for non-respondents. Also, our results show no linear patterns of differences in alcohol consumption between response waves. The question arises whether it may be more economical to draw a larger initial sample rather than sending reminders in order to enhance the response rate.

To estimate non-response bias it is advisable to do an intensive non-response follow-up on a small representative sample of non-respondents which

allows more time for each individual subject, rather than an extensive follow-up on a large sample of non-respondents.

References

Dalecki MG, Whitehead JC and Blomquist GC (1993). Sample non-response bias and aggregate benefits in contingent valuation: an examination of early, late and non-respondents. Journal of Environmental Management 38: 133-143.

Dillman DA (2000). Mail and internet surveys: the tailored design method. New York: Wiley.

Etter J-F and Perneger TV (1997). Analysis of non-response bias in a mailed health survey. Journal of Clinical Epidemiology 50(10): 1123-1128.

Green KE (1991). Reluctant respondents: differences between early, late, and nonresponders to a mail survey. Journal of Experimental Education 59: 268-276.

Hébert RG, Bravo N, Korner-Bitensky N and Voyer L (1996). Refusal and information bias associated with postal questionnaires and face-to-face interviews in very elderly subjects. Journal of Clinical Epidemiology 49(3): 373-381.

Kellerman SE and Herold J (2001) Physician response to surveys: a review of the literature. American Journal of Preventive Medicine 20(1): 61-67.

Kraus L and Augustin R (2001). Measuring alcohol consumption and alcohol-related problems: comparison of responses from self-administered questionnaires and telephone interviews. Addiction 96: 459-471.

Lahaut VMHCJ, Jansen HAM, Mheen van de D and Garretsen HFL. Non-response bias in a sample survey on alcohol consumption. Alcohol & Alcoholism 37(3): 256-260.

Larroque B, Kaminski M, Bouvier-Colle MH and Hollebecque V (1999). Participation in a mail survey: role of repeated mailings and characteristics of nonrespondents among recent mothers. Paediatric and Perinatal Epidemiology 13**:** 218-233.

Lin IF and Schaeffer NC (1995). Using survey participants to estimate the impact of nonparticipation. Public Opinion Quarterly 59: 236-258.

Rehm J and Arminger G (1996). Alcohol consumption in Switzerland 1987-93: adjusting for differential effects of assessment techniques on the analysis of trends. Addiction 91(9): 1335-1344.

Schmidt D and Fletcher J (2001). Measuring response bias in survey research: an analysis of age characteristics of early respondents and resistors. Chico, CA, California State University, Survey Research Center**:** 1-13.

Trinkoff AM and Storr CL (1997). Collecting substance use data with an anonymous mailed survey. Drug and Alcohol Dependence 48: 1-8.

Ullman JB and Newcomb MD (1998). Eager, reluctant, and nonresponders to a mailed longitudinal survey: attitudinal and substance use characteristics differentiate respondents. Journal of Applied Social Psychology 28(4): 357-375.

Goor van H and Stuiver B (1998). A wave-analysis of distributional bias, substantive bias and data quality in a mail survey among Dutch municipalities. Acta Politica 33(2): 179-196.

Voogt RJJ, Saris SE and Niemöller B. (1998). Non-response, and the gulf between the public and politicians. Acta Politica 33(3): 250-280.

Woodruff SI, Conway TL and Edwards CC (2000). Increasing response rates to a smoking survey for U.S. Navy enlisted women. Evaluation and the Health Professions 23(2): 172-181.

Chapter 4

Effects of survey sponsorship, questionnaire length, topic label and layout on response rate in mail alcohol surveys

Viviënne M.H.C.J. Lahaut, Harrie A.M. Jansen, Henk F.L. Garretsen, Dike van de Mheen. Effects of survey sponsorship, questionnaire length, topic label and design on response rate. *Submitted*

Abstract

Background
In order to increase response on a survey on risky habits we used Dillman's application of social exchange theory to manipulate survey characteristics.

Methods
Four characteristics of a mailed survey (sponsorship, questionnaire length, topic label and layout) were manipulated, resulting in twelve experimental conditions in order to investigate response effects.

Results
The response rates to the 12 questionnaires ranged from 38% to 63%. The municipal health service has the highest likelihood of response compared to a university and an addiction research institute. Other survey characteristics did not show a significant relationship.

Conclusions
Manipulation of survey factors has some effects on the response rates. Of the four factors, only sponsorship has a significant relationship with response rate.

Introduction

Despite the introduction of new technologies in population surveys, like computer-assisted telephone interviews (CATI), computer-assisted personal interviews (CAPI) and the internet, the mail survey still has some major advantages as a medium for survey research. It is a relatively cheap method, but also relatively insensitive to selection bias by inaccessibilty of respondents and to reporting bias by social desirability (de Leeuw, 1992). As with all other methods, however, mail surveys can suffer from a low response rate, leading to a smaller final number of observations and thus to less accurate population estimates and also enlarging the risk of selection biases. Therefore, it is important to minimise non-response as much as possible.

During the last two decades we (i.e. the Addiciton Research Institute Rotterdam) did several surveys on risky habits, with alcohol consumption and alcohol-related problems as most important subject. There was a trend of a decrease in response rate: whereas the 1980 survey had a response rate of 70%, the very similar 1994 survey got only 45%. This decrease may be a reflection of the general trend of decreasing response rates, but it may also be the effect of specific characteristics of the the 1994 survey. In order to increase response on the 1999 survey on risky habits we used Dillman's application of social exchange theory (Dillman, 2000) as a theoretical guideline for the choice of survey characteristics to be manipulated and to generate hypothesis to be tested. According to Dillman's application of social exhange theory, survey participation is determined by value assessments based on rewards, costs and trusts. For instance a phrase as "we very much appreciate your help" or monetary incentives can be a reward value. Costs are what a subject gives up or spends to obtain the rewards: e.g. time or physical and mental effort to fill in the questionnaire.Trust is the expectation that (in the long run) the rewards of participating will outweigh the costs. Respondents valuation of the survey sponsor may be a relevant factor in gaining trust.

In line with this theory we distinguish three major dimensions that may be involved in decisions to participate:

1. The subject matter. Here are two factors important: saliency and sensitivity.

Saliency: it is well known that people are not easily responding to surveys on habits which they do not have. For someone, an interesting survey topic can be a reward value.

Sensitivity: People who regard the survey topic as 'sensitive' are not very willing to participate. Sensitivity is for them a cost value. This may be one of the reasons why frequent excessive drinkers are low responders to surveys on drinking.

Sensitivity not only lowers the response rate but it also causes a selection bias (Lahaut, Jansen, van de Mheen & Garretsen, 2002).
In this present study we manipulated the survey topic by entitlement of the questionnaire: 'health survey' versus 'survey on alcohol consumption'. We expect higher response on 'health survey' than on 'survey on alcohol consumption'. Firstly because 'health' is generally valued as one of the most important aspects of life, secondly because 'alcohol' may be irrelevant to many people (especially to abstainers) and too sensitive to others (heavy drinkers or illegitimate drinkers –by age or religion).

2. The perceived social use. Apart from the subject matter, the perceived social use may depend for a large part to the prestige of the sponsor (Dillman, 2000). People are more likely to comply with a request if it comes from an authorative source, one who is legitimate to make such request (Cialdini, 1984). These differences in trust in sponsors may be an explanation why government-sponsored surveys achieved higher reponse rates than those sponsored by marketing research firms (Heberlein & Baumgartner, 1978). Because our survey concerns a health-related topic, we hypothesized that a well-known governmental health agency would yield a higher response rate than the university sponsorship and than an addiction research institute and the university would yield a higher response the addiction research institute.

3. The questionnaire. Here we may distinguish two aspects: a) The questionnaire length, which determines the perceived effort to be done. In general, a longer questionnaire decreases response rates because of higher cost values. However, shortening an already brief questionnaire may not improve response. There may be a limit to which shortening the questionnaire will improve response (Dillman, 2000; Lund & Gram, 1998). We hypothesize that the shorter questionnaire will have the highest response rate. b) The psycho-physical attractiveness of the questionnaire which depends on aspects like colours, layout, visibility of routing, wording etcetera. Response rate can be improved by a respondent-friendly design which is defined as a form that is easy to complete and allows respondents to feel neutral or positive about the form itself (Dillman, Sinclair & Clark, 1993). This form lowers the cost value. We hypothesize that a 'new-style' layout will yield a higher response rate.
In this study, all three of the aforementioned dimensions were manipulated in order to investigate response effects.

Methods

Experimental Design
In our aim to increase response on a survey on risky habits, four survey factors were manipulated: sponsorship, length, topic and design. We use the 1994 survey

as a benchmark.The features of this survey are: the questionnaire has 19 pages with 54 questions on health, alcohol, drugs and socio-demographics; the survey is conducted by a well-known health-related government agency, i.e. the local municipal health service, and the topic of the survey is presented as a health survey. In addition, two other sponsors were chosen to investigate the response effects of sponsorship: the local university and a research agency, the addiction research institute. The length of the existing questionnaire was also manipulated by shortening the questionnaire to 5 pages with 22 questions on health, alcohol, drugs and socio-demographics. Another short questionnaire was developed but with alcohol consumption presented as topic; this latter questionnaire had 3 pages with 7 questions on alcohol consumption and socio-demographics. Finally, this latter questionnaire on alcohol consumption also has been presented in two layout designs: the 'new style' questionnaire is a folded form and has more background fields and more symbolic language than the old-style stapled version.

To summarise: four different types of questionnaires were constructed: 1) long health survey; 2) short health survey; 3) short survey on alcohol consumption, old design; 4) short survey on alcohol consumption, new design. Each type of questionnaire was pesented by three sponsors: the Municipal Health Service, the University and the Addiction Research Institute. Therefore, 12 different experimental conditions were constructed (Table 1).

Table 1 Experimental survey design

	Sponsorship			
	Municipal Health Service	Addiction Research Institute	University	
Questionnaire type	N	N	N	Total N
Long Health Survey, old layout	3000 (existing survey)	125	125	3250
Short Health Survey, old layout	62	62	63	187
Short Survey on Alcohol consumption, old layout	63	63	62	188
Short Survey on Alcohol consumption, new layout	125	125	125	375
Total N	3250	375	375	4000

Procedure and Sample
Two random samples were drawn from the Rotterdam municipal registry. One sample consisted of 3000 subjects, aged 16-69 years; this sample received by mail the benchmark survey (i.e. long health survey, Municipal Health Service as sponsor). The second sample consisted of 1000 subjects, aged 16-69 years and this sample received (in subgroups) the other 11 questionnaires by mail. These latter subjects were randomly assigned to receive one questionnaire (Table 1). The long health survey was mailed to 250 subjects, the short health survey to 187 subjects, the short survey on alcohol consumption with old design to 188 subjects, and the short survey on alcohol consumption with new design to 375 subjects. The subjects for each type of questionnaire were equally allocated to one of the three sponsors. The reason for the difference in the distribution of the number of subjects was that the questionnaires were also used for other research purposes. The subjects in both samples were approached by mail in the same time period. For both samples, the same mailing procedure was applied: questionnaires were accompanied by an explanatory letter, two mailed reminders were sent, whereby the last reminder included another copy of the questionnaire.

Statistical Analysis
Response rates were calculated for each type of questionnaire. Logistic regression analysis was performed to calculate the odds ratios of response with 95% confidence intervals for the four factors: questionnaire sponsorship, length, topic and layout, corrected for the following variables: gender, age group (16-25; 26-35; 36-45; 46-69 years), civil status (married/registered partnership; unmarried; divorced/widowed) and experimental factor. The variable 'experimental factor' consisted of sponsor (municipal health service; addiction research institute; university), length (short/long), topic (health; alcohol), and layout (old layout; new layout).

Results

Table 2 gives response rates for each questionnaire type. Of the three sponsors, the municipal health service had the highest response on both the long and short health survey. The highest response rate on the short survey on alcohol consumption (old layout) was found for the municipal health service and the addiction research institute (both 50.8%). The municipal health service and the university had the highest response (both 48.0%) on the short survey on alcohol consumption (new layout). Comparison of the response rate of the short and long health survey shows that the short health survey yielded the highest response (62.9% vs 52.1%; 50.0% vs 35.2%; 42.9% vs 40.0%, respectively). Compared to the

short survey on alcohol consumption (old layout) the short health survey yielded the highest response rate for the municipal health service whereas the short survey on alcohol consumption yielded the highest response rate for the addiction research institute and the university. For both short surveys on alcohol consumption, the old layout yielded a higher response than the new layout. Table 3 gives the results of multivariate modeling showing the odds ratios of response and 95% confidence intervals for the four factors (i.e. sponsorship, length, topic and design). Of the four factors, only sponsorship was significantly related with response. Subjects receiving a questionnaire from the university and the addiction research institute responded less than those receiving a questionnaire from the municipal health service. No significant difference in the likelihood of response was found between the university and the addiction research institute. The results indicate that short questionnaires had a higher (nonsignificant) likelihood to be returned than long questionnaires. With regard to topic, no relationship was found between response and questionnaire topic (odds ratio of 1.00). Results indicated that the new layout yielded less response (but not significant) than the old layout.

Table 2 Response rates (%) of the four types of questionnaire

	Municipal Health Service	Addiction Research Institute	University
Long Health Survey, old design	N=3000 52.1	N=125 35.2	N=125 40.0
Short Health Survey, old design	N=62 62.9	N=62 50.0	N=63 42.9
Short Survey on Alcohol consumption, old design	N=63 50.8	N=63 50.8	N=62 48.4
Short Survey on Alcohol consumption, new design	N=125 48.0	N=125 37.6	N=125 48.0

Table 3 Odds ratios and 95% confidence intervals of response with manipulation for four factors: sponsorship, length, topic and design.

	Multivariate modeling[c]	
	OR[a]	95% CI[b]
Sponsorship:		
Municipal Health Service	1.00	(ref.)
Addiction Research Institute	0.60	0.46-0.77
University	0.66	0.51-0.84
Length:		
Short	1.00	(ref.)
Long	0.74	0.53-1.02
Topic:		
Alcohol	1.00	(ref.)
Health	1.00	0.66-1.53
Design:		
Old style	1.00	(ref.)
New style	0.74	0.52-1.07

[a] OR = odds ratio

[b] CI = confidence interval

[c] This model included the variables: experimental factor, gender, age and civil status.

Conclusions

Our results support the hypothesis that the municipal health service will yield a higher response on health-related research than a university. Subjects receiving a questionnaire from the municipal health service were significantly more likely to respond than those who received a questionnaire from the university. The hypothesis that the well-known municipal health service and the university would yield a higher response than the addiction research institute was only partly confirmed. The municipal health service questionnaires were more likely to be returned than those from the addiction research institute, but the university did not yield a significantly higher response than the addiction research institute.

Because the difference in the likelihood of response was not significant, our hypothesis that a short questionnaire may lead to a higher response than a long questionnaire is not be confirmed. However, it was in the hypothesized direction (odds ratio of 0,74).

The hypothesis that a health survey may receive more response than a survey on alcohol consumption can be rejected (odd ratio of 1.00).

In contrast to our hypothesis, the old layout questionnaire seems to yield a higher response than the new layout questionnaire. However, the difference was not statistically significant.

Discussion

Before discussing the main findings, some methodological issues need to be addressed. First, in our experimental design not all combinations of the four investigated factors were included, implying that not all interaction effects could be explored. Therefore, we restricted our analysis to the investigation of main effects. Second, subjects participating in this study came from two random samples of different size ($N_1 = 1000$; $N_2 = 3000$). However, because both were randomly drawn from the same municipal registry and had the same mailing procedure, this could not introduce any bias.

