UNIVERSITY OF SOUTHAMPTON

FACULTY OF LAW, ARTS & SOCIAL SCIENCES

SCHOOL OF SOCIAL SCIENCES

Modelling Survey Participation

in Surveys Involving Multiple Phases of Data Collection

by

Petra Marjut Johanna LAIHO

Thesis for the degree of Doctor of Philosophy

September, 2006

ABSTRACT

UNIVERSITY OF SOUTHAMPTON
THE FACULTY OF LAW, ARTS & SOCIAL SCIENCES
SCHOOL OF SOCIAL SCIENCES
SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
MODELLING SURVEY PARTICIPATION IN SURVEYS INVOLVING MULTIPLE PHASES OF DATA COLLECTION
by Petra Marjut Johanna LAIHO

This Thesis aims to link the theory of response effects (Sudman and Bradburn, 1974), the conceptual theory of survey participation (Groves and Couper, 1996) and the route map of social exclusion developed (Atkinson, 1998), extending the survey non-response framework for studying the associations between social exclusion and non-response. In the empirical part, we examine the Finnish Health 2000 survey data with direct linkage to auxiliary information at individual level. In addition, we have conducted an interviewer perception survey amongst the interviewers, who participated to the fieldwork of Health 2000. We model survey participation behaviour of individuals in the presence of high response burden, analysing survey attrition across multiple data collection phases. Using multilevel sequential logit modelling, we incorporate the interviewer level information into the survey participation analysis.

We have found that the survey participation behaviour of individuals is greatly affected by their socio-economic circumstances, social capital, and social connectedness. People with affluent circumstances are more co-operative than people with any of the social exclusion risk factors. We demonstrate that a single model oversimplifies the survey participation in a survey with multiple data collection phases. We show that the interviewer effect in face-to-face interviewing survey may impact participation at further data collection components, which by survey design are independent from the presence of the interviewers. Finally, we illustrate that the survey estimates can be improved, if the survey non-response propensity weighting is accounted for depending also on the characteristics and perceptions of the interviewers. This finding shows that the interviewer effect can contaminate the obtained survey information not only at individual level, but also at the level population distributions for the survey estimates.

Key words: non-response analysis, survey participation, interviewer effects, interviewer perception, response propensity, non-response adjustment, inverse probability weighting

List of contents

List of tables

List of figures

Acknowledgements

I thank with deepest gratitude my supervisor Prof. Chris Skinner for the advice and guidance he has provided throughout the process of supervising my PhD studies. Secondly, I thank my advisor Prof. Ray Chambers for the spark of statistical inspirations he has given. In the process of the upgrade, Prof. John Micklewright has given valuable comments and suggestions that have greatly helped me to develop the research. I would like to express my sincere thanks for him. When examining the distribution of the random effects in multilevel models, presented in Chapter 5, Prof. Risto Lehtonen has given valuable advice. I would like to thank him for the guidance, co-operation and support.

I would like to recognise the essential support from my sponsors and express my gratitude for the sponsoring organisations. This research has been supported with ESRC grant [ref. 2000050429] for the academic year of 2000-2001, and since the second year by student grants of The Finnish Cultural Foundation [ref. 00020740, 00030746, 00040846]. In addition, my employer Statistics Finland has supported me with 6 months study leave.

I express my sincere thanks for Statistics Finland and National Public Health Institute of Finland for allowing me access to their rich data resources. Especially, I would like to thank the Health 2000 team, Professor Arpo Aromaa and Chief Physician Seppo Koskinen for allowing access to the Health 2000 data. The help of Sirkka Rinne and Pirkko Alha on constructing and providing the data sources with background information has been most valuable. I would also like to thank Director Jussi Simpura and Director Riitta Harala in Statistics Finland, for allowing me to link auxiliary data to my research data set. In addition, I would like to thank Director Jussi Simpura for allowing me and my colleagues to design and conduct an interviewer perception survey, and for finding the resources to fund this survey.

General director Heli Jeskanen-Sundström, the leadership of Statistics Finland, the Head of Research of the Statistical R&D Timo Alanko, and my colleagues have given their continuing support. I am most grateful for their support and understanding. The International Household Non-response Workshop, many statisticians and social researchers have encouraged me in my research. I am also most grateful for all my friends, who I had the privilege to meet at the Division of Social Statistics during my studies, and who created the inspiring research atmosphere. I would like to thank especially Nikos, Gabi, Gail, and Amos for their friendship, care and support. In addition, I would like to thank Christina Thompson, the IT support, and the administration for providing the excellent research facilities at the Division of Social Statistics, University of Southampton.

I am forever indebted to my husband for his everlasting love and support; and for the encouragement, understanding and endless patience expressed by him, my parents, my parents' in-law, and my friends.

Chapter 1. Introduction

1.1 Background to this research

The effects of survey non-response on data quality and reliability are very important for those commissioning, carrying out or using surveys. Particularly, non-response influences the errors of survey estimates, which are critical for the usability of the data. The survey organisations and the clients are extensively anxious about declining response rates (e.g. de Leeuw and de Heer, 2002) as non-response increases potential for biased estimates (e.g. Kish, 1965; Groves, 1989, Dillman et al. 2002). A general concern is that non-respondents differ significantly from the respondents, which leads to biased survey estimates even with standard non-response adjustment. Regardless the long traditions in survey methodology (Smith, 2002 referring to Hansen and Hurwitz, 1946 and Politz and Simmons, 1949; and Groves et al. 2004 referring to Hansen, Hurviz and Madow, 1953), little is known why sub-population groups differ in their tendencies to co-operate in surveys. The emphasis of the thesis is on studying how to improve survey estimates by detecting and exploiting the factors behind non-response. We focus especially on studying the non-response in surveys with multiple data collection phases.

Most studies developing survey non-response theories have concentrated on household surveys excluding the institutionalised population (e.g. Groves and Couper, 1995 and 1998; Lepkowski and Couper, 2002). However, the need of general population surveys is gradually increasing. For example, Riedel-Heller et al. (2000) argue that to make international comparisons on health conditions meaningful, the inclusion of institutionalised individuals is crucial. Furthermore, the ageing of the populations in many societies increases the information needs of the elderly, many of whom are living in institutions. In addition, several recent studies have attempted to determine the socio-economic factors influencing the non-response. However, only a few analyses have addressed the relationship between social exclusion and survey non-response, namely Johnson et al. (2002), Groves and Couper (1998), Couper et al. (1997), and Mathiowetz et al. (1991).

The survey participation theories suggest that non-response depends on the characteristics of individuals, their immediate social networks such as families, households and dwelling units as well as local neighbourhoods. Moreover, non-response is suspected to depend on the interviewers (f. ex. Groves and Couper, 1995; de Leeuw and Hox, 1996; Campanelli and Sturgis, 1997). In order to detect the underlying reasons of survey participation, Lyberg and Couper (2005) state the importance of using process data in the analysis. Also, Little and Rubin (1987) emphasise that knowledge of the mechanism that has led to certain values or units being missing, is a key element both in choosing appropriate analysis methods and in interpreting the results. Groves and Couper (1995) argue that the theoretical bases in non-response adjustment should lie in socio-psychological theories that specify influences on human behaviour affecting survey participation. However, the conventional non-response adjustment methods applied in official statistics ignore the survey participation modelling and reject the process data.

1.2 General aims of the research

One aim of the thesis is to detect the effects of non-response in surveys involving multiple phases of data collection, and describe these effects using statistical models. Couper and Groves (1995) suggested that distinguishing the contacting phase from assessing co-operation could be beneficial for the non-response adjustment as this may bring in specific knowledge, characterising the differences of the sequential processes. Based upon this approach, we emphasise that conceptual frameworks of survey participation are developed mainly for the cross-sectional surveys with single data collection phase. Traditionally, these frameworks lack an element for assessing non-response in multiphase surveys. However, individuals, who initially begin to respond, may later cease their co-operation to all subsequent phases of data collection and drop-out of the survey. This type of data loss is also called survey attrition, which is commonly investigated in panel surveys. Following this, the purpose of the survey participation models developed in this thesis is to detect characteristics affecting the non-response in multi-phase surveys. In addition to that, these models aim to provide an estimate of each individual's probability to respond. We explore how the conventional models simplifying the participation can be developed further, to describe better the survey non-response in multiphase surveys. Therefore, we will ascertain whether the statistical models built upon the socio-psychological framework, have a potential to be more informative and realistic than the conventional models.

A second aim is to explore the significance of interviewer effects on non-response in multiphase surveys. We narrow the research focus on surveys in which the initial interview is followed by other data collection components. These subsequent components are generally administrated by other than the interviewers, for example, by the respondents themselves filling self-completion questionnaires or by the medical field workers collecting measurements. We aim to develop models for studying how influential the interviewer effects are in relation to other factors associated with non-response in multiphase surveys. We extend the current models of interviewer effects on survey participation to multiphase context. Hence, we analyse how far the interviewers effects can be observable in multiphase context, especially, after the interviewer administration has ceased in data collection.

Thirdly, we investigate whether the survey estimation could benefit from the socio-psychological survey participation theories as hypothesised by Couper and Groves (1995). At present, the use of survey participation modelling in the adjustment of survey estimation is more of an idea than a generally accepted practise in statistical offices. Thus, we explore whether the standard non-response adjustments can be improved by incorporating more knowledge of the survey participation theories into the non-response adjustment methods. Subsequently, we explore how to correct for the observed interviewer effects in survey participation. Finally, we evaluate interviewer performance as well as the bias of survey estimates by analysing the non-response reduction efforts and the conversion rates. In this evaluation, we contrast the early respondents to both the late respondents and the remaining non-respondents. We aim to analyse the impact of boosting the response rates by estimating the bias of the survey estimates.

1.3 Specific objectives and hypothesis of the research

Currently, most social and health surveys in official statistics are restricted to the household population. However, household population surveys may provide inadequately representative information about the society as a whole. For example, solely the aging of population is expected to increase the proportion of institutionalised people in the future. This may affect the balance between household and general population. We consider the possible differences in survey participation between general population surveys and household surveys. One objective of this thesis is on studying the differences between household population, institutionalised people and those of which little *known* information exists in the auxiliary data sources.

We will use the Health 2000 survey from Finland as the empirical data for our analysis[1]. The Health 2000 survey is a general population multiphase survey, the contents of which are typical to many national health surveys (Aromaa and Koskinen, 2002). These types of surveys often collect data with mixed data collection methods and techniques. The diversity of the methods is required to capture the requested information from the respondents in most suitable manner to prevent measurement errors. Therefore, the results of this thesis can provide useful information in the field of non-response analysis and adjustment for other health surveys and possibly also for some social surveys using multiphase data collection.

We have chosen the data set of the Health 2000 survey from Finland in order to enrich the knowledge of survey participation in general population surveys as the data structure allows us to use direct matching to a rich source of auxiliary statistical and administrative registers. The auxiliary information exists at the level of individuals and dwelling, for which we have micro-level identification codes. Subsequently, we can aggregate the data to the level of register-derived families or households, dwelling units or geographical areas. The auxiliary information describes demographic, socio-economic, health and living conditions, for example. In addition, information is available on the interviewer's characteristics, and their professional attitudes, which were measured in interviewer perception survey.

A particular focus of this thesis is on studying whether social exclusion is a crucial factor affecting survey participation in health and social surveys. We are interested in exploring to which extent social exclusion can be relevant to the problem of survey non-response. Socially excluded people have been found to have lower level of participation in civil society (e.g. Aasland and Fløtten, 1999) and lower survey participation rates (e.g. Groves and Couper, 1998). Similarly, the lack of further education, for

[1] The Health 2000 survey was organised by National Public Health Institute of Finland. The interviewers of Statistics Finland conducted the health interviews. The medical examinations were organised by the National Public Health Institute.

example, has been found to be associated with level of response (e.g. Gray et al., 1996; Couper, 1997) and social exclusion (e.g. Walby, 2000). Socially excluded people have experienced diverse severe difficulties that have lead to isolation. In this context, negligible commitment to participation in governmental surveys may seem as a trivial problem. This can lead to a vicious circle if the policy evaluations of societies fail to capture the poorly perceived conditions and the trends in social dysfunction. Furthermore, this can lead to biased estimates or too narrow distributions in social and health surveys. For example, surveys may fail to estimate health or social inequalities. Based on the findings from the sociological studies we look into how the auxiliary information could be exploited for profiling those that have high risk for social exclusion. Subsequently, we study whether the risk groups identified by profiling predict non-response in multivariate context. Later we will ascertain whether this increases the problem of biased estimates.

In addition to the initial non-response, we are interested in studying the associations of survey attrition with increasing response burden (Sharp and Frankel, 1983) in cross-sectional multiphase surveys. In these surveys the method, the administration and the data collection technique vary across the data collection components. This implies diverse nature of the tasks within the survey request. Sudman and Bradburn (1974) have demonstrated that the nature of the task is a significant aspect of the response process. Therefore, we aim to explore different dispositions in response behaviour due to a set of background factors. Furthermore, we aim to take into account the observable self-selection in the remaining respondents exposed to the survey request. We suspect that in multiphase surveys, the use of various data collection components will cumulate perceived response burden, and this can escalate the estimation problems. In addition, the concurrently varying data collection techniques and administration methods are expected to lead into diverse nature of task and differential response structure for the key survey statistics. Consequently, the non-respondents and attritors can contain a very heterogeneous mix of members from different sub-populations, while those who co-operate fully can actually over-represent the average population. We aim to develop a modelling procedure suitable for multiphase surveys that estimate the level of co-operation for individuals sampled. We will study whether the increased response burden at later stages will increase the non-response of those who are in high risk of being socially excluded.

To evaluate the overall fieldwork performance, the survey participation models are extended to explore the presence and impact of possible interviewer effects in relation to other factors. The research on interviewer effects has also explored interviewer perceptions and professional attitudes in conjunction with their attributes. Our objective is to study how influential the interviewer attitudes are in relation to other factors associated with initial non-response in multiphase surveys. The professional attitudes were measured in an interviewer perception survey, conducted after the fieldwork of collecting our empirical data[2]. We replicated both the set of questions developed by Lehtonen (1996), and the ones analysed by Couper and Groves (1992). In addition, the questions partly coincide with those analysed

[2] The interviewer perception survey was designed by a team of survey methodologists in Statistics Finland: Nieminen, Laiho, Lehtonen and Vikki.

by Hox and de Leeuw (2002). We then model the relationship between interviewer factors on their field performance. We aim to explain whether the interviewer attitudes can predict the interviewer response rate in the initial data collection phase. Secondly, we compare whether interviewer attitudes influence the survey performance more than the interviewer characteristics. Thirdly, we are interested to investigate how significant the interviewer attitudes are in multiphase surveys, when the response models contain background information also at the level of individual, dwelling unit and geographical sub-population. Fourthly, we examine how to adjust for the non-response of the interviewers to their perception survey, in order to exploit the data correctly for the modelling of the impact of interviewer effects on the survey participation of the individuals sampled to the actual survey.

Standard non-response adjustment methods are based on simple implicit or explicit non-response models assuming that non-response can be adjusted via weighting for population totals, margins or cells. The starting point for the adjustment is generally the concern that non-response is not random. Non-response can be due to complex factors than can be described by limited set of variables in the conventional methods. Thus, we explore whether the adjustment methods incorporating the non-response models based on theories from social sciences can produce less biased estimates than the standard methods, as suggested by Groves and Couper (1995). We also aim to evaluate the magnitude of the survey bias and test how sensitive the survey estimates are in relation to the non-response in multiphase surveys. For this analysis, we use selected variables from the registers that correlate with the survey variables and prevalence information on long-term diseases. The weighting methods are then examined using theoretical comparison and quantitative indicators for weight comparison developed by Särndal and Lundström (2005), who also hypothesise that the use of auxiliary information improves the estimator more than the mathematical methods. We will investigate whether this holds in our data using the indicators for weight comparison.

We also evaluate the success in non-response reduction efforts carried out by other than the interviewers. The survey organisations and the clients tend to use the response rates as simple indicators of data quality. Thus, they are willing to invest additional resources to boost the response rates. However, the exploitation of data obtained by the non-reduction efforts is still relatively undeveloped. For example, the non-response reduction efforts often use shortened questionnaires and different data collection modes than the original data collection. This can cause problems with internal data comparability by introducing mode and context effects. In addition, the contradictory results indicate that boosting of the response rates can actually increase the bias of survey estimates (e.g. Stang and Jöckel, 2003). Even so, few non-response bias studies have used the late respondents even as a representation of the remaining non-respondents to estimate the non-response bias (e.g. Lin and Scaeffer, 1995; Lynn et al., 2002). However, this assumption can severely conflict with the survey participation theories. Therefore, we study from the survey participation theories the possible disconnection of the early respondents and the late respondents. Then using empirical data, simulate the impact of non-response reduction methods on survey estimates and estimate the survey bias using highly correlated proxy variables.

1.4 Statistical models

In this thesis, all non-response models are based on logistic type regression, which have been recommended for the non-response analyses by Little (1986). The simple logit model is compared to more disaggregated models. For example, the sequential logit (Mare, 1980) and the multinomial logit (e.g. McCullagh and Nelder 1989) can potentially capture different underlying factors in the survey participation steps. Furthermore, the conditional sequential modelling predicts the probability of success at each data collection phase. This allows for analysing the success and the maintenance of achieved co-operation through the data collection. The aims of these models are to explain and describe the non-response behaviour and to provide the means for the most appropriate non-response adjustment. Therefore, the flow of the survey request and the data gathering process is studied in order to understand the different mechanisms affecting the participation behaviour. In this context, the applicability of various other modelling techniques on survey participation analyses are explored, such as, the event history analyses (Allison, 1984; Cox and Oakes, 1984; Kleinbaum, 1996), and the path models (Jöreskog and Sörbom, 1989; Hatcher, 1994; Bentler, 1995). The usability of these models is dependent on the structure of non-response.

In addition, the use of multilevel modelling in survey non-response analyses has been beneficial, since the survey processes contain many hierarchical features (see e.g. De Leeuw and Kreft, 1986; Bryk and Raudenbush, 1992; Longford, 1993; Goldstein, 1995). For example, in surveys on individuals, non-response may occur at the individual or the regional level, which categorises the primary sampling unit. In interviewer surveys, the interviewer effect can be a significant factor in gaining co-operation, which the multilevel modelling techniques can control for. Especially in health surveys, both the interviewers and the medical staff can collect the data, which leads to an even more complex interaction of human factors affecting survey participation.

1.5 Scope of the thesis

We will focus on sampling surveys of individuals with multiphase data collection. We define the non-response as non-contact or non-cooperation with the survey request. More precisely, in multiphase surveys the non-response is defined as an observed failure to obtain adequately information at the initial interview or at any subsequent data collection phases. We assume a situation for survey estimation in which external factors such as a statistical act inhibits to impute for the missing data at individual level. Therefore, unless adjusted for the missing data, the survey attrition may cause severe problems for the analysts, as most of the multivariate regression methods omit the incomplete records from the analysis even due to a single missing item. In addition, the appropriate construction of weights and possibly compiling multiple weights in relation to different analysis situations of the final data are considered. The statistical models used in this thesis assume that there will be some auxiliary information available of the target population. Furthermore, the models can be used more efficiently, if this information can be linked to the survey data at individual level for the purpose of non-response adjustment.

1.6 Stages of research

In Chapter 2, we assess the current knowledge about survey response behaviour and aim to develop further the existing theoretical frameworks and reflect for the potential of auxiliary data in statistical modelling. The basis for the theoretical framework of this research lies on the response effects theory by Sudman and Bradburn (1974), the conceptual framework for survey co-operation by Groves and Couper (1995) and the predictor grouping by Lepkowski and Couper (2002). In Chapter 3, we will discuss surveys using multiple data collection phases and describe the survey-taking climate in Europe and Finland especially. Subsequently, we will give motivation for using the Health 2000 data in the empirical analysis of this thesis. We will present the structure of the auxiliary information available and the related interviewer perception survey. In Chapter 4, we will investigate to which extent the sequential modelling improves the performance of statistical models and our estimates on the response probabilities. These models benefit from direct matching, linking the sample data to the registers at individual level. Furthermore, we will assess how significant the social exclusion factors are in relation to other features that also diminish the survey response in multiphase surveys.

In Chapter 5, we will explore whether the respondent-interviewer interaction may affect the willingness of the respondents to continue with the survey co-operation after the initial interviewing phase in multiphase surveys. Therefore, we extend the current methods of analysing the interviewer effects on survey participation. To study the effects more precisely, we explore how the initial interviewer-respondent interaction, influence the maintenance of the co-operation at later data collection phases. The focus of Chapter 6 will be in examining whether the non-response adjustments of multiphase surveys can be improved by the non-response modelling. We will conduct a comparison on the selected weighting methods. In this comparison, we compare the performance of probability weighting to other more conventional methods. Finally, we will study the implications of the non-response reduction from two perspectives. Firstly, we investigate whether the conversion rates per interviewer vary significantly and whether the interviewer characteristics and the attitudes explain these differences. Secondly, we aim to determine whether the non-response reduction efforts decreased or increased bias of the estimates in our empirical data using the non-response adjustment methods developed.

1.7 Summary

This thesis analyses the survey participation behaviour, linking the statistical models to the survey participation theories. We extend these models to analyse the interviewer effects on survey participation in multiphase surveys. We then investigate the weighting methods that incorporate estimates of the non-response models. Subsequently, we explore whether these methods improve the survey estimates in comparison to the conventional non-response adjustment methods. Finally, we evaluate the impact of non-response reduction efforts on survey estimates. Our study focuses on the following questions:

- We build quantitative profiles of socially excluded people using auxiliary quantitative data to test whether people in high risk of social exclusion become also more likely non-respondents (Chapter 3).
- We explore whether the non-respondents in multiphase surveys can be characterised. Furthermore, we analyse if their profiles differ depending on the data collection phase at which they cease the co-operation. (Chapter 4).
- We determine the usability of sequential modelling in multiphase surveys and compare the method to other plausible modelling techniques (Chapter 4).
- We study whether the cumulating response burden lead to increased non-response at later data collection stages. Subsequently, we aim to quantify the associated factors and estimate the response probabilities for the purpose of non-response adjustment via weighting (Chapter 4).
- Our analysis extends the current interviewer effect models to study how far the characteristics, the attitudes and the assignment allocation of interviewers may influence the survey participation in multiphase surveys (Chapter 5).
- We explore if the response probability weighting can improve the survey estimates of the prevalence of selected long-term illnesses, and health related information on the use of social benefits. In addition, we investigate how the non-response structures can affect the estimates (Chapter 6).
- We investigate whether the non-response reduction efforts improve the quality of survey estimates or whether these efforts actually increase the survey bias. We assess the impact, the efficiency and the usability of non-response reduction efforts in terms of aiming for unbiased survey estimates. (Chapter 6)

2. Review on Survey Participation Theories

2.1. Introduction

The motivation for studying the non-response is often demonstrated by findings on the declining survey participation (e.g., Steeh, 1981; Goyder, 1987; Bradburn, 1992;Groves and Couper, 1998; de Leeuw and de Heer, 2002). The research activity in the area has increased, presumably, because of this trend or believed future continuity of the trend. In addition, the impact of non-response to the quality of statistics, and the intensifying use of surveys in policymaking has given motivation to study and report on the non-response. However, not all studies show an increasing trend on survey non-response. After comparing European response trends in 1980-1991, de Heer and Israëls (1992) indicate that there is no general negative trend in Europe caused by changing attitudes. Their findings are similar to the ones by Lievesley (1988) who de Heer and Israëls quote '*not the attitudes of people have changed so much, but the survey organizations have not adapted themselves enough towards changing circumstances*'. In addition, Lynn et al (2002a) have shown that while the refusal rates have increased, the non-contact rates have actually decreased in the United Kingdom. This may indicate that while the survey organisations have been able to exploit new techniques and methods for improving the level of contact, they have actually improved less the social skills of their interviewers than the skills needed for contacting people.

This Chapter reviews the participation theories and the findings from the applied survey research. We focus on applications in the area of social and health surveys. We emphasise surveys of general populations rather than household surveys, although the published research has more weight on the latter. The aim is to explore the elements that can provide useful for analysing the non-response in surveys with multiple data collection phases, which start with the interviewer-administrated phase. Our theoretical framework extends the theory of response effects (Sudman and Bradburn, 1974) to the survey participation. Sudman and Bradburn distinguished the role of the interviewee and the interviewer as well as the nature of the task, which typically differs in multiphase surveys. We merge their approach with the conceptual model of survey participation (Groves and Couper, 1995), which has been widely adopted in the survey literature. To capture the deeper meaning of characteristics of the respondents and the interviewers, we use the predictor grouping developed by Lepkowski and Couper (2002). Our approach permits to connect the survey design features and the survey performance data with the background characteristics in the presence of increasing response burden.

We study the reported differences of survey participants and non-respondents, and whether these findings are systematic across studies. We acknowledge that the comparisons across surveys and the findings from previous studies can be limited to some extent. For example, differences in survey design, target population, research hypothesis, use of auxiliary data, and whether the results are based on weighted or unweighted analysis can influence to what extent one can make generalisations and comparisons. Thus, the freedom to make generalizations of response behaviour can be more restricted than perceived in many non-response studies. We aim to take these factors of uncertainty into account in our comparative analysis. Subsequently, we will analyse the impact and the significance of the risk factors in multivariate context.

One section in this Chapter focuses on reviewing the survey participation and the topic saliency in health surveys especially. The main concern has been in assessing whether the health status of the individual is related with the survey response behaviour (e.g. Koponen and Aromaa, 2003; Cohen and Duffy, 2002). We will use health survey data from Finland for the empirical analysis in this thesis. We limit the review of survey experiences to cover studies from Northern America as well as Northern or Central Europe to prevent large cultural differences affecting the comparability of the results. We will also discuss partly the impact of survey environment and climate on survey participation in this Chapter, continuing in Chapter 3 where the empirical data is presented.

A special objective of this thesis is to investigate the survey participation in relation to social exclusion and relative well-being in the society. We will study how these concepts are defined in social psychology and review the related characteristics from the literature. Afterwards, we review how the relationship between social exclusion and survey non-response has been researched previously (e.g. Johnson et al. (2002), Groves and Couper (1998), Couper et al. (1997), and Mathiowetz et al. (1991)). Then, we aim to compile a set of profiles to indicate an increased risk for social exclusion. After that, we will review how these profiles coincide with the findings from the previous non-response studies.

Logistic regression has been widely used for the modelling of survey non-response (e.g. Goyder, Lock and McNair, 1992, Lehtonen 1996, and Groves and Couper 1998). We assess how logistic regression models have been applied in the literature for studying the patterns of non-response behaviour. We then investigate how the interviewer effects on survey participation have been studied previously, and what statistical methods have been found most suitable for this analysis. For example, the multilevel logistic models have been applied for hierarchical data to analyse the impact of interviewer assignments. In addition, we review how the professional attitudes of the interviewers have been measured in the past research. Following this, we describe how significant these attitudes have appeared to be in relation to other factors affecting the non-response.

We begin by defining how response is measured in surveys. Subsequently, we review the existing survey participation theories to build the theory of response behaviour in multiphase surveys. This framework is then applied to the analyses of survey participation and topic saliency in health surveys. After profiling those in high risk of being isolated in society, the theories are used to link the association between social exclusion and survey participation. We assess how the past research has analysed the interviewer attitudes, and which interviewer characteristics have been found to influence survey participation. Similarly, we describe the modelling techniques used in the applied literature for analysing non-response. Subsequently, we review the conventional non-response adjustment methods. To conclude, the aim of this review is to lay comprehensive foundations for the empirical work of this thesis. In the last section of this Chapter, we will discuss how the findings from the literature have given us guidance for formulating the research hypothesis of this thesis.

2.2. Defining survey participation and non-response in multiphase surveys

Non-response can be defined as a failure to obtain information from sampled and eligible units (Kalsbeek, 1980). An important distinction, especially for multiphase surveys, is to separate complete and partial response (AAPOR, 2000). de Leeuw, Hox and Huisman (2003) define partial non-response by time dependency; after a certain point in time all data is missing from the unit. Respondents may break off during the interview, or in between subsequent data collection components. Another example of partial non-response is panel attrition, which is by definition time dependent (Fitzgerald, Gottschalk and Moffitt, 1998). According to Lynn, et al. (2002a), in practise all surveys accept some degree of partial information. Hidiroglou, Drew and Gray (1993) classify a response as partial, if the respondents provide usable information for some items but not for others. Survey participation literature (e.g. Hidiroglou, Drew and Gray, 1993; Groves and Couper, 1998; Smith, 2002) has shown that survey participation can be a very complex process, and there is thus a need for detailed analysis. Response is often measured in response rates, as a proportion of successful units of the eligible units. However, basic response rates alone can hide differences or patterns of missing data, possibly, leading to incorrect conclusions and inappropriate actions. Further analysis and error correction methods are needed to provide adequate survey estimates. Non-response analysis can be seen as a tool to examine the successfulness of the data collection (Biemer and Lyberg, 2003).

Survey non-response has a holistic impact on decreasing the quality of statistics. Groves and Couper (1998) emphasise that in addition to reducing accuracy, non-response also affects timeliness, as the fieldwork period may lengthen when aiming for higher response rates. Besides, survey data processing time increases while developing suitable ad hoc methods for the treatment of missing survey data. Kalton and Kasprzyk (1986) have reviewed the demands of missing data for the extensive use of advanced methodology in the imputation and the weighting of survey data. Non-response may also impact relevance, comparability, and coherence of statistics. Groves (1989) argue that non-response rates are often mistakenly used as a measure of total quality of the survey statistics. Low response rate may incline low levels of achieved co-operation, or alternatively, strict survey policy, possibly accepting only cases with full response in comparison to other patterns. Similarly, a high response rate can indicate successful fieldwork performance, or alternatively, it can suggest loosely defined rules for acceptable response. Also the validity of the design, survey coverage, sampling, measurement, processing and adjustment can introduce errors to survey estimates (Groves et al., 2004).

Applying further non-response reduction efforts adds survey costs (Groves, 1989; Lessler and Kalsbeek, 1992). Cost and compliance cost i.e. the burden on respondents are not usually considered to be quality attributes, but they are components of the total quality (Eurostat, 2001). Mason, Lessler and Traugott, (2002) distinguish the costs for conducting a survey into several components: to fixed costs such as research staff, questionnaire design and the overhead for an interview facility as well as to variable costs that depend upon the sample size and survey rules on contacting sample members.

Comparisons between early and late respondents show that there can be significant differences in the distribution of survey estimates. Some studies have speculated that the late respondents could actually resemble the non-respondents, as the early and the late respondents differ significantly. Thus, the late respondents have been used as a representation of the remaining non-respondents to estimate the non-response bias (e.g. Lin and Scaeffer, 1995; Lynn et al., 2002b). In contrary, Stang and Jöckel, (2003) demonstrate that boosting of the response rates can actually increase the bias of survey estimates. In addition, the proportion of item-missing information can be higher among late respondents than among early respondents (Stang and Jöckel, 2003 referring to Helasoja et al. 2002). This finding is connected with the context and the mode effects introduced to the data, by shortened questionnaires and mixed use of data collection modes in non-response reduction efforts. These effects can actually deteriorate the internal data quality. In meta-analysis on mode effects, de Leeuw (1992) emphasises that the data collection modes differ both in response rates and in levels of item non-response.

Gray et al. (1996) studied survey participation in multistage Health and Life Style Survey in order to assess to what extent non-response may have been connected with the topic of the survey. Non-response was examined at various phases of the survey, first by demographic and socioeconomic characteristics of the sample and secondly through a selection of key health related and psychological variables. They emphasize that a detailed non-response analyses by non-response groups highlights how the bias incurred was different for each of these groups. The result suggests focusing the non-response reduction efforts to certain high bias categories, only some of which can be tackled through improved fieldwork procedures. The study of the difficult categories of respondents also shifts the focus from simply a matter of high response rates to one of bias reduction. Gray et al. (1996) emphasize that a greater effect will be gained by raising the response rate of a high bias category than by raising the response rate of low bias category by the same amount.

2.3. Theory of response behaviour in multiphase surveys

Groves and Couper (1998) emphasise the importance of viewing survey participation from the sampled persons' and households' perspective. In depth analysis and statistical modelling is needed to gain information on the underlying factors for missing data. In this section, we review factors that may affect survey participation tendencies. The research interest is on finding significant factors explaining the survey participation behaviour that could possibly be used in response probability modelling. According to Groves, Cialdini and Couper (1992), the development of descriptive tools for dealing with non-response has begun by Deming (1953) and Hansen, Hurwitz, and Madow (1953). However, there are some very early remarks and guidance on non-response in the early literature of statistics, for example, dealing with refusals in censuses by Pidgin (1888). The research of survey non-response started formulating in the emergence of polling in the 1930's, and since the 1940's has expanded to a wide research field combining cumulating knowledge from different fields (Smith, 2002).

Sudman and Bradburn (1974) emphasise that both the respondent and the interviewer play an important role in the success of their interaction in terms of survey participation or unit non-response.

Sudman and Bradburn (1974) studied the structured characteristics of the interviewing situation. They suggested that these contribute in important ways to the magnitude and variance of the response obtained and that it is the task of methodological research on response effects to study the nature and magnitude of these effects. Their theory on response effects was developed upon the earlier works of Hyman (1954), Kahn and Cannell (1957), and Scheuch (1967). This theory connects the survey design features with the respondent interviewer interaction theories suggesting that there are three conceptually distinct sources of variance, relating to the response effects in the given situation. These sources of variance can be projected via variables that are derived from: (i) the nature and structure of the task, (ii) the characteristics of the interviewers, and (iii) the characteristics of the respondent. Consequently, the task variables can be further divided into variables relating to the structure of the task and method of administration, problems of self-presentation on the part of the respondent, as well as to saliency of the task to the respondent. This approach may be very appropriate for analyses of non-response in multiphase surveys. We will thus separate in our models the data collection phases by the task and method of administration.

Sudman and Bradburn (1974) studied the response and interviewer effects in general. In contrary, Groves, Cialdini and Couper (1992) focused their socio-psychological research a set of conceptual developments and experimental findings that appear to be informative about causes of survey participation, which lead to the development of conceptual model for contacting sample households and conceptual framework for survey co-operation by Groves and Couper (1995). They present an integration of that work with findings from the more traditional statistical and survey methodological literature on non-response, and, given the theoretical structure, deduce potentially promising paths of research toward the understanding of survey participation. Groves, Cialdini and Couper (1992) suggest integrating the observed influences of socio-demographic characteristics and survey design factors, with the less observable impact of the psychological components of interactions between interviewer and respondent. This applies as well in psychological concepts relevant to survey participation like compliance with requests, helping tendencies, and opinion change theory. Factors influencing survey participation according to Groves, Cialdini and Couper (1992) are: (i) societal-level factors, (ii) attributes of the survey design, (iii) characteristics of the sample person, (iv) attributes of interviewer, (v) respondent-interviewer interaction and (vi) compliance with request.