This study explored the effects on response of sponsorship, length, topic and layout of a mailed questionnaire on alcohol consumption. The main finding is that sponsorship has a significant effect on response rate. Compared to the university and the addiction research institute, the municipal health service has a higher likelihood of response. There was no significant difference in the likelihood of response between the university and the addiction research institute. The higher likelihood of response for the municipal health service is probably due to its public health character. The general public considers this well-known organisation to be trustworthy and legitimate authority to conduct research. In comparison to a university and an addiction research the municipal health service probably conveys more trust that survey results will actually be implemented in public health policy. Therefore, a municipal health service may attract a broader public than the university and an addiction research institute. Governmental surveys achieved higher response than surveys from the private sector because subjects may value governmental authority as trust (Dillman, 2000; Cialdini, 1984; Heberlein & Baumgartner, 1978). The relatively low response rate to the Addiction Research Institute as a sponsor, may probably be due to the negative connotation of 'addiction'; people probably do not like to associate with addiction.

Contrary to our expectations we found no relationship between questionnaire length and response; this is also in contrast to the systematic review of Edwards et al. (2005). We could not find any logical explanation for this.

In contrast to other studies (Lund & Gram, 1998; Martin, 1994; van Kenhove, Wijnen & de Wulf, 2002) we found no relationship between topic and response. An explanation for this may be that our study samples come from the general population rather than from a specific population. For example, one study (Martin, 1994) explored a specific study population (i.e. amateur bowlers)

whereby one survey topic was about bowling whereas the other survey was about restaurants; understandably, the bowling survey is of greater interest to this sub-population. Participants may value completing a questionnaire (i.e. reward). Another study (van Kenhove, Wijnen & de Wulf, 2002) investigated the influence of topic on response among undergraduates, but they first investigated which topics would be most interesting for them. Our hypothesis that a health questionnaire will be of greater interest to the general public than a questionnaire on alcohol consumption is not confirmed.

Contrary to our expectations, no significant difference in the likelihood of response between the two layouts of questionnaires were found, and even more surprising, the direction of the difference was in favour of the old layout. One study (Dillman, Sinclair & Clark, 1993) also found no overall difference in response between two forms (one a respondent-friendly booklet form versus a more traditional form) but they did find that a subgroup of the general population (i.e. low response areas) was more likely to return the respondent-friendly booklet than the more traditional form. Different sub-populations may respond differently to a particular questionnaire style.

The present study investigated the effect of four survey factors on response and showed that only sponsorship has a significant effect on response. The municipal health service has the highest likelihood of response compared to a university and an addiction research institute. The response rates to the 12 questionnaires ranged from 38% to 63%. This large range also indicates that manipulation of survey factors has some effects on the response rates.

References

Cialdini RB (1984). Influence: the new psychology of modern persuasion. New York:Quill.

Dillman DA (2000). Elements of the tailored design method. In Mail and Internet surveys: the tailored design method, 2nd Ed.; New York: Wiley and Sons, Inc: 3-194.

Dillman DA, Sinclair MD and Clark JR (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. Public Opinion Quarterly 57*:* 289-304.

Edwards P, Roberts, I, Clarke M, Diguiseppi C, Pratap S, Wnetz, R, Kwan I and Cooper R (2005). Methods to increase response rates to postal questionnaires (Cochrane Methodology Review). The Cochrane Library, 2.

Heberlein TA and Baumgartner RM (1978). Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. Amererican Sociological Review, 43: 447-462.

Kenhove van P, Wijnen K, Wulf de K (2002). The influence of topic involvement on mail-survey response behavior. Psychology & Marketing, 19(3): 293-301.

Lahaut VMHCJ, Jansen HAM, Mheen van de H and Garretsen HFL (2002). Non-response bias in a sample survey on alcohol consumption. Alcohol & Alcoholism, 37 (3): 256-260.

Leeuw ED (1992). Data quality in mail, telephone, and face-to-face suveys. PhD thesis. Vrije Universiteit Amsterdam.

Lund E and Gram IT (1998). Response rate according to title and length of questionnaire. Scandinavian Journal of Social Medicine, 26(2): 154-160.

Martin CL (1994). The impact of topic interest on mail survey response behaviour. Journal of the Market Research society, 36(4): 327-338.

PART II

Measurement error

Chapter 5

Comparison of two formats of the weekly recall and quantity frequency alcohol measures

Viviënne M.H.C.J. Lahaut, Harrie A.M. Jansen, Dike van de Mheen, Henk F.L. Garretsen. Comparison of two formats of the weekly recall and quantity-frequency measures. Journal of Substance Use 2003; 8 (3): 164-169

Abstract

Aim
To compare data quality of two formats of the Weekly Recall measure (WR1 vs WR2) and the weekly Quantity-Frequency alcohol measure (QF1 vs QF2).

Design
Participants were randomly allocated to one of the four formats of the alcohol measures. On aggregate level, formats were compared for mean number of alcohol units/drinking days and for item non-response. Respondent 's problems with completing the questionnaire were assessed by cognitive interviewing.

Findings
No differences in alcohol consumption were found between WR1 and WR2, and item non-response was higher on WR2 than on WR1. QF2 yielded a higher mean number of drinking days/week than QF1 but no differences in number of units/week; QF2 had a higher item non-response rate than QF1. Most problems occurred in averaging and in reporting consumption according to the given instructions.

Conclusions
On an aggregate level there were no differences in alcohol consumption between WR1 and WR2, but WR2 had a higher likelihood of item non-response. According to the 'more is better' principle, QF2 is preferred, but also has a higher item non-response than QF1. Interviewing uncovered problems and misreporting that could not be revealed by comparing aggregate scores.

Introduction

Most general population surveys report lower estimates of alcohol consumption than estimates derived from sales data. Survey estimates generally cover 40 to 60% of the recorded alcohol sales data, although lower and higher rates have been reported (Leifman et al. 2000). This underreporting may arise from different sources, including selective response in the surveys. High non-response among heavy drinkers is frequently considered a causal factor for the underreporting in surveys. However, our earlier study on non-response bias (Lahaut et al. 2002) revealed strong evidence for an underrepresentation of abstainers and only weak evidence for an underrepresentation of frequent excessive drinkers. Similarly, Lemmens et al. (1988) reported that non-respondents generally do not drink more than respondents. Based on these results, we may question selective response as a source for underreporting. A second source of underreporting may be the quality of self-reports. Subjects may deliberately misreport their alcohol consumption due to shame or social desirability, or inadvertently misreport their alcohol consumption because of recall failure or because the required task to perform is too complicated or unclear. Instead of giving a false answer, subjects can also skip questions because the question(s) may be unclear, too difficult or too sensitive, causing item non-response.

In this recent study we focus on the quality of questions about alcohol in postal surveys. Much research has aimed to improve the quality of both postal and interview surveys on alcohol consumption, leading to the development of several measures to assess alcohol consumption. Until now, the most widely used alcohol measure is the Quantity-Frequency (QF) question which ask subjects about their drinking frequency and their average alcohol consumption per drinking occasion. A disadvantage of the QF measure is that subjects may find it difficult to report an average of their alcohol consumption (Lemmens et al. 1992). The Graduated Frequency measure has been developed to simplify this 'averaging' task; this measure asks subjects to report frequencies for each of a graduated series of consumption levels, starting with the highest consumption level. Another disadvantage of the QF measure is the long reference period, which is likely to reduce accuracy because subjects must recall details of events over a long time span.

To minimise the recall failure short-term recall measures have been developed, such as the Weekly Recall (WR) measure that asks subjects to report their alcohol consumption on each of the seven days preceding the survey. Other methods that try to minimise recall failure are the diary and the Timeline Follow Back (TLFB) method. The diary asks subjects to record each drink consumed by day, ideally shortly after consumption; to be effective, this method requires compliance of the respondent and intensive monitoring by the research

team. The Timeline Follow Back method (Sobell and Sobell 1992) presents subjects a calendar as a memory cue and asks them to recall their drinking on a daily basis retrospectively.

Besides developing alternative measures, attempts can be made to improve existing measures by re-phrasing or restructuring the questions. For example, questions can be specified according to the beverage type or by day, thereby providing subjects with extra memory cues which may reduce recall failure. Examples of improving an existing measure are the beverage-specific QF and beverage-specific QF with drink size (Serdula et al. 1999; Williams et al. 1994). In addition, Gmel and Lokosha (2000) experimented with open-ended and closed-ended question formats within a frequency question. Other studies have compared several alcohol measures with each other (e.g. Carney et al. 1998; Flegal 1991; Grant et al. 1995) but few studies have compared different formats of the same alcohol questions (e.g. Williams et al. 1994).

The current study focuses on improving two commonly used alcohol measures by rephrasing or restructuring the questions. Thus, we compared a widely used format of a Weekly Recall measure (WR1) with a new format (WR2) that provides more memory cues. Similarly, a widely used format of the Quantity-Frequency measure (QF1) was compared with a new format (QF2) that asks subjects to report day-by-day by means of open-ended questions.

The aim of the present study is to compare the four different question formats on the quality of the alcohol data. For these comparisons, we used two approaches. The first approach follows the widely-accepted 'more is better' principle which is derived from the fact that survey estimates of alcohol consumption are always lower than estimates from sales data. Thus, the question format that elicits the highest estimates on aggregate level will be considered better than the other format. We also investigated item non-response, which is a second quantitative indicator of quality on aggregate level. A high item non-response on a particular question may imply that the question is too difficult, unclear or too sensitive and will thus reduce the accuracy and reliability of population estimates. The second approach is cognitive interviewing, which we used to gain insight into problems/difficulties subjects experience whilst answering the alcohol questions in our survey. This approach enables us to evaluate the accuracy and reliability of the answers on an individual level.
The specific research questions in this study are: 1) Does WR2 elicit higher estimates of alcohol consumption than WR1? 2) Does QF2 elicit higher estimates of alcohol consumption than QF1? 3) Do WR1 and WR2 lead to different item non-response rates? 4) Do QF1 and QF2 lead to different item non-response rates? 5) What problems do respondents experience in answering the questions in the four different formats?

Method

Measures

In this study a 7-day weekly recall (WR) and a quantity-frequency (QF) measure were applied, with two different formats for each measure.

The WR measure asks subjects to list their alcohol consumption on a daily basis during 7 days retrospectively. The following two formats of this retrospective measure were constructed and investigated:

WR1) Please indicate for each day in the previous week how many units of all kinds of alcoholic beverages you drank. Start on the day that refers to yesterday. Please, calculate in standard units: a small bottle or can is 1.5 units.
Last Monday [...] units;
Last Tuesday [...] units;
...
Last Sunday [...] units.

WR2) Please, indicate for each day in the previous seven days how many units of beer, wine or spirits you drank. Please, calculate in standard units: a small bottle or can is 1.5 units.
What day is it today? □ Monday □ Tuesday ... □ Sunday
Yesterday was [...]day, I drank [...] units of beer, [...] units of wine, [...] units of spirits;
Two days ago was [...] day, I drank [...] units of beer, [...] units of wine, [...] units of spirits;
…
Seven days ago was [...] day, I drank [...] units of beer, [...] units of wine, [...] units of spirits.

The QF measure asks respondents to list their alcohol consumption on weekdays and on the weekend during a typical week; first the average number of days on which they consumed alcohol during a week and then the average amount of alcohol consumed on a typical drinking day. The following two formats of this summary measure were constructed and investigated:

QF1) On average, on how many of the four weekdays (Monday through Thursday) do you drink alcoholic beverages? Please, calculate in standard units: a small bottle or can is 1.5 units. The answer categories are: 4 days; 3 days; 2 days; 1 day; less than 1 day; I never drink on weekdays (→ skip to question ...)
If you drink on a weekday, how many units of alcoholic beverages do you drink on average? Please, calculate in standard units: a bottle or can is 1.5

units. The answer categories are: 11 or more units; 7-10 units; 6 units; 4-5 units; 3 units; 2 units; 1 units.
On average how many of the three weekend days (Friday through Sunday) do you drink alcoholic beverages? The answer categories are: 3 days; 2 days, 1 day; less than 1 day; I never drink on the weekend days (→ skip to question ...)
If you drink on a weekend day, how many units of alcoholic beverages do you drink on average? The answer categories are: 11 or more units; 7-10 units; 6 units; 4-5 units; 3 units; 2 units; 1 unit.
QF2) How much do you usually drink in a typical week? Please indicate for each day how many units of alcohol beverages you usually drink. Please, calculate in standard units: a small bottle or can is 1.5 units.
On Monday I usually drink [ ] units;
On Tuesday I usually drink [ ] units;
…
On Sunday I usually drink [ ] units.

Procedures and samples

Four different questionnaires were constructed whereby each questionnaire type contained one of the four formats of alcohol measures. Each format was part of a larger questionnaire which also included questions on lifestyle and health. Four samples were randomly drawn from the Rotterdam municipal registry. The questionnaire including WR1 were mailed to 625 subjects, WR2 to 188 subjects, QF1 to 250 subjects and QF2 to 187 subjects. The reason for the difference in sample sizes is that the questionnaires were also used for other research purposes that are beyond the scope of this study. Survey participation was voluntary and based on informed consent. The response rates for the four questionnaires were 45.6% (n=285), 44.7% (n=84), 37.6% (n=94) and 44.4% (n=83), respectively. All questionnaires also contained a form on which subjects could indicate their (un)willingness to participate in further research in the near future.

Our next step was cognitive interviewing to gain insight into possible problems that subjects may experience during answering the questions. Of all the subjects who had indicated their willingness to co-operate, potential respondents were phoned and asked whether they wanted to be interviewed at home. Beforehand, we stratified all subjects for educational level and drinking pattern. For the sample, we aimed for some variation in educational level and drinking pattern because these may be important factors of cognitive processes in answering questions on alcohol consumption. If a subject agreed to co-operate, an appointment was made for a face-to-face interview at home which is the natural setting of completing postal questionnaires. The interview consisted of two steps. First, the respondent was asked to complete the survey questions and to verbalize

his/her mental activity while completing the questions (thinking aloud). During thinking aloud we observed the respondent and, in the second step, we asked questions based on our observations and on the respondent's interpretation of the questions and words. Because this part of the study was aimed at an inventory of possible problems subjects may experience in answering the questions, the sample should be theoretically rather than statistically representative, which means that it should represent all possible problems. Therefore, we applied a procedure of constant comparison and theoretical saturation (Glaser and Strauss 1967) whereby after each interview it was ascertained whether the interview yielded new information. If this was not the case after some interviews, data collection was stopped. Saturation was achieved after 17 interviews for WR1, 11 interviews for WR2, 12 interviews for both QF1 and QF2.

Data analysis and statistical methods

Quantitative analyses were carried out using the program SPSS 9.0 for Windows. The data analysis was anonymous. Abstainers (i.e. subjects who reported no alcohol consumption in the last year) were left out of the analyses. Independent samples t-tests were applied to test differences in mean number of units/week and mean number of drinking days/week within both the WR and QF measure. The number of units/week was log transformed because its distribution was skewed. QF1 has closed-ended response categories; we used the midpoints of these categories for estimates of both the number of drinks and number of drinking days. The midpoint for the highest response category of '11 or more drinks' was defined as 13 drinks. For QF1, the total number of units/week of each individual was calculated by multiplying the number of drinks on a drinking day by the number of drinking days.

Accounting for possible confounders a Univariate Analysis of Variance procedure (General Linear Model, Univariate) also analysed these means according to four social demographic variables: gender (male/female), age (16-25; 26-35; 36-45; 46-69 years), education level (low, middle, high) and nationality (Dutch only, not-Dutch only). Logistic regression analysis was performed to calculate the odds ratios (OR) of item non-response within both measures (i.e. WR and QF). In all analyses a p-value <= 0.05 was considered statistically significant.