Groves and Couper (1995) formulated the conceptual structure and the theory of survey participation that we combine with the approach by Sudman and Bradburn (1974), presented in Figure 2.1. The effects on sample person's behaviour arise from multiple levels of aggregation of psychological and sociological phenomena. Factors affecting participation to surveys can be divided by whether they are under researcher's control or out of reach. The survey design and interviewers can be controlled for while the social environment and characteristics of the sampled individual or household are out of the control. Characteristics and behaviour of both the respondent and the interviewer impact to their interaction, which is the followed by a decision to co-operate or refuse by the respondent.

The limitation of the approach is that one has to satisfy with the measurable and comparable quantitative socio-demographic factors that have an association with the willingness to co-operate. Groves and Couper (1998) accentuate that socio-demographic factors should not be regarded as causes of the participation decision. Instead, these tend to produce a set of psychological pre-dispositions that affect the decision. These may also have an affect on the initial approach of the interviewer to the sampling unit. Goyder (1987) has emphasised the importance of taking into account past experiences of people, such as their previous participation to surveys. In multiphase surveys this can be extended to the experiences from past data collection components. Lepkowski and Couper (2002) have also included past survey experiences into division of survey co-operation by following components: socio-demographic and regional, community attachment, social and political integration, situational circumstances, survey experience and accessibility or willingness to be found.

When comparing non-response patterns in separate surveys the design features should be controlled for. The mode of the initial contact affects both the number of channels of communication between interviewer and respondent (Groves, 1978), the selection of persuasion strategies and the effectiveness of alternative strategies (Groves, Cialdini and Couper, 1992). Also, the topic of the survey may affect the respondents' level of interest in the survey. Recognising the survey design impacts is important, in comparative studies and in multiphase surveys with varying data collection modes across phases.

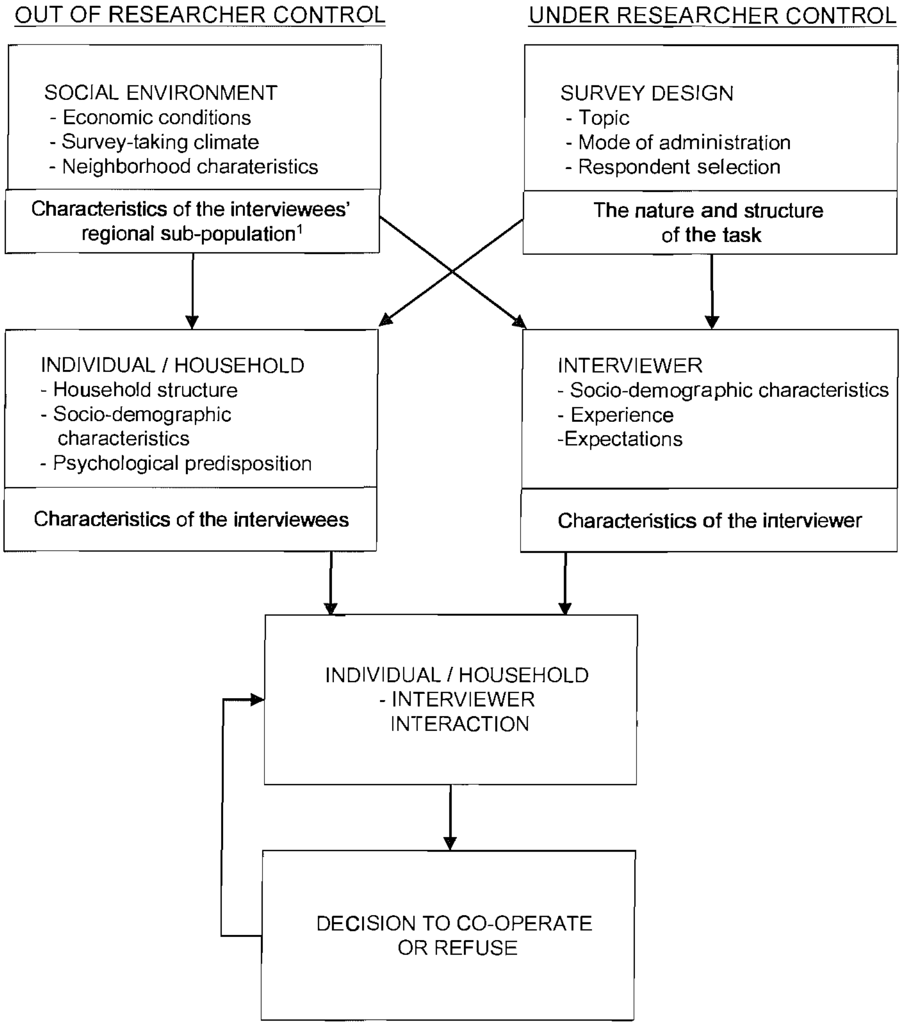
2.3.1. Characteristics of the sampled individuals and households

The aim of non-response analysis is to find causes of possible survey bias and to apply suitable adjustment methods to gain unbiased survey estimates. To develop surveys, it is important to characterise the risk groups of non-response. Initiated by Groves and Couper (1996), it has become popular to analyse the differences between non-contacts and refusals. Panel surveys have also studied differences between respondents to the first wave of data collection with the later waves (e.g. Gray et al., 1996, and Pickery and Loosveldt, 2000).

Demographic characteristics of individuals

Demographic characteristics, like age and sex, are among the most commonly studied factors of survey non-response. Influential associations between sex of individuals and non-response has been found, for example, by Groves (1989) referring to Smith (1979), DeMaio (1980), and Lindström (1983); as well as by Groves and Couper (1996), Couper (1997), Campanelli and O'Muircheartaigh, (1999), and Pickery and Loosveldt (2000). Findings indicate that women co-operate more than men in surveys, but it is not yet known clearly what causes this tendency. Also, little has been discussed on the informative manner to analyse the impact of age on the risk of survey non-response. Age has been used both as continuous and categorical variable. The categorisation of age can help comparison of research results across surveys with varying age restrictions. Categorisation can also introduce more power to study the response behaviour of young people versus middle aged versus elderly.

Figure 2.1 A conceptual framework for survey co-operation (Groves and Couper, 1995; links added to Sudman and Bradburn, 1974 (in bold))



NOTE: [1] Characteristics of the interviewees' regional sub-population is emphasized here as the characteristics derived from the people's living, economic and social conditions and their demographic features of people living in the same area with the interviewee tend to be more In survey participation than the regional division of geographical areas as such

Age has been found significant factor, for example, in Mercer and Butler (1967), Brown and Bishop (1982), Herzog and Rodgers (1988), Groves (1989), Groves and Couper (1996), Couper (1997), Campanelli and O'Muircheartaigh, (1999). Response has been found more difficult to achieve from the young and the elderly than the middle aged. Differences may result from underlying factors associated by age, such as the time use, physical and mental abilities, and civic duty. Groves and Couper (1995) emphasise that these factors can also conflict with each other as elderly people often have higher civic duty, but are less likely to participate, perhaps due to lower physical abilities and fear of crime. Groves and Couper refer to earlier studies by Miethe and Lee (1984) and Rucker (1990) showing that elderly people can be fearful for crime. Elderly may also be more concerned about their data protection, and consult their adult children how to react on the survey request. In the interviewer perception survey we conducted, we found that some adult children acted as gate keepers denying the participation of their elderly parent to respond to the survey. In the ageing societies, the surveys should take into account the possible reduction in capabilities and independency of some of the elderly population.

Groves and Couper (1996, 1998) and Gray et al. (1996) found the marital status of the individual to be a good predictor of the survey participation, controlling the findings for other factors in multivariate analysis. In assessment of panel data, Gray et al. (1996) found that single, widowed, divorced or separated had reduced contactability, and were less likely to be traced in between panels than people married or co-habiting. Groves and Couper (1998) found that the single had significantly lower estimated propensities to be contacted, and to co-operate with the survey request.

Previous research suggests the membership in language and ethnic minority groups to increase the risk of non-response. For example, Gray, et al. (1996), Campanelli and Sturgis (1997), Campanelli and O'Muircheartaigh (1999) indicate that members of ethnic minorities may have lower probability to participate to surveys in the United Kingdom. In Finland, the ethnicity has been studied in survey research using the definition of maternal language. Laiho (1998) and Lindqvist et al. (2001) have found that the Swedish speaking minority has slightly lower tendency to participate than the majority of Finnish speakers, while other foreign language speakers participated at significantly lower rate.

Socio-economic conditions of individuals

Studies of socio-economic conditions and survey participation have focused on the economic wealth and income, employment status, socio-economic class, and education, which are all interacted with each other. The income information can be available in studies analysing the panel attrition or in surveys with the possibility to link register, administrative or census information to the survey data. Direct linkage to the level of individuals sampled can provide reliable and rich information for the assessment of survey bias in terms of socio-economic characteristics. Brehm (1993), Campanelli and O'Muircheartaigh (1999), and Goyder (1987) have found the income to be associated with survey participation. Brehm (1993) assessed the impact of family income groups[1]. Over the years of 1978 to 1988 the co-operation had reduced amongst those earning less and increased amongst those earning more. Also, higher education has been found to be associated with increased survey participation, for example, in Brehm (1993), Gray et al. (1996), Couper (1997), as well as in Pickery and Loosveldt (2000). Similarly, the socio-economic group as well as the economic activity and experiences of unemployment have been observed to increase the propensity to respond in Gray et al. (1996), Couper (1997), Campanelli and Sturgis (1997), Campanelli and O'Muircheartaigh, (1999).

Some of these studies have exploited the regional information when the individual level information is not available. However, caution is needed to avoid the ecological fallacy when exploiting the regional information, which has not always received the attention needed. In the absence of auxiliary data, interviewer coding for characteristics of all sample members can be utilised, as suggested by Smith (1983). This can include information on housing by observations or information from neighbours as guided by Elliot (1991). Although, this enriches available information when auxiliary information is scarce, it may bear problems of reliability, subjectivity and measurement errors. Comparison of survey information to population totals is another plausible way to assess the associations of survey non-response with socio-demographic factors. This type of comparative analysis was conducted, for example, by Brehm (1993) in assessing the survey environment in the United States using data from the National Election Studies, the General Social Survey, and the Current Population Studies.

Characteristics and structure of households

The size of the dwelling unit, household or family has been hypothesised to increase the contactability as there are more people likely to be present during the contact attempts. The hypothesis has been supported by the findings of Kemsley (1976), Lievesley (1988), Gray, et al. (1996), Groves and Couper (1996), Groves and Couper (1997). Also the age of the household members have been found significant (e.g. Kemsley, 1976; Lievesley, 1988; and Groves and Couper, 1997). The age composition of the household members can reflect the dynamics of the household. For example, families with children have been found to be more co-operative than other households by Lievesley (1988), Groves and Couper (1997), and Couper (1997). The findings of Groves and Couper (1996, 1998) and Gray, et al. (1996) on the impact of marital status on the individual's propensity to respond is linked with the finding of Campanelli and Sturgis (1997) on the effect of the household type on survey response.

Characteristics of dwellings

The dwelling information has been found influentially explaining survey non-response. Non-response has been observed to be affected by the tenure (e.g. Gray, et al., 1996; Campanelli and Sturgis, 1997; Groves and Couper, 1997; Campanelli and O'Muircheartaigh, 1999). Also, the type of accommodation (in Campanelli and Sturgis, 1997; Campanelli and O'Muircheartaigh, 1999) and the number of rooms in accommodation (Campanelli and O'Muircheartaigh, 1999) have been found to be associated with the level of survey co-operation in multivariate context. Kemsley (1976) and Lievesley (1988) found also the quality and upkeep of housing to be directly associated with the survey participation.

2.3.2. Characteristics of the social environment

The social environment can be conceptualised at two broad levels. Firstly, societal-level conditions may facilitate or mitigate survey participation in a particular society. Secondly, local variations in the context of the community or local area level may shape the decision to participate or refuse (Groves and Couper, 1998). The survey participation is also affected by the general survey taking climate in the society (Groves, Cialdini and Couper, 1992). The survey taking climate can be dependent on the number of surveys conducted in society (the 'over-surveying' effect), the perceived legitimacy of surveys, and the public relations activities (de Heer and Israëls, 1992). In addition, the survey participation is dependent on the environmental influences, urbanicity effects and crime rates (House and Wolf; 1978) as well as on the social, economical and cultural climate (de Heer and Israëls, 1992). Following to the grouping of Lepkowski and Couper (2002) we review the characteristics of the social environment that have been found to be significantly associated with survey participation.

Social environmental factors relate to psychological reactions such as social isolation or fear of crime, which consequently may affect the level of co-operation. One is generally unable to directly measure

[1] The income limits in Brehm's study were restricted by the comparability issues across surveys. Three-level income variable contrasted the survey co-operation of those having income less than $10,000 and those with income above $25,000.

such attitudes or dispositions among non-respondents, a key factor underlying the intractability of the non-response problem. Even so, examining aggregate societal-level attitudinal variables and how they interact with non-response, may give us insight into the effect of such variables on non-response (Groves and Couper, 1998). The type of region and city, census or administrative records at small area level, interviewer description of the neighbourhood, and interviewer description of the dwelling unit can be used as aggregate level data of the social environment.

Voting behaviour and patterns of reported crimes in local areas reflect the social atmosphere linked with privacy and confidentiality concerns (e.g. House and Wolf, 1978; and Djerf, 2004). Level of urbanisation has been found to increase survey non-response (e.g. House and Wolf, 1978; Smith, 1983; Lievesley, 1988; Brehm, 1993; Gray, et al., 1996; Couper, 1997; and Groves and Couper, 1998). The urbanicity is a latent variable with underlying causes on the time use, and lower social connectedness, which is related to the population density, also found to reduce survey participation in Lievesley (1988), Gray, et al. (1996), Campanelli and Sturgis (1997), Groves and Couper (1998). Similarly, the size of the city is found to affect the survey response in Goyder, Lock and McNair (1992).

The demographic structure or the regional population may relate to survey participation behaviour. Lievesley (1988) studied the age structure of the local area, and found the proportion of people less than 20 years affecting the survey response. The urbanicity may be connected with the age structure and the ethnic composition of the area. The high ethnic composition of the ecological population has been associated with high non-response rates (House and Wolf, 1978; Goyder, Lock and McNair, 1992; Groves and Couper, 1998; Campanelli and O'Muircheartaigh, 1999).

Community attachment and social integration

Groves and Couper (1998) assessed the community attachment studying the impact of proportion of multiunit structures and owner-occupied homes on survey response. They hypothesised that single-person households and those recently moved into the area were less associated with the community. Also Campanelli and O'Muircheartaigh (1999) studied the proportion of flats in the residential buildings in the area. Research results support the hypothesis that deprivation and lack of social integration are associated with increased survey non-response in the area. However, it is denoted that when using small area information, these features characterise the ecological population living in the local area and do not characterise directly the non-respondents. Traditional family structures are considered to support the social integration. Campanelli and O'Muircheartaigh (1999) found the proportion of couples with dependent children increased survey co-operation, while Goyder, Lock and McNair (1992) found the proportion of single parents to increase survey non-response. In the United Kingdom, it has been found that the proportion of no car households in the local area reflects regional poverty and the variable acts as a good predictor of non-response (e.g. Campanelli and Sturgis, 1997; and Campanelli and O'Muircheartaigh, 1999).

Situational circumstances

House and Wolf (1978) studied the effects of urban residence on interpersonal trust and helping behaviour. House and Wolf (1978) relate trusting and helping behaviour to willingness to be interviewed and/or willingness to admit a stranger to one's home. Their approach is based on hypothesis that urban residents exhibit less trusting and helpful behaviour than non-urban residents do, but residents of different places differ little in trusting, helpful attitudes or dispositions. They studied social and ecological features of the cities and tested whether place of residence had a significant impact on the interpersonal trust of segments on the population who may be more vulnerable to harm or exploitation such as women, elderly and the poor. The analysis indicate that differences in refusal rates were largely due to variations in reported crime rates, rather than population size, density and heterogeneity, which have been the focus of traditional urban social psychology. House and Wolf (1978) conclude that the contextual effect imbedding response rates in larger cities can be termed social disorganisation.

Crime rates have been found influentially affecting the survey participation by many studies, for example, in House and Wolf (1978), Smith (1983), Brehm (1993), Gray, et al. (1996), and Groves and Couper (1998). However, these studies have not reported the association of type of crime and non-response more in detail. For assessing whether the total crime rate or the rate of violent crimeshave more influence on the survey non-response, we suggest that more detailed analysis by crime types is conducted in non-response research.

Willingness to be found

The willingness to be found can be mostly related to updating address information to registers, having a listed phone number (Groves and Couper, 1996) and to some extent to migration, which has been assessed in non-response analysis by Goyder, Lock and McNair (1992).

2.3.3. Attributes of interviewer and theory on interviewer effects

The concern of interviewer effects is always present in face-to-face surveys, because interviewers' characteristics, experience, behaviour and perception may influence the success of their assignments in terms of survey participation and measurement errors (e.g. Singer, Frankel and Glassman, 1983; Groves, 1989; de Heer and Israëls, 1992; Brehm, 1993; Lehtonen, 1996). Socio-demographic characteristics of the interviewer are believed to affect the 'script' evoked in the respondent's mind at the first contact (Groves, Cialdini and Couper, 1992). However, Groves and Couper (1998), Hox et al. (1991), and Hox and de Leeuw (2002) found no strong evidence between the interviewer-level response rates and personality factors. Interviewers may form a relatively homogeneous group. In addition interviewer training, tailoring, and other adoptive behaviours may have reduced the underlying differences (Groves and Couper, 1998). Vast research has focused on the impact of interviewer's experience and characteristics[2]. Interviewers' age has been found significantly associated with their work performance in Singer, Frankel and Glassman (1983), Lehtonen (1996), Campanelli and Sturgis (1997), Groves and Couper (1998), Campanelli and O'Muircheartaigh, (1999). Differences in work performance between male and female interviewers have been studied, for example, in Kane and

[2] E.g. Singer, Frankel and Glassman (1983), Groves and Fultz (1985), Lievesley (1988), de Heer and Israëls (1992), Lehtonen (1996), Campanelli and Sturgis (1997), Campanelli and O'Muircheartaigh, (1999).

McCaulay (1993) and Groves and Couper (1998). Singer, Frankel and Glassman (1983) have examined the effect of race and education of the interviewer.

Interviewers' attitudes have been studied widely[3]. Hox and de Leeuw (2002) report on international meta-analysis the influence of interviewers' attitude and behaviour on survey non-response using household data, interviewer background information and interviewer perception data, measured with a harmonised interviewer questionnaire. Interviewer's sex was not found strongly influential. The findings of Hox and de Leeuw (2002) suggest that although the age and experience are correlated, the interviewers' age counts more than their experience. This finding conflicts with the prevailing belief that interviewer experience is a critical factor for the performance. It can also indicate that more experienced interviewers are assigned with more difficult cases.

The interaction effect between the interviewer, the respondent and the social environment should be accounted for in survey participation analysis (Couper and Groves, 1992; and Campanelli and O'Muircheartaigh, 1999). Interviewers often work in different local areas, which constitute from different sub-populations, possibly affecting their work performance. In particular, Campanelli and O'Muircheartaigh questioned whether this difference in performance arises from differences among interviewers or differences among those areas allocated to the interviewers, or both.

2.3.4. Respondent-interviewer interaction

The strategies the interviewer employs to persuade the sample person are determined by the interviewer's own ability and expectations, but also by the survey design, immediate environment and broader society. Similarly, the responses that the interviewee makes to the request are affected by a variety of factors, both internal and external to the respondent, and both intrinsic and extrinsic to the survey request (Groves, Cialdini and Couper, 1992). Interviewers play a key role in how the interviewees are motivated to respond and supported to provide information in surveys. According to Singer, Mathiowetz and Couper (1993) and Singer, von Thurn and Miller (1995) the confidentiality assurance improves response, when data to be collected are of sensitive nature – although the effect of the confidentiality assurance was not found to be large in their study. In addition, some interviewees can feel discomfort in realising that their personal records are linked to administrative files. Subsequently, the further use of data relate to the data protection. Especially, in some health surveys the data providers may also be requested for consent to allow future medical records to be merged with the data in a longitudinal analysis. This consent is a priori consent that requests the responding person to allow provision of unknown information in the future, content of which they are mostly unaware, as the use of medication, illnesses or a cause of their death.

Morton-Williams (1993) emphasise the importance of interviewers' professional competence, tailoring introduction of the survey and maintaining interaction during the interview in gaining co-operation from the respondents, and stresses that social skills needed can be taught to the interviewers. Similar

[3] Singer, Frankel and Glassman (1983), Groves and Couper (1992), Lehtonen (1996) de Leeuw et al. (1997), Campanelli, (1997), Groves and Couper (1998), and by Hox and de Leeuw, (2002).

findings were found by Snijders, Hox and de Leeuw (1999) who also found in their study that productive interviewers were friendly, projecting an image of self-confidence and trust. These traits are partly trainable through social skills training; partly connected with building and maintaining work moral as well as with coping with stress and disappointments. They also emphasise that interviewer training should contain strategies for coping with refusals.

Interviewer attitudes and perception

Interviewers who, prior to the survey, are confident about their ability to elicit co-operation tend to achieve higher co-operation rates (Groves and Couper, 1998). Interviewer performance can also be related to the expectations of the interviewers (e.g. Sudman, et al., 1977; Singer and Kohnke-Aguirre, 1979; Singer, Frankel and Glassman, 1983; and Groves and Couper, 1998). Lehtonen (1996) studied the impact of the professionalism of the interviewers comparing the performance, ethical norms, and attitudes of professional interviewers with those of the public health nurses participating to data collection. A set of questions was developed to obtain information directly from the interviewers on their opinion about the interviewers' professional role, persuasion of respondents and acceptance of refusals[4]. The persuasion of respondents has also been analysed by Singer, Frankel and Glassman (1983). The interviewer completion rates were then analysed using the logistic regression for proportions with the attitudinal orientation of interviewer and interviewers' age group as predictors.

In Table 2.1 we compare the opinions of interviewers from four Northern European countries: Finland, England, The Netherlands and Sweden. The privacy of the respondents is highly respected, but the attitudes of the interviewers differ greatly across countries on acceptance of refusals. The level of persuasion of the professional interviewers is closest between the interviewers in Finland and in England. A softer interviewing approach was adopted by the Finnish nurses and by the interviewers in Sweden and the Netherlands. Interviewers' opinion about emphasising the voluntary nature of surveys differs largely. The largest difference is observed by Lehtonen (1996) between Finnish nurses and professional interviewers. The nurses have a higher tendency to accept a refusal from reluctant respondent than interviewers. The refusal acceptance is also relatively more acceptable amongst the English and Swedish interviewers in comparison to interviewers operating in Finland and in the Netherlands. This may indicate differences in interviewer training, organisational culture in survey organisations or difference in the survey culture in the four Northern-European countries

[4] The questions were originally developed for the purpose of studying the differences in completion rates in a Finnish Health Security Survey conducted in 1995. The survey data was collected both by professional interviewers working for the Statistics Finland as well as by public health nurses of local health centres.
Five attitude questions were formulated in a special interviewer questionnaire containing among other things five attitudinal questions measured on a 5-point Likert scale: 1 Strongly agree, 2 Agree, 3 Undecided, 4 Disagree, 5 Strongly disagree.

Table 2.1 Proportions of interviewers agreeing strongly or moderately with the arguments

Strongly agree with the statement:	Finland			England		The Netherlands	Sweden
	Nurses 1996, %[1]	Professional interviewers 1996, %[1]	Professional Interviewers 2000, %[2]	Professional Interviewers, SCPR 1997, %[3]	Professional Interviewers, NOP 1997, %[4]	Professional interviewers 1998, %[5]	Professional interviewers 1999, %[5]
Reluctant respondent should always be persuaded to participate	25	60	52	75	63	36	24
With enough efforts even the most reluctant respondent can be persuaded to participate	15	29	21	50	38	5	41
An interviewer should respect the privacy of respondent	99	96	97	100	94	100	85
If respondent is reluctant, refusal should be accepted	82	27	27	50	44	32	42
Voluntary participation should always be emphasised	87	35	32	0	75	9	30

1) Lehtonen (1996)
2) Interviewer perception survey of Statistics Finland on interviewers who conducted health interviews in the Health 2000 survey
3) Campanelli and Sturgis (1997); the professional interviewers of the SCPR. The interviewers were based in London and worked on a survey on political attitudes.
4) Campanelli and Sturgis (1997); professional interviewers of the NOP Research. The interviewers were based in London and worked on a survey on family resources and finances.
5) de Leeuw et al. (1998) Statistics Netherlands professional interviewers
6) Japec and Lundqvist, 1999

2.4. Survey participation and topic saliency in health examination surveys

2.4.1 Motivation for health examination surveys

For the empirical part of this study, we use data from the Health 2000, health interview and examination survey from Finland. The main concern in non-response analyses of the health surveys is to assess whether the non-response behaviour is related with the health status of the individual, and the emphasis is to analyse the possible bias of the results. The health of the individuals has been found to have associations with the response propensity. Further evidence on the possible source of non-response bias on health estimates have been found by Pennel (1990), Gray, et al. (1996), Campanelli and Sturgis (1997), and Cohen and Duffy (2002). Jones, Koolman and Rice (2006) have studied survivorship bias in British Household Panel Survey and European Community Household Panel Survey in which the attrition of initial respondents raises concerns on the bias introduced to the survey results with increasing response burden. They found that those with good health status were more likely to co-operate further in panel surveys than those with worse self-assessed health.

As mentioned previously in Section 2.3.1, the topic of the survey can have an impact on the level of response. The general view is that health surveys tend to gain higher response rates than social surveys. This is supported by the view that people are interested on their health and that they gain a free medical check-up in health examination surveys. Topic saliency has also been found to be a plausible explanation for earlier co-operation rates among seniors with older persons having more health-related incidents to report and therefore greater interest to the survey (Duhart et al, 2001). Thus following this hypothesis, one should also be concerned on the adverse associations with topic saliency on people with less health-related incidents having lower propensity to participate initially.

In the planning of the Canadian Health Measures Survey (CHMS) Tremblay (2005) has developed the conceptual model of measuring health of individuals, presented in Figure 2.2. The health of individuals is defined by the characteristics and behaviours associated with protecting the health, or increasing health risks together with the population health determinants. These factors can be further identified by individual and community level determinants, as shown in Table 2.2. Community level determinants can be traced to large extent from existing data sources, exploiting also GIS and environmental information. However, Tremblay emphasises that important health issues even at regional level, such as impact and exposure of environmental toxins, cannot be monitored without direct measures on individuals. The motivation for the health measurement surveys is that the health information collected through self-report surveys or administrative records may be incomplete or inaccurate many health factors and conditions cannot be assessed in the absence of direct physical measurements. In addition, directly measured variables can be reported on continuous scales, and they are more robust and objective.

Figure 2.2 Conceptual model of measuring health of individuals (Tremblay, 2005)

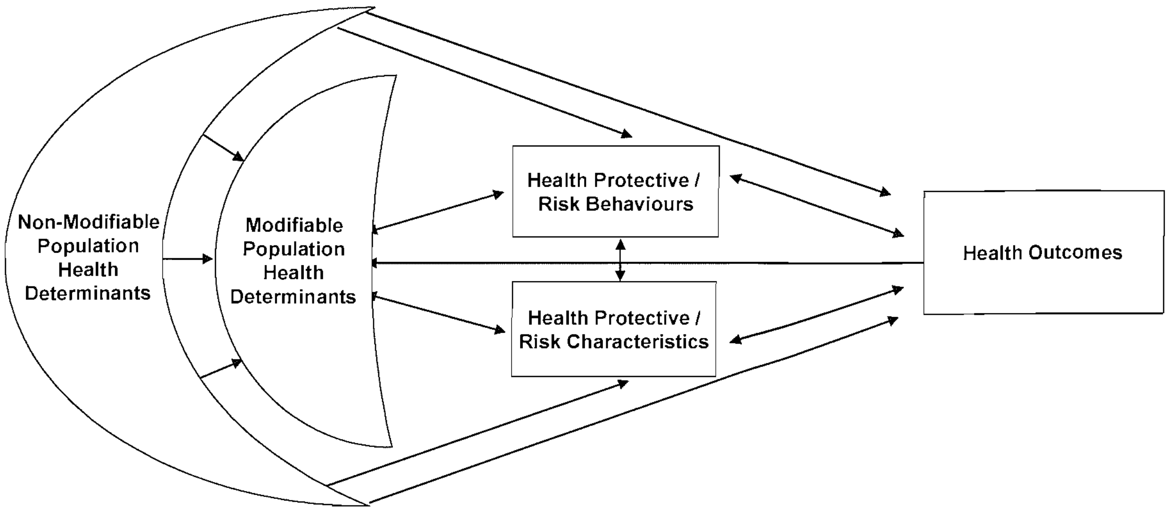


Table 2.2 Determinants affecting the health of individuals within communities by factors of the conceptual model of measuring health of individuals (Tremblay, 2005)

	Individual level	Community level
Non-Modifiable Population Health Determinants	– Age – Sex – Ethnicity – Genotype	– Geography – Culture – Climate
Modifiable Population Health Determinants	– Income – Education – Social environment – Physical environment – Health care system	– Social inequality – Social environment – Physical environment: → workplace health → school health
Health Protective / Risk Behaviours	– Physical activity – Nutrition – Alcohol / substance abuse – Smoking status – Immunizations – Medications – Sex practices – Stress exposure	– Air quality – Water quality – Food access – Local land use
Health Protective / Risk Characteristics	– Fitness: → Morphological → Metabolic → Muscular → Motor → Cardiovascular – Functional status – Stress reactivity	– Green space – Safety – Traffic – Health care – Population density
Health Outcomes	– Detectable disease – Health care contact – Disability	– Morbidity: → Prevalence → Severity → Distribution – Health care utilization

In Table 2.2, the main determinants affecting the health of individuals are presented at the level of the individual and their community i.e. the ecological population and the environment of the local area as defined by Tremblay (2005). The health of individuals is largely affected by their genotype, health behaviour, exposure to risk factors and environmental health. However, some determinants of health overlap with previously presented factors of survey non-response. If the response propensities and health variables varying by the same demographic characteristics, there is a risk for response bias. In addition socio-economic characteristics, may affect individuals' health, but have also previously shown to be associated with the risk of social exclusion (see Section 2.3.2). Thus the concern of the previous non-response studies in health surveys is well justified, and there is a need for gaining more knowledge on the possible sources of bias and suitable adjustment methods.

2.4.2 Comparison of health examination surveys

Although the health surveys are recognised widely to provide an important measurement tool for the health of nations, a harmonized health survey designs are still work in progress. The heterogeneous health survey designs inhibit the genuine comparison of survey results. Differences are due to varying focus of the surveys, concepts and definitions, question wording, contextual differences, data collection methods and instruments. In addition, the target groups of the surveys, especially age groups, may vary. Some health surveys also over-sample sub-population groups in order to increase the accuracy of minorities or difficult to get populations. In addition, the sampling units, and sample selection rules may differ. These differences add challenges also to the comparison of non-response definitions and rates as the definition of sampling units differ. In addition, differences can be observed in allowing proxy answers and accepting partial responses. The sampling units of health surveys can be either households or individuals. In household surveys many household members are sampled from the same household, imposing internal dependency structures into the survey data. Some surveys, impose restriction reducing the dependency of the units to a minimum level with the simultaneous cost-savings in the data collection. For example, in the Scottish Health Survey 1998, one adult was selected from each household and two children at maximum (Deepchand and Laiho, 1999). Dependency structures imposed by a sampling design should be dealt with weighting adjustment and/or using mixed models for the survey estimation. The Appendix 2.1 provides an insight to the varying health surveys conducted in Europe with differences in the survey contents that may also have an impact on the tendency to participation.

Koponen and Aromaa (2001 and 2003) have reviewed and evaluated sampling frames and survey protocols of health interview surveys (HIS) and health examination surveys (HES) in Europe to understand differences between surveys and to propose recommendations for sampling and fieldwork procedures (see Appendix 2.1). Non-participation is typically lower in health interview surveys (HIS), varying from 20 to 49 percent, than in health examination surveys (HES). These differences were assumed to arise to a large extent from different sampling frames and differences during fieldwork procedures. In some countries low response rates may also be due to factors outside the survey organisers' control, for example, due to survey fatigue among the population.

Koponen and Aromaa emphasize that as the problem of possible selection bias must be considered (referring to De Marco et al. 1994) it is also important to assess whether selection bias varies across different ages and whether the selection bias depends on the phenomena studied. Carter et al. (1991) found that the potential for health related selection bias is particularly critical in studies involving persons aged 65 or older because this age group has greater heterogeneity in health status and disease burden than any other. Even when the response rate is high, bias may be important when respondents and non-respondents differ systematically with respect to survey measures (Kessler et al. 1995, Novo et al. 1999, Koponen and Aromaa, 2003).

2.4.3 Non-response in health examination surveys

Table 2.3 compares the survey participation across health interview and examination surveys conducted in the United States, England, Scotland, and Finland. The non-response rates vary from 11 percentages in the Finnish Health 2000 survey to 33 percentages in the Health Survey for England. In comparison, to the surveys presented in Table 2.3 the total non-response rate of the first German National Health Interview and Examination Survey was 39 percentages, carried out between October 1997 and March 1999. However, the comparability of the non-response rates is severely reduced especially by differences in the sampling units, over-sampling, eligibility rules and clustering of sampling units. All surveys contain multiple data collection components. The table illustrates that all surveys suffer from data loss in the form of attrition and there is a reported tendency for mounting survey non-response with increasing response burden.

Table 2.3 Survey participation to health examination surveys by the phases of data collection

	Data collection phase for occurrence of non-response conditional on success at previous stage, % of adults					
	Sample type	**Health inter-view**	**Nurse visit (NV) or Symptom interview (SI)**	**Medical health examination**	**Blood samples**	**Self-completion question-naires**
Health 2000 (Finland)	Sample of individuals	10.9 % (ii)	6.7 % (ii)	6.6 % (ii)	Included to medical examination	9.2 % (ii)
NHANES III 1988-1994 (United States of America)	Sample of households	14.4 %		9.3 %	Included to medical examination	NA
NHANES 1999-2000 (United States of America)	Sample of households	18.1 %		6.9 %	Included to medical examination	NA
Health Survey for England 2001	Sample of households, all people aged 16-74 & maximum two children aged 0-15 selected	32.9 %	20.7 % (NV)	2.3 % (saliva sample)	22.4 %	NA
Health Survey for Scotland 1998	Sample of households; one person aged 16-74 & maximum two children aged 2-15 selected	23.6 %	17.6% (NV)	2.5 % (saliva sample)	15.0 %	NA

In most health surveys the data is collected by face-to-face interviewers. However, some health surveys are carried out using mailed self-completion questionnaires, such as the Survey of Lifestyles, Attitudes and Nutrition in Ireland (Kelleher et al., 2003). Many health surveys contain a medical examination; some health surveys do not carry out medical examinations, and some carry out them only for a part of the sample. In the Finnish Health Interview and Examination Survey 2000, the health examination was targeted to the adults aged 30+. Additionally, some surveys exploit self-completion questionnaires.