The interviews were transcribed by the interviewer. In the analysis of these interview data all problems observed and mentioned were collected and coded into three main categories: problems with regard to the comprehension of the question; problems with the instructions; problems in actually answering the question. The latter category was subdivided into: recall problems and calculation/estimation problems.

Results

Alcohol consumption estimates

Table 1 presents data on comparisons of self-reported alcohol consumption between the two formats of both the WR and the QF measures. The mean differences between WR1 and WR2 in number of units/week and number of drinking days/week were not significant. QF2 yielded significantly more drinking days/week than QF1 (3.7 vs 2.7); this difference remained significant after correcting for gender, age, education and nationality (data not shown). The mean difference in units/week between QF1 and QF2 was not significant.

Table 1. Comparison of alcohol consumption between the two formats of both the Weekly Recall and the Quantity-Frequency measures.

	Weekly Recall			Quantity-Frequency		
	WR1 (n=209)	WR2 (n=58)	P-value	QF1 (n=68)	QF2 (n=45)	P-value
Alcohol consumption Mean (±SD):						
Number of units/week	11.1±17.3	7.7±9.7	ns	10.8±13.4	11.6±16.2	ns
Log (number of units/week)	0.8±0.5	0.7±0.5	ns	0.8±0.5	0.9±0.4	ns
Number of drinking days/week	3.0±3.0	2.5±2.1	ns	2.7±2.2	3.7±2.2	*

* Independent samples t-test, P<0.05

Item non-response

Table 2 presents the results of univariate and multivariate modelling showing the OR of item non-response and 95% confidence intervals for both formats of the WR and QF measures. WR1 had a lower likelihood of item non-response than WR2 (OR 0.18); this result remained significant after correcting for gender, age, education and nationality. Both the univariate and multivariate modelling showed a lower item non-response for QF1 than for QF2.

Tabel 2. Odds ratios (OR) and 95% confidence intervals (CI) of item non-response with both the Weekly Recall and the Quantity-Frequency measures

	Univariate modelling		Multivariate modelling[a]	
	OR	95% CI	OR	95%CI
Weekly Recall				
WR1	0.22	0.05-0.85	0.17	0.03-0.94
WR2	1.00	(ref.)	1.00	(ref.)
Quantity-Frequency				
QF1	0.12	0.02-0.57	0.05	0.00-0.48
QF2	1.00	(ref.)	1.00	(ref.)

[a] This model included the variables: question format, gender, age. education and nationality.

Problems/difficulties during cognitive interviewing

The weekly recall measure: WR1 and WR2

The most problems experienced in completing WR1 were related to the instructions given to complete this question. The instruction that denotes to start on the day that refers to yesterday yielded many problems. The purpose of this instruction was to recall alcohol consumption backwards, which may facilitate the process of recalling. From our observations during the thinking-aloud, it appeared that most respondents did not start to list their alcohol consumption on the day that refers to yesterday but listed it from top (i.e. on Monday) to bottom (i.e. on Sunday). Some of the subjects did start on the day that refers to yesterday but did not continue to list backwards, or they listed the alcohol consumption of yesterday at the top of the list (i.e. on Monday) which was not the day that refers to yesterday. This resulted in two types of errors: 1) the alcohol consumption listed on a certain day did not belong to that day; and 2) halfway, subjects listed expected alcohol consumption for coming days instead of actual alcohol consumption. The other instruction causing problems was the request to convert a bottle/can into one and a half (1.5) units. Subjects converted a bottle/can into two units or did not make the conversion at all, or made miscalculations. Problems also arose with the answers to WR1. Further questioning on the response process showed that some subjects did not list the alcohol consumption

in the previous seven days, but listed their usual alcohol consumption. Explanations for this, given in the interviews, were that it was very difficult to recall alcohol consumption backwards or that the alcohol consumption in the last seven days did not represent their usual drinking pattern. Another problem was that some subjects did not recall their alcohol consumption on a particular day and skipped that day by listing zero units, or put a question mark on that day. Interviews also showed that subjects forgot to list some units of alcoholic beverages. During thinking-aloud, it appeared that subjects generally recalled their alcohol consumption on a particular day by thinking of activities that took place on that day. If subjects consumed many alcoholic beverages on a particular day, they tended to estimate their alcohol consumption by thinking of the pleasure and duration of the drinking occasion.

The WR2 measure yielded no problems with regard to the instructions. However, similar to WR1, subjects often forgot to list some units of their alcohol consumption because they were thinking of usual drinking occasions rather than of atypical occasions.

Quantity-Frequency measure: QF1 and QF2

The instruction in QF1 that denotes to convert a bottle/can into 1.5 units again yielded some problems. Subjects converted a bottle/can into two units or made no conversion at all. The other instruction that yielded problems was the navigational instruction; from our observations during the interviews, it appeared that subjects skipped navigational instructions or they used a navigational instruction not intended for them. The most common problem in answering QF1 was the difficulty in averaging the frequency and quantity of drinking. This resulted in two types of errors: frequencies of the drinking occasions were often not averaged, and the quantities of alcohol were sometimes not averaged. Subjects reported frequencies for drinking events without taking into account weeks/weekends in which they did not drink at all. These 'averaging' errors also occurred in QF2. The interviews revealed some difficulties in identifying typical alcohol consumption: e.g. one subject remarked that alcohol consumption was partly determined by sports activities which occur only in winter and therefore his alcohol consumption differed between seasons; another subject skipped QF2 because she sometimes did not drink for several weeks. A striking characteristic that applied to both QF1 and QF2 was that subjects listed different frequencies and quantities on the questionnaire during the interview than on the earlier identical postal questionnaire. This raises questions about the reliability of the self-reports.

Discussion

This study investigated two different formats of both a Weekly Recall and a Quantity-Frequency measure in order to compare the two formats of each measure on their data quality. For these comparisons, two approaches were used. The first approach follows the 'more is better' principle whereby the highest estimate is considered to be the best; another quantitative indicator for the quality of the alcohol data is item non-response. The second approach uses cognitive interviewing to reveal problems/difficulties whilst actually answering the questions.

To avoid learning effects, we used different samples to test WR1, WR2, QF1 and QF2. This may be a limitation because the samples may differ in their alcohol consumption. All samples were, however, randomly drawn from the municipal registry, and the analyses were corrected for differences in important socio-demographic variables.

The main findings of this study showed no differences in reported alcohol consumption between WR1 and WR2. Compared to WR1, the item non-response on WR2 was significantly higher. In contrast, cognitive interviewing showed that WR1 generated the most problems due to the instruction on how to complete the form. Based on these interviews, a higher item non-response would be expected on WR1. An explanation for the higher item non-response on WR2 may be that WR2 requires more filling-in and may therefore demotivate respondents. The finding that subjects firstly think about their activities during the day before they report their alcohol consumption supports the basis of the Timeline Followback method and the specific settings approach whereby calendar key dates or social situations serve as anchors for reporting drinking (Sobell and Sobell 1992; Single and Wortley 1994).

With regard to the QF measurement, according to the 'more is better' principle QF2 should be preferred. On the other hand, QF2 has a higher item non-response, which may reduce the accuracy and reliability of the alcohol consumption estimates. Cognitive interviewing showed that on OF1 and QF2 misreporting occurs: subjects did not average their drinking adequately, the frequencies were not averaged over all days but only over drinking days and were therefore overestimated. It is not clear to what extent this misreporting occurs and whether it occurs equally for both measures on aggregate level. Lemmens et al. (1992) also suggested that respondents have difficulty in summarising their drinking frequency over a long period of time. Although summary measures ask for mean quantities and frequencies, respondents probably provide modal frequencies and quantities (Room, 1990; Midanik, 1993). According to Gruenewald and Nephew (1994) self-reports of 'typical' use reflect a mixture of modal and average drinking levels. A striking result is that, like the

WR measure, the more specific format of QF (i.e. QF2) has a higher item non-response. This may be an effect of the requirement of more filling-in on specified questions, which may demotivate respondents, especially in self-administered questionnaires.

Therefore, we are confronted with a dilemma, especially in self-administered questionnaires, between cognitive quality and motivation: memory aids which are widely acknowledged to enhance the quality of self-reports may cause item non-response and thereby reduce the survey quality. Specification of questions apparently require too much repetitive filling-in which bores respondents; this may be particularly so for nondrinkers and light drinkers, who do not profit from the memory aids. Further studies on both cognitive interviewing and postal surveying are needed to see whether specific response to specified questions distorts the comparison of reported alcohol consumption on the aggregate level.

Conclusions

No mean differences in number of units/week and number of drinking days/week were found between WR1 and WR2. Thus, our hypothesis that WR2 would elicit higher estimates of alcohol consumption than WR1 was not confirmed. In contrast, alcohol consumption on WR2 was somewhat lower. The hypothesis that QF2 would elicit higher estimates of alcohol consumption than QF1 was not confirmed on the frequency domain of QF. On average QF2 yielded relatively more drinking days/week than QF1, which remained so after a univariate analysis by gender, age, education and nationality.

The question format of the WR measure seemed to play an important role in the extent of item non-response; the likelihood of item non-response is higher on WR2 than on WR1. The question format of the QF measure was also an important factor for the extent of item non-response. QF2 had a higher likelihood of item non-response than QF1.

The cognitive interviews revealed no problems with regard to the general comprehension of the questions on WR1 and WR2. Most recall problems in answering WR1 and WR2 occurred when subjects did not remember the drinking activities that took place. Most frequent problems were related to the instructions of WR, particularly of WR1. The instruction of WR1 that asks to start with the day that refers to yesterday created the greatest confusion among the subjects and therefore often led to misreporting. Subjects were confused or found it difficult to start listing alcohol consumption halfway through the list of days and/or to list events backwards instead of forwards.

The results of the cognitive interviews showed a serious problem that led to misreporting on both QF1 and QF2, i.e. that subjects did not average their

alcohol consumption. Particularly the frequency of drinking was not averaged over all days but only over the drinking days. Subjects who did not have a regular drinking pattern or who seldom drank alcoholic beverages had the most problems in answering QF1 or QF2.

References

Carney MA, Tennen H, Affleck G, Del Boca FK and Kranzler HR (1998). Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time "electronic interviews". Journal of Studies on Alcohol 59(4): 447-454.

Flegal KM (1991). Agreement between two dietary methods in reported intake of beer, wine and liquor. Journal of Studies on Alcohol 52(2): 174-179.

Glaser BG and Strauss AL (1967). The discovery of grounded theory: strategies for qualitative research. Chicago: Aldine Pub.Co.

Gmel G and Lokosha O (2000). Self-reported frequency of drinking assessed with a closed- or open-ended question format: a split sample study in Switzerland. Journal of Studies on Alcohol 61(3): 450-454.

Grant KA, Tonigan JS and Miller WR (1995). Comparison of three alcohol consumption measures: a concurrent validity study. Journal of Studies on Alcohol 56(2): 168-172.

Gruenewald PJ and Nephew T (1994). Drinking in California: theoretical and empirical analyses of alcohol consumption patterns. Addiction 89(6): 707-723.

Lahaut VMHCJ, Jansen HAM, Mheen van de D and Garretsen HFL (2002). Non-response bias in a sample survey on alcohol consumption. Alcohol&Alcoholism 37(3): 256-260.

Leifman H, Österberg E and Ramstedt M (2002). Alcohol in postwar Europe: a discussion of indicators on consumption and alcohol-related harm., University of Stockholm Finnish National Research and Development Centre for Welfare and Health Swedish National Institute of Public Health (NIPH)**:** 6-84.

Lemmens PHHM, Tan FES and Knibbe RA (1988). Bias due to non-response in a Dutch survey on alcohol consumption. British Journal of Addiction 83: 1069-1077.

Lemmens P, Tan FES and Knibbe RA (1992). Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. Journal of studies on Alcohol 53(5): 476-486.

Midanik LT (1993). Two random repeat recall methods to assess alcohol use. American Journal of Public Health 83(6): 893-895.

Room R (1990). Measuring alcohol consumption in the United States: methods and rationales. Research Advances in alcohol and drug problems 10: 39-80.

Serdula MK, Mokdad AH, Byers T and Siegel PZ (1999). Assessing alcohol consumption: beverage-specific versus grouped-beverage questions. Journal of studies on Alcohol 60(1): 99-102.

Single E and Wortley S (1994). A comparison of alternative measures of alcohol consumption in the Canadian National Survey of alcohol and drug use. Addiction 89: 395-399.

Sobell LC and Sobell MB (1992). Timeline Followback: A technique for assessing self-reported ethanol consumption. In: Allen J and Litten RZ (eds) Measuring alcohol consumption: psychosocial and biological methods. Totowa, NJ:Humana Press: pp.41-72.

Williams GD, Proudfit AH, Quinn EA and Campbell KE (1994). Variations in quantity-frequency measures of alcohol consumption from a general population survey. Addiction 89: 413-420.

Chapter 6

Measuring the year consumption of alcohol: the development of a questionnaire

Harrie AM Jansen, Viviënne MHCJ Lahaut, Henk FL Garretsen, Dike van de Mheen. Measuring the year consumption of alcohol: the development of a questionnaire. *Submitted*.

Abstract

Aim
To develop, test and validate a year consumption questionnaire for the general population as an alternative to the week based estimates of alcohol year consumption.

Design
A multi-mode test survey comparing the year consumption questionnaire with a typical week consumption questionnaire.

Participants
A sample of 101 drinkers among the general population.

Measurements
Two measures of year consumption. One based on the year consumption compiled by adding up daily habits, weekly habits and thirteen categories of cultural/personal drinking occasions over the previous year, and the other compiled by extrapolation from the report of alcohol consumption in an average week.

Findings
The year consumption questionnaire produces considerably higher reports of annual consumption, compared with the week-based estimates of year consumption. The year consumption questionnaire also yields a considerably higher prevalence of excessive drinking, regardless of gender and type of completion mode. Correlations between both estimates are very high, ranging from 0.78 to 0.92.

Conclusion
The year consumption questionnaire appears to achieve a more complete coverage of both of the alcohol consumption over the year and of the prevalence of excessive drinking compared with the weekly based reports. The correlations between the two instruments showed them to be almost equivalent in ordering respondents on the consumption continuum.

Introduction

General population surveys on alcohol consumption typically (at least implicitly) aim at representing the consumption of alcohol during one entire year to overcome seasonal bias, and to ensure comparability for both trend analysis and international comparison. Traditionally these surveys relied mostly on reports of drinking in average or typical weeks (TW) or quantity-frequency (QF) questioning for longer reference periods to establish year consumption. However, comparison of data from these surveys with tax-based sales data, revealed serious undercoverage of actual alcohol consumption at the aggregate population level by 30 to 60% (Knibbe & Bloomfield, 2001; Leifman, Österberg & Ramstedt, 2002) This under-coverage problem has been a major concern in studies aiming to enhance self-reports of alcohol consumption. Apart from technological innovations like the CAPI, CATI (de Leeuw, Hox & Snijders, 1995; Dawson, Makela, Midanik et al., 1998) and the "electronic interview" (Carney et al., 1998) most of this work has focused on facilitating both the cognitive and the motivational burden of respondents, by applying cognitive psychology in designing questionnaires (Jabine, Straf, Tanur & Tourangeau, 1984; Midanik & Hines, 1991; Sudman, Bradburn &Schwartz, 1996; Conrad & Blair, 1998; Routangeau, Rips & Rasinski, 2000). Compared to QF for typical weeks or months, concurrent daily reports in diaries have proven to be by far the best method to cover the person's alcohol consumption; therefore the diary is generally considered the 'gold standard' in alcohol survey methodology (Lemmens, Tan & Knibbe, 1992; Wyllie, Zhang & Casswell, 1994; Poikolainen, Podkletnova & Alho, 2002; Sobell & Sobell, 2005). However, in long-term applications the diary method has a serious validity problem, apart from the economic and motivational problems. Keeping a diary for long a period of time may produce significant bias due to self-monitoring effects that make a subject aware and reflective of his behaviour, which often leads to behavioural change. That is precisely why self-monitoring is a common tool in behavioural therapy. Therefore, other means have been sought to improve the coverage of alcohol consumption in surveys. Two main strategies have been derived from cognitive psychology (Menon & Yorkston, 2000) to enhance survey questioning.

The first strategy is to specify substantive or quantitative elements of the drinking behaviour itself to facilitate retrieval of information from memory. Examples are asking for seasonal variation, the exact number of drinks per occasion and the type of beverage (beer, wine, liquor, etc.).(Serdula, Mokdad, Byers & Siegel, 1999). The second strategy is to specify contexts of drinking, e.g. days of the week, moments of the day, cultural and family events or activities or even emotional states.

One common measure following the first strategy is to add a question on the frequency of atypical heavy drinking ('binge' drinking) to the summary report of drinking in an average or typical week. This may produce relevant additional information about the year drinking pattern. A recent initiative within the strategy of specification drinking itself, is the 'graduated frequency' (GF) questionnaire Midanik, 1994; Greenfield, 2000; Dawson, 2003). It starts with a question on the highest number of units a subject has ever consumed on one occasion and then asks for the frequency of drinking that number of units (e.g. 9) in the reference period (e.g. last year), the frequency of drinking one unit less (e.g. 8), etc. until the frequency of drinking only one unit. However, this measure is vulnerable to over reporting because of double counting of heavy drinking occasions in GF reports (Poikolainen, Podkletnova & Alho, 2002; Graham, Demers, Rehm & Gmel, 2004; Heeb & Gmel, 2005)

Another possibility is to ask how many alcoholic beverages are purchased instead of consumed. However, only moderate correlations between reported purchase and reported drinking have been found (Fitzgerald & Mulford, 1987). The main cognitive problem in survey questioning on alcohol consumption remains unsolved both in GF and the other types of questionnaires mentioned, i.e. the problem of the lack of saliency of drinking. Because drinking is generally a secondary activity it is not easily remembered without a cue to its social behavioural context (Lemmens, Tan & Knibbe, 1992; Loftus, Fienberg & Tanur, 1985; Tourangeau, 2000). Therefore, following the second strategy, memory of drinking may be enhanced by starting with memorizing the primary activities as a context, rather than asking for drinking without any meaningful context. This is done in the Timeline Followback (TLFB) questionnaire (Sobell & Sobell, 1992) as well as in the specific settings questionnaire (SS) (Single & Wortley,1994) and in the Inventory of Drinking Situations (IDS) (Annis & Graham, 1995). TLFB offers *temporal anchors* by a calendar to the respondent with dates of national significance (e.g. in the USA for January: New Year's Day, Epiphany and M. Luther King day). The questionnaire asks respondents to estimate daily consumption for each period in the last year (or even two years), as structured by these national events. In the 'helpful hints' provided with the questionnaire, this strategy of temporal anchoring is taken further in utilizing the biographical context that includes both cultural and personal elements (e.g. consulting an appointment book and memorizing parties) (Sobell & Sobell, 2000). It is doubtful, however, whether these hints provide enough memory aids for a complete report on drinking after more than three or four weeks (Searles, Helzer & Walter, 2000; Searles, Helzer, Rose & Badger, 2002). Apart from that it takes much time and motivation when it has to be filled in for a 12-month period; that makes it almost impracticable for self-administered general population surveys.

The SS questionnaire (Single & Wortley, 1994) provides a list of eleven categories of personal *leisure activities* e.g. 'spending quiet evening at home' and 'going to a bar or tavern'. Subsequently it asks the frequency of engaging and the mean number of drinks (if any) per occasion for each item; it is an activity based quantity-frequency summary. The IDS (Annis & Graham, 1995; Victorio-Estrada & Mucha, 1997) provides a 100-item list of *socio-emotional states* (e.g. feeling lonesome, having fun with peers, conflict with parent). This inventory is developed for diagnostic aims in alcohol addiction treatment; because of its length it is not very suitable for general population surveys, however.

The year consumption questionnaire (YC) that we discuss below, also follows the strategy of biographical contextualisation. It includes elements of SS (personal leisure events) but also annual national holidays (e.g. in the Netherlands Queens Day and Carnival) and also annual special periods (winter holidays and summer holidays) in the previous year. This assumes that people are able to give a reasonable estimate of e.g. the frequency of birthday parties they attended in the previous year, and the average number of alcoholic drinks consumed.

The Year Consumption Questionnaire (YC)

The idea to develop a questionnaire based on personal habits and events arose from our development of a three-step interview for testing a self-completion questionnaire (Jansen & Hak, 2005) with 'typical week' and 'binge frequency' questions. As a follow-up we asked subjects for details about their drinking patterns in an open interview – along the lines of daily, weekly and other habits r events throughout the year. It appeared that many instances of drinking were not accounted for by the written self-reports, which was later confirmed explicitly by the respondents.

The YC is basically a standardization of these follow-up interviews. In YC we aim to overcome the problem of the low saliency of drinking by focusing on the events and situations where people consume their drinks. To capture these occasions and link the personal drinking habits to these, we could have started by constructing a personalised calendar of the previous year. However, because this is time-consuming in interviews and hardly applicable in self-completion modes, we decided on a QF summary for those occasions that are not fixed on daily or weekly moments.

The questionnaire builds on the daily, weekly and yearly life rhythms in order to anchor drinking habits, and then adds a series of occasions that may occur frequently but without a priori fixed weekly or yearly rhythms. Many people have daily habits (e.g. drinking wine or beer with dinner, sherry or whisky when coming home from work) and/or weekly habits (e.g. having a drink

with colleagues on Fridays, having a beer after sports on Wednesday evening, etc.). These daily and weekly habits may be covered by typical-week reports. Serious undercoverage may be expected, however, when it comes to drinking on annual events (e.g. New Year's eve, vacations, birthday parties) or at social events which may occur often but without fixed intervals (e.g. visiting friends or family, going to the theatre, having a drink on a terrace, etc.).
YC first asks for the daily habits (fixed moment of the day) and continues with weekly habits (fixed day of the week). It concludes with a series of 13 occasions that are relevant as a drinking occasion for many people in the Netherlands; some are public yearly events (e.g. Queens Day), others are private yearly events (e.g. birthdays, vacations). The general format of the questions for these occasions is: "How often do you (participate in) X and how much do you normally drink at these occasions"? Although we assume that the 13 categories of drinking events will cover nearly all of the drinking occasions in the Netherlands, an open question is included to cover unforeseen occasions
(Appendix 1).

Research questions

On the basis of this questionnaire many aspects of the person's drinking habits may be calculated, such as the proportion of habitual daily drinking in the total year consumption or the proportion of family versus peer group drinking, etc. For this preliminary test of YC, however, we focus only on the total amount of alcohol consumed in a year and the prevalence of excessive drinking.
The first two questions to be addressed are:
1) Does YC yield a higher report of alcohol consumption compared with a typical week report (TW), and
2) Does YC yield a higher prevalence rate of excessive drinking?
Many general population surveys are less interested in prevalence estimates than in the search for correlates (determinants, meanings or consequences) of differences in alcohol consumption (e.g. to provide data for public health care or prevention). For such studies absolute consumption levels are less important than relative levels (i.e. the ordering of respondents on the consumption continuum). Studies comparing estimates based on summary measures (like QF, GF or TW) with daily diaries as a standard, have shown considerable underreport for the summary measures (except GF), but very high correlations with daily diaries (Poikolainen, Podkletnova & Alho, 2002). We explore whether this applies also for TW compared to YC. To judge the robustness of results, all comparisons will be calculated for sub-samples (gender, and mode of administration). Therefore two additional questions are addressed:

3) How strong is the correlation between TW-based year estimates and YC-based estimates?
4) Is this correlation stable when calculated within subgroups of gender and mode of administration (interview versus self-completion)?

Methods

Data collection
To test the applicability and quality of YC under various conditions, we decided on a multi-mode (face-to-face interview, paper self-completion and internet self-completion) format with a highly diverse sample of drinkers (i.e. a qualitative rather than a probability sample). Data were collected by interviews with 47 persons (21 males and 26 females, aged 18 - 69 years, mean age 39 years). Of these 14 were volunteers from a student association selected by an interviewer at their club, and 33 were respondents of a general population survey who indicated that they were willing to participate in future surveys. In addition we collected data by a self-completion questionnaire (paper & pencil version) from 12 volunteers (10 females, 2 males, aged 21-43 years, mean 34 years) from our research institute, and by an electronic version published on the IVO website. The website survey was completed by 42 'visitors' (22 females and 20 males, aged 16 – 69 years, mean 39 years). The total sample therefore comprised 101 subjects. Because this sample is not a random sample from the general population; this study doest not claim to be statistically representative. For the purpose of testing the validity and practicability of the questionnaire, this qualitative sample does however offer a relevant variety in age, gender, and mode of administration.

Measures
In the test survey questionnaire, apart from YC, a day-by-day version of TW was also included. Furthermore, it contained some basic questions about the subject's gender, education level, marital status and household composition, as well as an open question about their evaluation of the survey.
According to common research practice, we define excessive drinking for women as drinking more than 14 units per week (i.e. > 729 units per year) and for men more than 21 units per week (>1094 units per year).
The TW-based year consumption is calculated by multiplying the TW total by 52.
The YC-based year consumption estimate is calculated by summing the quantity*frequency products for daily habits, weekly habits and occasions without a fixed daily or weekly timing (see Appendix 2).

Results

Experiences in data collection
In the interview mode YC presented no serious problem. Furthermore, respondents confirmed the 'member validity' that was reported earlier (Jansen & Hak, 2005), stating that their reports in this interview were more complete than their foregoing reports in the self-completed TW questionnaire.
The paper & pencil version of the self-completion questionnaire also presented no problems; but the website version performed less well. Only 42 out of 53 responding visitors delivered a usable report (complete and serious answers at face value). This poor result seems mainly due to motivational factors, e.g. some respondents found this questionnaire to be "very boring". This is probably because respondents had to report the same behavior two or three times due to the cross-validation goal of this survey. In a regular survey, the questionnaire will be shorter (also on the website) and will take only five minutes to complete.

Year consumption estimates
The mean TW-based year consumption in the present sample is 638 units per year, compared with 876 units per year (.i.e.37% higher) in the YC-based estimate. Therefore, the answer to the first research question is positive: YC yields a considerably higher report of alcohol consumption compared to TW.

Table 1 gives data on year consumption estimates based on the TW and YC reports for the same respondents, in both the interview mode and the self-completion mode.

Table 1.TW-based estimates of year alcohol consumption compared with YC-based estimates for the same respondents (n=1010, in the face-to-face interview sample and in the self-completion sample.

	TW-based year consumption	YC-based year consumption
0 – 103 units/year (< 2 per week)	5%	5%
104 – 364 units/year (2 to < 7 per week)	35%	25%
365 – 729 (7 to < 14 per week)	30%	34%
730 - 1093 (14 to <21 per week)	17%	15%
1094 + (≥ 21 per week)	13%	22%
Total	100%	100%

Table 1 shows that the differences between distributions are concentrated in two consumption categories. TW generates a higher frequency in the category of 104 to 364 units, whereas YC generates higher frequencies in the highest category (> 1094 units); differences for other categories are relatively small or absent.

Table 2 shows that in this sample the prevalence estimates of excessive drinking for women are 21% on TW reports and 30% on YC reports compared with 21% and 29%, respectively, for men. Therefore the differences between TW and YC are relatively consistent.

Table 2. TW-based and YC-based estimates of the prevalence of excessive drinking among women and men.

	Women		**Men**	
	TW	**YC**	**TW**	**YC**
Light/moderate drinking	79%	70%	79%	71%
Excessive drinking *	21%	30%	21%	29%
Total	100%	100%	100%	100%
(N)	(56)	(56)	(44)	(44)

* Excessive drinking: Women >14 units/week, Men > 21 units/week

Accordingly, the second research question can also be answered positively: YC yields higher estimates of the prevalence of excessive drinking than TW. Table 3 presents percentages of excessive drinkers by gender and by completion mode.

Table 3. TW-based and YC-based prevalence of excessive drinking* among men and women in interview mode and self-completion mode

	Interview			**Self-completion**		
Gender	TW	YC	(N)	TW	YC	(N)
Men	10%	24%	(21)	30%	39%	(23)
Women	12%	27%	(26)	30%	37%	(21)

* Excessive drinking: Women >14 units/week, Men > 21 units/week

Although absolute frequencies in table 3 are very and samples are far from random, there is a high level of consistency in the differences in both gender and method between the interview sample and the self-completion sample, i.e on TW the percentages of excessive drinking for man and women are similar and the YC-based reports are substantially higher than TW-based reports in both modes. Apart from this, the prevalence of excessive drinking is significantly higher in the self-completion mode compared with the interview mode. This difference may be caused by self-selection of heavy drinkers in the website based self-completion sample.

Correlations

The overall Pearson's correlation between TW-based year consumption and YC-based year consumption is 0.88 in this sample. The difference between subgroups is small: i.e. interview mode 0.85 versus self-completion 0.88, women 0.78 versus

men 0.92. Therefore, research questions 3 and 4 may be answered positively: the overall correlation is very high, and stable over gender and administration mode.

Discussion

Traditionally, general population surveys on alcohol consumption rely on the previous week or typical week reports (TW) and mean quantity-frequency measures (QF) for a month in last six months or last year. All such surveys are used to estimate the year consumption.

In the present study we tested a 'true' year consumption questionnaire (YC), which is composed of QF questions for fixed daily habits, fixed weekly habits, and a series of common cultural and private drinking occasions (some are fixed yearly cultural occasions and others are personal or family-based occasions). The aim is to produce more complete reports of both yearly consumption and excessive drinking than are delivered by TW.

Because this test survey was not designed to represent the general population statistically, the results from this sample may not be interpreted statistically either, but rather qualitatively. Obviously, comparisons between YC-based and TW-based population estimates require random population samples with high response levels.

First the study shows that the YC questionnaire is practicable both in the interview and in the self-completion mode. However, the website version demands extra motivation because about 20% of the returned questionnaires were insufficiently completed. Secondly, YC produces more complete reports of year consumption. In the strategic (non-random) test sample of non-abstainers the mean year reports were 37% higher than year estimates based on TW. In the interview mode respondents generally confirmed that the YC report was more accurate than the TW report; one effect of this is a substantial raise (about 50%) in the prevalence of excessive drinking compared to TW, both for men and women. Despite the large difference in absolute alcohol consumption reported, the year consumption scores on the basis of YC show a very strong correlation with TW scores. This finding is consistent with other studies comparing different consumption measures (Poikolainen, Podkletnova & Alho, 2002). This implies that the new YC instrument is especially useful for prevalence estimates. For statistical studies on determinants or consequences of differences in total alcohol consumption, TW seems to be equivalent to YC. Finally, we suggest that the YC questionnaire may probably be very useful for studies comparing international data. The YC can easily be adapted to local culturally-based drinking habits and drinking occasions (without the problem of violating standardization rules), because its logic is contextually within itself.

References

Annis HM and Graham JM (1995). Profile Types on the Inventory of Drinking Situations: Implications for relapse prevention counseling. Psychology of addicted behaviors, 9.

Carney MA, Tennen H, Affleck G, Boca FKD and Kranzler HR (1998). Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time "electronic interviews". Journal of Studies on Alcohol, 59: 447-454.

Conrad F and Blair J (1998). From impressions to data: increasing the objectivity of cognitive interviews, Proceedings of The 1996 Annual meeting of the American Statistical Association. Alexandria, VA: The American Statistical Association.

Dawson DA (2003). Methodological issues in measuring alcohol use. Alcohol Research and Health, 27: 18-29.