2.5. Social exclusion and survey participation

The relationship between social exclusion and survey non-response has been hypothesized in survey non-response theories. However, the theoretical link between social exclusion and survey non-response is relatively undeveloped. Atkinson (1998) discusses the three-way relationship between poverty, unemployment and social exclusion. Atkinson emphasizes that these concepts are related but should not be equated, although poverty and social exclusion has been used also interchangeably. Woods et al. (2004) suggest that the groups most seen as at risk of cultural exclusion are those who are financially and socially disadvantaged, young, disabled, immigrants or refugees. They refer to firstly to Mayes et al. (2001) who emphasised the multidimensional concept involving poverty, unemployment, disability, poor health, and lack of rights; an secondly to de Haan (1999) who defines social exclusion as 'recurring patterns of social relationships in which groups are denied access to goods, services and resources that are associated with citizenship'.

Aasland and Fløtten (2001) have associated social exclusion with citizenship rights, participation in civil society, and exclusion from labour market and social arenas. Also Atkinson and Davoudi (2000) emphasise the multi-dimensional definition of social exclusion, and deprivation as part of social relations that can have economical consequences at individual level. Social exclusion embraces the aspects of poverty, health inequalities, homelessness, lack of access to information communication technologies, exclusion from active citizenship, social and cultural lives, employment opportunities, geographical and micro-level exclusion, community focused experiences, links to issues of social solidarity and social or national cohesion (Woods et al., 2004). Our empirical study for the analysis of associations between social exclusion and survey non-response focuses to Finnish data. The auxiliary data matrix we have constructed benefits from Woods et al. (2004), who identified groups at risk from social exclusion in Finland. They emphasised the risk factors associated, in particular, with long-term and recurrently unemployed, disabled, children living in unstable conditions, immigrants, the chronically ill, substance abusers, violence against women and prostitution, the over-indebted, the homeless, crime-related social exclusion, and minority ethnic groups (e.g. Sámi and Romany). However, not all of this information has been available to our study from auxiliary information.

The importance of the income poverty to the social exclusion may have a lesser importance in Finland than in many other countries due to the long term development of the nations' social policies that aim for re-distribution of income. Osberg (2000) has found significant differences comparing poverty across countries using the data from the Luxembourg Income Study. Comparing the poverty in

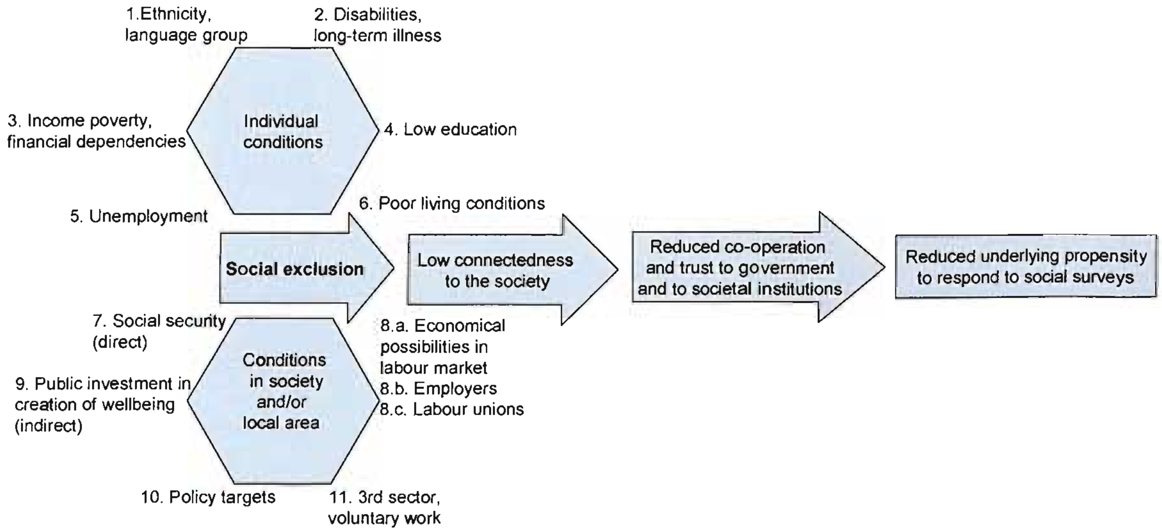
Canada, the United States and Finland, Osberg (2000) showed that the poverty intensity was significantly lower in Finland in the 1990's than in Canada or in the United States. However, as the low income levels have previously been associated with high non-response in Finland (Laiho, 1998; Lindqvist et al. 2000), we will examine the association of income poverty and survey non-response.

We define social exclusion as deprivation in term's of person's social wellbeing and weakened connectedness to their social environment and society. Using quantified data from registers we identify factors affected with causes or consequences of social exclusion. Thus restrictions imposed by long term illnesses, disabilities, low social or cultural participation, low use of services, and economic deprivation all can be classified as determinants or consequences of social exclusion. In addition, one risk factor on social exclusion is that when entitled to social benefits one has not applied for them. The supporting measures are targeted for deprived sub-population groups or for those in risk of deprivation.

We aim to profile quantified characteristics of social exclusion that have been possible to detect from the existing auxiliary data resources. These profiles contain records of multiple difficulties such as low education level, experiencing relative or absolute poverty in economical terms, having multiple social and/or health problems, being less active in the society and using less available services of the social safety net either due to the lack of information, interest or possibilities and negligible connectedness to the society. We study the associations between survey non-response and direct determinants of social exclusion such as the use of unemployment benefits, income support as well as care support and rehabilitation benefits for the disabled. In addition, indirect determinants of social exclusion can be measured from the ecological population, for example by the use of public social services. Owens et al. (2001) suggest that minority group respondents and members of less acculturated immigrant groups may have greater difficulties comprehending survey items that in most cases are developed by middle-class representatives of nation's dominant cultural group. Owens et al. (2001) continue hypothesising that the minority group members may also be less willing to reveal sensitive information during survey interviews in the United States. In Finland, the non-response studies have played attention to varying response propensities across ethnicity of people, defined by language groups (see Section 2.3.2).

In Figure 2.3 we present the theoretical framework for studying the associations of social exclusion and reduced propensity to co-operate in surveys. We have extended the route map of social exclusion developed by Atkinson (1998) for the purpose of studying the associations of social exclusion and survey non-response. The original route map is supplemented with components of low education and disabilities, as well as by public sector investments in creation of wellbeing and third sector voluntary work. In addition, we have replaced the component of companies with three-dimensional labour market dynamics that contain economical possibilities, employers and labour unions. The starting point of the theoretical framework is to recognise that both individual conditions and conditions of the society and/or the local area can have an impact on the increased risk for social exclusion. The theoretical model describes associations rather than causes and consequences.

Figure 2.3 Theoretical framework for studying the associations of social exclusion and reduced propensity to co-operate in surveys



The Figure 2.4 demonstrates that individuals' risk on social exclusion can be increased by any or a combination of the following factors: minority ethnicity or language group (1), disabilities and long term illnesses (2), income poverty and financial dependencies (3), low education (4), poor living conditions (including homelessness), which have also been hypothesised as risk factors for increased difficulties in finding employment (5). Social exclusion can also be affected by the dynamics of society and local areas. Social security aims to help directly people in economic deprivation and improving the mental and physical wellbeing (7). Economical possibilities in the labour market are greatly affected by the economic trends (8a). The job creation is largely affected by the employers, but also by the labour unions (8b & 8c). Voluntary organisations create possibilities for unpaid work and may provide activities for economically inactive people (10). Thus, the third sector creates possibilities for active life and enables mutual exchange of helping. Public investments into wellbeing and recreation improve the social and recreational circumstances, and create safety nets for individuals (9), and the policy targets allow specific resources targeted to deal with specific problems on exclusion. In relation to the survey participation behaviour, the social exclusion causes reduced connectedness to the society and the reduction in the co-operation with governmental organisations due to reduced trust (Groves, Couper and Cialdini, 1992). Therefore, the increased risk for social exclusion reduces the underlying probability to respond to social surveys.

In our analysis, the public investments into wellbeing (9) relate to the provision of services that aim to tackle or prevent social exclusion. We hypothesise these also these as plausible indicators of social exclusion or its prevention and study their impact on survey participation processes. These determinants contain namely targeted policy measures and public services that aim to reduce the inequalities in the society. Ministry of Social Affairs and Health (2002) has noted that the provision of child care services is one efficient way of preventing the social exclusion of children. The national legislation in Finland gives an obligation to local municipalities to provide day care for children. More precisely, every child aged three or over has a subjective right to full-time day care provided by the local municipality. The purpose of the child care system is both to support the participation of parents in the labour market and to provide activities to small children (Kerola et al., 2005). The provision of the childcare services utilises both private and public sector service providers in Finland, which has been described in detail in OECD background report (2000). Most commonly used service is a full-time day care[5]. The use of part-time childcare is most typical amongst households in which the supporter(s) are part-time employed, unemployed, self-employed, farmers or either of the parents is at home looking after children. Thus, the variable can imply a latent factor describing the proportion of these groups with small children in the area. However, the use of alternative day care arrangements depends largely on the selection of the services in local areas.

The research on the social exclusion and survey non-response has been concentrated on social connectedness, isolation and disengagement, related to the theory of social exchange. The social isolation theory suggests the opposite effects of socio-economic status than the social exchange theories (Groves and Couper, 1998). The lower socio economic groups would be alienated from the central institutions of society, and resentful of their dependence on the government. Members of the higher socio economic group may perceive themselves to hold an important place in society, and may as a consequence have a greater sense of civic obligation or recognise the value of survey data for the common good. This suggests a positive relationship between socio economic status and co-operation propensity. However, Groves and Couper found a negative relationship, and thus their data refute isolation theory application to survey co-operation, at least as indicated by socio economic status.

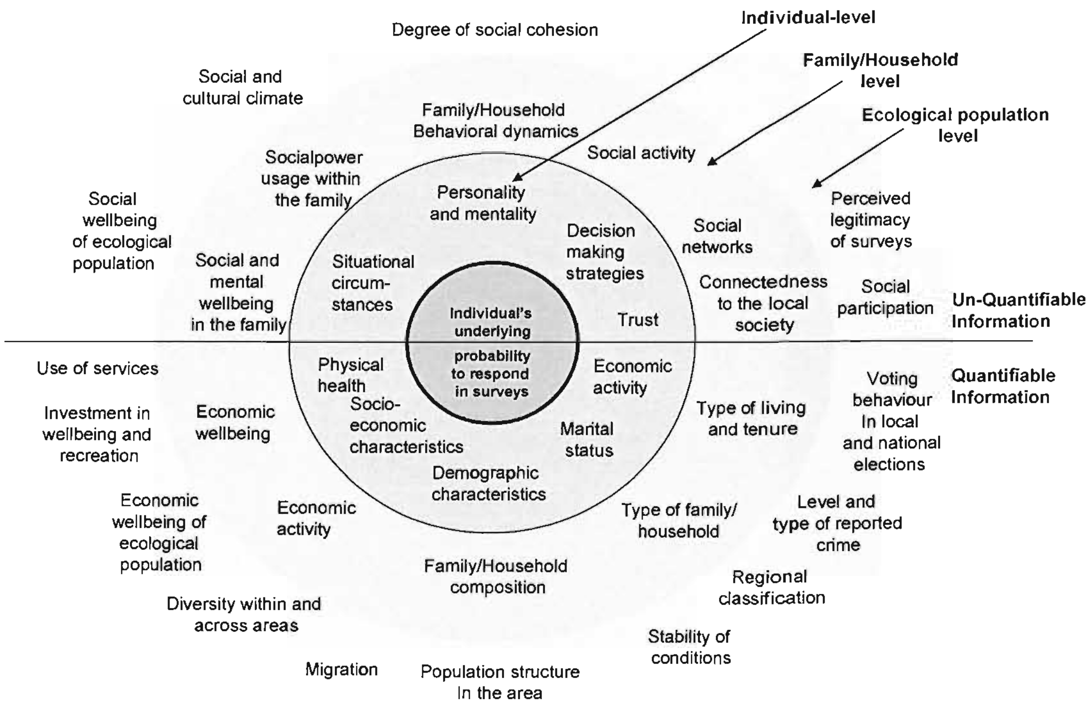
A long history of inequitable social exchange relationships between a subgroup and the larger society may lead to the development of a subculture that explicitly fails to include the norms of the larger culture. If a person feels cheated by larger society because of their membership in a sub-group, he/she might tend to ignore the norms of the larger society. This logic has been applied to findings of lower response rates among racial and ethnic subgroups as well as among the elderly (Groves and Couper, (1998), referring to Glenn, 1969 and Mercer and Butler, 1967). In addition, structural and social psychological aspects of alienation have been linked to social isolation, which are even more difficult to capture in the analysis of non-response. Some groups by virtue of their position in society may not be bound to the larger society to the same extent as others. This may be reflected both in –input alienation' (e.g. powerless, lack of political efficacy) and 'output alienation' (lack of trust in government or in the responsiveness of government institutions) (Groves and Couper (1998), referring to Southwell, 1985; Weatherford, 1991). In these circumstances the heuristic rules for compliance within the society would also indicate the difficulties of achieving co-operation, especially if the survey is conducted by a governmental agency (Groves, Cialdini, and Couper, 1992). This view equates survey participation with other acts of political or social participation such as voting (see Groves and Couper (1998), referring to Couper, Singer and Kulka, 1997; Mathiowetz, De Maio and Martin, 1991).

[5] Full time day care is defined by the length of the day care being between five and ten hours. Correspondingly, the part time day care is at most 5 hours per day.

In Figure 2.4 we present factors affecting the underlying response propensity of individuals. We have distinguished the levels of individual, family or household and ecological population, due to the nature of the dynamics these factors impose to the behavioural circumstances. At the household and level of ecological population, the survey non-response and social exclusion are expected to be associated by the degree of social cohesion, social and cultural climate as well as by social wellbeing of the ecological population. However, we are unable to measure the impact and dynamics between quantifiable and unquantifiable factors. Although using vast auxiliary information available from administrative records or registered information, we still face the limitations imposed by the nature of quantifiable and un-quantifiable information. The quantifiable factors associated with social exclusion and survey non-response can mostly be retrieved from a variety of administrative data records and registers. In contrary, the un-quantifiable factors cannot be represented directly in data analysis, but must be reflected via identifying quantifiable factors that are closely related as far as this is possible.

Woods et al. (2004) have compiled a trans-national comparison to analyse and identify cultural polices and programmes that contribute to preventing and reducing poverty and social exclusion. They have found that it is widely accepted that inclusion in cultural activities is an important stepping stone in preventing or addressing social exclusion. Thus we have identified need to study the preventive methods of social exclusion, such as the public investment in wellbeing and recreation and indicators of the use of services at the level of ecological population in relation to the survey participation analysis. The public investments can be traced from the regional accounts, which contain potentially a good source of auxiliary information to be used widely in non-response analysis.

Figure 2.4 Factors affecting the underlying propensity to respond in surveys by the level of information



2.6. Modelling survey participation

2.6.1. Logistic regression models

The aim of non-response modelling is to find the relationship between a response and explanatory variables, with a measure of the uncertainty of any such relationships. Logistic regression has been widely used in modelling survey non-response (e.g. Goyder, Lock and McNair, 1992, Lehtonen 1996, and Groves and Couper 1998). Mean regression models have been used to model response rates at small domains (e.g. Couper and Groves, 1992). In this section, we review how the models have been defined, used and tested in the literature when studying the patterns of non-response behaviour.

In the logistic regression type models, the dependent variable *R* refers to the outcome of the unit responding or co-operating. Therefore, the *R* is called the response indicator. The explanatory variables *x* refer directly to the characteristics of the unit *i* or their intermediate surroundings. The response indicator for sampled individual *i* is a binary variable and is defined as follows:

$$R_i = \begin{cases} 1 \text{ if outcome for unit } i \text{ is response} \\ 0 \text{ otherwise} \end{cases} . \tag{2.1}$$

Logistic regression models are used for both modelling the non-response and alternatively on survey participation depending how the dependent variable has been defined. The model allows the explanatory variables to be either continuous or categorical. The probability for successful survey outcome for unit *i* can be defined in a form of the probability π_i that R_i takes the value of unity:

$$P(R_i = 1) = \pi_i \tag{2.2}$$

Correspondingly the probability of a failure for unit *i* can be defined as:

$$P(R_i = 0) = 1 - \pi_i \tag{2.3}$$

By defining *r* as the observed value of the random variable *R*, where *r* is binomial, the probability distribution of the *R* can be seen as the Bernoulli distribution:

$$P(R = r) = \pi^r (1 - \pi)^{1-r} \tag{2.4}$$

In the formula above, each unit is expected to have the same probability π . In social surveys it is often accepted that the likelihood of successful outcome may vary across population domains. Thus in the modelling the assumption of a constant probability π must be relaxed and allow the variation of the response probabilities of each unit. However, if one can assume that all sampled units have the same response probability π , the random variables R_1, $R_{2,}$ ... ,R_n each have a Bernoulli distribution:

$$P(R_i = r_i) = \pi^{r_i} (1 - \pi)^{1-r_i} , \tag{2.5}$$

r_i being the observed binary response for the i^{th} unit in sample of n units, $i = 1, 2, \ldots, n$ (Collett, 2003).

Logistic regression models for survey response work around the probability of the binary (or categorical) event of responding (or reason of not responding) to occur. Let $\pi(x_i)$ be the probability that the individual i will respond to the survey. If the probability depends on p explanatory variables X_1, X_2, ... , X_p, the probability can be written as:

$$\pi(x_i) = E(R_i|x_i) = \Pr\{R_i = 1|x_i\} = \frac{e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}}} \tag{2.6}$$

$$g(x_i) = \log\left[\frac{\pi(x_i)}{1 - \pi(x_i)}\right] = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} \tag{2.7}$$

where the $x_{1i,}$ x_{2i}, ... , x_{pi} are the values of X_1, X_2, ... , X_p for the ith unit, the coefficients β_0, β_1, ... , β_p are unknown parameters, and the $g(x_i)$ is the link function. In other words, the $g(x_i)$ describes the log of the odds of survey response. The odds ratio describes the chance for particular specification of the sampled units to respond. The $g(x_i)$ can be of any real value, as by definition the $\pi(x_i)$ is constrained to be within (0, 1).

Groves and Couper (1998) used logistic models to predict the likelihood of contact among previously non-contacted households and co-operation amongst contacted households. They used direct linkage with the 1990 U.S. decennial census information and the National Survey of Health and Stress survey data, assessing subsequently the household level data over the local area estimates in non-response modelling. In Table 2.4, we present models on contactability and co-operation with the covariates of social environmental, housing unit, and household. The housing unit and household level variables are relatively more powerful in comparison to the regional data. Large urban areas show reduced contactability, also in the presence of household information. Contactability is estimated to be significantly higher for households with small children and for households with all members aged 70 or above, while single-person households have reduced contactability. Factors associated with social exchange are found to be important at the level of co-operation. The characteristics of the housing unit are excluded from the model as they relate to the accessibility of the household rather than to the theories of co-operation with survey request. The impact of household type remains the same in both survey models. Young households are more difficult to contact, but when contacted they have higher estimated response propensity than other households.

The limitations of this study are that the hierarchical data is used in simple logit model that cannot take into account the hierarchy in statistical tests and estimation. In addition, the study presents only the estimates and standard errors of the explanatory variables, but does not assess the model fit or prediction power. Thus it is difficult to judge the suitability of the model for describing the non-response, and to assess the completeness and sufficiency of the chosen auxiliary information in relation to the alternative non-response models from the available data sources.

Table 2.4 Estimated coefficients of logistic regression models on contactability and co-operation of households

Explanatory variables	Model estimates of Contacting: Basic model	Model estimates of Contacting: 1990 U.S. decennial census link	Model estimates of Co-operation: Household only	Model estimates of Co-operation: Environment and household
Constant	5.83**	4.51**	2.78**	2.72**
Social environment:				
Urbanicity				
- Central city	-1.53**	-0.49**		-0.27
- Balance of the CMSA	-0.96**	-0.27*		-0.14
- Other	-	-		
Population density (1000 people per square mile)	-0.01	-0.01		-0.02*
Crime rate (per 1000 people)	-0.01	-0.01		-0.01
% multiunit structures in block	-0.01			
% under 20 years old				0.01
Housing unit:				
Physical impediments to access	-0.67			
Large multiunit structure (10 or more units)		-0.41*		
Single-family home		0.32*		
Household (HH):				
Social exchange:				
Owner occupied			-0.10	-0.20
Monthly rent (in $100)			-0.06*	-0.04
House value (in $10 000)			-0.02**	-0.01
Social isolation:				
Race/ethnicity:				
- Black reference person (vrs other)			0.24	
- Hispanic reference person (vrs other)			0.39*	
Type of household:				
- Single-person HH (vrs other)		-0.57**	-0.37**	-0.36**
- Children <5 years in HH (vrs other)		0.50**	0.65**	0.63**
HH age structure:				
- All HH members <30 years old (vrs mixed ages)		-0.11	0.70**	0.67**
- All HH members >69 years old (vrs mixed ages)		0.59**	0.40*	0.42*

Survey: NSHS refers to National Survey of Health and Stress

CMSA refers to central cities of the large metropolitan statistical areas

* $p < 0.05$; ** $p < 0.01$

Models for analysing non-response at multiple data collection stages

As a comparison for the attrition in surveys with multiple data collection phases, the data loss in the Health and Life Style Survey and its modelling is reviewed in this section. Gray et al. (1996) studied the attrition in this survey that was conducted in Great Britain in 1984-85 (HALS1) and 1991-92 (HALS2). The survey was conducted over two rounds in which there were three phases of data collection[6]. The data loss in the panel survey was high. The wave non-response reduces the co-operative individuals from 7 to 18 percentages per wave, leading only 42 percentages of those initially interviewed to fully respond at all phases. The levels of survey attrition to be modelled were: (i) initial non-response to the HALS1, (ii) sample attrition from the HALS1 survey to the HALS2 survey, (iii) components of the sample attrition i.e. refusals, non-contacts, addresses, not traced and death, (iv) sample attrition between the interview and the measurement session, and (v) sample attrition between the measurement session and the self-completion form. Gray et al. (1996) used a series of logistic regression models to study the relationship of socio-demographic and -economic variables and survey non-response patterns. They used logistic regression models to demonstrate the odds of not responding to a stage of the survey in question holding the intercept constant.

Another study focusing on the panel attrition and refusals studied the response behaviour in Belgian Election Survey. Loosveldt, Pickery and Billiet (2002) tested whether same factors are causing unit and item non-response and if both kinds of non-response are related to each other. Logistic regression was applied to predict refusals at second wave in a panel survey based on the gender, education, item non-response for difficult and income questions. The hypothesis tested was whether item non-response to threatening or difficult questions, which were strongly related to the substantive topic of the questionnaire, were good predictors of unit non-response at later stage (Table 2.5).

Regardless of using data from a single national survey, the research provides results of wider interest and applicable approach. The key finding is the strong explanatory power of the item-nonresponse for income questions, when controlling for other factors. The limitation of the study is that the attrition is studied solely at the second wave of the panel survey. Thus there is no indication on how severely the attrition bias has been prior to the research situation for the first wave. However, the study results show clearly the impact of the item-nonresponse for five aggregated variables with significant bivariate association with the decision to participate or not. The refusal rate at the second wave is lowest, with 19 percentages, for those with no item-nonresponse in the previous wave and highest, i.e. 39 percentages, for those with three or more items missing out of five questions.

Table 2.5 Predicting refusal in wave 2 (Loosveldt, Pickery and Billiet, 2002)

Explanatory variables	Estimate	P-value	Odds ratio
Item non-response for difficult questions	0.088	0.00	1.03
Further education	-0.093	0.00	0.66
Item non-response for income questions	0.073	0.00	1.58
Gender (1=Female, 0=Male)	-0.058	0.04	0.81

We will continue assessing further the approach of Loosveldt, Pickery and Billiet (2002), as in Chapter 4 and 5 we will assess whether the item non-response to certain health survey questions has an impact on the survey participation at later data collection stages. The questions in our analysis relate to person's health and cannot necessarily be regarded as threatening. More likely they indicate individuals capabilities to co-operate under a high response burden. In addition, they link the health status of an individual to the participation to health survey with varying nature and structure of the task.

2.6.2. Assessing interviewer effects on survey participation

The previous research on interviewer effects can be divided into those studying the nature of interviewer effects in terms of (a) unit non-response, (b) item non-response or (c) reliability, consistency and accuracy of the data (Hox and de Leeuw, 2002). We focus on the former i.e. on studying the existence, significance and characteristics of interviewer effects on the survey participation behaviour. We aim to develop statistical models that could capture the impact of interviewer effects

[6] The first phase consisted of personal interview, the second of a measurement session carried by a nurse and the third phase was a self-completion component.

appropriately according to the behavioural models developed by Couper and Groves (1992) and Groves and Couper (1995).

In this section, we will focus on studies that have aimed to detect interviewer effects from survey participation data. We review the techniques applied in the previous research. Based upon these techniques we develop a method suitable for detecting the presence of the interviewer effects in surveys with multiple data collection phases in Chapter 5. The logit models presented previously are most commonly used for the analyses of survey non-response. However, some researchers have argued (e.g. Goldstein, 1995; Snijders and Bosker, 1999; Hox, 2002) that the hierarchy of the data should not be ignored but accounted for when using any hierarchical data. In face-to-face interviewing surveys the clustering by interviewer assignments is always present. The most commonly used method to study the existence of the interviewer effects on survey participation is to use multilevel modelling in which the clustering by the interviewer assignments is accounted for (e.g. Campanelli and Sturgis, 1997; Campanelli and O'Muircheartaigh, 1999; Pickery, Loosveldt and Carlton, 2001).

Pickery, Loosveldt and Carlton (2001) have studied interviewer effects on survey participation in panel surveys using the data from the Belgian Election Studies. They examined the survey participation in the second wave of a panel survey focusing on refusals. The refusals constitute 70 percentages of the non-respondents. The refusal for the second wave was found to be related to the characteristics of both the respondents and the interviewer of that wave. However, the key finding of Pickery, Loosveldt and Carlton (2001) was that the interviewer of the first wave had more impact on the participation of the respondent at the second wave than the interviewer of the second wave. On the other hand, this finding must be reviewed in the light of restricting the analysis to study the impact of the interviewer on refusing to participate from the second wave. There is a slightly undefined grey area between refusals and non-contacts as well as between refusals and non-respondents due to other reasons; some contacted people may use other reasons in their reasoning in order to be polite, and some non-contacted people may deliberately avoid the interviewer.

Multilevel modelling

Multilevel analysis is a methodology for the analyses of data with complex patterns of variability, with a focus on nested sources of variability (Snijders and Bosker, 1999). Therefore the problems connected with logistic regression analyses can be avoided in the multilevel context. Firstly, the multilevel models can be used for defining the source of variance. Secondly, more efficiency is gained using mixed effect models. The mixed effect models are statistical models in the analysis of variance and in regression analyses where it is assumed that some of the coefficients are fixed and others are random. In contextual modelling, the individual and the context are distinct sources of variability, which should both be modelled as random influences. (Snijders and Bosker, 1999).

The logistic multilevel model is defined as logistic regression before with the added random effect associated with an interviewer, denoted by:

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 x_{1ij} + \ldots + \beta_p x_{pij} + c_j \,. \quad (2.8)$$

The data is modelled by introducing random effects corresponding to the interviewer and the regional small area. Then the remaining factors and variables are modelled using fixed effects. The binary response variable is defined as R_{ijk} and π_{ijk} the corresponding probability of response for the unit *i* of the interviewer *k* at the local area *j*.

Hox and de Leeuw (2002) used multilevel modelling for analysing the relationships between survey non-response, interviewers' attitude and behaviour and interviewer background information in international context. The interviewer attitudes and behaviour have been measured with interviewer questionnaires. The findings on interviewer attributes is that older interviewer have a somewhat higher response rate than their younger colleagues but interviewer's sex does not have a strong influence on interviewer response rates. Although the age and experience are correlated, the experience counts less than age. Similarly as in previous studies (Groves and Couper, 1998, Hox et al. , 1991) also Hox and de Leeuw (2002) find that there is no strong evidence for relation between interviewer-level response rates and personality factors although interviewer experience and attitudes do have an effect, but they explain only a small part of the variation among countries (see Table 2.6). The most significant attitudinal factor was persuasion, which on the other hand had some striking differences across countries. We will discus the interviewer attitudinal and behavioural factors more in detail in Chapter 5.

Table 2.6 Multilevel model for interviewer response rates (Hox and de Leeuw, 2002)

Explanatory variables	Estimate	p-value
Intercept	0.80	0.40
Age (in years)	0.01	0.00
Sex (1=Female, 0=Male)	0.05	0.02
Experience (in years)	0.01	0.00
Factor scores:		
- Social value	-0.02	0.01
- Foot-in-door	0.01	0.01
- Persuasion	0.10	0.01
- Voluntariness	-0.02	0.01
- Send other	-0.01	0.01
Country	0.58	0.36
Survey	0.39	0.12

Multilevel modelling with crossed effects

The logistic regression models may help to gain understanding in the patterns of non-response behaviour and issues related to the survey participation. The approach of multilevel modelling can be exploited in survey participation modelling if the study design has a hierarchical or nested structure or influential auxiliary information is available at another logically and contextually important hierarchical level. One example of nested structure is the clustering within interviewer assignments in interviewer surveys. It is of great importance to study whether the interviewers have an impact to the survey participation patterns and whether participation rates differ significantly across interviewer assignment classes. Interviewer effects will be reviewed more in detail in Chapter 5 in the context of studying the efficiency of non-response reduction per interviewer assignment.

Cross-classified multilevel modelling has been used in survey participation analysis, for example, by Campanelli and Sturgis (1997), Campanelli and O'Muircheartaigh (1999), O'Muircheartaigh and Campanelli (1998), and by Martin and Beerten (1999). Households are seen to be nested within the cross-classifications of the interviewer assignments and primary sampling units (PSUs), which are in turn nested within the larger geographical pools (Campanelli and Sturgis, 1997). The use of cross-classified multilevel models is argued to provide proper modelling tools as analysing this type of data by using logistic regression would throw away valuable information about the hierarchical nature of the design. In addition, the crossed-effect models allow for distinguishing the random effects of the interviewer from the random effects of the local areas.

An analytical example of multilevel modelling with crossed effects is provided by O'Muircheartaigh and Campanelli (1998), who applied multilevel modelling in a classical situation that contained the interviewer assignment by PSU cross-classification within geographical area. They modelled the probability of response for binary response variable. In the model, the response is predicted by using geographic level information, interviewer coded non-response form variables (easy obtained information for both respondents and non-respondents) taking into account the cross-classification of the interviewer assignments and regions:

$$\log\left(\frac{\pi_{i(jk)l}}{(1-\pi_{i(jk)l})}\right) = \alpha + \beta x_{i(jk)l} + u_j + u_k + u_l \tag{2.9}$$

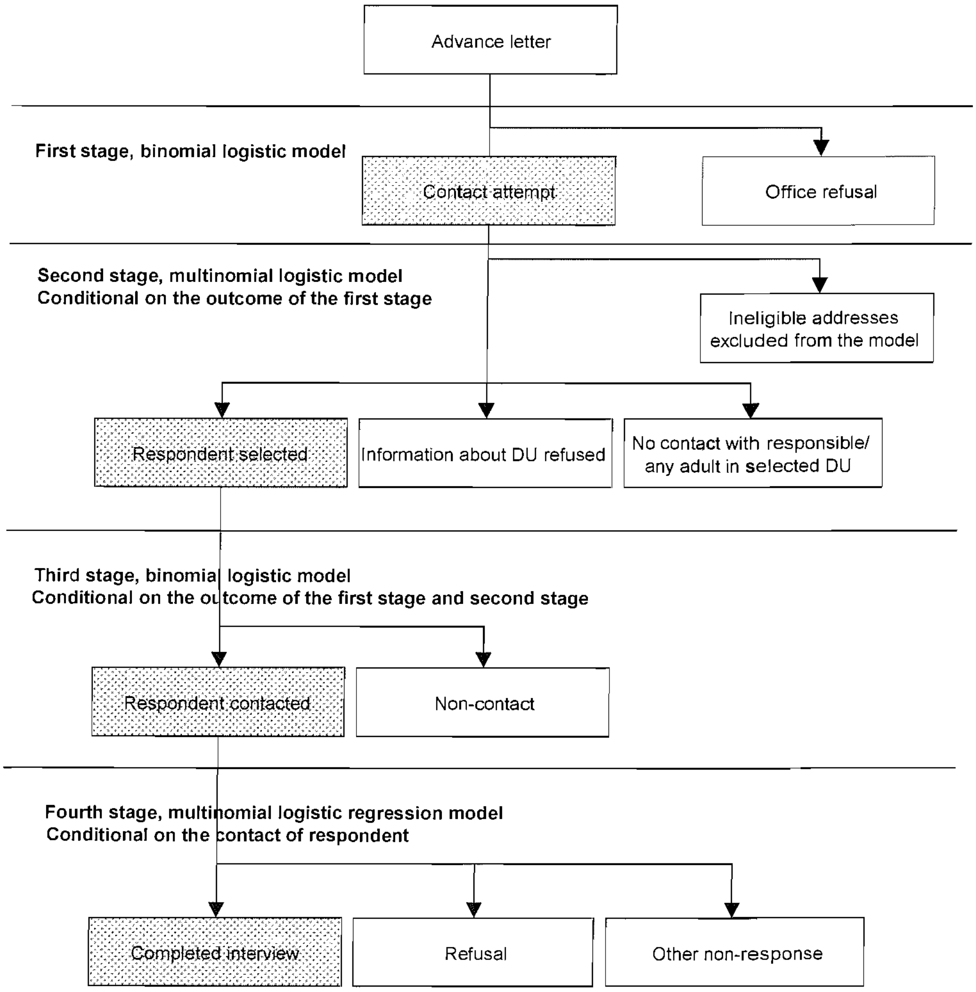
where i indicates the individual, j the PSU, k the interviewer and l the geographical pool. O'Muircheartaigh and Campanelli (1998) found that sample design effects and the clustering of individuals in interviewer assignments were comparable in impact. Thus they suggest that survey organisations should incorporate the measurement of interviewer effect on the variance-covariance structure of the observations in their designs. Alternatively, O'Muircheartaigh and Campanelli suggest that the interviewer workloads should be reduced. We consider that at least from the point of assessing the quality of the interviewing survey data, it would be beneficial to establish routines and systematically monitor these interviewer effects in survey organisations.

O'Muircheartaigh and Campanelli (1998) point out that in the application of crossed-effect multilevel models, the model assumption of independently and identically distributed (IID) observations are actually in most surveys with complex designs violated. Variances computed on the IID assumption do not take into account the effects of the sampling design and the clustering by interviewer assignments. This defect has gained little attention in other applications. O'Muircheartaigh and Campanelli (1998) discuss how to ensure that the effects on the univariate distributions would not contaminate the estimates of relationships between variables in the population, and whether one should control for these effects or to eliminate them.