Dawson D, Makela R, Midanik LT et al. (1998). Measuring alcohol consumption: limitations and prospects for improvement. Addiction, 93: 965-968.

Fitzgerald JL and Mulford HA (1987). Self-report validity issues. Jounal of Studies on Alcohol, 48: 207-211.

Graham K, Demers A, Rehm J and Gmel G (2004). Problems with the graduated frequency approach to measuring alcohol consumption: results from a pilot study in Toronto, Canada. Alcohol & Alcoholism, 39: 455-462.

Greenfield TK (2000). Ways of measuring drinking patterns and the difference they make: experience with graduated frequencies. Journal of Substance Abuse, 12: 33-49.

Heeb J-L and Gmel G (2005). Measuring alcohol consumption: A comparison of graduated frequency, quantity frequency, and weekly recall diary methods in a general population survey. Addictive Behaviors, 30: 405-413.

Jabine TB, Straf ML, Tanur JM and Tourangeau R (1984). Cognitive aspects of survey methodology: building a bridge between disciplines. Report of the advanced research seminar on cognitive aspects of survey methodology.Washington D.C.: National Academy Press.

Jansen HAM & Hak T (2005). The Productivity of the Three-Step Test-Interview (TSTI) Compared to an Expert Review of a Self-administered Questionnaire on Alcohol Consumption. Journal of Official Statistics, 21: 103 - 120.

Knibbe RA and Bloomfield K (2001). Alcohol consumption estimates in surveys in Europe: comparability and sensitivity for gender differences. Substance Abuse, 22: 23-38.

Leeuw de ED, Hox JJ and Snijkers G (1995). The effect of computer-assisted interviewing on data quality. Journal of the Market Research Society, 37: 325-344.

Leifman H, Österberg E and Ramstedt M (2002. Alcohol in post war Europe: a discussion of indicators on consumption and alcohol-related harm.Stockholm: Swedish National Institute of Public Health.

Lemmens PHHM, Tan ES and Knibbe RA (1992). Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. Journal of Studies on Alcohol, 53: 476-486.

Loftus EF, Fienberg SE and Tanur JM (1985). Cognitive psychology meets the national survey. The American Psychologist, 40: 175-180.

Menon G and Yorkston EA (2000). The use of memory and contextual cues in the formation of behavioral frequency judgments. In: Stone AA, Turkkan JS, Bachrach CA et al. (Eds.). The science of self-report. Implications for Research and practice.London:Lawrence Erlbaum: pp.63-80.
Midanik LT (1994). Comparing usual quantity/frequency and graduated frequency scales to assess yearly alcohol consumption: results from the 1990 US National Alcohol Survey. Addiction, 89: 407-412.

Midanik LT and Hines AM (1991). `Unstandard' ways of answering standard questions: protocol analysis in alcohol survey research. Drug and Alcohol Dependence, 27: 245-252.

Poikolainen K, Podkletnova I and Alho H (2002). Accuracy of quantity-frequency and graduated frequency questionnaires in measuring alcohol intake: comparison with daily diary and commonly used laboratory markers. Alcohol & Alcoholism, 37: 573-576.

Searles JS, Helzer J, Rose G and Badger G (2002). Concurrent and retrospective reports of alcohol consumption across 30, 90 and 366 days: interactive voice

response compared with the timeline follow back. Journal of Studies on Alcohol, 63: 352-362.

Searles JS, Helzer JE and Walter DE (2000). Comparison of drinking patterns measured by daily reports and timeline follow back. Psychology of addicted behaviors, 14: 277-286.

Serdula, MK, Mokdad AH, Byers T and Siegel PZ (1999). Assessing alcohol consumption: beverage-specific versus grouped-beverage questions. Journal of studies on Alcohol, 60: 99-102.

Single E and Wortley S (1994). A comparison of alternative measures of alcohol consumption in the Canadian Mational Survey of alcohol and drug use. Addiction, 89: 395-399.

Sobell LC and Sobell MB (2005). Alcohol consumption measures Assessing Alcohol Problems A guide for clinicians and researchers. US department of health and human services.

Sobell LC and Sobell MB (2000). Instructions for Filling out the Timeline Alcohol Use Calendar. http://www.nova.edu/gsc/online-files. Accessed 18 September 2005.

Sobell LC and Sobell MB (1992). Timeline Followback: A technique for assessing self-reported ethanol consumption. In: Allen J. and Litten RZ (Eds.). Measuring alcohol consumption: psychosocial and biological methods.Totowa, NJ: Humana Press: pp. 41-72.

Sudman S, Bradburn NM and Schwartz N (1996). Thinking About Anwers. The application of cognitive processes to survey methodology.San Francisco: Jossey-Bass: pp.305.

Tourangeau R, Rips LJ and Rasinski K (2000). The psychology of survey response.Cambridge: Cambridge University Press.

Tourangeau R (2000). Remembering what happened: memoriey errors and survey reports. In: Stone A A, Turkkan JS, Bachrach CA et al. (Eds.). The science of self-report. Implications for Rresearch and practice.London: Lawrence Erlbaum: pp. 28-29.

Victorio-Estrada A. and Mucha R (1997). The Inventory of Drinking Situations (IDS) in current drinkers with different degrees of alcohol problems. Addictive Behaviors, 22: 557-565.

Wyllie A, Zhang J-F and Casswell S (1994). Comparison of six alcohol consumption measures from survey data. Addiction, 89: 425-430.

Appendix 1

The year consumption questionnaire on alcohol consumption (YC)

Question 1.
This question regards the kinds of alcoholic beverages you consume.

Which alcoholic beverages did you drink once or more often in the past twelve months?
Please indicate all the boxes that are applicable:
beer (not alcohol-free beer)
wine, sherry, port, vermouth
liquor, blackcurrant jenever, advocaat
gin, rum, cognac, whisky, vodka, other spirits
breezers, shooters, other mixed drinks with alcohol

none, I did not drink any alcoholic beverage in the past year → go to Q 5.

Question 2.
Your daily drinking habits in the past twelve months.

During the past year, did you have fixed daily habits of drinking alcoholic beverages?
Yes, after work I usually took […] drinks (please fill in the number)
Yes, with dinner I usually took […] drinks
Yes, in the evening I usually took […] drinks

Did you perhaps have one or more other daily habits not mentioned above?
yes, namely ……
At these moments I usually took […] drinks

No. I had no daily drinking habit

Question 3.
Your habits on fixed days of the week over the past year.
In the past year, did you have any habitual activities on fixed days of the week on which you drunk alcoholic beverages?
Please fill in the number of drinks you had on average on these fixed weekly activities, <u>apart from</u> the daily drinks that you filled in on Question 2.

Every MONDAY, I had […..] drinks (please write the number)
Every TUESDAY, I took […..] drinks
Every WEDNEDSDAY, I had […..] drinks

Every THURSDAY, I had […..] drinks
Every FRIDAY, I had […..] drinks
Every SATURDAY, I had […..] drinks
Every SUNDAY, I had [….] drinks

Question 4.
Special days or periods in the past year (twelve months).

Below are some special occasions at which many people drink beer or wine or other alcoholic beverages.
Please indicate for every type of occasion how often you participated in the past twelve months, and how many drinks you usually had on this type of occasion.

Occasion type	Number of times per **month OR per year**	**Number of alcoholic drinks**
Bar or society	[……..] times per month year	[……] drinks per visit
Disco/club	[……..] times per month year	[……] drinks per visit
Sports canteen	[……..] times per month year	[……] drinks per visit
Restaurant	[……..] times per month year	[……] drinks per visit
Birthday party	[……..] times per month year	[……] drinks per party
Other party with friends or family	[……..] times per month year	[……] drinks per party
Reception	[……..] times per month year	[……] drinks per visit
Party/reception at work	[……..] times per month year	[……] drinks per party
Winter holidays	[……..] days	[……] drinks per day
Summer holidays	[……..] days	[……] drinks per day
Carnival	[……..] days	[……] drinks per day
Queen's Day/night		[……] drinks
New Year's eve		[……] drinks
Other occasions Please specify: …… ……..	[……..] times per month year	[……] drinks per day

Question 5.
Your comments.
What do you think of this questionnaire? Please write your comments or suggestions below.
Thank you very much for completing this questionnaire.

Appendix 2

Calculating the year consumption

The raw year consumption is the sum of all products of frequency (F) * typical number of drinking units (Q) for all occasions. It is composed of three sections: I) fixed daily habits (365*Qi), II) fixed weekly habits (52 * ∑Qi) and III) drinking occasions without a fixed daily or weekly timing (∑Fi*Qi). This raw year consumption may overestimate the actual year consumption due to overlap between fixed weekly habits (section II) and occasions from section III with a frequency ≥ 52. Thus for respondents who report weekly habits, all drinks on occasions with frequencies ≥ 52 will be ignored in the corrected year consumption measure (YC). YC = 365 * Qi daily + 52 * ∑Qi weekly + ∑ Fi*Qi for other occasions with a year frequency Fi< 52.

Chapter 7

General Discussion

Introduction

To arrive at an effective alcohol policy the Dutch government, municipalities and/or (local) health authorities often use postal surveys to gain insight into the prevalence of alcohol use, drinking patterns and possible problems related to alcohol drinking. Surveys are a relatively cheap method to reach a large population sample in a short time period. However, to interpret the survey results correctly, potential sources of error must be taken into account. For postal alcohol surveys particular attention should be paid to nonresponse error and measurement error due to self-reports, as both types of errors can affect the alcohol survey data. The studies described in the previous chapters of this thesis focus on these two potential sources of error. In this chapter we summarize and discuss the main findings and address some methodological issues of these studies. Finally, we make some recommendations for future research and discuss some practical implications of our findings.

Main findings

This thesis consists of two main parts: Part I addressed questions on nonresponse error and Part II addressed questions on measurement error due to self-reports. Here, we summarize the main findings for each research question.

Part I: Nonresponse error

1. Is there a bias in the alcohol survey results due to nonresponse, and is there evidence for the theory of a 'two-tailed nonresponse' pattern?

The results of our nonresponse follow-up survey showed a bias in the alcohol data. In the survey results abstainers and moderate drinkers were overrepresented. We also tested the theory of the 'two-tailed nonresponse' pattern which implies an overrepresentation of nonrespondents in the lowest and highest extreme of the survey variable of interest (in our case alcohol use). According to this theory a relatively higher nonresponse in the lowest extreme (abstainers) may be explained by lack of topic saliency whereas Social desirability may explain the overrepresentation of nonresponse in the highest extreme. Our results showed no evidence to support this theory. The only evidence was found for the 'lowest tail' of nonresponse pattern: i.e. abstainers were overrepresented among nonrespondents. There was no significant overrepresentation of frequent excessive drinkers.

2. Is there evidence for the 'continuum of resistance' model: Can alcohol estimates of late respondents be used as proxies for estimates of

nonrespondents? Are there differences in alcohol estimates between the different response waves?

Based on the results of our three surveys we could not confirm the 'continuum of resistance' model, which may imply that alcohol estimates of late respondents can not be used as proxies for estimates of nonrespondents. The results showed no significant pattern of difference in alcohol estimates between the different response waves.

3. What are the response effects of different sponsors, questionnaire length, topic label and layout?

Of the four manipulated survey characteristics (sponsor, questionnaire length, topic label and layout), only sponsorship had a significant relationship with response rate. Three agencies carrying out the survey were compared: an addiction research institute, a university, and a municipal health service. Results of our study showed that the municipal health service as sponsor reached the highest response rate, the university came next, and the addiction research institute had the lowest response rate.

Part II: Measurement error

4a Could formats including more memory cues be better alternatives of the commonly-used formats of the 'weekly recall' and 'quantity-frequency' alcohol measure for enhancing self-reports?

4b. What problems do respondents experience in answering the different formats of the 'weekly recall' and 'quantity-frequency' alcohol measure?

We found no difference in alcohol estimates between the two different formats of the weekly recall measure. This was also the case for the quantity-frequency measure, although the mean number of drinking days was found to be higher among the format including more memory cues. A striking result was the higher item nonresponse among the formats including more memory cues.

The most noteworthy findings from our cognitive interviews were:

- Subjects generally recalled their alcohol consumption on a particular day by thinking of activities that took place on that day. They also estimated their alcohol consumption by thinking of the pleasure and duration of an activity.
- The most common problem in the quantity-frequency measure was the task to average the quantity and frequency of drinking. It was often not averaged.
- Particularly subjects who did not have a regular drinking pattern or who seldom drank alcoholic beverages had difficulties in identifying their typical alcohol.
- Atypical drinking occasions were often forgotten.

- The instruction given for answering the commonly-used format of the weekly recall measure created the greatest confusion among the subjects which led to misreporting.
- The request of converting a bottle/can into 1.5 units, which was included in all question formats, was not followed. This led to miscalculations.

5. Could a newly-developed year consumption questionnaire be a better alternative for a typical week measure in enhancing year alcohol consumption?

The results of our explorative study showed that the newly-developed year consumption questionnaire produced year alcohol estimates that were about 35% higher than the typical week measure.

Methodological issues

Specific limitations of our studies have already been discussed in the previous chapters of this thesis. Before discussing the main findings, we would like to address some general methodological issues.

First, it should be noted that our studies are based on a study population randomly drawn from the municipal registry of Rotterdam. One limitation of this municipal registry is the absence of the homeless, illegal inhabitants, incorrect addresses for e.g. students and sub-tenants, etc. These groups may not be covered by our sample.

Secondly, there is evidence for a direct relationship between survey participation and social isolation, social exclusion and lack of social involvement. This means that people with low and marginal positions in the society may be underrepresented, such as ethnic minorities, people who stay in institutions, homeless people, unemployed or low-educated people (Groves & Couper, 1998; van Goor & Rispens, 2004; Stoop, 2005). Additionally, it is known that socially marginalized groups may also include relatively many people who drink heavily (Poznyak, Saraceno & Obot, 2005; Room, 2005). A striking example that supports the relationship between low social position and alcohol use is a finding from a Dutch alcohol study among Turks and Moroccans in Rotterdam that feelings of discrimination predicted higher alcohol use (Dotinga, 2005).
In case of our alcohol studies, this may imply that socially marginalized groups are more likely to refuse participation, which may result in underestimation of alcohol consumption.

Thirdly, to get valid information on a possible nonresponse error it is necessary to achieve a high response rate in the nonresponse follow-up survey. In our nonresponse follow-up survey we contacted 83% of our sample of

nonrespondents, but only 52% of these nonrespondents answered all of our alcohol questions. This implies that our results are based on the assumption that the respondents of our nonresponse follow-up survey did not differ in their alcohol estimates compared to the nonrespondents of our nonresponse follow-up survey. The question arises whether this assumption is correct. It may be that the 'hard-core' nonrespondents are those who differ most from respondents in the prevalence of excessive drinking. Regarding the positive relationship between social marginalization and survey refusal one may assume that those who have low social and marginal positions in the society are more likely to be heavy and problematic alcohol users.

In absence of a 'gold standard' criterion to evaluate the validity of alcohol measures, alcohol researchers assume that those measures that yield the highest estimates are the most valid, also called the 'more-is-better' rule (Del Boca & Noll, 2000). This rule is derived from the fact that survey estimates are always lower than estimates from the alcohol sales data. In Chapter 5 we also apply the more-is-better rule to evaluate alternative question formats of two alcohol measures. However, this rule may be applicable to the total volume of drinking on the aggregate level, but is questionable for drinking on the individual level. (Room, 1990). Undercoverage on the individual level is related to the type of questioning, respondent's drinking pattern and motivation. The daily reports in the diary are generally considered the 'gold standard'. The assumption that daily diary reports are more valid is based on the role of memory processes. More frequent reporting of alcohol consumption is assumed to enhance accuracy by reducing forgetting. However, diaries may not be very practicable for application in large population samples. Additionally, keeping a diary for a long period of time may produce bias due to self-monitoring effects. Another alternative 'gold standard' is the use of biochemical test. However, like the diary, this method is not practical for large population samples and is also limited by a short detection time period.

Finally, the last methodological issue we addressed is the mode of data collection. For our main alcohol survey we used a postal alcohol survey, whereas the nonresponse follow-up survey was predominantly by face-to-face and telephone interviews. These differences in mode may influence the reporting of alcohol consumption. Due to the greater anonymity, it has been suggested that postal surveys result in fuller reports of alcohol than face-to-face and telephone interviews (Beebe et al. 2005; Kraus & Augustin, 2001, Gmel, 2000). However, after several reminders with a copy of the postal survey, it may be necessary to approach nonrespondents with a different survey mode.

Discussion of results

The study of Pernanen (1974) showed, that in many surveys, subpopulations with a higher proportion of heavy drinkers tend to show higher nonresponse rates; his point was that heavy drinkers are harder to locate and more likely to refuse a survey. The study of Wilson (1981) confirmed this statement. However, since 1983 several nonresponse surveys have been carried out and the evidence of overrepresentation of heavy drinkers among nonrespondents has been weaker (Garretsen, 1983; Lemmens, Tan & Knibbe, 1988). Indeed, in line with these surveys we found strong evidence for overrepresentation of abstainers among nonrespondents and no evidence for overrepresentation of frequent excessive drinkers.

Overrepresentation of abstainers among nonrespondents could be explained by topic salience; people who do not drink alcohol are not interested in a survey on alcohol consumption. This was also confirmed by our interview experience in the nonresponse follow-up survey where nonrespondents were also asked for their reasons for non-participation. Remarks such as "I do not drink, why should I fill in this survey?" and "This survey is not meant for me because I do not drink" were often made. Results of a systematic review carried out by Edwards et al. (2005) also showed that people are more likely to respond if the questionnaire is of interest to them.

The lack of evidence for an overrepresentation of frequent excessive drinkers may be questioned. Within our nonresponse follow-up survey sample the difference in percentages of frequent excessive drinkers between respondents and nonrespondents was relatively large (3.8% versus 6.3%). However, this difference was not significant, perhaps due to the small sample size. Another explanation for not finding evidence may be that excessive drinkers are not reached in a nonresponse follow-up survey. Until now, most nonresponse follow-up surveys (Garretsen, 1983; Lemmens, Tan & Knibbe, 1988; van Dijck & Knibbe, 2005) have had low response rates. In our nonresponse follow up survey we reached a contact rate of 83%; however, the response rate to our alcohol questions was 52%. As described in the paragraph 'Methodological issues' people with low social and marginal positions are more likely to be underrepresented in a survey. These people (i.e. homeless people, people who are institutionalized, or mobile people) are difficult to reach and are more likely to refuse a survey. In most alcohol prevalence studies (Garretsen, 1983; Bongers, 1998, van Dijck & Knibbe, 2005) problem drinkers are overrepresented among people who have low social and marginal positions. Additionally, social marginalization may be related to excessive drinking. Consequently our nonresponse follow-up survey may underestimate frequent excessive drinking. In order to get valid information on alcohol use among the difficult-to-reach nonrespondents, future alcohol

nonresponse surveys should invest considerable resources into collecting alcohol information from this group.

We found that people who filled in the questionnaire after several reminders are not different in their alcohol consumption than people who participated immediately. Therefore, the 'continuum of resistance' model could not be confirmed. This implies that reminders are not necessary to reduce a possible nonresponse bias; nevertheless, reminders are very effective in enhancing the response rate. These results are in line with the study of Stoop (2005). Dependent on financial resources a choice can be made between taking a larger initial sample size or sending reminders in order to enhance the response rate.

As described in Chapter 4, compared to the university and addiction research institute, the municipal health service had a higher likelihood of response. This finding could be explained by the social exchange theory of Dillman (2000). A municipal health service may convey more trust because people consider this organisation as a legitimate authority to make the survey request and may be more convinced that the survey results will be used for social ends, in this case local public health. The 'trust' may outweigh the 'costs'. The addiction research institute had the lowest response rate. An explanation could be that an addiction-related sponsor elicits some concerns that people may be stigmatised as addicts. These findings confirmed our hypothesis based on the declining response rate of the 1994 alcohol survey, sponsored by the Addiction Research Institute, compared to the response rate of the 1980 alcohol survey, sponsored by the municipal health service. However, this decline may also partly be explained by the general trend of decreasing response rates. In contrast to other studies, we found no relationship between topic label and response rate (Lund & Gram, 1998; Martin, 1994; van Kenhove, Wijnen and de Wulf, 2002). An explanation for these differences in results may be the survey population. We had a general heterogenous survey population whereas the other studies had a more homogenous population (i.e. bowlers, students). In these latter studies it may be more predictable which topic is of more interest. We found no relationship between response and questionnaire layout. It is difficult to compare studies on this subject because there is no standard definition of a 'respondent friendly' layout. There are several ways to operationalize this concept. Perhaps there were minimal perceived differences in respondent friendliness between our layouts and therefore no effects were found. Contrary to our expectations we found no significant relationship between questionnaire length and response; this is also in contrast with findings from the systematic review of Edwards et al. (2005). We could not find any logical explanation for this.

Formats including more memory cues of the 'weekly recall' and 'quantity-frequency' alcohol measures did not lead to 'better' measures. This means that these formats did not reach higher alcohol estimates, although a higher mean of drinking days was found in the quantity-frequency format. A remarkable finding was that more memory cues lead to more item nonreponse. More filling in because of the memory cues may demotivate respondents. For the weekly recall measure, the format including memory cues produced less confusion about the way of filling in and therefore reduces misreporting. To overcome item nonresponse, it is interesting to test these formats in other survey modes, such as web surveys and telephone surveys. These modes are suitable to minimize item nonresponse and calculation errors (i.e. converting a bottle or can into 1.5 units). It is also useful to gain insight into the consequences of misreporting on the total alcohol consumption on the aggregate level.

We experienced cognitive interviewing to be a good method to get insight into the question-answering process. The finding that subjects did not average their alcohol consumption and that atypical drinking complicates this averaging task were in line with other studies (Lemmens, Tan and Knibbe, 1992; Room, 1990; Midanik, 1993; Gruenewald and Nephew, 1994). Also, the finding that subjects try to remember their alcohol consumption by thinking about certain activities is acknowledged by others (Single and Wortley, 1994; Sobell & Sobell,1992). These authors developed several measures: for example the Timeline Followback method of Sobell & Sobell (1992) that presents a calendar as a memory cue, or the specific settings approach of Single & Wortley (1994) that first asked how often people are involved in common activities and subsequently posed the question how often they drink at these activities.

Based on our experience with cognitive interviews and the specific settings approach of Single & Wortley (1994), we developed a year consumption questionnaire which asks people for their personal fixed daily and weekly habits, and series of common cultural and private drinking occasions. Our results are promising: our questionnaire yielded considerably higher year alcohol estimates than a typical week measure. An advantage of this questionnaire is its flexible format. The format may easily be adapted to the cultural context. However, a potential weakness may be that not every situation in which people consume alcohol is covered by the composite list of common activities. It is important to know in what social settings alcohol is used within the study population.

Recommendations

The results of our studies provide recommendations for areas of future research and practical implications for organisations supporting alcohol policy.

Areas of future research
No evidence was found for the theory of the 'two-tailed nonresponse' pattern. We only found evidence for one 'tail': i.e. overrepresentation of abstainers among nonrespondents. Relatively more excessive drinkers among nonrespondents was not proven. Nevertheless, we recommend more research on the validity of the theory of the 'two-tailed nonresponse' pattern, especially focused on the factor social desirability. In order to further explore the 'two-tailed nonresponse pattern', it is advisable to carry out subgroup analyses because interest in topics and prevailing norms on survey topics can differ within subgroups (e.g. alcohol use among students, youth).

Another area for future research is our newly-developed instrument for measuring year consumption. The results of our pilot study are promising, but cause of its explorative character we did not test the questionnaire in a statistically representative sample. The first step is to conduct the same study in a statistically representative sample. Additionally, we restricted our analysis to comparisons of year estimates of our instrument with those of a typical week measure. For health purposes it is also important to compare the two measures on drinking patterns. The next step is to test different modes of the questionnaire: e.g. face-to-face survey versus postal questionnaire.

Practical implications
For alcohol prevalence and problem drinking studies it is important to carry out more intensive nonresponse research. For this type of research a high response rate is essential. Therefore it is important to invest resources in nonresponse research and to allocate these resources in advance. We experienced our nonresponse follow-up survey as a time-consuming process in which interview techniques are very important. It is advisable to employ well-trained and experienced interviewers for nonresponse research. An example of a procedure of an intensive follow-up is described by Stoop (2004a). It consists of several steps: 1) collecting information on eligibility of sample; 2) reissuing no contacts and temporary refusals by a second wave; 3) reissuing the remaining no contacts and temporary refusals by a third wave. In the follow-up of nonrespondents a multi-mode approach (choice between face-to-face, telephone, internet or self-completion) and the possibility of giving (monetary) incentives may be useful. In the case of limited resources we recommend to restrict the nonresponse follow-up

study to a small representative sub-sample allowing more attention to be paid to the individual subject.

Municipal health services can play an important role in getting more insight into nonresponse and measurement errors of alcohol and/or health surveys. It is advisable to collaborate within this field and to standardize the procedure of nonresponse follow-up surveys. The following elements should be standardized: the sample of nonrespondents, alcohol questions, and the total minimal number of contact approaches per nonrespondent. Standardisation of nonresponse follow-up surveys enables comparability of the results and, consequently, more definite conclusions can be drawn.

It is already known, but cannot be stressed enough that a survey should be first be tested in a small sample of the survey population. This can be done by cognitive interviewing, which is a good method to reveal problems which subjects experience in answering the questions.

Finally, this thesis has contributed to but has not ended the search for the 'missing drink' in alcohol surveys. More nonresponse alcohol research focusing on 'hard-core' nonrespondents, and further exploration of the use of memory cues and contextual cues in alcohol measures may help to find the 'missing drink'.

References

Beebe TJ, McRae JA, Harrison PA, Davern ME and Quinlan KB (2005). Mail surveys resulted in more reports of substance use than telephone surveys. Journal of Clinical Epidemiology, 58(4): 421-424.

Bongers IMB (1998). Problem drinking among the general population: a public health issue? PhD thesis. Erasmus University Rotterdam.

Del Boca FK and Noll JA (2000). Truth or consequences: the validity of self-report data in health services research on addictions. Addiction, 95(3): S347-S360.

Dillman DA (2000). Mail and Internet surveys: the Tailored Design Method. New York:Wiley.

Dotinga A (2005). Drinking in a dry culture. Alcohol use among second-generation Turks and Moroccans: measurement and results. PhD thesis. Erasmus University Rotterdam.

Dyck van D and Knibbe RA (2005). De prevalentie van probleemdrinken in Nederland. Een algemeen bevolkingsonderzoek. Universiteit Maastricht.

Edwards P, Roberts I, Clarke M, Diguiseppi C, Pratap S, Wentz R, Kwan I and Cooper R (2005). Methods to increase response rates to postal questionnaires (Cochrane Methodology Review). The Cochrane Library, 2.

Garretsen HFL (1983). Probleemdrinken: prevalentiebepaling, beïnvloedende factoren en preventiemogelijkheden. Theoretische overweging en onderzoek in Rotterdam. PhD thesis. Katholieke Universiteit Nijmegen.

Gmel G (2000). The effect of mode of data collection and of non-response on reported alcohol consumption: a split sample study in Switzerland. Addiction, 95: 123–134.

Goor van H and Rispens S (2004). A middle class image of society. Quality & Quantity, 38: 35-49.

Groves RM and Couper MP (1998) Nonresponse in household interview surveys. New York: Wiley.

Gruenewald PJ and Nephew T (1994). Drinking in California: theoretical and empirical analyses for alcohol consumption patterns. Addiction, 89(6):707–723.

Kenhove van P, Wijnen K and Wulf de K (2002). The influence of topic involvement on mail survey response behavior. Psychology & Marketing, 19(3): 293–301.

Kraus L and Augustin R (2001). Measuring alcohol consumption and alcohol-related problems: comparison of responses from self-administered and telephone interviews. Addiction 96: 459-471.

Lemmens PHHM, Tan FES and Knibbe RA (1992). Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. Journal of Studies on Alcohol, 53(5): 476-486.

Lemmens, PHHM, Tan FES and Knibbe RA (1988). Bias due to non-response in a Dutch survey on alcohol consumption. British Journal of Addiction, 83: 1069-1077.

Lund E and Gram IT (1998) Response rate according to title and length of questionnaire. Scandinavian Journal of Social Medicine, 26(2): 154-160.

Martin CL (1994). The impact of topic interest on mail survey response behaviour. Journal of Market Research Society, 36(4): 327-338.

Midanik LT (1993). Two random repeat recall methods to assess alcohol use. American Journal of Public Health, 83(6): 893-895.

Pernanen K (1974). Validity of survey data on alcohol use. In: Gibbens R, Israël Y, Kalant H, Popham R, Schmidt W and Smart RG (eds). Research Advances in alcohol and drug problems John Wiley & Sons: New York: p 355-374.

Poznyak V, Saraceno B and Obot IS (2005). Breaking the vicious circle of determinants and consequences of harmful alcohol use. Bulletin of the World Health Organization, 83(11): 803-804.

Room R (2005). Stigma, social inequality and alcohol and drug use. Drug Alcohol Reviews, 24(2): 143-155.

Room R (1990). Measuring alcohol consumption in the United States: methods and rationales. Research Advances in Alcohol and Drug Problems, 10: 39-80.

Single E and Wortley S (1994). A comparison of alternative measures of alcohol addiction in the Canadian National Survey of alcohol and drug use. Addiction, 89:395-399.

Sobell LC and Sobell MB (1992). Timeline Followback: A technique for assessing self-reported ethanol consumption. In: Allen J and Litten, RZ (eds). Measuring alcohol consumption: psychosocial and biological methods. Humana Press: Totowa, NJ: p 41-72.

Stoop IAL (2005). Nonresponse in sample surveys. The hunt for the last respondent. PhD thesis. The Hague: The Social and Cultural Planning Office of the Netherlands.

Stoop IAL (2004a). Surveying nonrespondents. Field Methods, 16(1): 23-54.

Wilson P (1981). Improving the methodology of drinking surveys. The Statistician, 30: 159-167.

Summary

Summary

Survey research is often used by Dutch authorities and community health services to gain insight in the population's general health situation. In survey research a random sample of the population is requested to answer a questionnaire, often sent by post. This kind of research is a relatively cheap and quick method of determining a population's health situation and tracing possible risk groups. Besides these advantages, the survey method also has its drawbacks that may affect research results. For example, low response is one of the threats to the survey method. In the Netherlands (as well as internationally), a trend towards lower response rates can be discerned. Nowadays, a 50% response result can be judged as a high response rate. Non-response constitutes a problem if there is a selection of people who do respond to the survey and those who do not, based on the central research variable. An example would be that only healthy persons would fill in the questionnaire and return the envelope and that ill or disabled persons would give no reaction. As a result of this, based on the survey data, the general population health would be judged better than it actually is. Another threat to the survey method is the quality of self-reports. In self-reports, people report their behaviour by themselves. Frequently, much is demanded from the cognitive process: people must recall distinct behaviour over a certain time span. For example, how often they have drunk alcohol during the last half year. It is questionable to what extent people are able to recall this accurately and to estimate this adequately. Another factor affecting the quality of self-reports is that people are not always willing to give a valid answer. This can be the consequence of feeling ashamed about their own behaviour, and/or people's inclination to conform to generally accepted norms and values, and to answer in socially accepted ways because of this.