2.6.3. Sequential logistic regression modelling

Laiho and Lynn (1999) have examined the survey participation process from the respondents' point of view. They emphasise the importance of measuring the success of the complete survey process and modelling the multiple steps of the process. More precisely, they aimed to gain knowledge on how successful were the field areas and/or the interviewers at (i) contacting households and sampled persons, (ii) gaining co-operation once contact made, (iii) avoiding refusals, and (iv) achieving complete interviews. They constructed a hierarchical approach to non-response modelling, consisting of multiple modelling stages. The survey response is then presented in a series sequential and conditional logistic regression models, which can be binomial or polytomous (Figure 2.5). The process of survey response is thought as a series of stages, each of which has both successful and unsuccessful outcomes. The explanatory variables contained Census small area estimates, sampling frame information and interviewer coded address level information. The models are clustered by interviewer at all stages. The approach allows the models of later stages to incorporate variables obtained during the data collection. The availability of this extra information has the potential to increase the explanatory power of the models and, therefore, of any non-response weighting procedures based upon them. The sequential model is described further in Chapter 4.

Figure 2.5 The sequential approach to modelling non-response (Laiho and Lynn, 1999)[7]



[7] In the Figure 2.5 the DU denotes for the dwelling units

2.7. Estimation in the presence of non-response

We look into the possibilities of survey weighting to adjust for the non-response and improve the estimation in the presence of non-response. The importance of exploiting the use of auxiliary information for non-response adjustment has been emphasised in the past (e.g. Särndal and Lundström, 2005; Bethlehem, 2002; and Kalton and Flores-Cervantes, 2003). For non-response adjustment, the main consideration in choosing the variables for the auxiliary vectors is generally that they would have good prediction power on survey response. However, it is also important to make the survey estimates conform to external sources of data, often also ensuring the comparability of estimates over time. In addition, benchmarking to closely related external data sources, increase the precision of the survey estimates based on those variables.

The survey weights are often developed in series of stages to compensate first for the differential inclusion probabilities and non-coverage errors when the design weights are derived and subsequently compensating for the non-response and non-sampling fluctuations from known population values (Brick and Kalton, 1996). Constructing survey adjustment methods can be divided into those that are fully model based, and those that are based on the design features of the complex survey designs. A classic example of modelling response probabilities for inverse probability weighting is given by Ekholm and Laaksonen (1991). They presented a method that employs the auxiliary information available for both respondents and non-respondents as they model the response propensity by logistic regression on explanatory factors such as household structure, urbanism, region and indicator of capital income. Another viewpoint is the restrictions imposed on the weighting, varying from designs based on simple self-weighting methods to calibration that restricts the weighted distributions to follow known population distributions and population totals. Deville (1988), Deville and Särndal (1992) and Deville, Särndal and Sautory (1993) have developed calibration, i.e. a method of re-weighting sample weights in order to reduce the effect of non-response and to produce as accurate estimates as possible on the most important survey variables.

In the case of surveys with multiple data collection phases, a single weighting method may not provide adequate result for all possible research situations. Therefore, a sequence of non-response adjustments may be employed in panel and multistage surveys where non-response may occur at each successive stage of data collection as suggested, for example, by Clayton, et al. (1998). Iannacchione (2003) applied a multi-hierarchy for constructing survey weights. The weights were based on the inverse of the inclusion probabilities that were then divided by the probability of success at previous stages. However, the study does not compare the performance of these weights in comparison to other weights and does not assess the weighted survey estimates.

Särndal and Lundström (2005) have proposed indicators to measure how well the auxiliary vector explains the response influence and an alternative indicator to measure how well the auxiliary vector explains the target variables. According to Särndal and Lundström (2005) the most ideal situation for improving the estimation is to be able to construct an informative auxiliary vector with rich source of information related to non-response. Their hypothesis is that the non-response bias will be most efficiently reduced if the auxiliary variables exploited in weighting are strongly associated with the survey non-response. In Chapter 6 we will apply these indicators and assess the performance of alternative weighting methods that aim to adjust for survey non-response.

2.8. Conclusions and motivation for the analysis of this research

The co-operation of sampled individuals and households may differ greatly across data collection phases by their background factors. Clearly observable patterns of data missingness for some data collection components should be distinguished from occasional item-nonresponse. Thus the survey policy must define and examine when the partially co-operating cases can be treated as respondents, and when they must be treated as non-respondents. The situation becomes more complex in multipurpose surveys used for different analysis, which may exploit different combinations of the collected data. Thus the structure and impact of missing data may impose diverse impacts on the research data sets. The completeness of the data in multiphase surveys can be compared to the situations arising in panel surveys. The assessment of the attrition of respondents is important as the survey participation behaviour may be directly or indirectly related with the phenomena the survey is attempting to study.

It is generally accepted fact that surveys may differ greatly in the definitions of the target populations, their geographical coverage, sampling units and data collection methods. However, the comparison of the results across survey non-response studies can be affected by factors of sample and survey design such as the survey topic, sampling frame, and mode of data collection. Therefore, the sample and survey design features are not trivial for drawing conclusions from the survey non-response analyses. The impact of the survey design is rarely assessed when reviewing the factors found to be associated with survey non-response behaviour. The results from non-response studies tend to be generalised and compared across studies without assessing carefully the limitations and key assumptions that can actually severely limit the comparisons. On the other hand, the researchers are generally concerned on the impact on the non-response to reduce the comparability of the survey estimates.

Another general weakness of non-response studies is scarce auxiliary information or lack of exploiting existing resources. Thus it is not always possible to compare testing hypothesis across studies. For example, the theoretical frameworks distinguish the impact of individuals and social environments, but the auxiliary data may exist only at the regional level which can actually have a very vague connection with response behaviour at individual level. Also, when information is linked at address level from previous census and used as direct information about the occupants the data can actually contain outdated information about the previous people living in the address. Some research settings concentrate on analysing later waves of panel surveys where auxiliary information exists at individual level from the previous waves. Using the individual level data obtained at initial waves results tend to be generalised to non-response ignoring the self-selection to respond to initial waves. In contrary, the contacting and persuasion to the initial wave should be accounted for if the results are to be compared with other findings.

Survey literature gives grounds to suspect that surveys that use multiple data collection modes for the same sample/panel at subsequent data collection can experience different type of survey attrition. It needs to be studied further how much the varying characteristics of the individual sampled, their household and regional sub-population contribute to the survey attrition at differential data collection stages. Direct matching to auxiliary data sources on all sample members is still relatively rare due to available data resources. Similarly the use of metadata (or paradata) of the survey process is often limited. Also, the justification of the models for the non-response weighting in comparison to other weighting techniques is limited. For example, Gray, et al. (1996) recommend multiple weights for non-response adjustment. Also, the modelling of non-response has included analysis of interviewer effects for the non-response modelling. However, statistical non-response models are still bound to simplify the process and this has not been discussed in-depth in the published research.

Some of the studies reviewed for this Chapter, did not have a critical or informative approach on the statistical properties of the estimated models. This follows a general tendency of the published non-response studies to focus on reporting, analysing and interpreting the effects of the explanatory variables, while the studies seldom provide analysis on the distributions of the predicted probabilities with regard to the fit of the models. In addition, information about the statistical tests and their results, as well as odds ratios are given rarely. The lack of focus to assess the statistical properties of the estimated non-response models has in some situations also lead to disregard the presentation of the mathematical formats of the estimated models. In addition, the model assumptions of the statistical models are rarely discussed or reflected to the theoretical behavioural model applied in survey non-response studies.

In Chapter 3, we present the auxiliary information retrieved from vast information sources based on the suggestions of previous non-response studies, theories of social exclusion, and health determinants reviewed in this Chapter. In Chapter 4, we aim to develop further the non-response modelling based on the theories reviewed and the theoretical framework for studying the associations of social exclusion and reduced propensity to co-operate in surveys. In Chapter 5, we assess the impact of the interviewer on the survey participation behaviour based on the findings from the literature and the need for further empirical research. We focus, in particular, on assessing the benefits and model improvement of multilevel modelling in comparison to the logit models constructed in Chapter 4. In Chapter 6, we will apply and compare the performance of different weighting methods by applying alternative plausible weighting strategies and use of various sets of auxiliary data.

3. Description of the data of the Health 2000 survey in Finland

3.1. Introduction

The social and health policy development needs complex compilations of the statistical information and forecasts based on registers and surveys. Some surveys have access to registers or censuses that can be used as sampling frames or auxiliary information. Under the data protection acts and the user right protocols these surveys may have the ability to link auxiliary information to the survey data for the purpose of producing statistics or conducting research. For example, health surveys can benefit from records containing information on individual's use of medical services. However, these data resources rarely contain representative information on direct medical measurements or self-assessed health conditions. In addition, the national legislation may impose limitations for the use of health register information. For example, due to the data protection and sensitivity of the information combining health registers with other statistical registers can be prohibited. Therefore, health surveys remain vital source of information for the public health research, decision-making and policy evaluations. For example, the age specific prevalence estimates can be used for predicting the future health conditions of people and the need for health services. In addition, health survey data allow micro level analysis on people's health conditions and behaviour. In this Chapter, we will review how survey design can benefit from auxiliary data. We will also examine how the survey participation theories developed in Chapter 2 can be linked to empirical research and available auxiliary data. Subsequently, we construct explanatory factors that relate to the survey participation theories. We then conduct explanatory analysis using this merged data.

We present a case-study using an empirical data from the Health 2000 (Aromaa and Koskinen, 2002). This survey is a health interview and examination survey conducted in Finland by the National Public Health Institute. Having the aim to enrich the knowledge of survey participation and to challenge the survey participation theories tested with less refined data previously, we chose to use the Health 2000 survey data together with the related interviewer perception survey. The survey contents are typical to most national health surveys including, for example, a health interview, medical measurements as well as self-completion questionnaires. The data structure allows us to use direct matching to a rich source of statistical and administrative registers. The auxiliary information exists at the level of individuals and dwelling, for which we have micro-level identification codes. In addition, we have geographical point coordinates for the dwellings, which have enabled further linkage as well as aggregation of the data. Subsequently, the data can be aggregated to the level of dwelling units, register derived families or households, and geographical small areas. In addition to auxiliary data, we exploit data from an additional interviewer perception survey. All combined, these sources of auxiliary information provide good grounds for testing hypothesis of the survey participation theories. Thus, the results can be useful for other national health surveys, and possibly, for social surveys using multiphase data collection.

This Chapter specifies the problem of survey non-response in multiphase surveys. Health and social surveys typically use multiple data collection phases in order to collect the information in most appropriate manner in terms of reliability and measurability. The mode of administration and the data collection technique differ across the phases in cross-sectional surveys. Different survey components may increase the risk of break-offs, i.e. survey attrition during the fieldwork. In the presence of attrition, the dilemma for the analysts is to balance between both simplifying the response process and maximising the use of collected data. The latter option would complicate the analysis and lead to different compositions of research data sets, depending on the combination of variables used and the response structure. The problem is generic to health interview and examination surveys (Korn and Graubard, 1995). Similar dilemma challenges also social surveys, such as household expenditure and time use surveys, where the data is obtained using interviews, diaries and self-completion questionnaires. Conventional approach simplifies the response process or uses imputation for the missing values. However, in a situation where the imputation cannot be used, the plausible methods are scarce unless the estimation method accounts for the probabilistic nature of the response structure.

Based on the findings of the review on the survey participation theories in Chapter 2, we aim to relate the quantified profiles to the definitions of social exclusion. These profiles contain records of multiple difficulties. We aim to outline and interpret the relevant information related with social exclusion. For this purposes, we apply the framework of Lepkowski and Couper (2002) for the predictor grouping of non-response factors. Subsequently, we structure the available auxiliary data according to this grouping. Social exclusion is associated with citizenship rights, exclusion from labour market, participation in civil society, and exclusion from social arenas (Aasland and Fløtten, 2001). We aim to construct a set of social exclusion factors in addition to indicators such as relative poverty, low education and social benefits as a main source of income. Relating to this, we explore non-response patterns in multiphase context. We assess at which data collection phase the failures of survey participation largely arise.

We first present the motivation for using the data from the Health 2000 survey. Secondly, we present the survey and the sampling design of the Health 2000. In this context, we also discuss the coverage issues in register-based sampling. Consequently, we describe the structure of the auxiliary information available from the administrative and the statistical registers, which we have linked with the survey data for the purpose of our analysis. We profile the social exclusion using auxiliary information and conduct descriptive analysis of the survey participation examining whether the non-response is associated with the social exclusion profiles. Subsequently, we examine the importance of the auxiliary data in simple logistic regression models for survey participation. We then present the interviewer perception survey conducted after the fieldwork of the Health 2000 survey. This interviewer perception survey measured professional attitudes and perceptions of the interviewers collecting data for the Health 2000 survey. This interviewer survey data is exploited especially in Chapter 5 that focuses on studying the interviewer effects. Finally, we discuss the survey taking climate in Finland and benchmark the fieldwork outcome of the Health 2000 survey to other surveys conducted in Finland.

3.2. Survey and sampling design

The Health 2000 Survey (Aromaa and Koskinen, 2002) was carried out in the year 2000 by a consortium lead by the Public Health Institute, Finland. The target population of the main part of the survey covered the general population aged 30 years or over resident in Finland. The data were collected from a sample of individuals using mixed data collection methods for specific survey components (Aromaa, Koskinen et al., 2004). Face-to-face interviews were conducted for obtaining the health behaviour, health history and background, self-assessed health, and symptoms of the individuals. Clinical examinations were used for obtaining medical measurements and clinical tests. In addition, self-completion questionnaires were used for obtaining further information on functional capacity, quality of life, vaccination history, health behaviour, health experiences, alcohol consumption and sexual behaviour, for example.

Statistics Finland developed a sampling design utilising two-stage stratified cluster sampling for the Health 2000 survey. The first sampling stage consisted of drawing some health centre districts and the second stage of sampling individuals within the previously selected districts, i.e. clusters. The purpose of multi-stage sampling was to select a geographically representative sample of the population of whole Mainland Finland and the aim of clustering was to improve the cost-efficiency of the sample from the viewpoint of fieldwork, or of interviewers and health examinations, i.e. the organisation of clinical data collection points. The essential criteria for clustering were good land transport connections to the clinical data collection point and that the distances would not be too long. In addition, during the planning of the sampling design attention was given to that the workload of each interviewer would be allocated optimally. The purpose of stratification, on the other hand, was to guarantee adequate achieved sample by age and sex groups within regions also for the elderly people.

Stratification and clustering

The sampling design was enforced to include regional clustering due to clinical examinations and cost-efficiency. The sample design was a two-stage clustered PPS sample in which the sample frame was stratified regionally according to five university hospital districts[1] using relative allocation in proportion to the size of population. Within stratification 80 health centre districts were selected to the sample out of the total number of 249[2] at the first stage of sampling. In the second sampling stage, the Population Information System was used as the sampling frame for the people resident in the selected health centre districts. Residents were allocated proportionally to the strata and sorted by their region and age. In order to improve the accuracy of the health information on elderly people, the persons aged 80 or more had twice as high selection probability than the persons in younger age groups.

Fifteen largest towns were selected with the probability of 1, while in other clusters sample sizes were calculated according to the design above so that the sample size of the strata corresponded to the

[1] University hospital districts of Helsinki, Turku, Tampere, Central Finland and Oulu.

[2] First, the 15 largest health centre districts were selected with probability 1. The remaining 234 health centre districts were then stratified by the five university hospital regions. This was followed by the selection of 65 health centre districts.

requirement of relative allocation[3]. After the initial checks the gross sample contained 8,028 persons aged 30 or over and 1,894 young adults aged 18 to 29. More detailed sampling design description is given in Finnish in Laiho and Nieminen (2004) and a concise description in English in Aromaa and Koskinen (2004) or in Lehtonen et al. (2003).

Each person was to be interviewed by a professional interviewer employed by Statistic Finland. The data was collected both in Finnish and Swedish depending on the preference of respondents. Both native Finnish and Swedish speaking interviewers and nurses were involved in the data collection and questionnaires were available in both languages. First, the health interviews were conducted for those aged 30 or over, after which they were asked to participate to medical health examination. After this, the fieldwork focused on interviewing young adults on their health, who did not have health examination. This study will focus solely on the main part of the survey, i.e. for those aged 30, because of dual reasons. First, the multiphase nature of the data collection allows us to examine in-depth survey participation and attrition in the presence of high response burden. Secondly, we are interested in analysing the precision of survey estimates of long term illnesses, which prevalence is negligible for young adult population.

3.3. Assessing the survey coverage and coverage errors

In practise, the target population definition refers to people registered as permanent residents in Finland at the time of drawing the sample[4]. The target population covers the household population, the homeless and people living in the institutions. The need of general population surveys and related knowledge on survey non-response has been gradually increasing. For example, Riedel-Heller, Busse, and Angermeyer (2000) argue that to make international comparisons on health conditions meaningful, the inclusion of institutionalised individuals is crucial. Also, the observed ageing of populations increase the information needs of the elderly, many of whom are living in institutions. However, the survey non-response theories have been developed based on the findings from household surveys excluding the institutionalised population (e.g. Groves and Couper, 1995 and 1998; Lepkowski and Couper, 2002).

The Population Information System is used as the sampling frame for most individual and household surveys of official statistics by Statistics Finland (Statistics Finland, 2002). However, the Population Information System was not allowed to be used as a sampling frame in this particular case, due to the decision by the Ethical Committee of Statistics Finland. The arguments were based on the interpretation of data protection of sensitive information defined in the Finnish Statistical Act, and because of the health survey data to be released would be linked to information from external health related registers by an external organisation. Although, the sampling design was planned in Statistics Finland, the sample was drawn in the Social Insurance Institution of Finland from the Social Insurance Register. Like the Population Information System, the Social Insurance Register contains information

---

[3] In the main survey group (persons aged 30 or over) the smallest cluster-specific sample size was 50 and the largest 100. In each health centre district people aged 80 or over were sampled with a double inclusion probability to include a sufficient number of elderly people in the Health 2000 survey. The gross sample was subtracted by people registered to live permanently abroad as well as by diplomats and employees of the Finnish embassies

[4] The sample of Health 2000 survey was drawn reflecting the population resident in Finland on the 31 July, 2000.

of all permanent residents in Finland and their demographic data. In addition, it contains the personal identification number and address for each individual, which allow for linking data with other administrative records.

In terms of people living permanently in Finland the coverage of the Population Information System is generally considered to be complete. The updates of the Population Information System are consequently updated to the Social Insurance Register. Therefore, the coverage of the Social Insurance Register and the Population Information System can be regarded identical. Myrskylä (1991) has claimed that all births and deaths are officially recorded in Finland. In addition, Myrskylä (1991) claims that the registration of immigrants living permanently in Finland can be regarded as all-inclusive, "because a person cannot earn an income or receive educational, health or other public services without social security number given after registration to the Social Insurance Register."

However, the population registers encounter problems of quality that have recently been recognised in governmental information needs. The changes in the open societies, increased mobility of people, and changes in the trust and authority of public sector impose new challenges for maintaining and developing the quality of information systems. Firstly, in the opening of the labour markets in the European Union (EU), the registers do not necessarily capture people rotating across countries, unless they register themselves to live in the country, are employed by a Finnish employer and/or seek for the social security. Currently, the temporary labour force of foreign companies providing rented employees to Finnish companies can in practise remain unrecorded even though the employees would extend their contracts over a long period of time or re-new multiple short term contracts. In addition, unregistered population movement has increased. Also in Finland, people can earn income without Finnish social security numbers in black labour market. Thus, there can be weaknesses of relying solely on register based systems which reflect only the registered reality.

Frame errors appear to be reasonably low in the Health 2000 data. The proportion of non-contacted people was 1.5 per cent of the net sample. Some of the over coverage was detected before the fieldwork, such as people whose records were updated (i.e. deaths and emigration) in between drawing the sample and issuing the sample to the field. In contrary, under coverage is more difficult to assess. Under coverage arises from those people who are permanently resident in Finland, but are not registered or their information has not yet been recorded to the Population Information System. Even though emigration to Finland is very low (less than 0.5 per cent of the population in 2000), international migration is dependent on factors that are difficult to estimate. By definition, the target population covers adult people who are permanently resident in Finland. Thus, the problem of under coverage is ignorable. However, a critical question, in terms of assessing the coverage of sampling frames, is how the requirement for an individual or household to be permanently resident in one country coincides with free movement of labour force in EU and increasing mobility of people. This would need much deeper consideration in social statistics within EU both at national and at EU level, so that the official statistics would not miss the people rotating between EU countries.

Because of the regional clustering in the sampling design, lags in updating the internal migration across primary sampling units (PSUs) may add to coverage errors. Although, the overall impact of the coverage error of the sampling frame may be minor, it may be biased towards the young mobile population moving after employment and education opportunities due to timing of drawing the sample during the summer season. Consequently, the patterns of net and gross flows may vary by the type of regions. In particular, the selected sample contains the largest cities, which may differ from each other and especially from the rest of the country.

Although national surveys are required to have a full geographical coverage, the survey organisations are forced to consider exceptions due to practicalities and cost efficiency. In the Health 2000 survey, the geographical restriction was to include only the Mainland of Finland and islands with road connections. The decision was based on avoiding expensive fieldwork arrangements due to missing road connections to scarcely populated and remote islands in the outer archipelago. Due to the climate conditions and seasonal changes, there are periods during which the sea is not open and the ice is not strong enough to walk or drive on. In addition, the Autonomous Territory of the Åland Islands was excluded, because the survey was regarded too burdensome to the small population of the Åland Islands. Statistics Finland has adapted an omission rule in official statistics according to which sampled persons cannot be re-sampled to a survey in too frequent intervals[5]. Had the Åland Islands been included, the remaining population for future surveys of official statistics would have decreased considerably.

It is difficult and costly to investigate the frame errors precisely for the purpose of individual surveys. Therefore, the general frame quality assessments serve as the best available estimates for most surveys. These assessments have been carried out routinely in two year intervals, but they also encounter problems with non-response. It has been estimated that domicile data are erroneous for about three per cent of the population in the Population Information System (Ruotsalainen, 2002). However, incomplete address information is also counted as erroneous data. To deal with this problem, interviewers are trained to discover correct addresses from incomplete address information and trace people who have moved elsewhere. Another case of erroneous data in the register is related to its timeliness. The address information is updated with a time lag of a few weeks from a notice of change of address. In recent frame quality assessment it was estimated that 1.1 per cent of the addresses were erroneous, 0.2 percent of them were missing completely, and 0.2 per cent were out of date due to the time lag in updating the change of address into the Population Information System (Ylitalo, 2002).

Although the problem with the erroneous data is not proportionally large it may be biased towards certain population groups and cannot be ignored when describing the accuracy of any survey estimates. Instead of using only register based sampling, one could reach further improvements by implementing coverage error models suggested by Wolter (1986), strategies for assessing errors and total error model suggested by Mulry and Spencer (1991), or dual frames to minimise the problem of under coverage.

[5] Persons who have been sampled to surveys conducted by Statistics Finland in the last 5 years are subtracted from the frame population in order to reduce cumulative response burden on individuals.

3.4. Structure of the data for participation analysis

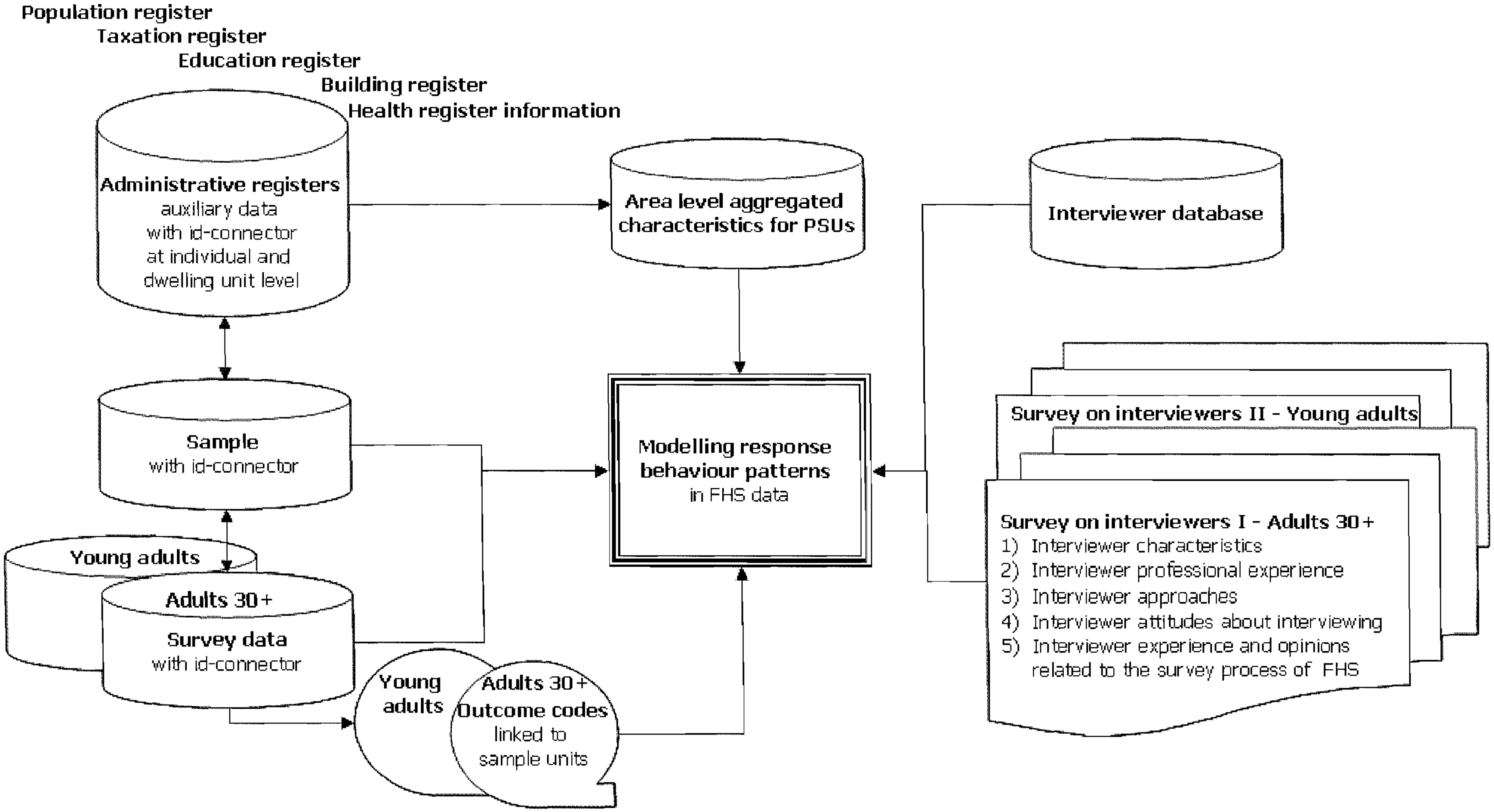
The auxiliary data sources available for the analysis contain information at individual, dwelling unit and regional level for the whole target population. This information can be directly linked to the sample data. Also, the survey outcome data, interviewer performance, interviewer characteristics, interviewer experience and attitudes can be linked to the analysis and modelling of the survey participation of the target persons. Thus, the data resources available for the survey participation analyses consists of:

(i) sample data (for sampled individuals),
(ii) survey outcome data (for sampled individuals),
(iii) administrative records (for sampled individuals),
(iv) health interview data (for responding set of sampled individuals),
(v) interviewer database (for interviewers),
(vi) interviewer perception survey (for responding set of interviewers),
(vii) operational fieldwork information, and
(viii) interviewer assessed information on the reliability of the provided information (coded by the interviewer for all respondents).

Operational information regarding the fieldwork can also be used in the survey participation analyses. For example, number of interviews conducted prior the new interview attempt indicates how well the interviewer is familiar with the survey contents and procedures. It also indicates the amount of experience the interviewer has gained on motivating interviewees to participate to the survey.

Figure 3.1 demonstrates the links between the data sources of auxiliary information and survey data. The data of the Health 2000 was selected for the analysis as it has rich and diverse sources of auxiliary information and a very complex data structure allowing for differential experimental analysis and testing and re-development of survey participation models (see also Appendix 3.1-3.3). The contents of the auxiliary information selected are described more in detail in Section 3.5, in which the selection criteria of the auxiliary information to be analysed in connection of the non-response behaviour and social exclusion is discussed. Auxiliary data consist of information from the Population Information System, taxation records, building register, unemployment register, and register of completed education and degrees. Data can be linked using the personal identification number of individuals. In addition, information can be aggregated at the dwelling unit and small area levels and to derive totals for the whole population. In the following analysis, population distributions are derived according to the target population definitions. Auxiliary information will be used to derive basic categorical variables, to construct sampling weights, for non-response analysis, and to calibrate expansion weights.

Figure 3.1 Data structure of the Health 2000 survey



Auxiliary health register information

Health register data has been linked directly to the individuals sampled to simulate estimation of the prevalence of diabetes mellitus, chronic cardiac insufficiency, connective tissue diseases, chronic asthma and similar chronic obstructive pulmonary diseases, chronic hypertension, and chronic coronary heart disease from the survey data. This health register information is based on diagnosis and entitlement to reimbursement for the medical expenses caused by the treatment of the illness. The medicine reimbursement system covers all permanent residents of Finland, regardless their age, wealth or place of residence (National Agency for Medicines and Social Insurance Institution, 2005). The system is administrated by the Social Insurance Institution as a part of the national social insurance.

There is an implicit difference between the survey variable "diagnosed with a specific long-term illness" and with the register variable that assumes in addition to the diagnosis, that the individual has been informed on the reimbursement possibilities, the individual has applied for reimbursement and that the application has been approved by the health authorities. There can be a time lag between the time of diagnosis of the long-term disease and the time of the official decision for approving the coverage of medical expenses. However, the time lag of five years between conducting this analysis and the survey allows us with sufficient correction time.

3.5. Quantitative profiles of social exclusion

In the survey participation analysis, we focus on studying specifically whether social exclusion is the main or major aspect reducing the obtained survey participation. In Chapter 2 we discussed the definitions of social exclusion and the factors associated to increase its risk for individuals in the population. We also discussed the problem of un-quantifiable and quantifiable data in terms of explaining situational human behaviour. We believe that the information structure exploited by the survey non-response studies can be greatly improved to reflect the socio-psychological theories. In this section, we review the available auxiliary information associated with social exclusion at the level of individuals or ecological population that will be used for survey participation analysis. Although many survey non-response studies have suggest that non-response may be dependent on social exclusion the association has not been studied in-depth. In addition, the information structures used for the analysis of the association have been relatively weak, and incomplete from the perspective of social exclusion. One can argue that the availability of information concerning the individuals can be limited, but nevertheless there is also a lack of focus in the construction of proper regional indicators in the non-response literature.

Traditionally non-response studies use mainly census-type information at regional level. We emphasise that the survey non-response research can benefit from the regional accounts of the municipalities or local areas for obtaining information on the use and provision of public services, which reflect the needs, behaviour and connectedness of the ecological population. This area has not been explored in detail in survey non-response studies previously. While the Census information describes the population composition and dynamics in demographic terms, the regional accounts contain information on the dynamics of local economics and the level of investment into various social and health services, and their use. Thus the regional accounts contain also information on the creation of possibilities that support people's health and social wellbeing in the local area. More importantly, regional accounts present the proportion of people who are using services targeted, for example, to disabled people or people with low income, who are in higher risk of social exclusion than others. In addition, regional accounts can tell us about the use of public service, such as visits and loans from the libraries and young children's participation to day care or part time play groups, as well as public investments to green and recreational areas. The variable transformations are generally restricted to describe volumes, proportions or rates instead of regional deviations from the average as suggested by Snijders and Bosker (1999), in conjunction of multilevel modelling. Centring values simplifies the interpretation of the survey non-response models, where negative values of the explanatory variables indicate values below the national average, and positive indicate values above average.

In Table 3.1, we present the data structure of survey participation analysis by predictor grouping (Lepkowski and Couper, 2002). Variables are mainly derived and based on compilations from administrative records or registers for the survey year. Table 3.1 represents individual and household data resources exploited for the analysis with selected examples of the regional data. In Appendix 3.4, we describe all regional information used for exploring associations of survey participation. To analyse the plausible connection between the survey non-response and social exclusion, we have gathered information on characteristics suggested by social exclusion studies, survey non-response literature, and information we believe is connected with increasing risk of social exclusion and survey non-response. The auxiliary information is presented in Table 3.1 structured according to the main factors of the conceptual framework for survey co-operation developed by Groves and Couper (1998) and Lepkowski and Couper 2002.

Groves, Cialdini and Couper (1992) emphasised the link between the perceived social responsibility of the individuals sampled and the survey participation. We believe that voting behaviour is a good indicator of social responsibility and social connectedness to the society. As the voting behaviour is not possible to obtain at an individual level as auxiliary information for the non-respondents we suggest using the regional polls during the local and/or national elections prior to the fieldwork for non-response analysis. The polls are behavioural indicators of the local communities of the individuals sampled, and reflect the regional sub-populations compliance for request when it comes for influencing in the society. The polls are also relatively easily available for many surveys and they can be aggregated from electoral districts to the level of local areas used in the sampling design and/or in the non-response analysis.

Table 3.1 The data structure of survey participation analysis by predictor grouping

Predictor grouping for survey participation (Lepkowski and Couper, 2002)	Type of information in the conceptual framework for survey so-operation (Groves and Couper, 1998)	Data available and the significance assessed in the explanatory non-response analysis and response probability modelling of the Health 2000 data
Socio-demographic / Geographic	Ecological population	Local area, population size, urbanicity, population density, dependency ratio, % of certain age groups living in the area, % distribution of language groups, % distribution by education level, average size of dwelling units (DU), average salary of DUs, % distribution of socio-economic status, number of cars
	Household/Dwelling unit	Size of the dwelling unit (DU), size of the register derived household (HH), age and sex structure of DU&HH, income
	Individual	Age, sex, maternal language, education level, income (level and structure), social benefits, relative income, socio-economic status
	Interviewer	Age, sex, maternal language, education level, geographic area of living and geographic spread of interviewer assignments
Community attachment	Ecological population	% of people renting flats, % of people in owner occupied housing, unemployment rate, structure of the unemployment, condition and level of equipment of housing, % of people living in cramped dwelling,
	Household/Dwelling unit	Children in the household, age of the youngest child, tenure status, type of living, institution, hostel, oversized DU or unknown), level of equipment in dwelling, condition of dwelling below standard, living space in dwelling, number of rooms
	Individual	Type and amount of social benefits received (home care of young children, long-term sickness, maternity, unemployment, pension), % of benefits of total income, experience and length of unemployment spells, homeless or does not belong to household population
	Interviewer	-
Social and political integration	Ecological population	% voted in the previous parliamentary and local elections and related voting information, distribution of votes by parties, % distribution of people in employment in economical areas, % distribution by family structure
	Household/Dwelling unit	Family type of register derived family
	Individual	Marital status, family status, savings to additional private pension
	Interviewer	-
Situational circumstances	Ecological population	Reported crime rate; type of reported crimes
	Household/Dwelling unit	Type of family, relative income poverty, income decile of the HH
	Individual	Activity in the labour market, experienced unemployment spells, weeks unemployed, any long-term illness with sought right for reimbursement for medical expenses
	Interviewer	Size of the interviewer assignment
Survey experience	Ecological population	Survey outcomes at PSU level
	Household/Dwelling unit	-
	Individual	Item non-response information, interviewer coding on individual's capacity to respond, use of proxy responses, and reliability of the information
	Interviewer	Interviewer perception survey, years employed as interviewer, worked in the previous national health survey
Accessibility / Willingness to be found	Ecological population	% of people moving in the area, net change of the population
	Household/Dwelling unit	Telephone number available, correctness and whether address information updated to the population register, living temporarily away
	Individual	- " -
	Interviewer	-

3.6. Descriptive analysis of survey participation

The success of the final survey outcome and the level of the response rate may show a discrepancy based on the definition of the response indicator. In a complex survey containing multiple data collection phases, the definition of the response indicator is essential for the interpretation of the results. In addition, in multi-purpose surveys the information structure used for further analyses by different researchers is likely to vary. The set of survey respondents can differ by the needs of the individual studies. The criteria for an acceptable response depend on the requirements for providing acceptable response to key analysis variables or alternatively responses at pre-defined data collection phases. Allowing partial responses, the nominal response rate is improved but the overall quality of the accepted survey data may be reduced as the amount of item-missing data is increased.