In this thesis, the two above-mentioned threats (non-response, and the quality of self-reports) to survey research will be investigated more closely. The rationale for this investigation is that the Addiction Research Institute (Instituut voor verslavingsonderzoek, IVO) has been involved in carrying out several surveys on alcohol and drugs (as part of general population health), and more specifically, problem drinking and related problems.

In comparison to an earlier alcohol survey in 1981 (by Garretsen, 1983), a comparable alcohol survey in 1998 (by Bongers, 1998) had to contend with quite a relatively lower response rate: 70% as compared to 44%. Several explanations were sought for this lower response rate: a general national trend of declining response rate or a different sponsor of the questionnaire (the municipal health service versus the Addiction Research Institute IVO). In addition, it was unsure in

what way this low response rate could affect the survey results. Moreover, there was an increasing concern about the quality of self-reports: would it be possible to refine alcohol measures in order to enhance the quality of self-reports?

This thesis has two objectives:

- To gain more insight in estimating non-response error and in enhancing response rates.
  To this end, two studies were carried out in which non-response error was measured by a non-response follow-up survey, and in which the 'two-tailed nonresponse pattern' theory and the ' continuum of resistance' model were tested. An additional study was carried out to test the effects of manipulated survey characteristics on response rates.
- To gain more insight in the possibilities of enhancing the quality of self-reports.
  To this end, two studies were carried out in which memory cues were added to alcohol questions, and in which a new questionnaire was developed to measure year alcohol consumption to ease the cognitive process.

This thesis consists of two parts: the first part addresses nonresponse error (Chapters 2, 3 and 4), and the second part focuses on measurement error due to self-reports (Chapters 5 and 6). For each chapter the main findings are summarized below.

Chapter 2 presents a nonresponse follow-up study that was carried out to investigate if the survey alcohol estimates were biased due to nonresponse. In this study unexpected home visits were made among a sample of nonrespondents. During these visits nonrespondents were asked for their reasons for nonparticipation and about alcohol consumption. Results of the nonresponse follow-up survey showed an overrepresentation of abstainers and moderate drinkers among the nonrespondents. In this follow-up study the theory of the 'two-tailed nonresponse' pattern was tested. According to this theory nonresponse is overrepresented in the two extremes of the survey variable of interest because of two aspects: topic saliency and social desirability. People who are not interested in the survey topic may not participate. Prevailing norms in a society play an important role in the aspect of social desirability. People, who do not fit in these norms, may be inclined to give a socially-desirable answer or may not participate. In case of our alcohol survey, this theory implies that people who do not drink may not participate because they are not interested in alcohol as a survey topic (lowest extreme). The highest extreme represents the excessive drinkers. In the Dutch society drinking alcohol is generally accepted but only in

moderation, which leads to the hypothesis that excessive drinkers are overrepresented among nonrespondents. Results of the nonresponse follow-up study showed strong evidence for overrepresentation of abstainers among nonrespondents. However, no evidence was found for an overrepresentation of excessive drinkers. Based on these findings we could not confirm the theory of the 'two-tailed nonresponse' pattern.

In Chapter 2 we applied the direct method of estimating nonresponse error, which means that direct information is gained from the nonrespondents themselves. In **Chapter 3** we tested the 'continuum of resistance' model, an indirect method to estimate nonresponse error and to correct for it. According to the 'continuum of resistance' model two extremes are located on the continuum: people who always participate in a survey and people who never participate in a survey. Based on this model, estimates of late respondents may be used as proxy for estimates of nonrespondents because they have a similar position on the continuum. We tested this 'continuum of resistance' model in our alcohol survey and in two larger similar alcohol surveys. Results showed that there were no differences in alcohol consumption between the three response waves (wave 1: people who respond immediately; wave 2: people who respond after one reminder; wave 3: people who finally respond after the last reminder). Based on these results we could not confirm the 'continuum of resistance' model. Sending reminders may not lead to less biased alcohol estimates. However, it appeared to be an effective strategy to enhance the sample size: two reminders in our alcohol survey contributed to about 23% of the total response rate (about 43%).

Obviously it is better to prevent nonresponse error and to apply response-enhancing strategies to the survey. In **Chapter 4** we manipulate some survey characteristics in order to enhance the response rate. The underlying theory is the social exchange theory of Dillman in which three elements of survey participation are essential: reward, costs and trust. An interesting survey topic or a chance to get a gift by participation can be a reward. Costs are, for example, time spent on answering the questionnaire or returning the completed questionnaire by post. An example of trust is survey sponsorship, which the respondent considers as a legitimate authority to do a survey request. To foster survey participation, reward and trust have to outweigh the costs. Based on Dillman's theory we manipulated four survey characteristics: sponsorship, questionnaire length, topic label and layout.

The survey was sent to three different sponsors: a municipal health service, a university, and our institute (Addiction Research Institute). We hypothesized that the municipal health service may reach the highest response rate. Results confirmed this hypothesis. The municipal health service was more likely to have

the highest response rate, the university came next and the Addiction Research Institute was more likely to have the lowest response rate.

Long questionnaires (19 pages) and short questionnaires (3 to 5 pages) were compared: it was hypothesized that short questionnaires were more likely to have a higher response rate. The burden on the respondent is obviously lower with a short questionnaire, which leads to lower 'costs' than for a respondent who has to invest more time and energy to fill in the long questionnaire. However, the results of our study showed no difference in response rates between the long and short questionnaire.

The third manipulated survey variable was topic label. Questionnaires were represented as an alcohol survey or as a health survey. The underlying assumption was that people are more interested in a health survey. However, the results also showed no difference in response rates.

Two layouts were used: one layout was the commonly used traditional layout. The 'new' layout was in a bookletform, has more symbolic language, and background colours which may appear more attractive to fill in. No difference in response rates was found between these two layouts.

To summarize: only one survey variable, i.e. sponsorship, had an effect on response rates.

In **Chapter 5** alternative formats of two commonly-used alcohol measures ('weekly recall' and 'quantity-frequency'), which included memory cues, were tested on enhancing self-reports. Both the weekly recall and the quantity-frequency had two formats: the 'traditional' format and the alternative format including more memory cues. To test these formats on enhancing self-reports the 'more-is-better'- rule was applied. This rule is derived from the 'gold standard' of alcohol sales data: Survey-based estimates cover only 40-60% of the sales-based estimates. We also compared the item nonresponse of these two formats and tested the formats by cognitive interviewing. The cognitive interview consisted of two parts: the respondents were asked to fill in the alcohol questions while thinking aloud. After this, based on the observations of the thinking-aloud phase, the respondent was asked about the interpretation and wording of the alcohol questions.

Results showed no difference in alcohol consumption between the two formats of the 'weekly recall'. This was also the case for the 'quantity-frequency' measure, except there was a difference in the mean number of drinking days: the alternative format led to a higher mean number of drinking days. A striking result was a relatively higher item nonresponse among the alternative formats. The presence of memory cues in the alternative formats means that people have to fill in more, which may lead to demotivation.
Results from the cognitive interviews revealed a number of problems:

- People often recalled their alcohol consumption by thinking of activities that occurred on a specific day. Estimating the number of alcohol units was dependent on the duration and pleasure of an activity.
- The most frequent problem on the quantity-frequency measure was that of averaging alcohol consumption; averaging was often not done.
- Especially infrequent drinkers has difficulties in estimating alcohol consumption in a typical week.
- Atypical drinking behaviour was often forgotten.
- The filling in of the traditional format of the 'weekly recall' led to great confusion among respondents; this led to many errors.
- The request of converting a bottle or can into 1.5 unit was not followed and/or led to many calculation errors.

In **Chapter 6** we explored a newly-developed questionnaire to measure year alcohol consumption. In earlier pilot studies cognitive interviews showed that people recalled their alcohol consumption by thinking of activities. This finding led to the development of alternative alcohol measures, such as the timeline follow-back method of Sobell & Sobell which gives respondents a calendar as memory cue, and the specific-settings approach of Single and Wortley that relates drinking alcohol to a social context. In our new questionnaire the social activities are central. Respondents were asked for daily and weekly habits, followed by a list of personal (e.g. birthdays) and public events (e.g. Queens Day). First, the respondent is asked how many times an event occurs and subsequently the respondent is asked how many alcohol units they drink during that event. In the explorative study the year alcohol estimates of the newly-developed questionnaire were compared with the year alcohol estimates based on a typical week measure. Results showed that the new questionnaire yielded about 30% higher alcohol consumption reports than the typical week format. These findings imply the need for further development and testing of this questionnaire.

Finally, the thesis ends with a general discussion, presented in **Chapter 7**, in which the main findings of Chapters 2-6 are summarized and discussed. Then, some methodological issues of our studies are addressed and the general discussion ends with some recommendations for future research and possible practical implications.

Samenvatting

Samenvatting

Om een beeld te krijgen van de algemene gezondheidssituatie onder de bevolking gebruikt de Nederlandse overheid en/of gemeentelijke gezondheidsdiensten vaak enquêteonderzoek. Bij enquête onderzoek wordt een random steekproef in de bevolking gevraagd een vragenlijst te beantwoorden. Vaak wordt de vragenlijst per post opgestuurd. Deze vorm van onderzoek is een relatief goedkope en snelle methode om de gezondheidssituatie van een populatie in kaart te brengen en eventuele risicogroepen op te sporen. Naast deze voordelen heeft de enquête methode ook nadelen, die de onderzoeksuitkomsten kunnen beïnvloeden. Een lage respons is bijvoorbeeld één van de bedreigingen van de enquête methode. In Nederland (en ook internationaal gezien) is een trend te zien van steeds afnemende respons. Een respons van 50% is tegenwoordig een relatief hoge respons. Non-respons vormt een probleem indien er een selectie plaatsvindt van mensen die wel en niet op de vragenlijst reageren op basis van de centrale onderzoeksvariabele. Een voorbeeld hiervan zou zijn dat alleen gezonde mensen de gezondheidsvragenlijst invullen en terugsturen en de ongezonde mensen niet reageren. Een gevolg hiervan is dat op basis van de enquête uitkomsten de algemene gezondheid beter wordt geschat dan daadwerkelijk het geval is. Een andere bedreiging van de enquête methode is de kwaliteit van de zelfrapportages. Mensen rapporteren zelf hun eigen gedrag. Vaak wordt veel van het cognitieve proces gevergd: mensen moeten zich bepaald gedrag herinneren over een bepaalde tijdsperiode, bijvoorbeeld hoe vaak ze afgelopen half jaar alcohol gedronken hebben. De vraag is in hoeverre mensen zich dit goed kunnen herinneren en een adequate schatting kunnen geven. Een andere factor die een rol speelt bij de kwaliteit van zelfrapportage is dat mensen niet altijd bereid zijn om het juiste antwoord te geven. Dit kan doordat mensen zich schamen voor hun gedrag en/of dat mensen geneigd zijn zich te conformeren naar de algemeen geldende normen en waarden en hierdoor sociaal wenselijk antwoorden.

In dit proefschrift worden deze twee bovenstaande bedreigingen (non-respons en kwaliteit van zelfrapportage) voor het enquêteonderzoek nader onderzocht. Aanleiding voor dit onderzoek is dat het Instituut voor VerslavingsOnderzoek (IVO) betrokken is geweest bij het uitvoeren van een aantal enquêtes over alcohol en drugs (als onderdeel van de algemene gezondheid), en dan met name specifiek gericht op probleemdrinken en gerelateerde problemen rondom dit gedrag. Vergeleken met een eerdere alcoholenquête in 1981 (door Garretsen, 1983) kampte een vergelijkbare alcoholenquête uitgevoerd in 1998 (door Bongers, 1998) met een aanzienlijk lagere respons (70% versus 44%). Verschillende verklaringen werden voor deze lage respons gezocht: komt het door de algemene nationale trend van afnemende respons of door een andere afzender van de vragenlijst (GGD versus IVO,

Instituut voor VerslavingsOnderzoek). Ook was de vraag in hoeverre deze lage respons de enquête uitkomsten kan beïnvloeden. Daarnaast was er een toenemende aandacht voor de kwaliteit van de zelfrapportages: is het mogelijk om alcoholmaten verder te verfijnen met het doel de kwaliteit van de zelfrapportages te vergroten?

Het doel van dit proefschrift is tweeledig:

- Het verkrijgen van meer inzicht in het meten van de nonresponsfout en in het vergroten van de respons:
  Hiervoor werden een tweetal studies uitgevoerd waarin de nonresponsfout gemeten werd aan de hand van een nonrespons follow-up onderzoek en waarin de 'tweezijdige nonresponspatroon' theorie en het 'continuüm van weerstand'model werden getoetst. Daarnaast werd een studie uitgevoerd waarin verschillende kenmerken van de enquête werden gemanipuleerd om te toetsen wat voor effect dit heeft op de respons.

- het verkrijgen van meer inzicht in mogelijkheden om de kwaliteit van de zelfrapportages te vergroten:
  Hiervoor werden twee studies uitgevoerd waarin geheugensteuntjes in alcoholvragen werden toegevoegd en waarin een nieuwe vragenlijst om de jaarlijkse alcoholconsumptie te meten werd ontwikkeld om het cognitieve proces te vergemakkelijken.

Het proefschrift bestaat uit 2 delen waarbij in het eerste deel de nonresponsfout centraal staat (Hoofdstukken 2, 3 en 4) en het tweede deel zich richt op meetfouten door zelfrapportages (Hoofdstukken 5 en 6). Hieronder worden per hoofdstuk de belangrijkste bevindingen samengevat.

In **Hoofdstuk 2** wordt door middel van een nonrespons follow-up onderzoek onderzocht of de alcoholschattingen uit de enquête vertekend zijn door non-respons. Hiervoor hebben we een steekproef uit alle nonrespondenten onverwacht thuis bezocht en korte vragen over redenen van niet-deelname en over alcoholconsumptie gesteld. Uit dit nonrespons follow-up onderzoek kwam naar voren dat geheelonthouders en matige drinkers oververtegenwoordigd zijn onder de nonrespondenten. Vervolgens is de theorie van het tweezijdig nonresponspatroon getoetst. Volgens deze theorie is de nonrespons oververtegenwoordigd in de extremen van de centrale onderzoeksvariabele vanwege twee twee aspecten, namelijk interesse in het onderwerp en sociale wenselijkheid. Mensen die niet geïnteresseerd in het onderwerp van de enquête zullen niet reageren. Bij sociale wenselijkheid spelen de algemeen geldende

normen en waarden van het onderzoeksonderwerp een rol. Mensen die niet voldoen aan deze normen en waarden zullen geneigd zijn sociaal wenselijk te antwoorden óf zullen de vragenlijst niet beantwoorden. Deze twee aspecten van het onderzoeksonderwerp bevinden zich aan beide extremen van de onderzoeksvariabele. Passen we de theorie van het tweezijdig nonresponspatroon toe op de alcoholenquête, dan zullen de mensen die niet drinken niet reageren, immers zij zullen niet geïnteresseerd zijn in het onderwerp alcohol. De andere kant van de onderzoeksvariabele is excessief drinken. In onze samenleving wordt alcohol drinken in het algemeen geaccepteerd mits dit met mate gebeurt. De hypothese is dan ook dat onder de nonrespondenten zich relatief meer excessieve drinkers bevinden. Resultaten van de nonrespons follow-up onderzoek laten zien dat relatief meer geheelonthouders zich onder de nonrespondenten bevinden, echter er is geen verschil in het percentage excessieve drinkers tussen de respondenten en non-espondenten. Hiermee hebben we geen overtuigend bewijs gevonden voor de theorie van het tweezijdig nonresponspatroon.

In hoofdstuk 2 hebben we de directe methode van het schatten van nonresponsfout toegepast, dat wil zeggen informatie over nonrespondenten komt rechtstreeks van de nonrespondenten zelf. In **Hoofdstuk 3** testen we het continuüm van weerstandsmodel, een indirecte methode om de nonresponsfout te schatten en vervolgens te corrigeren. Het model van het continuüm van weerstand gaat ervan uit dat er op het continuüm van weerstand twee uitersten liggen: mensen die altijd reageren en mensen die nooit reageren. Op grond hiervan kan gezegd worden dat schattingen van late respondenten (mensen die pas na de laatste herinneringsoproep reageren) lijken op schattingen van nonrespondenten. Immers, de groepen late respondenten en nonrespondenten hebben een soortgelijke positie op het continuüm van weerstand. Dit model hebben we in ons eigen onderzoek en in twee grotere soortgelijke alcoholonderzoeken getest. Resultaten laten zien dat er geen verschillen in alcoholconsumptie zijn tussen de drie verschillende respondentgroepen (groep 1: mensen die direct reageren; groep 2: mensen die na eerste herinneringsoproep reageren; groep 3: mensen die pas na de tweede herinneringsoproep reageren). Vergelijking van het percentage geheelonthouders tussen vroege respondenten en non-respondenten laat wel een verschil zien: er zijn relatief meer geheelonthouders onder de non-respondenten. Op basis van de resultaten kunnen we het model van het 'continuüm van weerstand' niet bevestigen. Het sturen van herinneringsoproepen leidt niet tot minder vertekende resultaten. Het blijkt wel een effectieve strategie om de steekproefgrootte te vergroten: de twee herinneringsoproepen in ons alcoholonderzoek leiden tot ongeveer 23% van de totale respons (ongeveer 43%).

Het is duidelijk dat het beter is om een vertekening van de resultaten door nonrespons te voorkomen en vooraf al een aantal responsverhogende strategieën toe te passen op de vragenlijst. In **Hoofdstuk 4** manipuleren we een aantal kenmerken van de vragenlijst met het doel het effect op de respons te meten. De achterliggende theorie is de sociale ruiltheorie van Dillman waarin drie elementen voor deelname aan onderzoek belangrijk zijn, te weten beloning, kosten en vertrouwen. Een interessant onderwerp of kans op een cadeautje bij deelname is een beloning. Kosten zijn bijvoorbeeld de tijd die iemand kwijt is aan het beantwoorden van de vragenlijst of de ingevulde vragenlijst per post weer terugsturen. Een voorbeeld van vertrouwen is bijvoorbeeld de afzender van de enquête waarvan de respondent het een legitieme organisatie vindt om een onderzoek uit te voeren. Om deelname aan het onderzoek te bevorderen dienen beloning en vertrouwen zwaarder te wegen dan de kosten. Op basis van Dillman's theorie hebben we in totaal vier kenmerken van de vragenlijst gemanipuleerd: afzender, lengte, titel van het onderwerp op het voorblad van de vragenlijst en tot slot de lay-out.

De enquête is verstuurd door drie verschillende afzenders: een gemeentelijke gezondheidsdienst, een universiteit en het eigen instituut (Instituut voor verslavingsonderzoek). Op basis van eigen ervaring binnen het instituut was de hypothese dat de gemeentelijke gezondheidsdienst de hoogste respons zou opleveren. De resultaten bevestigden de hypothese. De gemeentelijke gezondheidsdienst had een hogere kans op respons, gevolgd door de universiteit. Het eigen instituut had de laagste kans op respons.

Lange vragenlijsten (19 pagina's) en korte vragenlijsten (3 à 5 pagina's) zijn vergeleken waarbij de veronderstelling is dat korte vragenlijsten een hogere kans op respons geven. De respondent van de korte vragenlijst wordt minder belast, waardoor de kosten lager zijn om te responderen dan een respondent die meer tijd en energie kwijt is om de lange vragenlijst in te vullen. Echter, de resultaten laten geen verschil in kans op respons zien tussen de lange en korte vragenlijsten.

Als derde variabele was de titel van het onderwerp op het voorblad van de vragenlijst gemanipuleerd. Vragenlijsten werden aangekondigd als alcoholonderzoek óf als gezondheidsonderzoek. De achterliggende gedachte was dat gezondheidsonderzoek een hogere kans op respons geeft omdat meer mensen meer affiniteit hebben met het onderwerp gezondheid. Ook hier laten de resultaten geen verschil in kans op respons zien.

Twee lay-outs zijn gebruikt: de ene lay-out is de traditionele lay-out die tot nu toe werd gebruikt en de andere lay-out is in boekvorm, heeft meer symboliek (bijv. pijlen en kaders) en achtergrondkleuren waardoor de vragenlijst op het eerste gezicht aantrekkelijker is om in te vullen. Vergelijking van beide lay-outs levert geen verschil in kans op respons op.

Concluderend, slechts één variabele, afzender van de vragenlijst, heeft effect op de respons.

In **Hoofdstuk 5** wordt getoetst of het verfijnen van twee alcoholmaten ('weekly recall' en 'quantity-frequency') door middel van het inbouwen van geheugensteuntjes leidt tot het vergroten van de kwaliteit van zelfrapportage. Zowel de 'weekly recall' als de 'quantity-frequency' hebben twee formaten: het 'traditionele' formaat en het 'aangepaste' formaat met meer geheugensteuntjes. Om te toetsen welke formaat leidt tot vergroting van de kwaliteit van zelfrapportages wordt gebruik gemaakt van de 'more-is-better' regel. De 'gouden standaard' bij deze regel zijn de alcoholschattingen gebaseerd op verkoopcijfers. Alcoholschattingen afkomstig van enquêtes dekken slechts 40-60% van de schattingen uit de verkoopcijfers. Daarnaast hebben we de itemnonrespons van de vragen met elkaar vergeleken. Tot slot hebben we de alcoholvragen ook getoetst aan de hand van cognitieve interviews. Het cognitieve interview bestond uit 2 delen: de respondent werd gevraagd de alcoholvragen hardop denkend in te vullen. Daarna werden op basis van observaties uit het eerste deel vragen gesteld over de interpretatie en formulering van de alcoholvragen.

Resultaten laten zien dat er geen verschil in alcoholconsumptie was tussen de twee formaten van de 'weekly recall'. Dit gold ook voor de 'quantity-frequency' maat, behalve dat er een verschil was in het aantal gemiddelde drinkdagen: het alternatieve formaat leidde tot relatief meer drinkdagen. Opvallend resultaat was dat er relatief meer itemnonrespons was bij de 'alternatieve' formaten. Door de geheugensteuntjes in de maten moesten mensen meer invullen, wat wellicht heeft geleid tot demotivatie.
Uit de cognitieve interviews kwam een aantal problemen naar voren:

- Mensen herinnerden hun alcoholconsumptie vaak door aan activiteiten te denken die op een bepaalde dag plaatsvinden. Het schatten van het aantal glazen alcohol was afhankelijk van de duur en gezelligheid van een activiteit.
- Het meest voorkomende probleem bij de quantity-frequency maat was het middelen van de alcoholconsumptie. Het middelen werd vaak niet gedaan.
- Het beantwoorden van de alcoholvraag wat er in een typische week werd gedronken ervoeren met name niet-reguliere drinkers als lastig.
- Atypische drinkgelegenheden werden vaak vergeten.
- De invulinstructie bij het traditionele formaat van de weekly recall leidde tot grote verwarring bij de respondenten wat weer leidde tot veel fouten.
- Het verzoek om een flesje of blikje om te rekenen naar 1.5 unit leidde tot veel rekenfouten.

Concluderend: De 'alternatieve' formaten leidden niet tot vergroting van de kwaliteit van zelfrapportages; het inbouwen van geheugensteuntjes leidde zelfs tot meer itemnonrespons. De cognitieve interviews gaven meer inzicht in het cognitieve proces van de respondenten.

In **Hoofdstuk 6** verkennen we een nieuw ontwikkelde vragenlijst om de jaarlijkse alcoholconsumptie te meten. Cognitieve interviews in eerdere pilot studies lieten zien dat mensen hun alcoholconsumptie herinneren aan de hand van activiteiten. Deze bevinding heeft al geleid tot alternatieve alcoholmaten, zoals de timeline followbackmethode van Sobell & Sobell (1992) die een kalendar als geheugensteuntje verschaft en de specific settings approach van Single & Wortley (1994) die alcohol drinken in een specifieke sociale context plaatst. De vragenlijst die wij ontwikkeld hebben stelt ook de sociale activiteiten centraal. De respondent wordt gevraagd naar de dagelijkse, wekelijkse gewoonten en daarna volgt een lijstje van persoonlijke (verjaardagen, vakanties) en publieke (bijv. koninginnedag) gelegenheden. De respondent wordt eerst gevraagd hoe vaak zo'n gelegenheid voorkomt en vervolgens moet de respondent aangeven hoeveel glazen alcohol er bij zo'n gelegenheid wordt gedronken. In de exploratieve studie worden de gerapporteerde jaarlijkse alcohol consumptie van de nieuwe vragenlijst vergeleken met de gerapporteerde jaarlijkse alcohol consumptie gebaseerd op de 'typical week' alcoholmaat. Resultaten laten zien dat de nieuwe vragenlijst ongeveer 30% meer alcoholconsumptie oplevert dan de 'typical week' en geven aanleiding om deze vragenlijst nader te ontwikkelen en uit te testen.

Tot slot eindigt het proefschrift met een algemene discussie, beschreven in **Hoofdstuk 7**, waarin de belangrijkste bevindingen uit de Hoofdstukken 2-6 worden samengevat en bediscussieerd. Daarnaast wordt een aantal methodologische tekortkomingen van onze studies genoemd en eindigt de algemene discussie met een aantal aanbevelingen voor toekomstig onderzoek en mogelijke praktische implicaties.

Dankwoord

Dankwoord

"Hoe gaat het nu met jouw proefschrift?" is één van de meest gestelde vragen die ik de afgelopen drie jaar op feestjes en ontmoetingen met anderen moest beantwoorden. Langzamerhand werd deze vraag ófwel vermeden door anderen ófwel door mij gruwelijk gehaat. Maar nu… ik moet eerlijk bekennen dat ik intens geniet van de enigszins verbaasde, ongelovige en bewonderende gezichten van mensen wanneer ik antwoord "Ik ben klaar".

Het afronden van het proefschrift heb ik niet in mijn eentje gedaan; anderen zijn er natuurlijk direkt en indirect betrokken bij geweest. Hiervoor wil ik een aantal mensen in het bijzonder bedanken.
]
Dike, mijn copromotor, allereerst wil ik jou bedanken. Door een onverwachts telefoontje van jou heb ik mijn besluiteloosheid over het al dan niet afronden van het proefschrift om kunnen zetten in een definitief besluit. Je hebt me tijdens deze weg ernaar toe zowel op inhoudelijk vlak als praktisch goed gesteund. Dank hiervoor. Daarnaast Henk, mijn promotor, ook jou wil ik bedanken voor de steun aan mij. Jij zorgde voor het optimisme en schroomde niet om op z'n tijd mij complimenten te geven voor het werk wat ik had gedaan. Dit gaf mij extra energie om door te gaan met waar ik mee bezig was. Tot slot, Harrie, mijn directe begeleider, het was soms moeizaam tussen ons, maar ik wil je bedanken voor de tijd en energie die je besteed hebt aan het laatste artikel en aan het geven van inhoudelijk commentaar op al 't geschrevene.

Mijn oud-collega's van het IVO wil ik bedanken voor de plezierige werkomgeving en in het bijzonder wil ik Aafje en Renske noemen. Aafje, lange tijd heb ik met jou een werkkamer bij het IVO gedeeld. Dit was ontzettend plezierig, mede omdat we in hetzelfde aio-schuitje zaten. Renske, jouw aanwezigheid bij onze borrels in Rotterdam zijn altijd heel gezellig. Jouw openheid en betrokkenheid waardeer ik zeer.

Ik werk nu ruim drie jaar samen met fijne collega's bij ZonMw, team Kwaliteit en Doelmatigheid. Dank voor jullie belangstelling afgelopen tijd voor mijn proefschrift-traject. Elsbeth en Liesbeth, bedankt voor jullie adviezen. Wij zaten met z'n drietjes in hetzelfde schuitje en hebben alledrie nu het proefschrift geschreven. Wat goed van ons! En natuurlijk Karen, ook jij bedankt voor de adviezen en de lieve woorden dat het allemaal goed zal komen.

Nu bevind ik me meer op persoonlijk terrrein. Allereerst ga ik terug naar mijn 'roots'. Kim en Marie-José, mijn trouwste vriendinnnen oet Mestreech. Bedaank

veur alle kaffee-en wandel tochte door Mestreech, zoetot ik noets vergeet wie sjoen Mestreech is. Ik kin noe get dèkser aofkoume. Janice en Martijn, Henriëtte en andere vrienden, bedankt voor jullie vrienschap.

Michelle, Vera en Karlijn, het begon bij de summer school van de Nihes opleiding en nog steeds hebben we de traditie van gezellige etentjes bij elkaar thuis in ere gehouden. Laten we dit vooral blijven doen.

Ik wil mijn schoonfamilie bedanken voor hun interesse en betrokkenheid bij het afronden van het proefschrift. Suzan, heel erg bedankt voor het ontwerpen van het mooie voorkantje van het proefschrift. Onze brainstorm in de Oude Marenpoort was zeer geslaagd! Dit bewijst dat een bepaalde hoeveelheid alcohol zeker de creativiteit niet in de weg staat.

Natuurlijk wil ik mijn eigen familie bedanken voor de nodige afleiding en gezellige weekendjes in Mestreech. Pap, volgens mij heb je altijd stiekem gedacht dat ik het proefschrift nog zou afmaken. Daarvoor ken je me toch te goed. Gaston en Bonnie, voortaan kunnen we we weer wat vaker afspreken; genieten van bloedstollende enge filmpjes in de bioscoop.

Tot slot wil ik mijn dankwoord richten tot een heel belangrijk persoon in mijn leven. Lieve Mathijs, ik heb je veel geplaagd met mijn dankwoord… wat je niet allemaal hebt moeten doen om die felbegeerde plek in mijn dankwoord te krijgen. Je weet natuurlijk wel dat je die klusjes niet hoefde te doen om in mijn dankwoord te komen want ik zou je vanzelfsprekend noemen in mijn dankwoord. Waarom? Omdat je mijn allerliefste maatje bent met wie ik samen een bijzonder leven deel. Bedankt voor jouw steun tijdens mijn vrolijke maar ook moeilijke momenten. Daarnaast ben ik je dankbaar voor jouw onweerstaanbare kunst om mij af te leiden; dit gaf mij weer de nodige ontspanning. Op het moment dat ik dit dankwoord schrijf, is er weer veel gebeurd…lieve Thijs... op naar ons volgend gezamenlijk ‘life event’!

Curriculum Vitae

Curriculum Vitae

Viviënne Lahaut was born on April 12, 1974 in Maastricht, The Netherlands. She graduated from secondary school in 1993 at the "Stedelijke Scholengemeenschap" in Maastricht.

In the same year she started her study in Health Science at the University Maastricht. There, she followed the school of Health Organisation, Policy and Economics. During the final year of her study she conducted a pilot study on integrated day care of children with mental handicap within a regular day care in Eindhoven.

In May 1999 she started as PhD student at the Addiction Research Institute in Rotterdam on nonresponse and data quality in mail surveys on alcohol consumption, as described in this thesis. In 2001 she obtained a master's degree in Health Services Research at the Netherlands institute for health sciences (Nihes).

Since April 2003 she has been working as a programme officer for the Health Care Efficiency Research Programme of the Netherlands Organisation for Health Research and Health Innovation (ZonMw) in The Hague.