The aggregated data collection phases are presented in Table 3.2. We use initially the response indicators defined by the National Public Health Institute and aggregate them by the order of the original survey design and nature of the task (described in Appendix 3.1). We define strictly that the full response to the survey requires acceptable co-operation at each phase. We assume that phases can be ordered according to the original data collection procedure in the survey design and subsequently that the order can be used as an ordinal variable. In reality, some exceptions were made in relation to the predefined order to prevent a break-off or a drop-out to minimise the amount of missing data. It is also reasonable to relax the rule of returning the first self-completion questionnaire prior carrying out the symptom interview, as the information of self-completion was not required or used in the symptom interview for routing or screening. We choose a modelling path that majority of the sampled individuals would have followed. There were 7951 eligible sampled units[6] in the data out of 8028, and 45.3 percentages of the eligible sample were males. The survey gained partial co-operation from over seven thousand individuals and total 5608 individuals gave an acceptable response at all phases, out of which 44.9 percentages were males. The conditionality assumption across the data collection phases is defined as follows:

1. Sampled individual contacted (conditional on being sampled),
2. Responded to the health interview (conditional on being contacted),
3. Responded to the symptom interview (conditional on responding to health interview),
4. Participated to the health examination (conditional on responding to symptom interview),
5. Returned all self-completion (I, II and III) and nutrition questionnaires (conditional on the success at all previous phases).

The event of non-response at some data collection phase is also referred with *phase-non-response, wave-non-response* or *drop out.* The group of people surviving until the phase when they drop out can also be referred with the term of *drop-out cohort.*

---

[6] The gross sample was reduced by 52 dead people, who had died in between the last update of the frame and the fieldwork and 30 individuals sampled appeared to live permanently abroad. Finally in the additional non-response reduction efforts 5 non-respondents were found to have censored during the fieldwork due to their death.

Table 3.2 Data collection phases and drop-out patterns in the Health 2000 survey

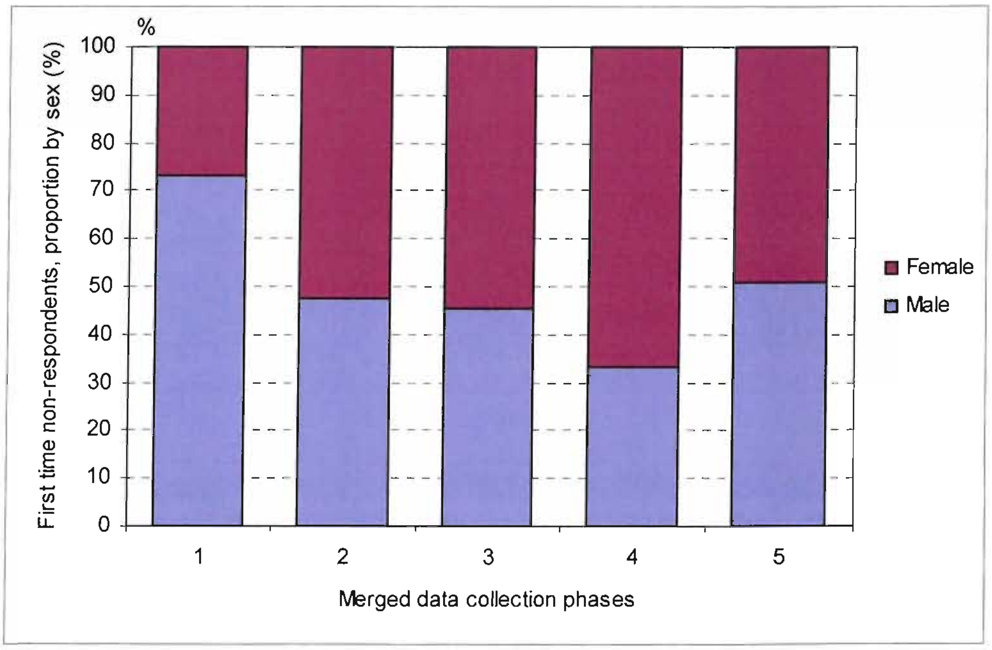
					Merged phases									
					Scheme I					Scheme II				
Data collection phase	Administrated by	# of coded successful[1)]	% males out of successful	Censored		# of cases available for model	# of cases successful	Dropping out	% males dropping out		# of cases available for model	# of cases successful	Dropping out	% males dropping out
1. Contacting target person	The interviewer	7835	44.9		1	7951	7835	111	73.0	1	7946	7835	111	73.0
2. Health interview (long or short)	The interviewer	7087	44.6	5	2	7835	7087	748	47.6	2	7835	7087	748	47.6
3. Self-completion questionnaire 1	The respondent	6736	44.5		3	7087	6718	369	47.7	5				
4. Symptom interview (long or short)	Health team	6630	44.6		4	6718	6535	183	41.0	3	7087	6611	476	45.2
5. Anthropometrical measurements	Health team	6351	45.3		5					4				
6. Other measurements including spiometry and bioimpedence	Health team	6339	45.3		5					4				
7. Laboratory tests	Health team	6711	44.6		5					4				
8. Dental health	Health team	6335	45.3		5					4				
9. Ability tests	Health team	6329	45.3		5					4				
10. Clinical health	Health team	6326	45.4		5	6535	6122	413	32.5	4	6611	6174	437	33.0
11. Self-completion questionnaire 2	The respondent	6734	44.5		6					5				
12. Self-completion questionnaire 3	The respondent	6269	44.0		6					5				
13. Nutrition self-completion questionnaire	The respondent	6005	44.3		6	6122	5608	514	50.8	5	6174	5608	566	50.7

[1)] Coded by the National Public Health Institute

Survey participation patterns by demographic characteristics

Our findings lend some support from the theory presented by Groves and Couper (1998) that there are tendencies for males to be at home less frequently than females, who are more easily contacted by the interviewers as they accept more responsibility on household duties and childcare. Also we find that although men are more difficult to contact, once contacted males do not have significantly lower co-operation rate than females. Similar results have been also been obtained by Smith (1983), and Lindström (1983). Especially the later data collection phases following the initial face-to-face interview add perceived response burden for elderly females reducing their participation after the health interview, in comparison to others. For men, the median age was highest amongst those who cooperated fully until the medical measurement phase, but did not participate to the health examination. Male non-respondents have narrower age distribution than females. In Figure 3.2 we demonstrate that the men had higher risk than women of dropping out of the study initially, and women had slightly higher risk at the health examination phase.

Figure 3.2 First occurrence of non-response by the data collection phase and gender[7]



The gender differences on survey participation can also be observed in multivariate context. In Figure 3.4.1 (Appendix 3.4), the proportion of those registered to live at their parental home are presented by data collection phases. In total there are fewer females than males living at parental home. This suggests that amongst those living at parental homes, females were more difficult to contact than males. For males living at parental home, the attrition at the phase of health interview was

[7] According to the classification presented in Table 3.2 the merged data collection phases are:

1 Contacting target person
2 Health interview (long or short)
3 Symptom interview (long or short)
4 Medical measurements and tests
5 Self-completion questionnaire 1, 2, 3 and the nutrition questionnaire

proportionally even higher than at contacting phase. Instead, for females the risk of non-response reduces significantly once contacted. To assess the presence of gender differences most of the explanatory data analyses in this chapter have been carried out by sex. However, some related characteristics, such as the prevalence of diabetes, are more associated with individuals' age than with their sex. This can mean that the high values of drop-out can ultimately be explained by the impact of high age reducing the likelihood of contact or alternatively the high age and poor health affecting the motivation and capabilities of the respondent to co-operate at further phases, where the response burden is very high (Figure 3.4.3 in Appendix 3.4). However, the prevalence of diabetes seems to increase the risk of non-response substantially only at the medical measurements phase for both sexes.

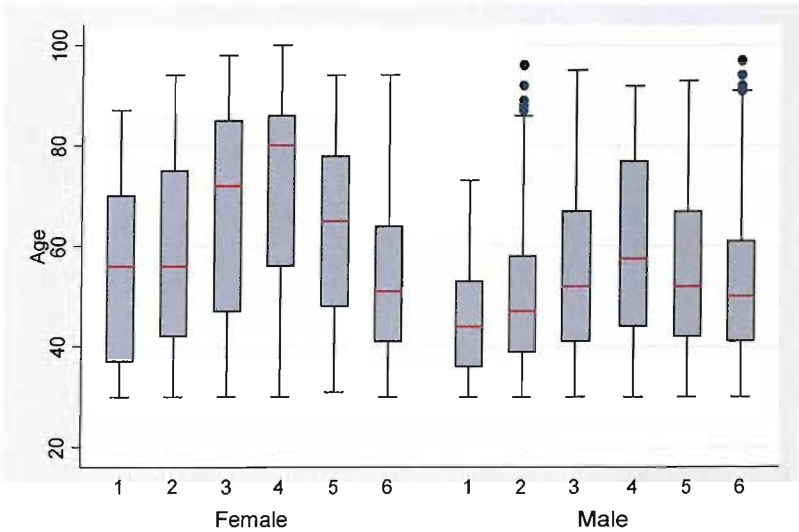
The non-response behaviour for females and males living (or registered to live) at their parental home can be affected by many reasons. For example, there can be differences in time use in comparison to people who run their own households. On the other hand, some living at parental home may look after their elderly parents. Some may also actually live temporarily or unofficially elsewhere, which increases their probability to remain un-contacted. It would be important for the survey participation analysis to be able to test whether the group has reduced compliance for request and/or avoid deliberately the contact with the interviewer. The latter has been recognised as the grey area between non-contacts and refusals. However, drawing any further conclusions on the behaviour of adults living at their parental home is not possible based on the limitations of the data.

The Figure 3.4.2 (Appendix 3.4) shows that the proportion of people not belonging to household population is very low in the obtained sample but high for non-contacted males. In addition, these people have relatively high prevalence of dropping out prior the symptom interview or the medical measurements, regardless whether initial co-operation has been gained. The variable "not household population" indicates a heterogeneous group of people with deprived living conditions. They are to large extent, elderly people living in institutions or working age people who have no permanent residence.

From Figure 3.3 we can observe that the median age for women experiencing non-response at any data collection phase was higher than for women co-operating fully. When comparing the age distribution between male and female respondents the main difference is in the level of age of those experiencing non-response. In general, non-responding men have younger age distribution than women at all phases. Individuals experiencing non-response for the first time at medical examination (phase 4) have higher median age than other response groups for both sexes.

The age seems to be connected with the survey participation not only directly, but also to some extent by age composition of the household as noted by Kemsley (1976) and Lievesley (1988). In Figure 3.5.1 (Appendix 3.5), we demonstrate that non-contacted females have on average older children than non-contacted men. Non-contacted men have younger children than males co-operating at further phases, or than females in general. Women having older children tend to work similar hours than men and their penetration in full time employment is very high in the Finnish labour market. However, the time use and patterns of being at home may differ by men and women, and by whether they have dependent children. The data also suggests that especially in the transition to symptom interview phase, females having young children have high perceived response burden or difficulties of making and keeping with the appointment to the health examination. The median age of the youngest child in the household is lowest for those females non-responding at symptom interview, which was conducted at the health examination, required booking of an appointment, and additional travelling arrangements.

Figure 3.3 Box plot of age by sex and the 1st occurrence of non-response the data collection phase



Note: The 6th stage refers to those participated at an acceptable level at all data collection stages

Many non-response studies have found the size of the dwelling unit, household or family to correlate positively with survey participation. Register information on the size of the dwelling unit indicates the number of people registered to live in the same dwelling. In our data the size of the dwelling units does not have a clear impact on survey participation, although non-contacts are more likely to be single-person households than in other non-response cohorts (see Figure 3.5.2 in Appendix 3.5). Also the multivariate analyses presented later in this Chapter will indicate that the single-person household is important characteristic of survey participation.

In Figure 3.5.3 (Appendix 3.6), we present survey participation by decile of register derived disposable income. It can be observed that income levels are lower in the non-responding groups. Those groups fully co-operating represent people with higher income than the original sample did. Graphical presentation suggests that the income is lowest for the non-contacted females and males. Thus there is a risk for response bias if the health variables studied and health inequalities are correlated by the income levels of the individuals.

The dependency ratio presented in the Figure 3.5.4 (Appendix 3.5) reflects the impact of the age structure of the regional sub-population on the survey participation. The dependency ratio seems to be relatively low in the areas of those non-contacted and those not responding to the health interview, when comparing to those with whom some level of co-operation has been achieved. Both Lievesley

(1988) and Gray, et al. (1996) found that non-response is connected with the population density in the local area. Population density can also be related with the crime rate which impact has also been analysed previously in non-response studies. In Health 2000 data, the range of population density is high for non-contacted females and males, but overall the population density of the local area was not strongly associated with survey participation behaviour.

As suggested by House and Wolf (1978), regional crime rate seems to be associated with survey participation. From Figure 3.5.5 we detect that high regional crime rates are associated with non-contacting men and women. Unemployment rate is another regional deprivation factor. In Figure 3.5.6, we show that the regional unemployment rate is slightly lower for those, who were not contacted in comparison to other response groups, within which the variation of the unemployment rate is very small. We have also found the proportion of self-employed is to some extent lower in the local areas of those people dropping out in the initial survey participation phases, see Figure 3.5.7. Another indicator of initiative and social participation is the voting behaviour in the local area. Interestingly when studying the effects between social exclusion and survey participation it was found that the poll in the last local elections are slightly connected with the level of survey co-operation. Non-contacted tend to live in areas where the election poll is low (Figure 3.5.8).

3.7. Predicting survey participation of individuals with varying levels of information

Examining survey participation hypothesises solely with local area and address level auxiliary information bares the risk for drawing conclusions that would fall into the ecological fallacy. Our aim is to assess how the use of individual level information changes the model interpretation and performance in survey participation analysis. We distinguish the auxiliary information into four categories based on the object the information is describing: local areas, dwellings, household units (based on register information), and the individuals. Our hypothesis is that more precise information, characterising directly the individual, would improve the prediction power of the survey participation model. The explanatory power of auxiliary information for predicting the full survey participation is tested by exploring rich sources of auxiliary data that reflect the theories of survey participation, social exclusion, as well as economic and social deprivation presented in Chapter 2. We compare the interpretation and efficiency of four survey participation models with increasing level of detail in the information structure. The models excluding the individual level data can be characterised as focusing into the effects of the social environment on the individuals' response behaviour.

We begin the modelling with a simple model and gradually move into more complex analysis of survey participation in multiphase surveys in Chapter 4. The simple logit model acts as a reference model to which the performances of more complex models are compared to. The model uses logit link to predict the likelihood of full survey co-operation of the individual by explanatory variables. The response of an individual is denoted by response indicator R_i defined in (2.1) and estimated via a logit link defined in (2.7). The model assumptions of the logistic regression model have been described in

Chapter 2. The explanatory variables can be continuous such as the regional crime rate, dummy indicators such as sex, or categorised versions of continuous variables such as the quintile of taxable salary income. The β coefficients indicate the magnitude and direction on how each explanatory variable impact the estimated response probability.

Comparison of fit of the models

The model exploiting individual level data allow us to construct more precise estimates for survey participation than less accurate information as shown in Table 3.3. Models utilising solely local area information had the poorest fit in comparison to other models. Dwelling information gave a relatively good improvement for the model. But the model performance was remarkably improved by incorporating both the household and the individual level information. Simple logit model for the full co-operation implies that the poor living and social conditions reduce the survey participation, and that the socio-economic wellbeing improves the likelihood of co-operation.

We compare the performance of the logit models by examining the predicted probabilities and statistical tests for model fit. Figure 3.4 presents the distribution of predicted response probabilities by the proportion of individuals who responded fully. The predicted probabilities are grouped into equal intervals of 0.1 width. There are two criteria to judge the model performance using this graphical presentation. First, a good model performance is indicated by linearly increasing proportion of respondents with the increasing predicted probabilities. Secondly, the prediction power of the model is associated with its ability to use the full range of probabilities. The distributions of predicted probabilities are most narrow for model with local area information only. The distribution of predicted probabilities exploiting individual level data are most close to the diagonal, when assessed against the proportional increase of individuals who co-operated fully within each class of predicted probability. Based on graphical judgement, the model performance of the individual level data performs best. When the response propensity models are enriched with individual, household or dwelling unit characteristics, the model performance improves. This indicates that in Health 2000 survey, the non-response behaviour has been more dependent on the characteristics of individuals rather than the areas where they live. This may be explained to some extent by the fact that there are relatively low regional differences in Finland at the level of the ecological population.

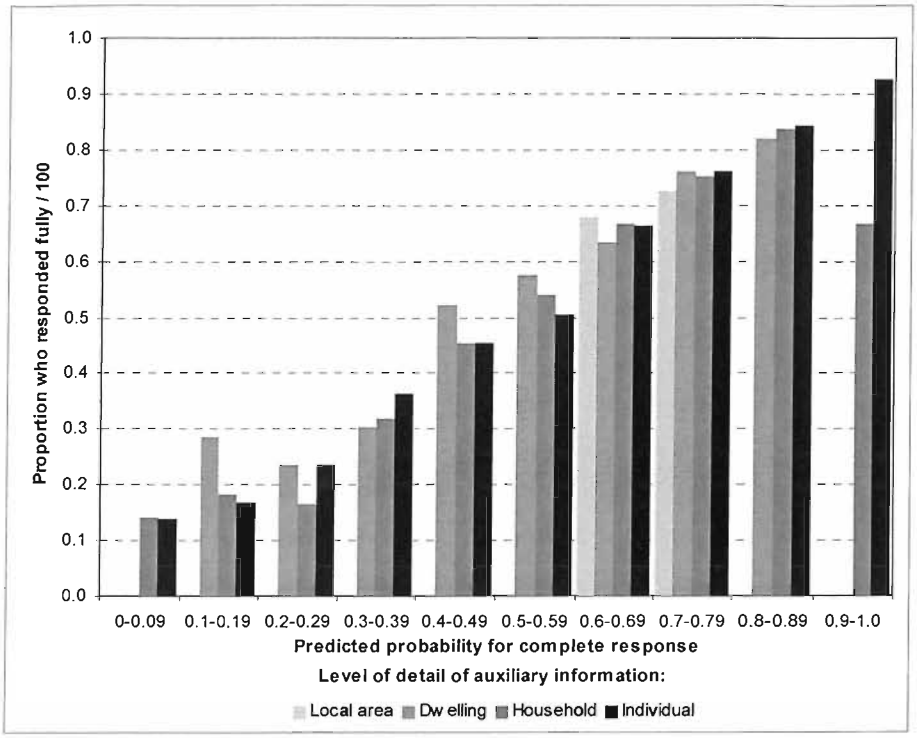
Table 3.3 Parameter estimates for simple logit model with increasing information levels

Explanatory variables	Ecological population		Dwelling data incorporated		Household data		Individual data incorporated	
	Odds	p-value	Odds	p-value	Odds	p-value	Odds	p-value
Intercept	2.42	0.00	1.22	0.00	1.80	0.00	0.77	0.10
Individual characteristics:								
Age of the individual:								
- 30-44 years (vrs 45-79)							1.04	0.48
- 80+ years (vrs 45-79)							0.67	0.00
Female vrs male							1.29	0.00
Maternal language:								
- Northern Sámi (vrs Finnish)							0.97	0.92
- Swedish (vrs Finnish)							1.39	0.03
- Baltic languages or Russian							1.15	0.64
- Other language							0.38	0.00
Pensioner (vrs other socio-economic group)							0.63	0.00
Further education (vrs basic education only)							1.46	0.00
Income quintiles of register derived disposable income:								
- 1st quintile or no income (vrs 2nd - 4th quintile)							0.82	0.01
- 5th quintile (vrs 2nd - 4th quintile)							1.08	0.19
Received rehabilitation support (disabled)							1.40	0.02
Received care support (pensioners or disabled)							0.53	0.00
Experience of unemployment (ue) in 2000								
- 1-25 weeks unemployed (vrs no ue-spells in 2000)							1.08	0.36
- 26-52 weeks unemployed (vrs no ue-spells in 2000)							0.84	0.03
Household information:								
Household experienced income poverty in 2000					0.69	0.00	1.08	0.54
HH member recived income support					0.71	0.00	0.66	0.00
HH member recived capital income					1.41	0.00	1.31	0.00
Household type:								
- Children in the HH (vrs Other)					1.47	0.00	1.22	0.00
- Individual sampled adult child in parental home (vrs Other)					0.54	0.00	0.57	0.00
- Single person HH (vrs Other)					0.86	0.02	0.96	0.57
- Couple with no children (vrs Other)					1.23	0.00	1.28	0.00
Not household population					0.13	0.00	0.29	0.00
Household with more than 2 adult members					0.78	0.00	0.74	0.00
Dwelling information:								
# of rooms in dwelling:								
- 1-2 rooms (vrs 3 rooms)			0.91	0.10	0.86	0.04	0.92	0.26
- 4+ rooms (vrs 3 rooms)			1.73	0.00	1.32	0.00	1.35	0.00
- # of rooms unknown (vrs 3 rooms)			0.47	0.00	0.81	0.24	0.70	0.05
Type of housing:								
Detached, semi or terraced house (vrs flats)			1.38	0.00				
Type of housing unknown (vrs flats)			0.60	0.00				
Tenancy:								
Rental housing, type A[1] or type B[2] (vrs other type of tenancy)			0.72	0.00	0.77	0.00		
Level of equipment in housing above average (vrs other)			1.41	0.00	1.26	0.00		
Type B: Rental housing[1]							0.58	0.02
Local area information:								
Crime rate [2]	0.97	0.00	0.98	0.00	0.97	0.00	0.97	0.00
Children in part-time child care [2, 3]	0.77	0.00	0.83	0.01	0.78	0.00	0.73	0.00
Elderly receiving services for their care [2, 4]	0.73	0.00	0.73	0.00	0.72	0.00	0.80	0.03
Mortality of 30 - 64 year olds [3]	2.13	0.01	3.04	0.00	4.23	0.00	3.21	0.00

[1] Type B: Rental housing constructed with the support by governmental interest rate subsidies
[2] Measured as a deviation from the national estimate per 100 inhabitants
[3] Part-time child care arranged by the local municipality nurseries
[4] Elderly care financed by the local municipality

Note: SAS PROC LOGISTIC procedure was used for the model-based estimation and SAS PROC SURVEYLOGISTIC for the design-based model.

Figure 3.4 Proportion of individuals sampled who fully responded by the product of their predicted probabilities to respond in the simple logit models by level of auxiliary information



Characteristics of ecological population associated with survey participation

A variety of regional factors were tested on their associations with response behaviour of individuals. The analysed factors are associated with social exclusion and survey non-response, such as the level of urbanicity, demographic and household structures, socio-economic wellbeing, economic deprivation, use of public services in social services, health care, recreation as well as public investment to these services and the neighbourhood. To avoid the ecological fallacy, the meaning of the ecological population[8] information cannot be treated as an indication of the individual-level characteristics nor behaviour. In contrary, the significant local area level variables are rather latent variables of the local population, lifestyle, social and economical wellbeing and possibilities.

The response probability of individuals is mainly associated with the deviation from the national average at crime rate[9], number of elderly people receiving care, mortality of adults aged 30 to 64, and proportion of children in part-time day care in their local area. Although being statistically significant, the impact of the crime rate to the response propensity is very small in multivariate analysis. In comparison, the high use of part-time day care of children in the region seems to be significantly associated with survey non-response of individuals living in that area.

[8] The ecological subpopulation represents the socio-economic environment of individuals and is defined by the local areas of individuals. This geographical division of local areas was also used as primary sampling units in the sampling design.

[9] The crime rate contains solely the crimes reported to the police. Crime rate has been found a good predictor of non-response in a number of studies, for example, in House and Wolf (1978), Smith (1983), Brehm (1993), Gray, Campanelli, Deepchand and Prescott-Clarke (1996), as reviewed previously in Chapter 2.

Examining the effect of the type of crime at the local area in depth, we found that the crime rate describing the total number of reported crimes is most significant. We also analysed separately the impact of crime directly threatening people i.e. murders, physical attacks and sexual crime in comparison to other type of crime, for example, traffic or alcohol offences. Against our hypothesis, we found out that the survey non-response was not significantly related to the reported physical violence against people in the local area. In contrary, we found that the total crime rate constituting of all crimes reported to the police, was the most robust factor for explaining regional variation in the survey participation.

The association of part-time day care and low survey non-response can be explained by latent impact that non-response is higher in areas in which the socio-economic diversity is larger between families with children, than on average. This can indicate how well the local municipalities have reacted to the needs of the local population in the provision of diverse public sector services, namely in child care. Similarly, the number of elderly people receiving care is found to have reducing impact on survey participation at the local area. This can imply that survey co-operation is more difficult at areas where there are more elderly people using public services provided for them. This may indicate that elderly use more public services provided for them in areas with weakened social networks. The survey response is increased in areas with proportionally high mortality of the working aged adults. One plausible association is the topic saliency of the health surveys as amongst population with worse health the interest to participate to health survey is higher. However, more plausible explanation is that the mortality of the adult population is higher on less affluent regions, in which also the lifestyle and time use of the individuals favour survey participation, increasing contactability and co-operation.

Dwelling information

When the dwelling information is incorporated into the survey participation model, the effect and significance of the regional information remain stable. The information related to the size, type, equipment, and tenancy of the dwelling bare important associations with survey participation. The data suggests that the likelihood of survey participation is higher for people living in larger dwellings, in detached, semi-detached or terraced houses, and in dwellings that have a better equipment level than dwellings on average. In contrary, the response propensity is significantly lower, if there is no information on the size and type of dwelling in the building register. This suggests that people had a lower tendency to participate to this survey if living in buildings not registered for housing, in temporary buildings or in buildings built without building permission. This finding is potentially interesting to be explored more in detail across surveys with access to building register information.

Household structure information

The register households have been constructed when merging the information from the population register with the building register information at dwelling level as we can identify people living in the same address. Subsequently, we can construct register households that represent the presumed household of the individuals based on logical rules developed for the population statistics. People who do not belong to household population have very low propensity to participate to surveys in comparison to people who belong to the household population. Individuals are not seen as part of the household population, if they are homeless, live in shelter homes, oversized dwelling units or institutions. The association of deprivation and social exclusion can also be seen with the low predicted likelihood to respond, when people belong to households experienced income poverty, live in small dwellings or rent from social housing.

The household type variable describes the social connectedness of the sampled individual in their family life. This latent variable compresses information of the social circumstances of individuals, relationships and demographic life-cycle. We have aggregated the household type variable into categories that distinguish differences in survey participation patterns most efficiently in our data. We bring forward four household categories: households with children, adult child in parental home, single persons, and couples without children contrasting them against other remaining household types. Confounding with the survey non-response literature, we find people in households with children as well as people living with their partners having a higher response propensity in comparison to single-person households, or to those living with their parents in their adulthood (aged 30 or over).

Demographic individual level factors

The individual information is highly informative when incorporated in the modelling of survey participation. Although, most characteristics of the ecological population, dwellings and households remain significant, the individual level information seems to have a relatively large impact on the model. These new variables describe namely the demographic characteristics and socio-economic conditions of individuals, as well as, physical health and connectedness to the society via specially targeted social security benefits.

When looking at the demographical variables, clear pattern arises. Men, elderly people aged 80 or over, and pensioners are less likely to participate fully than others. The prediction of the survey co-operation is higher for individuals with further education in comparison to basic education or unknown education level. The individual level results imply that full survey response in a health survey with relatively high response burden is lower amongst people with lower cognitive skills, and possibly also by reduced physical capacity affected both by ageing and by more difficult living conditions of the older age cohorts. The physical capacity of the elderly people has been affected by the direct and indirect effects of the Second World War, like harsh living conditions during the war and in post-war era. The psychological and behavioural impact of the experience of war has not gained attention in non-response studies literature. Similar effect of age reducing the survey participation has been found by Groves and Couper (1998), Campanelli and O'Muircheartaigh, (1999), and by Couper (1997).

Social exclusion factors at individual level

Examining the effect of the maternal language gives us interesting insight to the differences across population groups linked with risks to social exclusion. As discussed in Chapter 2 the social and economic possibilities to actively participate into society can be restricted by unequal opportunities, for example, civil rights, work permission, and also differences in language skills that greatly enable first or second wave immigrants to adapt oneself to the new society. We can observe from the model that in comparison to Finnish speaking majority the predicted likelihood to fully respond was higher amongst people whose maternal language was any of the Swedish, Northern Sámi, and other Scandinavian languages. We do not have reliable results for the Russians and Baltic people due to their small representation in the sample. In contrary, people whose maternal language was "other", namely African, Arabic, Asian, English, French, German, or Spanish, had lowest probability to participate. This may indicate that the skills in the survey languages, social responsibility and connectedness to society can be lowest in these minority population groups. However, our classification indicates only the maternal language of the individuals, and we do not have auxiliary information on people's skills on survey languages: Finnish or Swedish. However, for the minority groups of other languages there is an increased probability of lower skills in survey languages, which may also increase the risk that people are less integrated to the society and thus feel less obliged to respond to national surveys.

Findings in social exclusion research suggest that lack of civil rights and work permits impose an increased risk for social exclusion (e.g. Aasland and Fløtten, 2001). When originating from other Nordic or EU countries immigrants do not need work permits or residence visa in Finland, but most other immigrants do. Our model predicts the lowest response rates for people with any other foreign maternal language than those minority groups from close geographical reach. This may indicate that people who do not belong to large immigration groups, such as Russians and Baltic people in Finland, or whose cultural origin are very different from the prevailing local culture, are more likely to be excluded from surveys do to non-response. These people may also find it most difficult to adapt to the social and cultural landscape. The result imposes concerns that people resident permanently in Finland, who originate outside of neighbouring countries, can be under represented in this health survey. To find out whether this indicates a single incident of under representation, or whether it is a general phenomenon in surveys conducted in Finland, there should be an in-depth investigation across social and health surveys using meta-analysis.

Income poverty has been linked with social exclusion in many studies (see Chapter 2). We have found economic deprivation of the household affecting the survey participation in our models. When incorporating individual level income information, the results on economic deprivation from the household level information model remain similar. Looking into the individual level income information, individuals receiving income support to compensate their low income levels, have a lower prediction to participate. At the same time, individuals receiving capital income were more likely to participate fully than others to the survey. Taking into account the low predicted participation of the pensioners, the model suggest that the economical independency may be associated with the improved likelihood to co-operate with surveys.

The main finding of the individual level model is the impact of the nature of social benefits on survey participation of individuals. In Chapter 2 we noted that disabilities are linked with increased risk for social exclusion. From our data, we find that disabled people receiving rehabilitation support are more actively participating to the survey than people on average. This gives indication that rehabilitation programs aiming to support people to actively participate to the society and help their entry or return to labour market, can also increase people's social participation and connectedness to the society in the form of responding to surveys. In contrary, people with disabilities receiving long-term support and people receiving benefits such as income support have increased risk not to co-operate fully.

In general, we find the impact of social benefits negative or neutral on the survey participation, if the benefit provides monetary help for existing conditions without aims on improving peoples' social or economic conditions in the long term. This indicates weaker connectedness to the society and increased risk for social exclusion. When the social benefits aim to increase the activity of the individual in the society and in the labour market, the impact on survey participation was found to be positive or neutral. This supports our theoretical framework was presented in Chapter 2, the associations of social exclusion are reflected by reduced propensities to co-operate in surveys and can be to some extent detected by quantifying the social welfare policies. We did not find the length of the unemployment and receiving unemployment benefits associated significantly with the non-response. In contrary, income support aimed for individuals in economical deprivation increased the risk of non-response. Care support for the disabled is a benefit which is based on health conditions, while rehabilitation support depicts a policy aiming to improve the labour market situations of disabled and was found to increase also the survey participation. The care support was also found to increase significantly the risk of non-response.

The survey participation is increased by factors indicating increased social capital, such as living in households with children or living with a spouse. In addition human capital, measured by a simple indicator i.e. having further education, increases the odds of full participation. Similarly, relatively good economical conditions increased the survey participation. Highest income quintile and those with capital income were estimated to have higher response propensity than people with less income. The capital income and the dwelling information are latent variables indicating the cumulated wealth of income in the past that the individuals have been able to save and invest in terms of improved housing conditions, or increase financial wealth that has yield returns in survey year 2000. Thus the factors found increasing the survey participation are projections of economic and social wellbeing, whilst factors reducing the participation are associated with poor living conditions, lower social and economic status.

3.8. Interviewer perception data

After the fieldwork of Health 2000 survey, the interviewer perception survey was collected to measure interviewer attitudes and experiences related to the health survey and the fieldwork arrangements. The main purpose was to evaluate the importance of interviewer experience and attitudes together with characteristics of the target person, dwelling unit and the small area on survey participation. The contents of the interviewer perception survey have been described in Appendix 3.2. The questionnaire applied also questions developed by Couper and Groves (1992) and Lehtonen (1996). In order to reduce the response burden in the interviewer survey, the demographic characters of the interviewers were obtained from the interviewer database maintained by Statistics Finland. Due to data sensitivity issues, grade and salary information of the interviewers were not available for this analyses.

Out of 157 interviewers, 12 interviewers did not respond to the interviewer perception survey. Subsequently, 525 eligible individuals of the sample were assigned to these non-responding interviewers. Thus without data editing only 7423 individuals sampled out of the 7946 in-scope units could be used for the modelling of survey participation with interviewer characteristics because of the missing information for the perception of the interviewer. Our assumption is that the results of the interviewer effect analysis may not be representative if the respondents whose interviewer had not responded to the interviewer perception survey are excluded from the analysis. Therefore, the categorical variables of the interviewer perception survey were recoded. To denote for the non-response of the interviewer, a new category was added. This allows us to analyse whether interviewers' participation to their perception survey had any connection with their fieldwork performance in the multivariate analysis.

The interviewer performance

The purpose of studying the interviewer characteristics in conjunction with the completion of interviewer assignments is to broaden the understanding of possible underlying causes and relations influencing the outcome of interviewer-respondent interaction and survey participation. We begin by examining the success of the fieldwork by the interviewer assignments. The fieldwork performance is assessed against the interviewer characteristics and perceptions. The fieldwork performance is measured by the proportion of achieved interviews within the interviewer assignments. In addition, we have assessed the performance of interviewers separately for the contacting of individuals and persuading the contacted to participate to the survey.

Figures 3.5 and 3.6 illustrate the unbalanced design of the Health 2000 data. Figure 3.5 shows how the 157 interviewer assignments were spread across the 80 health centre districts (HCD) in the sample. Majority of the interviewers were interviewing individuals at 2 to 7 health centre districts. Figure 3.6 reveals that the number of interviewers operating in the same health centre district varied considerably. In majority of the health centre districts, there were 4 to 10 interviewers aiming to interview individuals. The survey design imposed the overlapping between interviewer assignments and regional health centre districts. The survey design imposed that the health interviews were to be carried out prior the mobile health examination centre would begin the medical measurements of the respondents at the area. Therefore, there was an unexceptional clustering of the interview assignments and a tight schedule for obtaining the health interviews prior the medical examination began in the local areas.

Figure 3.5 Number of health centre districts at which interviewers operated

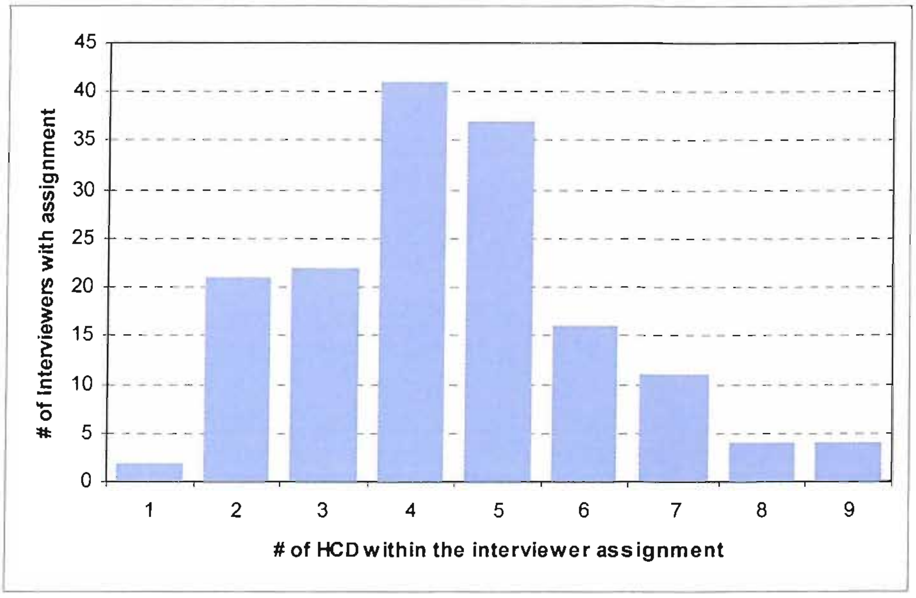
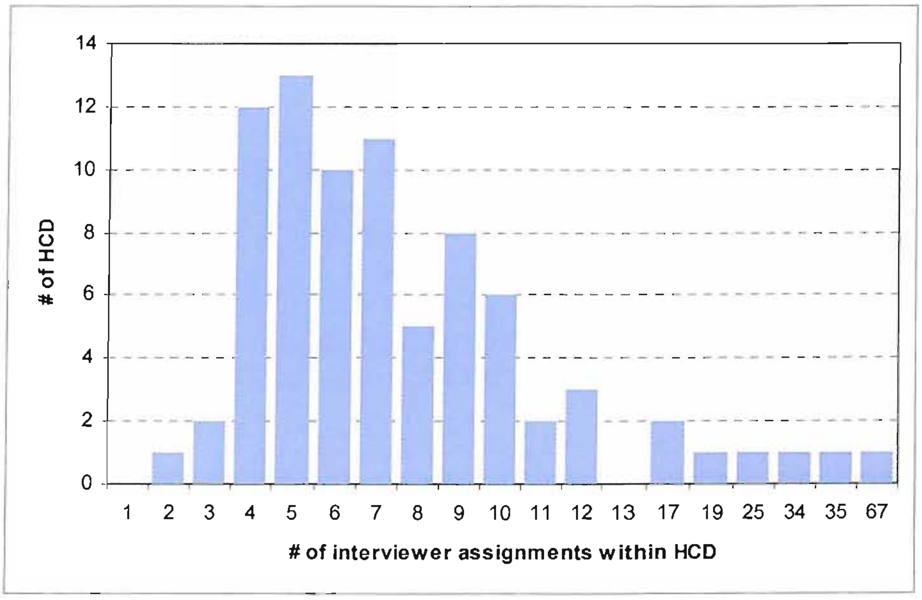


Figure 3.6 Number of interviewers operating in different health centre districts



The interviewer perception survey focused on the experiences, the professional attitudes, and the work motivation of the interviewers. Out of 157 interviewers who worked in the Health 2000 survey, 145 responded to the interviewer perception survey. Thus, the response rate for interviewers was 91.8% and non-response rate 8.2%. Out of the recently recruited interviewers, 13.7% did not respond. From Figure 3.7 we can observe that the participation of the interviewers to the perception survey is not affected by their fieldwork performance[10]. Actually, all interviewers either with very high or relatively

[10] The response rate per interviewer assignment, i.e. completion rate, has been calculated as an unweighted proportion, based on the success of achieving accepted health interview amongst those in the original interviewer assignment. The completion rate is defined as a proportion of individuals interviewed to the health interview by the interviewer out of

low response rates participated to the perception survey. The lowest proportion of participation was amongst the interviewers whose achieved response rate varied between 70–79 % of individuals sampled. From Figure 3.8 we can compare the proportional distribution of all interviewers to those interviewers who participated to the perception survey by their achieved response rates. We can observe that the impact of interviewer non-response is very small in the groups were interviewers had achieved cooperation with 70–79 or 80–89 % of individuals. Otherwise, there is no observable difference in the proportional distributions. Thus, the success of the fieldwork of the interviewers does not have a significant impact whether they participated to the interviewer perception survey.

Figure 3.7 Proportion of interviewers participating to the perception survey by their achieved completion rates within the interviewer assignments

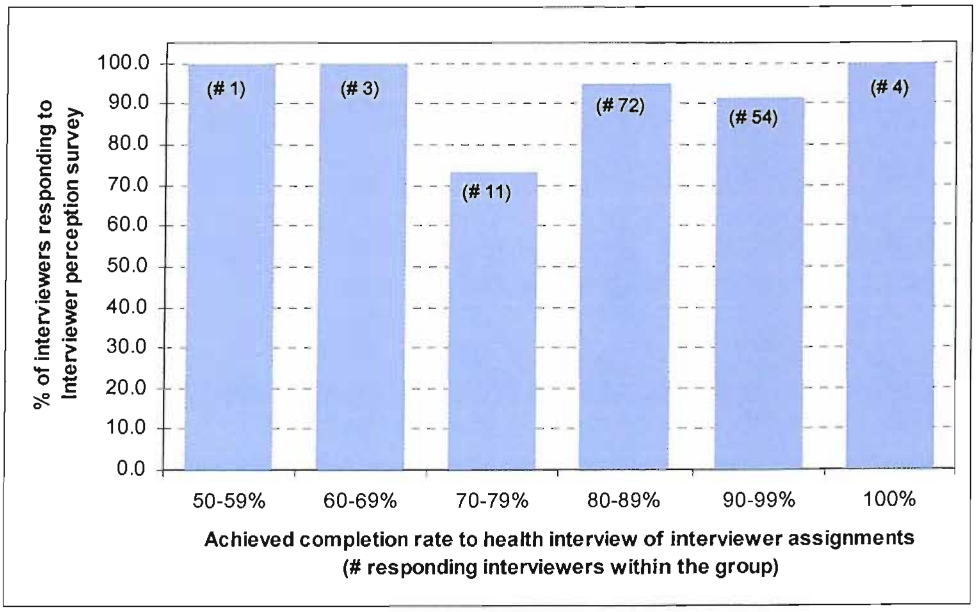
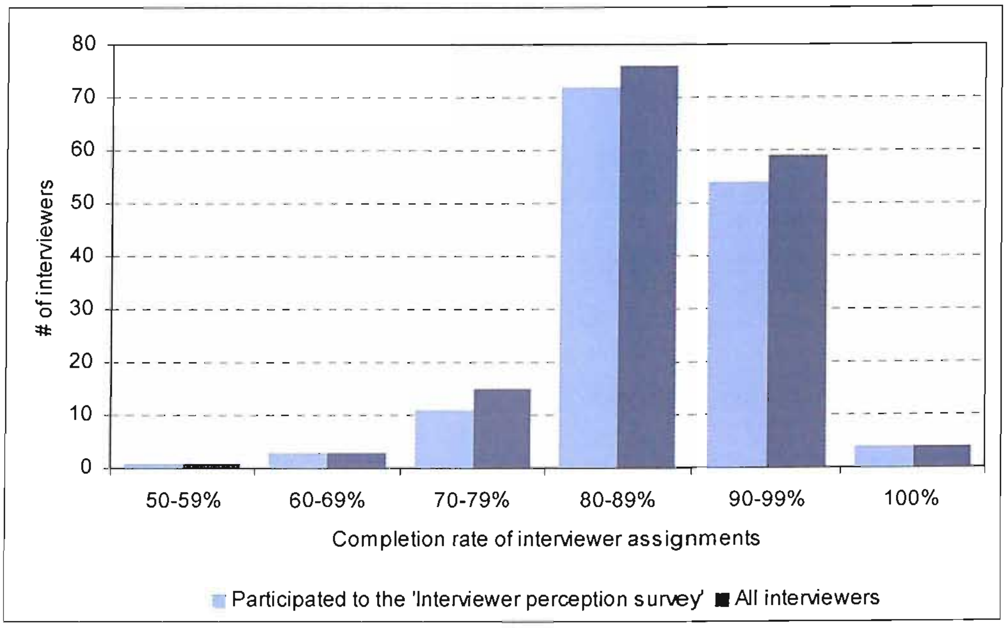


Figure 3.8 Proportional distribution of interviewers by their completion rates



the number of individuals allocated to the interviewer. The completion rates to the health interview varied between 55 and 100 percentages across the interviewer assignments.

3.9. Survey taking climate

The general trust to the authorities and to the government supports the survey environment in Finland. The public has accepted the wide use of personal identification codes, merging the individual level registers for administrative and statistical purposes in the society. The personal level information is protected by laws such as the Finnish Statistics Act (280/2004). However, the response rates have a declining trend for many surveys, as observed in the Health Behaviour Survey (Helakorpi, et al., 2005). Between the survey years of 1978-79 and 2000, the response rate has declined from the 83 % for men down to 64 %, and from 84 % for women to 75 %. As referred in the previous Chapter, Djerf (2004) has analysed the connection between the voting behaviour and survey response of the Labour Force Survey in Finland. He has found confounding declining trends in both behaviours of social participation. In Table 3.4 we can compare the response rates achieved in variety of social surveys conducted in Finland around the survey period of Health 2000, which has the response rate of 80 %. Highest response rates were gained in telephone surveys, and lowest in postal surveys. We conclude that the response rate gained in Health 2000 survey can be regarded satisfactory in comparison to face-to-face surveys in general, to other national health surveys and to other face-to-face surveys in Finland. In Statistics Finland, the target for response rates in face-to-face surveys has generally been set to be above 70 % of eligible sample units.

Table 3.4 Achieved response rates in selected health and social surveys in Finland 1998-2002

Survey	Survey Year	Achieved Response rate	Responding unit	Target population in Finland	Data collection method
Alcohol Consumption Survey (Mustonen, Mäkelä, Metso, and Simpura, 2001)	2000	78 %	Individuals	15-69 residents	CAPI
Adult Education Survey (Iisakka, 2004a)	2000	74 %	Individuals	18-79 residents	CAPI
Household Budget Survey (Iisakka, 2004a)	2001-2002	63 %	Households	All resident households	CAPI[11], diaries, and administrative registers
Time Use Survey (Iisakka, 2004a)	1999-2000	61 %	Individuals	All 10+ aged household members	CAPI of households and individuals, and diaries
Labour Force Survey (Iisakka, 2004a; Djerf, 2004)	2000	86 %[12]	Individuals	15-74 residents	CATI[13]
Income Distribution Statistics (Statistics Finland, 2002)	2000	83 %[14] 95 %[15]	Households	All resident households	CATI and administrative registers
Wealth Survey[16] (Iisakka, 2004a)	1998	66 %	Households	All resident households	CATI and administrative registers
Health Behaviour Survey (Helakorpi, et al., 2005)	2000	70 %	Individuals	15-64 residents	Postal survey
Social Security Barometer (Iisakka, 2004b)	2000	67 %	Individuals	18-74 Finnish or Swedish speaking residents	Postal survey
International Social Survey Programme (ISSP) (Iisakka, 2004b)	2001-2002	59 %	Individuals	15-74 residents	Postal survey

[11] CAPI: Computer aided personal interviewing, i.e. face-to-face interviewing
[12].Yearly average of monthly survey (Djerf, 2004)
[13] CATI: Computer aided telephone interviewing
[14] Response rate for 1st year panel (Statistics Finland, 2002)
[15] Response rate for 2nd year panel (Statistics Finland, 2002)
[16] Supplementary survey, conducted jointly with the Income Distribution Statistics

3.10. Conclusions

We have used a significant amount of derived variables from auxiliary data resources to explain response behaviour of individuals in the Health 2000 survey, described in this Chapter. The variables chosen are related to the survey participation theories, social exclusion, or both. Our findings of the survey participation, using simple logit analysis, are confounding with the survey non-response literature. However, our key finding relates to the association between social benefits by their policy targets on survey participation of individuals. We have found some evidence that social benefits aiming to increase the activity of people, who are in the risk of social exclusion, may be associated with increased survey participation. In contrary, benefits providing support without intension to improve the economical independency of the individuals in the long term, seem to decrease the response propensity. For example, people with disabilities receiving long-term support, and people receiving benefits such as income support, have increased risk not to co-operate fully. At the same time, rehabilitation programs supporting people to actively participate to the society, and help their entry or return to labour market, were found to be associated with increased likelihood to respond to surveys.

The correctness of the classifications' characteristics are essential, especially, when the information is used for modelling and estimating people's behaviour. An individual may have a different view on their status than the general classification rules imply, which can also contribute to their survey behaviour. In addition, one must be cautious on the interpretation and implications of the results, if there is a risk of an underlying classification error. In our analysis, we have preferred, as far as possible, objective and norm-free variables. For example, the size of the dwelling unit indicates the size of the group of people that are closely related to each other and live together in the dwelling, which may differ from the size of the register derived household and family size. For example, the register family consists of only two generations. In addition, two adults of the same sex cannot form a family in the family classification used for register data. It is evident that the conventional classifications are controversial if used as such in modelling behaviour, as they may differ largely from perceptions of the population as a whole, and from the perception of the individuals in the data.

The large number of explanatory variables in the estimated models suggests complex underlying processes. In fact, the full survey participation is likely to consist of multiple processes for which the individuals sampled are exposed during the data collection. In our empirical study the survey imposed a huge response burden consisting of many survey components with varying modes and a transition from interviewer administrated part to a health examination centre. For gaining more thorough knowledge of these processes, we need to examine when the possible non-response is likely to occur in a survey with multiple data collection components, and whether the sub-population groups differ in their response behaviour. We suspect that deeper analysis will reveal differential survey response behaviours in the sample. We will study whether there are large differences in survey non-response regarding the data collection phase and mode of data collection in relation to individual level characteristics. In Chapter 4, we explore the magnitude of survey non-response at various data collection phases. In Chapter 5, we incorporate the interviewer characteristics and perception information into survey participation analysis.

4. Sequential modelling of survey participation

4.1. Introduction - Purpose of the models and their use

In this Chapter, we focus on modelling the survey participation in multiphase cross-sectional surveys on individuals where the data loss can be due to initial non-response or subsequent survey attrition. Multiphase surveys are commonly used in health and social surveys, which gather information using both interviews and additional data collection components. The patterns of non-response can then differ significantly across the data collection phases. For example, the increasing response burden, the varying nature and the topic of the data collection components can influence the occurrence of non-response. Groves and Couper (1998) have emphasised the importance of studying the non-response separating the event of contact from the event of complying with the survey request. In our approach, the latter will be extended into studying the success of gaining the initial response and the maintenance of co-operation across all data collection phases. Applying the theory of response effects (Sudman and Bradburn, 1974) on the survey participation, we combine the original data collection components by their nature into aggregated phases. In addition, Riedel-Heller et al. (2000) have analysed the impact of cognitive skills of respondents in health surveys. We merge and extend these approaches to explore whether the non-response in multiphase surveys occurs mostly at those phases where the task structure is cognitively most demanding. This is relevant, especially, in analysing the completion and the return of self-administered questionnaires.

The specific objective of this chapter is to incorporate the theoretical framework of the survey participation in multiphase surveys and the social exclusion into statistical modelling. The purpose of the response probability models are firstly to detect the influential characteristics that affect the survey non-response, and secondly to provide an estimate of individual's probability to response. We will study the predicted response probabilities in the empirical part of this Chapter and assess the goodness of fit of the estimated models. We will benchmark the conventional logistic regression models against the more complex modelling. Generally, non-response models are based on logistic type regression for the binary outcome. These models have been recommended for the non-response analyses by Little (1986). Traditionally non-response modelling uses simple logit models for detecting the significant factors of the non-response. Alternatively, contacting and participation are modelled separately, as recommended by Groves and Couper (1998). In order to construct informative models, we assess methods for exploring the complete flow from contact to co-operation and the possible survey attrition.

One attempt to improve modelling of survey non-response was proposed by Laiho and Lynn (1999). Their approach involved identification of all the necessary survey phases in the process leading to the non-response. Acknowledging that each phase is conditional on the outcome of the previous one, they used the sequential logit models, which Mare (1980) introduced for modelling school continuation decisions. The survey participation process can be represented as a series of models for co-operation at subsequent survey steps, each of which is conditional upon the success at the previous one. In other words, the participation process is seen as a sequence of successive survey events, which are all prone to non-response occurrence. Thus we will also explore models that allow more flexibility in the model than the simple logit model. We will assess the usability of the multinomial and the cumulative logit models (e.g. Fienberg, 1980, McCullagh and Nelder 1989; Agresti 1990). We will also discuss the advantages and disadvantages of the model choice. In practise, more restricted model approach, such as simple logit model, can be more favourable than sequential logit models, as more complex models can be labour intensive in the model construction and testing phase.

In addition to logistic type regression modelling, we will investigate whether methods developed for event history analysis (Diggle et al., 2002; Allison, 1984; Cox and Oakes, 1984 and Kleinbaum, 1996) and especially, the discrete-time hazards (e.g. Singer and Willett, 1993), could be utilised successfully for modelling the non-response in multiphase surveys. Longitudinal methods are traditionally used for analysing the development of the same phenomena in time. Although the survey task can vary across the time points, we bring into the debate whether the outcome variable indicating the success or the failure at each phase could be analysed in longitudinal context. In other words, we assess how event history analysis could be exploited for exploring and modelling the probability of response or non-response occurring at certain data collection phases. Event history analysis, such as the survival and the hazard functions, are applied to locate those data collection phases and sub-populations that experience significant loss of data. Subsequently, the discrete-time hazard models enable us to evaluate survey attrition in multivariate context across data collection phases.

As emphasised above, one of the purposes of non-response modelling is to detect the influential characteristics that affect the survey non-response. We will investigate whether social exclusion is a crucial factor affecting survey participation. Social exclusion has often been associated with survey variables in social and health surveys. Conditions, such as poor health, low education level, inactivity in labour market and economic deprivation, are quantitative indicators of social exclusion. However, these factors are also found to be associated with survey non-response (e.g. Gray et al., 1996; Couper, 1997). This raises a question whether socially excluded people are poorly represented in survey estimates describing the society. Thus in our analysis, covariates of the survey participation models are chosen for testing whether social exclusion is a significant factor of non-response.

Social exclusion is in most cases a consequence of multiple severe difficulties. Therefore, developing prevention and helping policies may seem more important than studying the missingness of the special sub-population group from official statistics. However, the suspected data loss from socially excluded people can cause a vicious circle if the policy evaluations fail to capture the range of perceived conditions and their trends. Subsequently, the surveys may fail to provide reliable estimates for measures of health or social inequalities. The association between social exclusion and survey non-response is of particular interest of ours in modelling the survey participation. Using the quantified profiling of social exclusion from Chapter 3, we apply the theory of response behaviour for modelling non-response in multiphase surveys. These models will be used later for non-response adjustment in Chapter 6, when we aim to develop improved methods for obtaining less biased survey estimates.

The Finnish Health 2000 data is used for the empirical analysis in this Chapter. The survey was chosen as it contains a rich set of information with links to auxiliary register information both at micro and macro level. This enables us to test survey participation hypothesis using characteristics of individuals, households, and local areas. Previously in non-response studies, the hypotheses have been mainly analysed using area aggregated census data. Alternatively many studies using auxiliary information at individual level, test the research hypotheses actually for panel data based on the information obtained from previous panels. That kind of setting excludes the step of initial survey participation due to lack of data at individual level for that phase. Thus, the results of our application with direct linkage can bring new knowledge about the phenomena and be useful in the planning or re-designing similar surveys to prevent non-response in special sub-populations.

We begin by applying the event history analysis for multiphase surveys. In order to demonstrate the benefits of breaking the survey participation analyses into phases, we use the survival and the hazard functions to evaluate the loss of data. We then proceed with the modelling by applying the sequential logit models for survey response that recognise partial co-operation. The sequential logit model allows for a diverse set of explanatory variables and the coefficients are unrestricted across data collection phases. Later on, we return to the single model approach and examine the usability of the multinomial logit model allowing for varying coefficients across data collection phases. We then explore the usability and the statistical assumptions of the discrete-time hazard model and the cumulative logit model. Both of these are fixed models, allowing only the intercepts to vary across data collection phases. We conclude the modelling by compiling plausible response probability models and testing whether these compilations have any advantages in comparison to a single model approach. We then compare the statistical properties and the goodness-of-fit of the modelling techniques. Finally, we discuss the interpretation of the models, and bring into debate our research findings on whether social exclusion is significantly explaining survey participation behaviour.

4.2. Exploring survey participation in multiphase surveys using event history analyses

In this section we demonstrate the usefulness of the methods event history analyses for analysing the non-response and survey attrition of cross-sectional surveys with multiple data collection phases. The survey participation is separated into phases that follow closely the process of original survey design, described in Chapter 3. The event to be studied is the first occurrence of non-response across. The gradual break-off or dropout of some sampled units ceasing to co-operate is also called survey attrition.

4.2.1 Applicability of event history analysis for studying survey participation

Event history models (Allison, 1984; Cox and Oakes, 1984; Kleinbaum, 1996) allow for time component of ordered events. The event history models are generally applied to analyses of event occurrence or conditions that may change over the time when exposed to certain prevailing or changing factors. This logic can be transferred for the analysis of survey participation. Data collection phases can be treated as ordinal time points, and the non-response (or co-operation obtained) at each time point is the studied event of occurrence. Non-response can be seen as a repeatable binary event of response to multiple data collection phases at discrete time points. The event history analysis allows for differentiating the initial unit non-response and survey attrition occurred at later phases of data collection. Patterns of non-response and survey attrition can be explored using life tables, hazard functions and Kaplan-Meier survival function, which do not make strong distributional assumptions.

The data collection of health surveys consist generally of multiple phases of data collection components. Depending on the survey design, these components can be collected in predefined order at the same time, or they can be collected at various time points. The life table indicates the first data collection phase at which non-response occurs. The life table information consists of the number of individuals that are exposed to the risk of the first occurrence of non-response by data collection phases. It also describes the total number of individuals experiencing non-response, and subsequently the derived values of the survival function. The survival function $S(t)$ indicates the probability of respondent co-operating fully up to the data collection phase *t*. This function formulates the probability as a product of proportions of fully co-operating out of the eligible sampled individuals across subsequent phases. The probability is conditioned so that non-response to a survey component has not occurred prior data collection phase *t*. In other words, the probability of maintaining co-operation until phase *t* can be expressed as a survival probability that the time point T_i, when the co-operation of the individual *i* will cease, lies either at the time point *t* or later in the ordinal time axis:

$$S(t) = \mathrm{P}(T_i \geq t) = P\left(T_i > t-1\right). \tag{4.1}$$

where the polytomous dependent variable T_i^* indicates the outcome for individual sampled i and is defined using the data collection phases as before:

$$T_i^* = \begin{cases} 1 \text{ Not contacted} \\ 2 \text{ Contacted, not health interviewed} \\ \vdots \\ 6 \text{ Complete response} \end{cases} .$$

The survival function may be estimated by

$$\hat{S}(t) = \prod_{h=1}^{t-1} \left(\frac{n_h - d_h}{n_h} \right) \tag{4.2}$$

and the standard error of the survival function is estimated by (Greenwood, 1926):

$$\hat{s}_t = \hat{S}(t) \sqrt{\sum_{h=1}^{t-1} \frac{d_h}{n_h (n_h - d_h)}}, \tag{4.3}$$

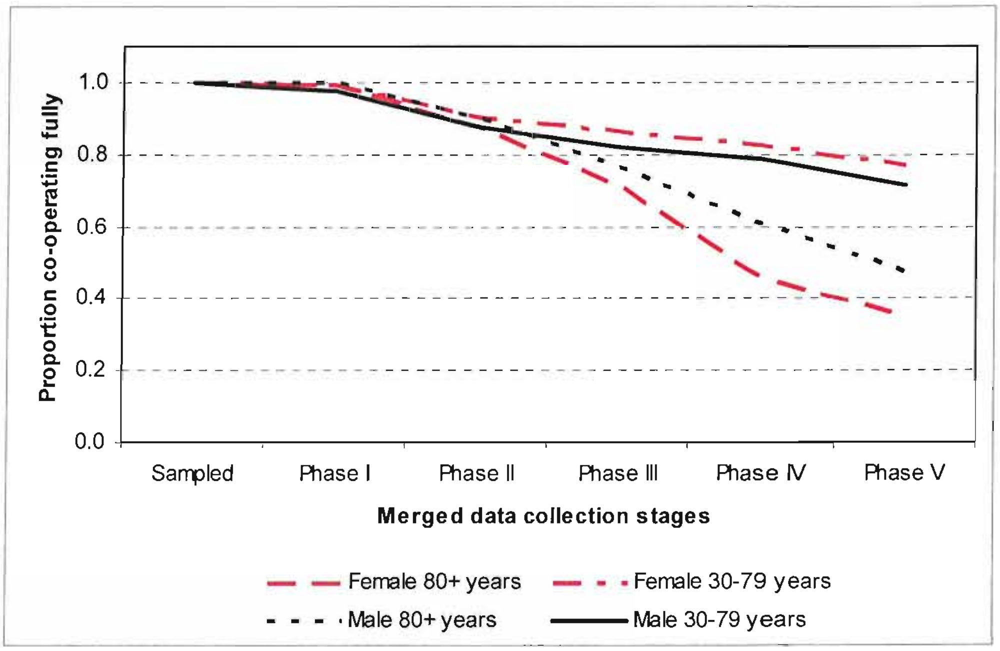
where $t = 1, \ldots, t_{\max}$ (in our analysis the $t_{\max} = 5$) and T_i takes possible values of $1, \ldots, t_{\max}$. The discrete time point indicator T_i is conditional on the individual i experiencing non-response. The number of individuals that are exposed to the risk of non-response occurring are denoted by n_h, and d_h indicates the number of individuals who cease to co-operate at that time point.

4.3.2 Event history analysis of survey participation to Health 2000

Using the event history analysis, we can observe risk groups in which the survey attrition is high across data collection phases, and assess at which phase the attrition is most severe. We apply the methods using ordered and aggregated data collection phases similarly to usage of time points in longitudinal data analysis. Table 4.1 presents the unweighted and weighted survival function for the achieved participation rate in the Health 2000 data. The estimation of the weighted survival function is explained in the Appendix 4.1. It emerges from the table that the participation rate decreases from 99 per cent of contacted to 70 percent with full response. Out of 7951 eligible individuals sampled 7840 were interviewed on their health, 6611 co-operated fully with medical measurements after being interviewed on their health and symptoms, and 5608 participated fully to all survey components. The relative importance of non-respondents who cease co-operation after the contacting phases is less for the weighted than for the unweighted function. The relative difference is largest at the phase of medical measurements and tests which had a high response burden. This suggests that non-respondents have on average lower design weights than respondents. This is due to the high attrition amongst those aged 80 or above, who also had twice the inclusion probability than younger people, as shown in the survival plot in Figure 4.1. Otherwise, the results from the weighted and unweighted survival function are almost identical, and from now on we focus on unweighted analysis.

The significance of age group on attrition appears to be more significant than the sex of individuals. Proportionally a significant number of elderly drop out of the study at the medical measurement. Our findings suggest that younger people are more difficult to contact, but gaining and maintaining the co-operation with increasing response burden is most difficult amongst the elderly. This result is confounding with previous findings of Groves (1989) also referring to Cobb et al. (1957), who have found evidence that high refusal rates among the elderly were a particular problem for studies of health conditions. Similarly, men are more difficult to contact than women, but once contacted and achieving co-operation at the initial data collection phase, men are more likely to stay in the survey than women, who may be more affected by the response burden.

Figure 4.1 Unweighted estimated survival function by age group and sex[1]



[1] According to the classification presented in Table 4.1 the merged data collection phases are:
- 0 Eligible sample member (implicitly assumed in table 4.1)
- I Contacting target person
- II Health interview (long or short)
- III Symptom interview (long or short)
- IV Medical measurements and tests
- V Self-completion questionnaire 1, 2, 3 and the nutrition questionnaire

Table 4.1 The estimated life table for sampled units in the Health 2000

Data collection stage	Phase	Sampled individuals requested of co-operation	First occurance of stage-non-response	Censored	Partici-pation rate achieved	Standard error	Participation rate [95% Conf. Int.] Lower	Upper
Unweighted								
1. Contacting target person	I	7951	111	0	0.986	0.0013	0.983	0.989
2. Health interview (long or short)	II	7840	748	5	0.892	0.0035	0.885	0.899
3. Symptom interview (long or short)	III	7087	476	0	0.832	0.0042	0.824	0.840
4. Medical measurements and tests	IV	6611	437	0	0.777	0.0047	0.768	0.786
5. Self-Complition questionnaires*	V	6174	566	0	0.706	0.0051	0.696	0.716
Full completion		5608	0	5608				
Weighted by inclusion probabilities								
1. Contacting target person	I	7951	112	0	0.986	0.0013	0.983	0.988
2. Health interview (long or short)	II	7839	728	3	0.894	0.0034	0.888	0.901
3. Symptom interview (long or short)	III	7107	441	0	0.839	0.0041	0.831	0.847
4. Medical measurements and tests	IV	6666	363	0	0.793	0.0045	0.784	0.802
5. Self-Complition questionnaires*	V	6303	550	0	0.724	0.0050	0.714	0.734
Full completion		5754	0	5754				

* Self-completion questionnaire 1, 2, 3 and the nutrition questionnaire

The connection between survey non-response and social exclusion is analysed at individual level via their prevalence of unemployment spells, indicator for person's type of living being unknown, whether the native language is other than the official languages in Finland, low education level, low income, proportion of social benefits of the total income and an indicator of not belonging to a family. In Chapter 2 these factors and their interactions were found to increase the risk of social exclusion. Some rather drastic drop-out patterns can be observed via the decline of the survival function by indicators of social exclusion (see Figures 4.2.1 to 4.2.8 in Appendix 4.2.).

There are strong indications on the non-ignorability of survey non-response in relation to the social inequalities and risk factors for social exclusion. In Figure 4.2.1, we present the estimated survival rate of sampled individuals by type of living and experience of unemployment. We find that the type of living is closely related to the risk factors of social exclusion, identifying those being outside household population and living in shelter-housing, oversized dwelling units or whose type of living remains unknown. People who are part of the household population have relatively stable survey attrition across data collection phases regardless whether they had experienced unemployment spells or not. In contrary, individuals outside the household population have a drastic reduction in survey co-operation from the contacting phase if they have had unemployment spells. This group has also a high risk for being socially excluded. If however, people outside the household population have not experienced unemployment their initial co-operation to the health interview is at the same level with members of the household population, but their co-operation reduces drastically at later data collection phases.

In Figure 4.2.2 we assess the survey participation by family status. The survival rate drops drastically for people whose family status is unknown after the health interview. The survival rate is also lower than on average for adults living still in their parental home and for single people not belonging to families. In Figure 4.2.3 (Appendix 4.2), we investigate further the associations of family background and maternal language on survey participation[2]. Instead of studying the response behaviour across ethnic population groups based on the census information, as is done in many studies in United States and United Kingdom (f. ex. Gray et al., 1996; Campanelli and Sturgis, 1997), it is common in the Finnish survey research, to analyse the response behaviour by the maternal language of individuals[3]. The languages are classified into Finnish and Northern Sámi, Swedish and other languages i.e. minority group languages. It can be observed that the co-operation and survey response was higher for the Finnish speaking majority and lower for the Swedish speaking minority. More importantly people speaking other language than Finnish, Sámi or Swedish as their native language had the lowest survival estimates at all data collection phases. They were hard to contact, whilst contacted one third did not participate to the health interview and significant number of health interviewed did not visit the health centre for the latter survey phases. This can indicate that the respondent perceived higher response

[2] As described earlier in Chapter 3, the health survey was bilingual as health interview was carried out both in Finnish and Swedish. In addition, the health centres had bilingual nurses in the regions with relatively high proportion of Swedish speaking minority.

burden, and the level of demand to participate was higher for those who could not use their native language in responding and who did not have further education. In comparison, the co-operation level is high throughout the data collection for people living in families. This may indicate that single people with minority language are more difficult to contact than other people in Finland. The underlying reasons should be investigated more in detail.

In Figure 4.2.4 we review the survival rate by educational background and native language. Our findings on the education level are consistent with Gray et al. (1996) who found that people with no further education have higher risk of becoming non-respondents at cross-sectional non-response analysis as well as have a higher risk to drop out of the longitudinal study in UK. In our Finnish data the education level has the same consistent pattern for all language groups, people with further education have higher participation rate than people who have only basic educational background. But even with further education, people in the smallest language minority group have lower survey participation than people in the main language groups with no further education.

In Figures 4.2.5 and 4.2.6 the survival rates the survey attrition are strongly connected with the self-assessed health status and physical mobility of the respondents. The better the self-assessed health or physical mobility is ranked in the health interview the more likely the respondents co-operation is maintained at later data collection phases. Respondents with worse rankings are likely to drop out at later phases. In Figures 4.2.7 and 4.2.8 we use health register information and examine the associations between long term illnesses, diabetes and asthma, to the survey participation. It seems that the probability of contact and success of gaining response is not strongly connected with having both diagnosis of diabetes or asthma and being entitled for reimbursements on medical expenses arising from their treatment. However, the survey attrition is greater at later data collection phases for respondents with diabetes (with the reimbursement entitlement) than for other respondents, which may at least partly be explained by the fact that the prevalence of diabetes increases by age.

4.3.3 Hazard rate indicating the risk of occurrence of non-response in multiphase surveys

The hazard rate indicates the risk or probability for an event of interest to happen in condition it has not happened before (Cox, 1972; and Brown, 1975). We will use the hazard rate for studying the event of first occurrence of non-response by aggregated data collection phases defined previously. More precisely the hazard rate measures the probability of a non-response to occur for an individual sampled i at a particular data collection point t given that the individual has not experienced non-response prior phase t. The hazard function provides an informative tool for non-response analyses. It measures the average risk of the non-response occurring for the first time at each phase among those respondents that are still eligible for full response co-operating at an acceptable level on all previous phases. Thus the hazard measures the probability of failure and indicates the phases at which the individuals are most liable for failure or alternatively, whether they are likely to respond fully.

---

[3] Language groups are derived from the registered maternal language in the population information system based on the

Following the presentation given in Yamaguchi (1991) the discrete-time hazard function $\lambda(t)$ for the risk of non-response occurring at data collection phase t conditional that the individual will experience non-response at t or later during the fieldwork can be written as a relative probability:

$$\lambda(t) = P\left(T_i = t \middle| T_i \geq t\right) = \frac{P(T_i = t)}{P(T_i \geq t)} = \frac{P(T_i = t)}{S(t)}, \tag{4.4}$$

where $S(t)$ has been defined earlier in (4.1), $t = 1,\ldots,t_{\max}$ and if non-response occurs for i then T_i takes possible values $1,\ldots,t_{\max}$. Conditional on the full co-operation prior to the phase t the hazard is a proportional probability of the non-response occurring for the first time at phase t or later without ruling out the possibility that the individual co-operates throughout all of the data collection phases and the event of non-response will never be measured during this period of the fieldwork. This indicates that those co-operating fully at the survey are still eligible of experiencing non-response if we were to monitor and model the participation to a follow-up survey at later phases. Using the definition of the hazard function $\lambda(t)$ we can re-write the survival function turning the sequence of successful events into a sequence of the possible failure not happening:

$$S(t) = P(T_i \geq t) = \left(1 - \lambda(t-1)\right)\ldots\left(1 - \lambda(1)\right) = \prod_{h=1}^{t-1}\left(1 - \lambda(h)\right). \tag{4.5}$$

The estimated hazard function can be used for explanatory non-response analyses as the only strong assumption made is the ordering of time points (data collection phases) and that the hazard can be defined by the probabilities of the event occurrence, which lie between 0 and 1. The maximum likelihood estimate of the hazard function is calculated by:

$$\hat{\lambda}(t) = \frac{d_h / n_h}{\left(1 - (d_h / n_h)/2\right)} \tag{4.6}$$

and its standard error is estimated by:

$$\hat{s}_{\hat{\lambda}_h} = \hat{\lambda}_h \sqrt{\frac{1 - (\lambda_h / 2)^2}{d_h}}. \tag{4.7}$$

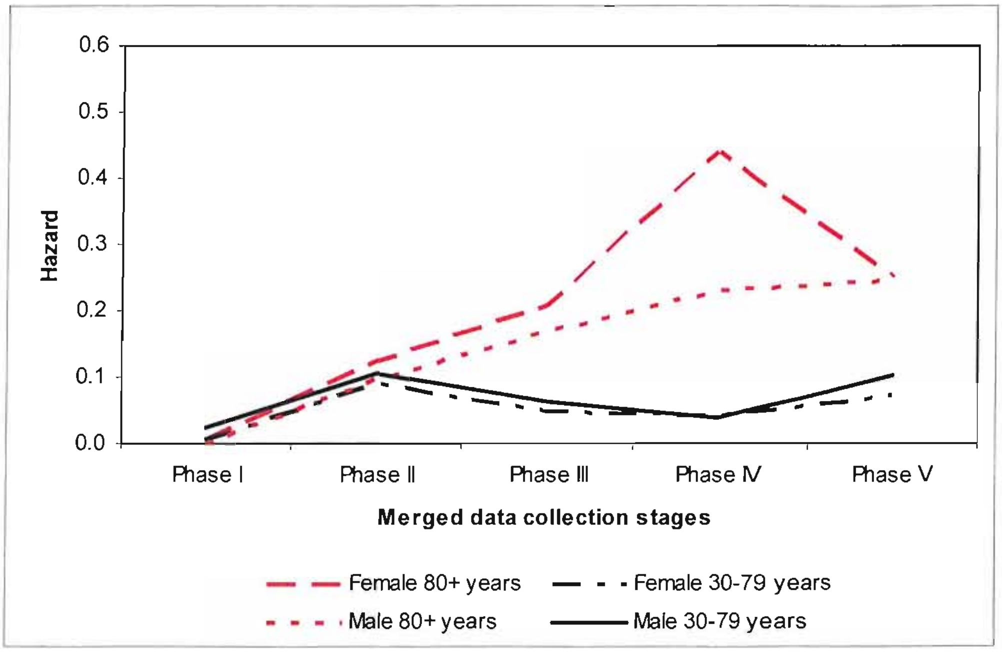
As for the survival function, the survey attrition and the risk for the first occurrence of non-response using the unweighted and weighted estimated hazard functions are almost identical, shown in Table 4.2. The estimation of the weighted hazard function is presented in Appendix 4.1. The interpretation of the hazard functions, presented in Figure 4.2 and Appendix 4.3, are the same as for the survival function. However, the advantage of the hazard function is that it underlines the data collection phases at which the drastic drop-out occurs.

---

information that the person (or parents) have informed as the person's maternal language.

The application has shown that the event history analysis can be more informative on survey participation processes than simple logit model presented in Table 3.4, which aggregates the survey outcome into one binomial variable. Although being significant factors of survey participation, some background characteristic constitute small sized sub-population groups, with low statistically significance in multivariate context. However, if the drop-out patterns are strongly related to sub-population groups and studied health status or social conditions the estimates may become biased in small domains. Therefore further analyses is needed and life tables, survival and hazard functions are very informative way of analysing the non-response as they allow us to study the sequential occurrence of non-response contrasting the population into groups of special populations. In the next section we will use discrete-time hazard modelling to study how meaningful these drop-out patterns are for the whole data.

Figure 4.2 The estimated hazard function by age and sex[4]



[4] As defined previously in the life table 4.1 the merged data collection phases are:
- I Contacting target person
- II Health interview (long or short)
- III Symptom interview (long or short)
- IV Medical measurements and tests
- V Self-completion questionnaire 1, 2, 3 and the nutrition questionnaire

Table 4.2 The estimated hazard and the cumulative failure in the Health 2000 by aggregated data collection phases

Data collection stage	Phase	Total at begin-ning	Cumulative Failure	Standard Error	Hazard	Standard Error	[95% Conf. Int.] Lower	Upper
Unweighted								
1. Contacting target person	I	7951	0.014	0.0013	0.014	0.0949	0.012	0.018
2. Health interview (long or short)	II	7840	0.108	0.0035	0.100	0.0365	0.069	0.084
3. Symptom interview (long or short)	III	7087	0.168	0.0042	0.069	0.0458	0.063	0.079
4. Medical measurements and tests	IV	6611	0.223	0.0047	0.068	0.0478	0.062	0.078
5. Self-Complition questionnaires*	V	6174	0.294	0.0051	0.096	0.0420	0.087	0.107
Full completion		5608						
Weighted by inclusion probabilities								
1. Contacting target person	I	7951	0.014	0.0013	0.014	0.0943	0.012	0.018
2. Health interview (long or short)	II	7839	0.106	0.0034	0.097	0.0370	0.069	0.084
3. Symptom interview (long or short)	III	7107	0.161	0.0041	0.064	0.0476	0.058	0.073
4. Medical measurements and tests	IV	6666	0.207	0.0045	0.056	0.0525	0.050	0.065
5. Self-Complition questionnaires*	V	6303	0.276	0.0050	0.091	0.0426	0.085	0.104
Full completion		5754						

4.3. Multivariate event history analysis for survey non-response by applying discrete-time hazard modelling

Previously in Chapter 3, we analysed survey non-response using a simple logit model with binary response indicator. However, descriptive event history analysis in the previous Section demonstrated that in surveys with multiple data collection phases, co-operation of individuals is lost due to differential risk factors influencing at different phases. To examine and test the significance of these differences, we apply discrete-time hazard modelling, which models the varying level of risk across data collection phases. Discrete-time hazard models have been applied to analyse, for example, the occurrence of mortality (Allison, 1982; Xie 1994) or change of employment (Singer and Willett, 1993). We test the usability of these models for modelling the non-response when data is collected at multiple data collection phases with varying response patterns.

Our strongest assumption is that the data collection phases can be ordered, and that they are discrete time points. We restrict the conditional probability of the hazard to be in between 0 and 1, as in other logistic regression models. The model allows us to study which of the predictors have significant explanatory power in multivariate analysis. Based on Cox (1972), Allison (1982), and Singer and Willett (1993), we use individual level data to model the discrete-time hazards for each individual for their probability to experience non-response at any data collection phase, given that no previous non-response event occurrence has been observed. The overall discrete time function defined in (4.2) is extended to be defined at individual level and conditioned on the characteristics of the sampled individual and their local area:

$$\lambda_{it} = P\left(T_i = t \middle| T_i \geq t, \mathbf{X}_i\right), \tag{4.8}$$

where $\mathbf{X}_i$ refers to the background information available for all individuals sampled. The information formulated via the $\mathbf{X}_i$ matrix contains information for i individuals on p background variables. In our analysis all background information is time invariant, so we can change into vector notation to indicate the auxiliary data for the ith individual by $\mathbf{x}_{ip}$. The t resembles the duration variable indicating the number of $t-1$ phases at which full co-operation has been achieved, when the individual i has been exposed to non-response. Subsequently, we denote the probability that sampled individual does not experience non-response at phase t conditional upon that no previous event has occurred previously by $(1-\lambda_{it})$. Following the earlier notation for logit models and response indicator in Chapter 2, we use the response indicator R_{it} to denote the response for individual i at data collection phase t:

$$R_{it} = \begin{cases} 1 \text{ if response for individual } i \text{ at data collection stage } t \\ \qquad\qquad 0 \text{ if non-response for } i \text{ at } t \end{cases}. \tag{4.9}$$

Indicating that co-operation is obtained at phase *t* we write $R_{it} = 1$, which is conditional on response at previous subsequent phases i.e. $\sum_{h=1}^{t-1} R_{ih} = t-1$. The time point for the plausible first occurrence of non-response is then $T_i > t$. Subsequently, the probability of obtaining co-operation from the individual *i* at the time point *t* can be denoted by:

$$P(T_i > t) = P(R_{it} = 1) = S(t+1). \tag{4.10}$$

This implicitly indicates that the first failure of co-operation will happen after the phase t, if it will happen at all during the fieldwork. Thus we can re-write the probability of non-response λ_{it} as a product of conditional probabilities:

$$\lambda_{it} = P(T_i = t | T_i \geq t, \mathbf{x}_i) = P\left(R_{it} = 0 \middle| \prod_{h=1}^{t-1} R_{ih} = 1, \mathbf{x}_i \right), \tag{4.11}$$

where $t = 1,\ldots,t_{\max}$, $h = 1,\ldots,t$ and if for any t $R_{it} = 0$ then T_i takes possible values $1,\ldots,t_{\max}$.

The probability of the non-response occurring the first time is modelled using the logit link function for the odds of the hazard. In Chapter 2, the logit transformation of the log of the odds of survey response was defined in (2.6) and (2.7). Similarly, the hazard of non-response occurring λ_{it} is constrained to be within $[0,1]$. In discrete-time hazard model, the intercept β_0 can be further divided into two essential attributes: into a baseline profile of risk defined by the α_t coefficients specific for *t* phases and into a shift parameter that captures the effect of the predictor on the baseline profile (Singer and Willet, 1993). The α_t coefficients represent the base value of logit hazard at the data collection phase *t*, and the shift parameters estimate the effect of one unit difference in the explanatory variables on the risk of the non-response across all phases. Thus the model has separate constants for each data collection phase and the $\beta_1,\ldots,\beta_P$ coefficients of the explanatory variables are constant over the phases. The model further assumes that we have information available on all sampled individuals. The information is formulated via the $\mathbf{x}_i$ vector. Subsequently, the discrete-time hazard model estimates the probability of first-time non-response for individual *i* at phase *t* by:

$$\text{logit}(\lambda_{it}) = \log\left(\frac{\lambda_{it}}{1-\lambda_{it}} \right) = \left(\alpha_1 d_{i1} + \ldots + \alpha_{T_i} d_{iT_i} \right) + \left(\beta_1 x_{i1} + \ldots + \beta_P x_{iP} \right), \tag{4.12}$$

where $\left[d_{i1},\ldots,d_{iT_i} \right]$ are a sequence of dummy variables indexing time periods, T_i refers to the last time period when sampled individual *i* was eligible for modelling the probability of non-response occurring for the first time, and each intercept parameter $\alpha_1, \alpha_2 \ldots \alpha_{T_i}$ represent the value of logit hazard at *t* for individuals in the baseline group.

Interpretation of the estimated discrete-time hazard model

The discrete-time hazard models were estimated after transforming the person-event data into person-period data and the parameter estimates are presented in Table 4.3. The final discrete-time hazard model finds the most powerful factors explaining the occurrence of non-response at aggregated data collection phases from the person level characteristics such as age group, sex, maternal language, education, family life cycle, income and benefits. The main findings of the model are to some extent in line with the simple logit model (see Chapter 3). However, some variables such as the indicator of not belonging to families and number of weeks been unemployment are predicting the opposite way in the descriptive event history analysis. The plausible reason for this is the dominance of the person-event data structure.

Table 4.3 shows that when assessing the proportional improvements in the model using the measures introduced in (3.2) and comparing the model fit to the simple logit models presented in Table 3.4, the discrete-time hazard model has very high model fit. For the respondents, the model predicts seemingly correctly high response probabilities for all responding units. The range of the estimated response probabilities varies between 0.93 and 0.99. In comparison, the model predicts low probabilities for non-contacted individuals, varying between 0.24 and 0.79. For non-respondents to the health interview the distribution of predicted response probabilities narrows ranging from 0.34 to 0.75. The predicted response probabilities are at lower level for the elderly aged 80 or over for all non-responding groups. In Figure 4.3, we look at the proportion of respondents by decile groups of predicted response probabilities. The proportion of respondents is greatest at the tenth decile. Thus the discrete-time hazard model is conservative in prediction of failure.

The discrete-time hazard models restrict the modelling into a single model where both the estimated covariates and the set of explanatory variables are fixed for all t phases allowing the variation across the phases only to be incorporated via the intercepts, which can vary from across phases. The discrete-time hazard model seems to be a reasonable tool for analysing the factors behind the non-response after the fieldwork, but it is not suitable for predicting survey response behaviour, because the model assumptions require the knowledge on person-time event data. Therefore these models are limited in analysing the factors behind the non-response.

Table 4.3 Parameter estimates for discrete-time hazard model predicting survey response

Explanatory variables	Estimate	Standard Error	Odds Ratio	Pr > t-value
Intercept:				
Phase I - Contacting	-20.14	181.30	0.00	0.91
Phase II - Health Interview	-4.93	0.09	0.01	0.00
Phase III - Symptom interview	-3.97	0.09	0.02	0.00
Phase IV - Health examination	-3.64	0.09	0.03	0.00
Phase V - Self-completion questionnaires	-3.36	0.08	0.04	0.00
Characteristics of the individual:				
Age group:				
- Age 30 - 44 years (vrs 45 - 79)	0.24	0.05	1.27	0.00
- Age 80+ years (vrs 45 - 79)	-0.42	0.06	0.66	0.00
Female (vrs male)	0.43	0.05	1.53	0.00
Maternal language:				
Northern Sámi (vrs Finnish)	-2.03	0.19	0.13	0.00
Swedish (vrs Finnish)	1.61	0.11	4.98	0.00
Baltic languages or Russian (vrs Finnish)	-1.51	0.19	0.22	0.00
Other languages (vrs Finnish)	-0.21	0.19	0.81	0.25
Further education (vrs basic education only)	0.59	0.06	1.80	0.00
Income quintiles of register derived disposable income:				
- 1st quintile or no income (vrs 2nd - 4th quintile)	-0.12	0.07	0.89	0.08
- 5th quintile (vrs 2nd - 4th quintile)	-0.15	0.06	0.86	0.01
Pensioner (vrs other socio-economic status)	0.35	0.08	1.42	0.00
Experience of unemployment:				
- 1 - 25 weeks in 2000 (vrs no unemployment spells)	-0.29	0.07	0.75	0.00
- 26 - 52 weeks in 2000 (vrs no unemployment spells)	-0.25	0.07	0.78	0.00
Type of living: shelter homes, over sized DU's or unknown	3.52	0.23	33.63	0.00
Household (HH) information:				
Received income support (vrs none)	0.59	0.10	1.80	0.00
Received capital income (vrs none)	0.37	0.06	1.45	0.00
Savings to additional private pension scheme (vrs none)	0.38	0.15	1.46	0.01
Household experienced income poverty in 2000	0.79	0.11	2.20	0.00
Family life cycle:				
- Adult-child in parental home (vrs 'Other type')	0.31	0.06	1.36	0.00
- Single person HH (vrs 'Other type')	-1.04	0.11	0.36	0.00
- Couple with no children (vrs 'Other type')	0.51	0.06	1.66	0.00
- Living in HH with children (vrs 'Other type')	0.32	0.07	1.37	0.00
# of adults 3+ in HH/DU	0.67	0.09	1.96	0.00
Dwelling information:				
Type of housing:				
Detached, semi or terraced house (vrs flats)	0.47	0.05	1.60	0.00
Type of housing unknown (vrs flats)	-0.85	0.08	0.43	0.00
Number of rooms in dwelling:				
1-2 rooms (vrs 3 rooms)	0.45	0.06	1.57	0.00
4+ rooms (vrs 3 rooms)	0.35	0.06	1.42	0.00
Number of rooms unknown (vrs 3 rooms)	-1.17	0.13	0.31	0.00
Local area information:				
Crime rate [1]	0.03	0.01	1.03	0.00
Mortality of 30 - 64 year olds [1]	-1.49	0.30	0.23	0.00
Voting ratio of men over women in EU elections [1]	5.25	0.83	190.26	0.00

[1] Measured as a deviation from the national estimate per 100 inhabitants

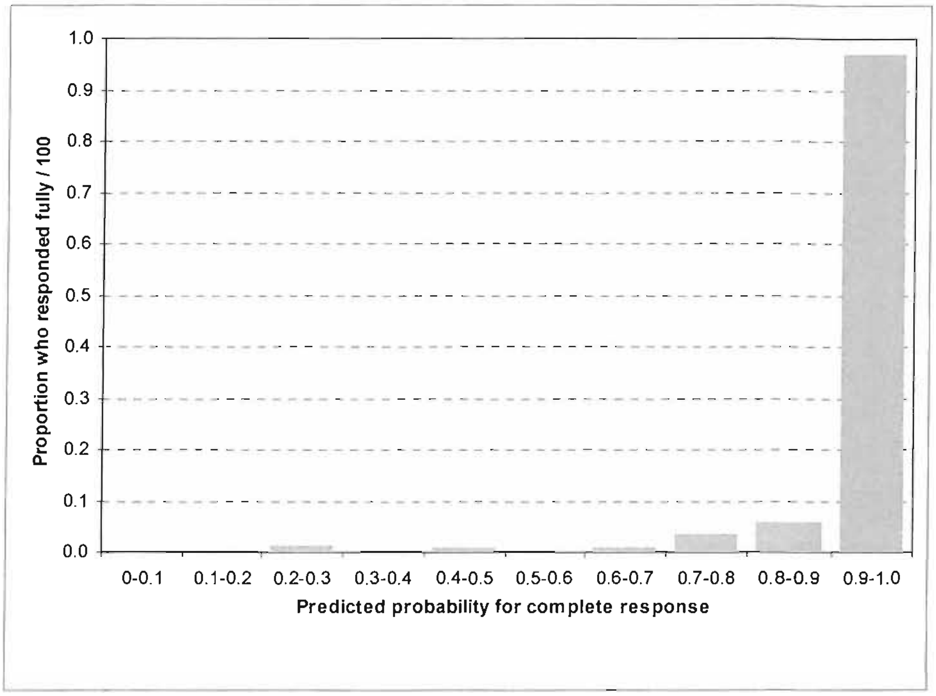
Abbreviations:

DU - Dwelling unit; HH - Household; EU - European Union

The SAS LOGISTIC procedure was used for modelling

The aggregated data collection phases, used for constructing the person-event data, are illustrated in Table 3.3 from which we have used the Scheme II.

Figure 4.3 Distribution of predicted probabilities for full survey participation by discrete-time hazard model



4.4. Assessment of cumulative logit models for modelling participation to multiphase surveys

The simple logit model and the discrete-time hazard model showed that the survey participation behaviour is dependent on the background of individual, household, dwelling and ecological population related to the demographic and socio-economic structure as well as participation to the society and situational circumstances. The assumptions of the simple logit model are however violated as discrete-time hazard model has shown that the intercepts for different data collection phases differ significantly from each other. However, the discrete-time hazard modelling is based on the event history data which reduces the usability of the model for predictions. Thus we examine further, whether we can bring the estimated model closer to the reality, when we wish to estimate the response behaviour prior to the survey. Cumulative logit model may provide an improvement. Similarly to discrete-time hazard model, it models the response probabilities for event data. But unlike in the discrete-time hazard model, the survey outcome does not impact the construction of the analysis data set. Thus the cumulative logit model may give a further advantage towards more flexible survey non-response modelling in comparison to previous models.

The modelling approach of cumulative logit model defines ordered probability models for individual-level data. The data collection phases are treated as ordinal responses in the model building. Thus the ordinality and structure of the outcome variable is related partly to the pre-defined data collection procedure, and partly to the level of successful participation. Depending on the model definition the cumulative probabilities are used for estimating the success or the failure at subsequent data collection phase. As in applying the event history analysis, the data collection phases are seen as a sequence of events arising from subsequent survey requests beginning from the initial contact attempt leading subsequently to the last survey component. Alternatively, when modelling the survey participation, we could consider a reverse order of events. We would then begin by modelling those, who completed the survey fully, relaxing gradually the definitions for participation. Thus, we would take into account those partially accepted respondents, who have given adequate response at previous data collection phases.

The cumulative logit models are used for assessing the probability of a group of ordered outcome variables. The cumulative probabilities for the ith sampled individual co-operated at the most up to the tth data collection phase are defined as a function of background variables $\mathbf{x}_i$ (Agresti 1990):

$$F_t(\mathbf{x}_i) = P(T_i \leq t) = 1 - P\left(T_i > t\right), \tag{4.13}$$

and thus

$$F_{t-1}(\mathbf{x}_i) = P(T_i \leq t-1) = 1 - P\left(T_i > t-1\right) = 1 - P\left(T_i \geq t\right) = 1 - S\left(t\right). \tag{4.14}$$

Using the survival function, we can write the probability of cumulative survey participation as follows:

$$\pi_{i[1,t]} = \frac{1 - S\left(t+1\right)}{S\left(t+1\right)}. \tag{4.15}$$

Subsequently, the estimated cumulative probability is of the form

$$\pi_{i[1,t]} = \frac{P\left(T_i \leq t \middle| \mathbf{x}_i\right)}{P\left(T_i > t \middle| \mathbf{x}_i\right)} = \frac{e^{\alpha_t + \beta_{t1}x_1 + \ldots + \beta_{tp}x_p}}{1 + \sum_{h=1}^{T-1} e^{\alpha_h + \beta_{h1}x_1 + \ldots + \beta_{hp}x_p}}, \tag{4.16}$$

where the values $1,\ldots,h,\ldots,T,\ldots,t$, correspond to ordered responses i.e. to the data collection phases indicating the level of co-operation achieved with the individual i. $P(T_i \leq t)$ is the probability that the individual i ceases the co-operation at T_i prior t (Powers and Xie, 2000), and the cumulative logit for $T = t$ is unity (Agresti, 1990). In connection to survey participation modelling, this can be interpreted as the probability of loosing the co-operation of the individual i during the first t phases relative to the probability of maintaining the co-operation at later phases. We use the last fifth phase of full co-operation as the reference group. The cumulative logit model is then defined as (Agresti, 1990):

$$L_t = \text{logit}\left[F_t(\mathbf{x})\right] = \ln\left[\frac{F_t(\mathbf{x})}{1 - F_t(\mathbf{x})}\right] = \ln\left[\frac{P(T_i \leq t | \mathbf{x}_i)}{P(T_i > t | \mathbf{x}_i)}\right] = \ln\left[\frac{\sum_{h=1}^{t} \pi_{ih}(x)}{\sum_{h=t+1}^{t_{\max}} \pi_{ih}(x)}\right], \tag{4.17}$$

and

$$P\left(T_i \le t \middle| \mathbf{x}_i\right) = \begin{cases} F\left(\alpha_t + \boldsymbol{\beta}'\mathbf{x}\right) & t = 1 \\ F\left(\alpha_t + \boldsymbol{\beta}'\mathbf{x}\right) - F\left(\alpha_{t-1} + \boldsymbol{\beta}'\mathbf{x}\right) & 1 < t \le T - 1. \\ 1 - F\left(\alpha_{T-1} + \boldsymbol{\beta}'\mathbf{x}\right) & t = T \end{cases} \tag{4.18}$$

The estimated probabilities are constrained so that $\sum_{h=1}^{t_{\max}} \pi_{ih} = 1$ and therefore the $\pi_{it_{\max}} = 1$ for all *i*, which means that the cumulative logit model allows that $(J-1)$ of the α_j parameters can act as a separate intercepts that correspond to the ordered categories of the dependent variable. The model uses all the information available on individuals and all ordinal outcomes (Berridge, 1992).

Model interpretation of the estimated cumulative logit model

The estimated cumulative logit model is presented in Table 4.4. Findings are confounding with the previously presented simple logit model (Table 3.4). People aged 80 or over, as well as males, are less actively participating than other. People with foreign maternal language, other than minority groups originating from neighbouring countries, have very high survey attrition. Socioeconomic factors indicating deprivation or exclusion decrease the survey participation, and factors indicating beneficial circumstances, increase the propensity to co-operate. Of the ecological population, proportion of self-employed people in the area is increasing to some extent the likelihood of survey participation. Crime rate and children in part-time child dare are reducing the survey co-operation, similarly as in the simple logit model in Table 3.4.

As the cumulative logit model estimates the level of co-operation for each individual without taking into account the realised survey outcomes, it can be used for predicting survey participation prior the fieldwork. The distributions of the predicted response probabilities are shown in Figure 4.4. In comparison to the discrete-time hazard model, the predicted distributions are more spread, indicating a better fit of the model. We can observe that the proportion of respondents within each decile of predicted response probabilities increases almost linearly. The cumulative logit model not only provides realistic tools for survey weighting, but it could also be used to identify risk groups in future surveys and target specific fieldwork operations for those estimated with low response probability.

In Table 4.5, we present the Score-test statistics for testing the proportionality assumption in the cumulative logit model. The proportionality assumption refers to the effects of $\boldsymbol{\beta}$ that will be the same across different data collection phases. The model definition for estimating the response propensity implies that the model has the same effects $\boldsymbol{\beta}$ for each logit, but the intercepts $\{\alpha_t\}$ are allowed to be decreasing with survey attrition occurring at phase *t*, since the $P\left(T_i \le t \middle| \mathbf{x}_i\right)$ decreases in *t* for fixed $\mathbf{x}$, and the logit is a decreasing function of this probability (Agresti, 2002). The same restriction was given previously in the discrete-time hazard modelling. An even stronger restriction was made in simple logit model, when assuming that also the a single intercept could be used for modelling as it did not make any distinction by data collection phases at which non-response could have occurred. Table 4.5 reveals that the Score test for the proportional odds assumption rejects the null hypothesis of proportionality for the complete model and all other the single variable cumulative logit models, except for the savings to private pension and rehabilitation support. To demonstrate the violation of the proportionality assumption, we will compare graphically the predicted probabilities to co-operate with survey by the age of the individual sampled later, when examining the multinomial model in forthcoming Figure 4.6.

The analysis of the Health 2000 survey participation data suggest that the odds ratios of the covariates and the significance of the covariates vary significantly across the data collection phases. Thus we conclude that the proportional odds model is not a sufficient summary of the odds ratios and the survey participation behaviour is affected by differential associations by background characteristics. This causes the violations of the proportionality assumption. Thus generalized logistic regression should be used instead to study in detail the impact of covariates across data collection phases. As the Score-test rejected the proportionality assumption for most variables, we will study in the following models, which allow the explanatory variables to have variable coefficients across data collection phases. We will next assess the use of multinomial logit models followed by sequential logit models.

Figure 4.4 Proportion of individuals sampled who co-operated fully by their predicted probability estimated by the cumulative logit model

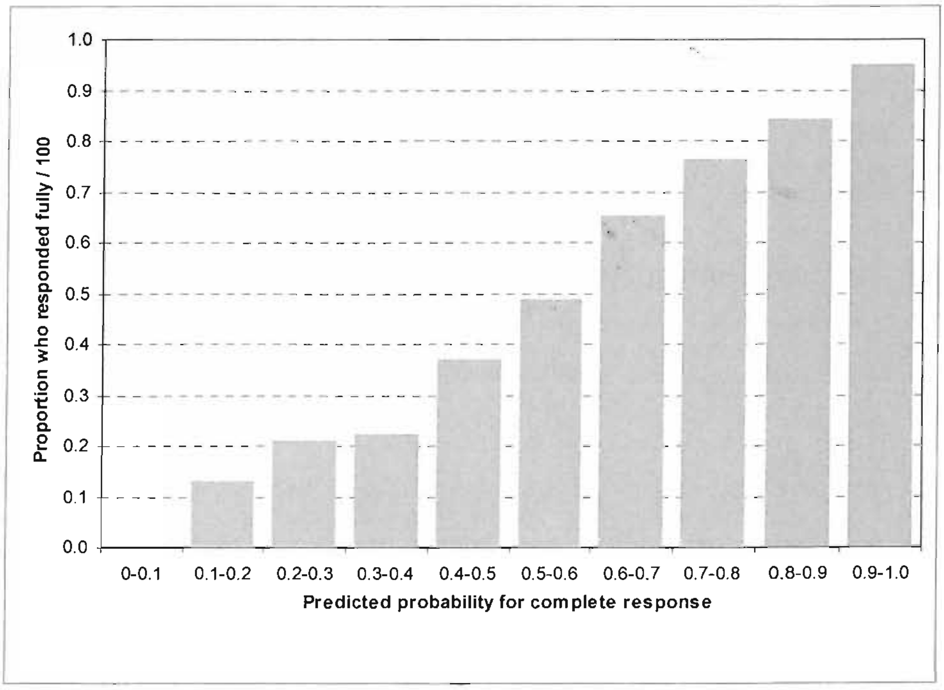


Table 4.4 Parameter estimates for cumulative logit model estimates for survey attrition

Explanatory variables	Estimate	Standard Error	Odds Ratio	Pr > t-value
Intercept:				
Phase I - Contacting	3.02	0.16	20.49	0.00
Phase II - Health Interview	0.84	0.14	2.31	0.00
Phase III - Symptom interview	0.30	0.14	1.35	0.03
Phase IV - Health examination	-0.09	0.14	0.91	0.49
Phase V - Self-completion questionnaires	-0.52	0.14	0.59	0.00
Characteristics of the individual:				
Age group:				
- Age 30 - 44 years (vrs 45 - 79)	0.01	0.05	1.01	0.78
- Age 80+ years (vrs 45 - 79)	-0.30	0.06	0.74	0.00
Female (vrs male)	0.27	0.05	1.31	0.00
Maternal language:				
- Northern Sámi (vrs Finnish)	0.00	0.29	1.00	0.99
- Swedish (vrs Finnish)	0.23	0.13	1.26	0.09
- Baltic languages or Russian (vrs Finnish)	0.12	0.27	1.12	0.67
- Other languages (vrs Finnish)	-0.95	0.22	0.39	0.00
Further education (vrs basic education only)	0.41	0.06	1.50	0.00
Income quintiles of register derived disposable income:				
- 1st quintile or no income (vrs 2nd - 4th quintile)	-0.18	0.05	0.84	0.00
- 5th quintile (vrs 2nd - 4th quintile)	0.13	0.05	1.13	0.01
Received rehabilitation support (disabled)	0.30	0.14	1.35	0.03
Received care support (pensioners or disabled)	-0.51	0.10	0.60	0.00
Experience of unemployment:				
- 1 - 25 weeks in 2000 (vrs no unemployment spells)	0.10	0.08	1.10	0.20
- 26 - 52 weeks in 2000 (vrs no unemployment spells)	-0.20	0.07	0.82	0.01
Household (HH) information:				
Received capital income (vrs none)	0.22	0.06	1.25	0.00
Savings to additional private pension scheme (vrs none)	0.47	0.13	1.60	0.00
Family life cycle:				
- Adult-child in parental home (vrs 'Other type')	-0.02	0.06	0.98	0.68
- Single person HH (vrs 'Other type')	-0.66	0.12	0.52	0.00
- Couple with no children (vrs 'Other type')	0.30	0.06	1.35	0.00
- Living in HH with children (vrs 'Other type')	0.33	0.06	1.39	0.00
Type of living: shelter homes, over sized DU's or unknown	-0.70	0.23	0.50	0.00
Dwelling information:				
Number of rooms in dwelling:				
1-2 rooms (vrs 3 rooms)	-0.11	0.07	0.90	0.10
4+ rooms (vrs 3 rooms)	0.24	0.07	1.27	0.00
Number of rooms unknown (vrs 3 rooms)	-0.21	0.16	0.81	0.18
Local area information:				
Crime rate [1]	-0.02	0.01	0.98	0.00
Children in part-time child care [1]	-0.23	0.08	0.80	0.00
Proportion of self-employed [1]	0.02	0.01	1.03	0.03

[1] Measured as a deviation from the national estimate per 100 inhabitants

Abbreviations:

DU - Dwelling unit; HH - Household

SAS PROC LOGISTIC-procedure was used for modelling with cumulative link function LINK=CLOGIT.

Aggregation of data collection phases is illustrated in Table 3.3 (Scheme II).

Table 4.5 Score test for the proportional odds assumption

Model	Chi-Square	DF	Pr > ChiSq
Full model	657.7	108	0.00
Separate models for each parameter			
Age group	270.9	8	0.00
Female	70.5	4	0.00
Maternal language	40.6	16	0.00
Further education	114.3	4	0.00
Income quintile	117.8	8	0.00
Received capital income	23.1	4	0.00
Savings to additional private pension scheme	2.6	4	0.62
Type of living: shelter homes, over sized DU's or unknown	127.0	4	0.00
Family life cycle	117.3	16	0.00
Received rehabilitation support (disabled)	1.4	4	0.84
Received care support (pensioners or disabled)	122.5	4	0.00
Experience of unemployment	16.9	2	0.00
Number of rooms in dwelling	137.8	12	0.00
Crime rate [1]	11.0	4	0.03
Children in part-time child care [1]	15.0	4	0.00
Proportion of self-employed [1]	29.8	4	0.00

4.5. Allowing the impact of covariates to vary by data collection phases by applying multinomial logit model

Exploring for a flexible model, enabling us to take into account the complexity of the survey participation behaviour, the previously assessed models have appeared to have too strong model assumptions. In our empirical data, using the Finnish Health 2000 survey, the survey non-response occurring at various data collection phases requires further relaxation of the model assumptions than the cumulative logit can provide for. Firstly, the level of non-response varies significantly across the phases. Secondly, the covariates explaining survey non-response do not have the same impact across all phases. To add more flexibility to the survey non-response modelling, we will allow the impact of explanatory variables to vary across data collection phases, which is the main advantage of the model. In addition, we will observe and assess whether the relationship of the individual variable and the survey participation process is changing over the data collection phases. The multinomial logit resembles the discrete-time hazard and the cumulative logit models presented previously as the set of explanatory variables is restricted to be the same for all modelled phases and that the modelling is done using single-model approach.

The conditionality of the outcome and the longitudinal characteristic of the survey outcome have been predominant characteristic in previous models but are disregarded in the multinomial logit approach, which typically treats the outcome of survey steps as a uni-dimensional polytomous variable. The model allows the response variable to define the outcome groups to which the outcome of each data collection phase and complete non-response can be related to. This ordering of the dependent categorical variable also reflects to the modelling interests of the Health 2000 data. The model allows testing, whether there is a significant difference between data collection phases or whether the response behaviour is similar, and should be captured into one general model. The multinomial logit model estimates the probability of the values of the polytomous dependent variable T_i to occur while T_i indicates the outcome for individual sampled i as follows:

$$P\left(T_i = t \middle| \mathbf{x}_i\right) = P\left(T_i \le t \middle| \mathbf{x}_i\right) - P\left(T_i \le t-1 \middle| \mathbf{x}_i\right) = \frac{\exp(\beta_t x_{it})}{1+\sum_{h \ne t} \exp(\beta_h x_{ih})}. \tag{4.19}$$

The highest category of data collection phases i.e. the full co-operation is used for the baseline probability in order to contrast the impact of the explanatory variables on the probability of failing to response at certain phases versus the success of responding fully. The probability of full response is defined as

$$P\left(T_i = t_{\max}\right) = \frac{1}{1+\sum_{h \ne t} \exp(\beta_h x_{ih})}. \tag{4.20}$$

The last ordered outcome category for the complete acceptable response at all data collection phases is used as a baseline category. Multinomial models can be used for both modelling the survey participation and alternatively on the non-response processes, depending how the dependent variable and the purpose of the analyses have been defined. As before, the modelling allows the explanatory variables x to be either continuous or categorical.

Model interpretation

The model estimates are presented in Table 4.6 for estimating the likelihood of survey attrition across data collection phases. The relatively good fit of the model is shown in Figure 4.5, with the linearity and good spread of the predicted probabilities for the respondents. Although the model has been constructed independently and the significance of all available auxiliary variables has been tested separately, the set of explanatory variables and their estimates are in line with the previous models. The multinomial modelling allows for comparing the impact of fixed set of explanatory variables at all data collection phases – as far as the effects are significant their effects are confounding with other models. We find that both the significance and coefficients of the explanatory variables vary across phases. We can also observe that when there are dramatic changes in the coefficient, there is also a change in the significance of the variable. Assessing both the odds ratio and the significance of the variables, we observe the impact of the explanatory variables confounding across the phases when the variable has a significant impact to the model. From Figure 4.6 we observe that when allowing the coefficients to vary, also the estimated probabilities to co-operate may vary between the phases. The differences can be increasingly large for the elderly people. We also find that the proportionality assumption does not hold, when the model does not force the estimates, as in the cumulative logit model.

The multinomial logit model provides an aggregated, to some degree a simplified, single-model approach for complex participation process modelling. It allows coefficients and the significance of the variables to vary – even drastically – across the data collection phases. On the other hand, this uniform model restricts the set of covariates to be fixed. This can be seen as a weakness, especially for those variables that are potentially significant only at few data collection phases.

Figure 4.5 Proportion of individuals sampled who responded at all data collections phases by their predicted probability to co-operate fully by the multinomial logit model

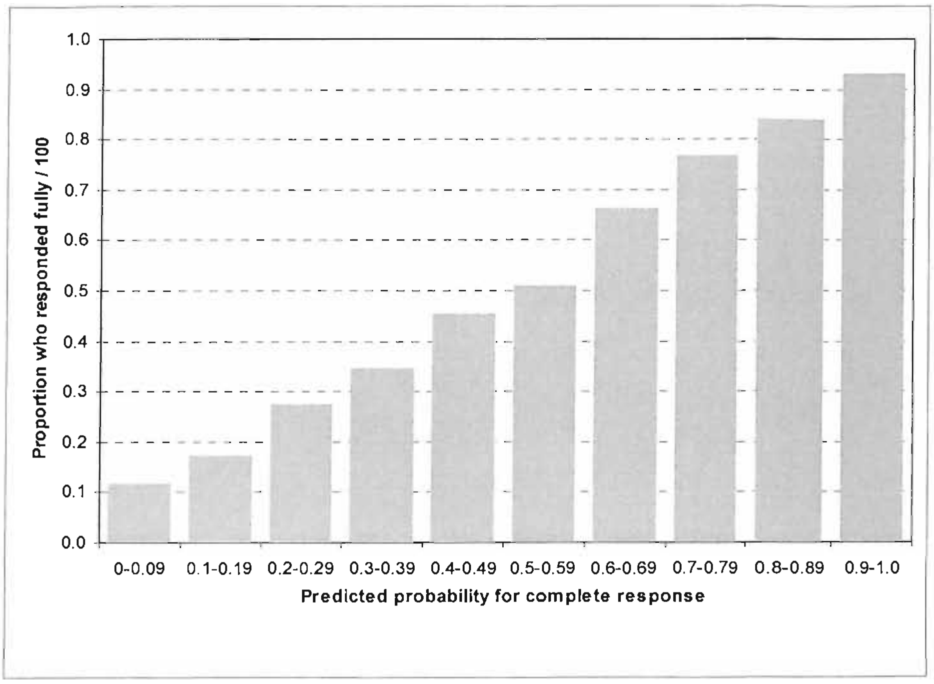


Table 4.6 Parameter estimates for multinomial logit model for survey attrition

	Estimate and (Standard Error)				
Explanatory variables	Phase I	Phase II	Phase III	Phase IV	Phase V
Intercept	-3.17	-1.15	-1.30	-2.44	-1.39
	0.44	0.19	0.20	0.29	0.21
Characteristics of the individual:					
Age group:					
- Age 30 - 44 years (vrs 45 - 79)	0.72	0.08	-0.12	-0.03	-0.13
	0.26	0.09	0.11	0.13	0.10
- Age 80+ years (vrs 45 - 79)	-0.86	0.23	0.54	0.61	0.27
	0.43	0.11	0.11	0.11	0.11
Female (vrs male)	-1.07	-0.17	-0.27	0.04	-0.42
	0.23	0.08	0.11	0.12	0.09
Maternal language:					
Swedish (versus Finnish or Northern Sámi)	-0.68	-0.01	-0.49	0.12	-0.09
	0.38	0.15	0.18	0.25	0.17
Other language (versus Finnish or Northern Sámi)	1.33	0.35	1.08	0.08	0.47
	0.34	0.22	0.22	0.42	0.25
Further education (vrs basic education only)	-0.32	-0.31	-0.51	-0.44	-0.46
	0.22	0.09	0.12	0.13	0.11
Income quintiles of register derived disposable income:					
- 1st quintile or no income (vrs 2nd - 4th quintile)	0.34	0.07	0.16	0.40	0.20
	0.18	0.08	0.09	0.10	0.08
- 5th quintile (vrs 2nd - 4th quintile)	-0.10	-0.05	-0.02	-0.22	-0.04
	0.21	0.08	0.11	0.13	0.09
Received care support (pensioners or disabled)	-0.70	0.24	0.81	1.13	0.29
	0.77	0.19	0.17	0.16	0.19
Pensioner (versus other social class)	-0.07	0.06	0.41	1.01	0.74
	0.32	0.12	0.16	0.18	0.13
Household information:					
Received capital income (vrs none)	-0.66	-0.29	-0.24	-0.15	-0.12
	0.28	0.09	0.12	0.13	0.10
Received income support (vrs none)	0.92	0.09	0.75	0.57	0.54
	0.26	0.15	0.16	0.20	0.16
Savings to additional private pension scheme (vrs none)	-0.22	-0.59	-0.37	0.14	-0.58
	0.54	0.21	0.28	0.29	0.25
Family life cycle:					
- Adult-child in parental home (vrs 'Other type')	0.88	0.82	0.65	0.32	0.17
	0.41	0.17	0.24	0.34	0.27
- Single person HH (vrs 'Other type')	0.59	-0.01	0.09	0.13	-0.11
	0.22	0.10	0.12	0.14	0.12
- Couple with no children (vrs 'Other type')	-0.76	-0.12	-0.36	-0.18	-0.32
	0.29	0.09	0.13	0.15	0.12
- Living in HH with children (vrs 'Other type')	-0.72	-0.36	-0.08	-0.24	0.19
	0.26	0.10	0.13	0.17	0.12
Household with more than 2 adult members	0.60	0.41	0.40	-0.03	0.20
	0.35	0.13	0.17	0.21	0.15
Type of living: shelter homes, over sized DU's or unknown	2.79	1.06	1.61	1.18	-0.27
	0.80	0.46	0.39	0.40	0.50
Dwelling information:					
Number of rooms in dwelling:					
1-2 rooms (vrs 3 rooms)	0.21	0.21	-0.02	-0.05	0.08
	0.26	0.12	0.12	0.13	0.12
4+ rooms (vrs 3 rooms)	-0.49	-0.19	-0.42	-0.13	-0.38
	0.29	0.12	0.12	0.13	0.12
Number of rooms unknown (vrs 3 rooms)	0.13	-0.04	0.62	0.56	0.51
	0.59	0.30	0.26	0.28	0.29
Local area information:					
Crime rate [1]	0.10	0.03	0.01	0.01	0.03
	0.03	0.01	0.01	0.01	0.01
Mortality of 30 - 64 year olds [1]	-1.85	-1.76	-0.26	-1.35	-0.69
	1.38	0.50	0.57	0.62	0.52
Children in part-time child care [1]	0.32	0.32	0.02	0.04	0.45
	0.34	0.13	0.16	0.17	0.15
Proportion of self-employed [1]	-0.14	-0.04	-0.02	-0.03	0.02
	0.06	0.02	0.02	0.02	0.02

[1] Measured as a deviation from the national estimate per 100 inhabitants
Abbreviations: DU - Dwelling unit; HH - Household

Table 4.6 Continues from previous page

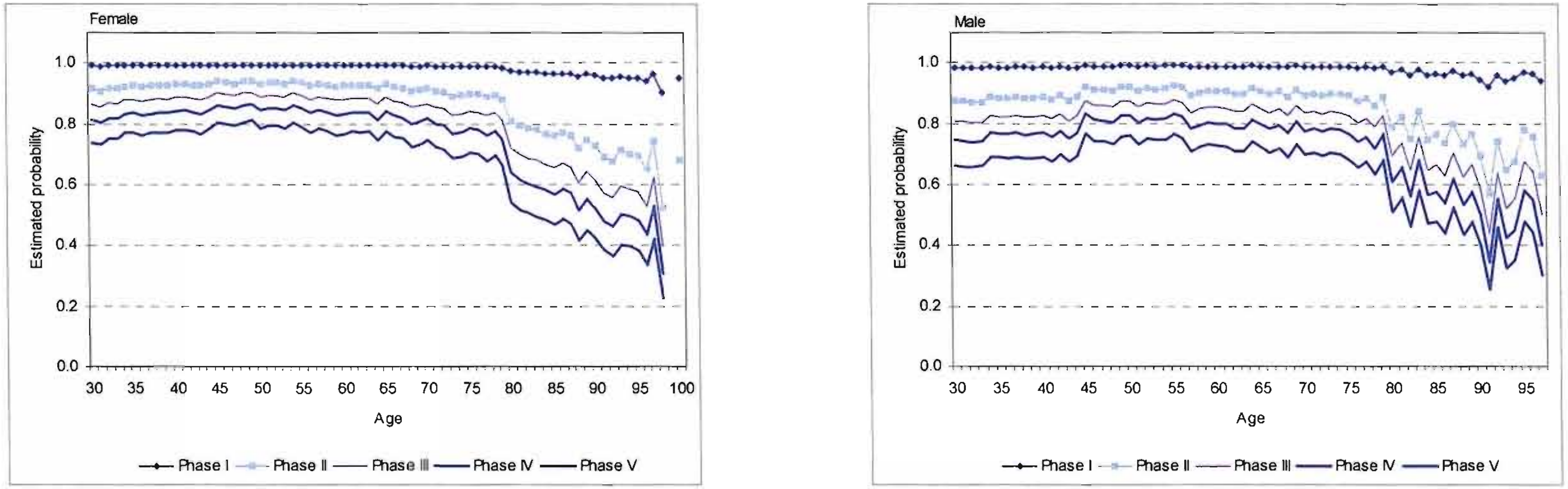
	Odds ratio and (Pr > Chi Square Test)				
Explanatory variables	Phase I	Phase II	Phase III	Phase IV	Phase V
Intercept	0.04	0.32	0.27	0.09	0.25
	0.00	0.00	0.00	0.00	0.00
Characteristics of the individual:					
Age group:					
- Age 30 - 44 years (vrs 45 - 79)	2.06	1.08	0.89	0.97	0.88
	0.01	0.36	0.28	0.79	0.20
- Age 80+ years (vrs 45 - 79)	0.43	1.26	1.71	1.84	1.31
	0.05	0.03	0.00	0.00	0.02
Female (vrs male)	0.34	0.85	0.77	1.04	0.66
	0.00	0.04	0.01	0.72	0.00
Maternal language:					
Swedish (versus Finnish or Northern Sámi)	0.51	1.00	0.61	1.13	0.92
	0.08	0.97	0.01	0.62	0.61
Other language (versus Finnish or Northern Sámi)	3.78	1.41	2.95	1.08	1.60
	0.00	0.12	0.00	0.85	0.06
Further education (vrs basic education only)	0.73	0.74	0.60	0.65	0.63
	0.16	0.00	0.00	0.00	0.00
Income quintiles of register derived disposable income:					
- 1st quintile or no income (vrs 2nd - 4th quintile)	1.41	1.08	1.17	1.49	1.22
	0.05	0.36	0.09	0.00	0.02
- 5th quintile (vrs 2nd - 4th quintile)	0.91	0.95	0.98	0.81	0.96
	0.63	0.53	0.87	0.10	0.66
Received care support (pensioners or disabled)	0.50	1.27	2.26	3.10	1.34
	0.36	0.21	0.00	0.00	0.11
Pensioner (versus other social class)	0.93	1.07	1.50	2.75	2.09
	0.83	0.60	0.01	0.00	0.00
Household information:					
Received capital income (vrs none)	0.52	0.75	0.79	0.86	0.89
	0.02	0.00	0.05	0.27	0.24
Received income support (vrs none)	2.50	1.10	2.12	1.77	1.71
	0.00	0.55	0.00	0.00	0.00
Savings to additional private pension scheme (vrs none)	0.80	0.56	0.69	1.15	0.56
	0.68	0.00	0.19	0.63	0.02
Family life cycle:					
- Adult-child in parental home (vrs 'Other type')	2.41	2.28	1.91	1.37	1.19
	0.03	0.00	0.01	0.35	0.52
- Single person HH (vrs 'Other type')	1.80	0.99	1.09	1.14	0.90
	0.01	0.95	0.48	0.35	0.36
- Couple with no children (vrs 'Other type')	0.47	0.88	0.70	0.84	0.73
	0.01	0.18	0.00	0.22	0.01
- Living in HH with children (vrs 'Other type')	0.49	0.70	0.92	0.78	1.21
	0.01	0.00	0.52	0.15	0.12
Household with more than 2 adult members	1.83	1.51	1.49	0.97	1.22
	0.09	0.00	0.02	0.88	0.21
Type of living: shelter homes, over sized DU's or unknown	16.27	2.88	5.03	3.24	0.77
	0.00	0.02	0.00	0.00	0.60
Dwelling information:					
Number of rooms in dwelling:					
1-2 rooms (vrs 3 rooms)	1.24	1.23	0.98	0.95	1.08
	0.41	0.08	0.87	0.71	0.51
4+ rooms (vrs 3 rooms)	0.61	0.83	0.66	0.88	0.68
	0.09	0.11	0.00	0.33	0.00
Number of rooms unknown (vrs 3 rooms)	1.14	0.96	1.86	1.74	1.66
	0.83	0.90	0.02	0.05	0.08
Local area information:					
Crime rate [1]	1.10	1.03	1.01	1.01	1.04
	0.00	0.02	0.62	0.59	0.00
Mortality of 30 - 64 year olds [1]	0.16	0.17	0.78	0.26	0.50
	0.18	0.00	0.65	0.03	0.18
Children in part-time child care [1]	1.38	1.38	1.02	1.04	1.57
	0.35	0.01	0.92	0.83	0.00
Proportion of self-employed [1]	0.87	0.96	0.98	0.97	1.02
	0.02	0.03	0.42	0.16	0.33

[1] Measured as a deviation from the national estimate per 100 inhabitants

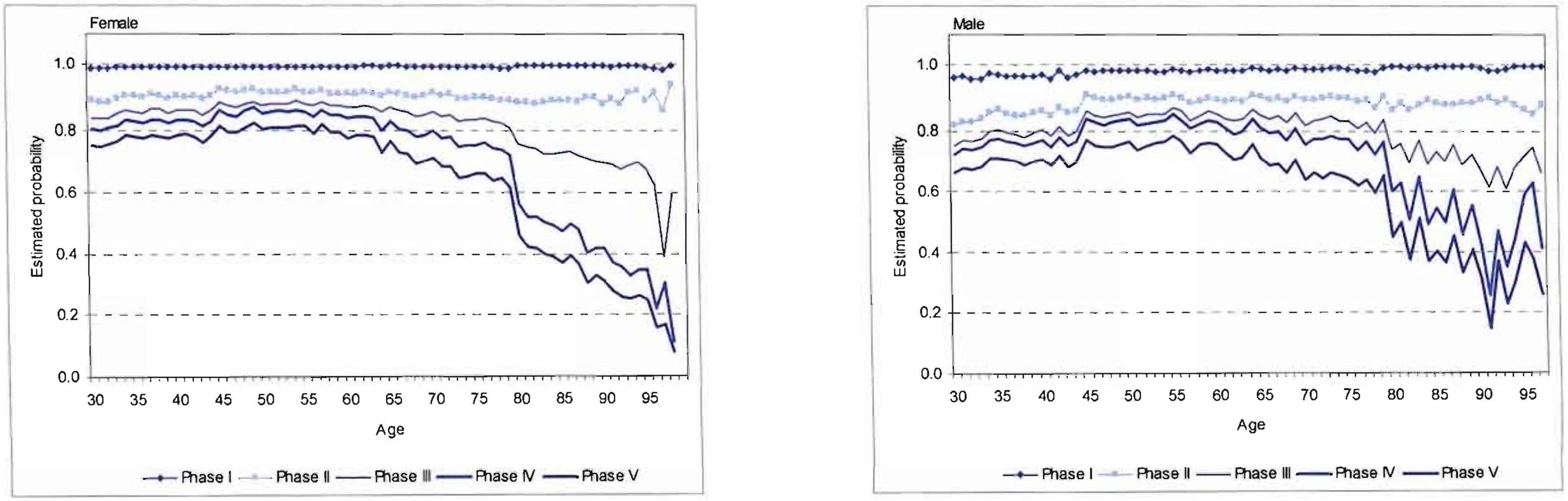
Abbreviations: DU - Dwelling unit; HH - Household

Figure 4.6 **Comparison of estimated propensity to co-operate by age, sex and data collection phase in multinomial and cumulative logit models**

(a) Cumulative logit model estimates



(b) Multinomial logit model estimates



4.6. Relaxing the covariate structure to vary across data collection phases in sequential logistic regression models

Previously presented models have relaxed to some level their model assumptions, but assumed that a survey participation could be modelled using a single-model approach. One can ambiguously aim to model the complex phenomena of survey non-response using solely one single model, this may not however provide an optimal solution. Although the analysis can be informative for comparing the effects of the same covariates across data collection phases, the models do not exploit fully the cumulating information structure from the data collection. If the non-response processes are significantly different at various data collection phases, a single model approach can be too general. In the analyses of our empirical data, we observe that the various data collection phases differ in terms of the data collection methods, instruments and also by the response burden. Therefore the probability of the survey non-response occurring is analysed separately for each phase, but conditioning the analysis only for the eligible sub-sample at each point. The advantage of the sequential logit models comes forward in providing the tools for assessing the significance of various variables at each conditional phase separately.

Sequential logit models have been developed initially by Mare (1980), who modelled school continuation decisions at six schooling levels. Using sequential logistic regression models Mare (1980) described the polytomous responses by a sequence of binary models. Persons who had not made the transition to the next schooling level from the previous level were excluded from the analysis. Similarly, the survey participation process can also be seen as a sequence of successive survey events at any of which the first occurrence of non-response may occur. The models of this process can be represented as a sequence of models for co-operation at sequential survey steps, each of which is conditional upon the success of participation in the previous one. These models should take into account the specific features of the survey design, data collection mode, respondent selection and the fieldwork efforts.

In sequential approach, the models of later phases may incorporate those survey variables that are not available at the earlier phases. For example, a significant amount of health data is obtained in the health interview. These variables can then be used in the modelling at later phases. Such partial information could not be utilised in uni-dimensional model of non-response. This is seen as one clear advantage of the sequential approach. The usability of the survey information has the potential to increase the explanatory power of the models and, therefore, of any non-response weighting procedures based upon them (Laiho and Lynn, 1999). Also Lepkowski and Couper (2002) studied non-response in a panel survey using similar sequential approach. They modelled survey participation of the second wave based on the success of participating at the previous wave conditional on being contacted.

Health surveys and some other social surveys such as time use surveys contain multiple data collection phases. Therefore, the phenomena of the survey participation may become very complex. For simplicity, the survey participation could be studied based only on few key survey variables. However, this might not give a complete picture of the phenomena. Therefore, a sequential participation model has been postulated. In this approach, the survey participation has been divided into sequential logit models indicating the success at each phase conditional on the success at the previous phase (Figure 4.7):

Phase 1: Contacting target person

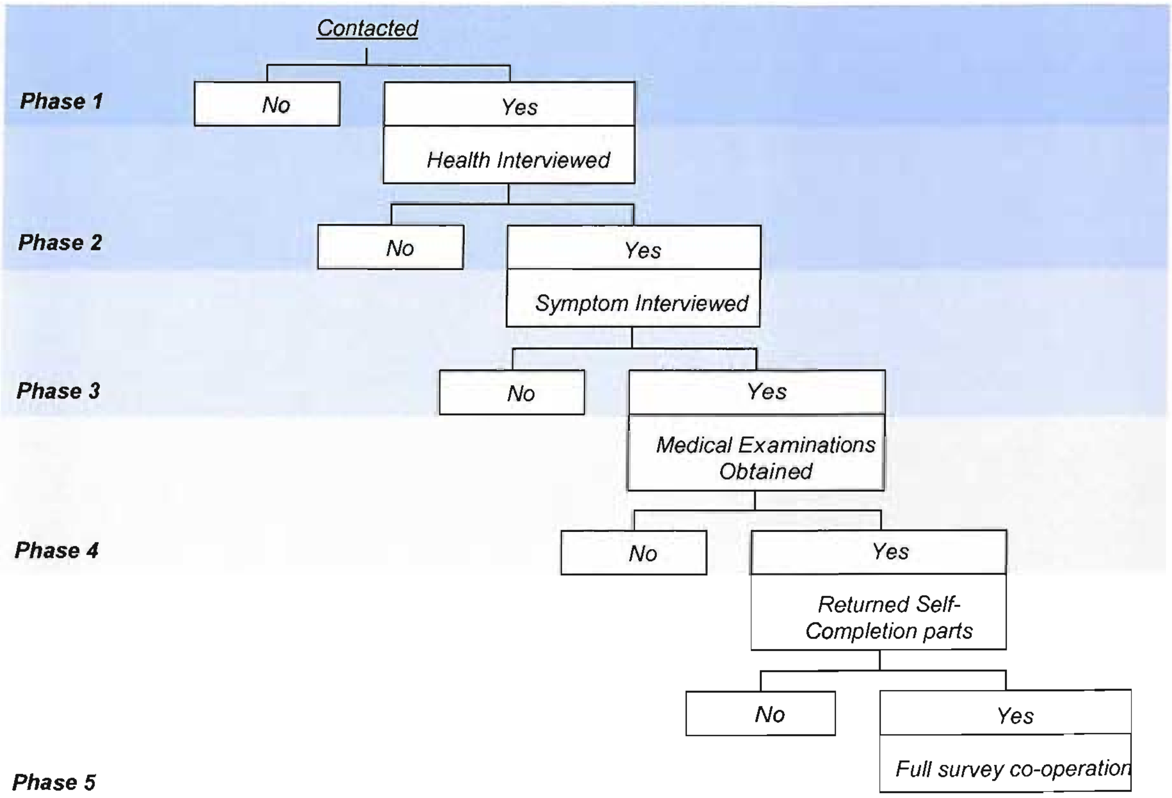
Phase 2: Health Interview

Phase 3: Symptom Interview

Phase 4: Medical measurements

Phase 5: Self-Completion 1, 2 and 3 as well as nutrition self-completion questionnaires

Figure 4.7 The path of sequential survey participation in the Health 2000 survey



Let S_k denote for the set of sample cases successfully passing phase t of the survey participation process ($t = 1,\ldots,5$). After the initial phase for modelling the probability of contact, the outcome at one phase is always conditional on successful outcome on the previous phase. As one proceeds successfully from one data collection phase to another also the information matrix about the sampling units increases. Due to simplicity it is assumed that the response pattern is monotone and response will be ignored at any later phases after the first occurrence of survey non-response as the cases will be excluded from the following models. For the probability of contact we use the simple logit model defined earlier and π_{it}, denote this probability of contact when $t = 1$. The parameter of interest for any following phase is the conditional on success the previous phase denoted by $\pi_{it|t-1}$ and the conditional probability of the failure is consequently $\left(1-\pi_{it|t-1}\right)$ when $t = 2,\ldots,t_{\max}$. The conditional probability for individual *i* with background characteristics $\mathbf{x}_i$ to be contacted is estimated by

$$\pi_{i1} = P\left(T_i > 1 \middle| \mathbf{x}_i\right) = P(R_{i1} = 1 | \mathbf{x}_i)$$

and more generally the probability of co-operation at any later phase *t* is defined with the binary response indicator R_{it} for the *i*th individual at the *t* data collection phase as follows:

$$\pi_{it|t\text{-}1} = P\left(T_i > t \middle| T \geq t, \mathbf{x}_i\right) = P(R_{it} = 1 | R_{i(t\text{-}1)} = 1, \mathbf{x}_i) = P(R_{it} = 1 \left| \prod_{h=1}^{t-1} R_{ih} = 1, \mathbf{x}_i \right.). \tag{4.21}$$

These estimates are the exact opposite of the probability modelled in the discrete-time hazard. Thus we can write

$$\pi_{it|t\text{-}1} = 1 - \lambda_{it} \tag{4.22}$$

where λ_{it} has been defined in (4.11). Subsequently, we can make the comparison between the discrete-time hazard model defined earlier in (4.12) and the sequential logit model emphasising the close connection between these two types of models:

$$\operatorname{logit}\left(\pi_{it|t-1}\right) = -\operatorname{logit}\left(\lambda_{it}\right) = \mathbf{x}_{it}\boldsymbol{\beta}_t\,. \tag{4.23}$$

Actually, the sequential logit model could also be constructed in terms of modelling the conditional failure after subsequent series of successes at previous phases according to the definition of estimated probability in the discrete-time hazard model:

$$\lambda_{it} = P(R_{it} = 0 \left| \prod_{h=1}^{t-1} R_{ih} = 1, \mathbf{x}_i \right.).$$

As stated earlier the differences of the sequential logit and the other models are in the modelling of the probabilities. When using the sequential logit modelling the set of covariates $\mathbf{x}_i$ can vary from identical sets to completely different set across *t*. The sequential logit regression enable the non-response analyses to exploit both the auxiliary data as well as the cumulatively additive survey data is obtained during the fieldwork, prior the phase for which the response models are been estimated.

The survey outcomes are defined so that they are mutually exclusive and exhaustive, in the sense that they cover a full range of possible outcomes. For example, the probability of response to the health survey is conditional on the success at the contacting phase:

$$\pi_{i2} = \pi_{i1}\pi_{i2|1} = P\left(R_{i2} = 1 \middle| R_{i1} = 1\right),$$

where $i = 1,\dots,n_1$ refers to the contacted individuals that are subject to the request to participate in the health interview. The probability of participating to the symptom interview is conditional on the probability of participation to the health interview at the previous phase:

$$\pi_{i3} = \pi_{i2}\pi_{i3|1} = \pi_{i1}\pi_{i2|1}\pi_{i3|2}$$

and thus we can write the conditional probability as a product of subsequent probabilities

$$\pi_{it|t-1} = \pi_{i1}\prod_{h=2}^{t_{\max}}\pi_{ih|h-1} = \frac{e^{\beta_{0t}+\beta_{1t}x_{1it}+\dots+\beta_{pt}x_{pit}}}{1+e^{\beta_{0t}+\beta_{1t}x_{1it}+\dots+\beta_{pt}x_{pit}}}. \tag{4.24}$$

In our data, all x variables of the auxiliary information are time invariant. However, the vectors of explanatory variable $\mathbf{x}_t$ may differ from those at $\mathbf{x}_{t+1}$ or at any other phase $t = 1,\dots,t_{\max}$. The most powerful explanatory variables are used as predictors of the response probability. The response probability can be interpreted as the probability of maintaining co-operation successfully during and in between data collection phases. As noted in Chapter 2 (Section 2.7) the binomial response indicator R_i follows the Bernoulli distribution. Thus the maximum likelihood estimates of $\hat{\beta}$ maximise the log likelihood:

$$\sum_{i=1}^{n}\left(R_i \ln\left(\pi\left(x_i\right)\right)+\left(1-R_i\right)\ln\left(1-\pi\left(x_i\right)\right)\right).$$

The interpretation of the estimated sequential logit model on survey participation

The model diagnostic of the sequential logit models illustrate that differential factors affect the probability of success at each aggregated data collection phases, shown in Table 4.7. The sequential logit models are dissimilar from each other with varying sets of explanatory variables proving to be significant at different phases. To control for the estimated effects, we have constructed the sequential models with a fixed and variable part. All sequential models are constrained to contain variables describing the age, sex, maternal language, education, and income quintile. The fixed part allows us to contrast the findings of the variable part against the most typical theories and findings of the non-response research. The variable part has no restrictions and exploits fully the auxiliary data, although some factors appear to be significant at many phases. We find that there is a large variation of the significance of individual, household, dwelling, and local area characteristics across data collection phases.

The survey participation theories reviewed in Chapter 2 speculate with the impact of the survey topic on the survey participation behaviour. We have confounding findings to these theories, presented in Table 4.7. Some health information obtained at the health interview phase is significantly related to the survey participation at later phases when the response burden increases. Respondents with very low BMI (less than 18) are less likely to participate to the subsequent symptom interview (following the health interview) and the medical examination. In contrary, people with higher BMI values (above 25) are more likely to participate. In addition, respondents who have not visited a doctor in the last 12 months were more likely to participate to the medical examination, providing all necessary samples, than those with recent doctor visits. People with less cognitive skills were in higher risk of dropping out in the middle of the data collection, especially at cognitively demanding phases. The cognitive skills were valued by the interviewer after collecting the health interview; interviewers were asked, for example, whether the respondent had any cognitive problems in understanding questions and instructions that the interviewer had observed during the interview. Individuals with cognitive problems had higher propensity to participate to the symptom interview and lower propensity to participate to the medical examination than others. When it comes to increased response burden across data collection phases, and demanding cognitive burden in self-reporting via the self-completion questionnaires, respondents with further education appear to co-operate at higher levels than respondents with only basic education.

Predicted response propensity is also related with self-assessed working capacity at later data collection phases. People, who assessed themselves partly incapable to work, had higher response propensity to participate to the medical examination and returning the self-completion questionnaires, in comparison to those who assessed themselves fully capable to work. In contrary, those fully incapable to work have reduced propensity to return all self-completion questionnaires. The results in Table 4.7 show that the risk of non-response at later data collection phases is increased in the presence of item non-response at health interview. In addition reduced co-operation with the interviewer, such as, not showing the Social Insurance card to the interviewer, has significant prediction power on the co-operation at later phases. Especially those respondents who had not provided information on their height and weight for the BMI calculation had very low odd ratio on the co-operation at later phases. As the response burden is increasing by the data collection phases this may indicate that the item-non response and response burden are predictors of survey break-off. Thus respondents with item-nonresponse are in the risk of dropping out from the survey and maintaining their co-operation needs special attention and perhaps even additional intervention during the fieldwork period. Thus further monitoring should be developed so that non-response reduction methods could be built into the routines of the actual fieldwork period. Results of the predictive power of item non-response on survey non-response at later phases are confounding with the findings of Loosveldt, Pickery and Billiet (2002).

In addition to the highly significant health information, the co-operation can be related at all phases with covariates associated with social exclusion and poverty or with affluence. The survey participation differs by the family life-cycle, type of living and household, and socio-demographic information of the local ecological population. However, the activity in the labour market and experience of unemployment were not significant when other factors were controlled for. The sequential model indicates that the importance of demographic, household and housing conditions is higher at the initial phases, and the relative income and health status at later phases. When information from the health interview can be incorporated to the modelling, the significance of the health related variables is drastic in comparison to other information.

Confounding with previous models, significant factors of the ecological population related with demographics, are general mortality rate and mortality rate of the 30 to 64 year olds, which were positively associated with the survey participation. An increase in the crime rate and the children in the part-time day care reduce the survey co-operation. Ratio of institutionalised men over women, and ratio of men over women voting in EU elections are factors representing the situational circumstances and participation to the society. Ratio above the national average in ratio of institutionalised men over women, increases the odds of co-operation to the symptom interview by 1,4. The local activity on the participation to elections with relatively low poll has dramatically significant explanatory power on predicting returning all self-completion questionnaires. The impact of increase in voting ration from the national average increased the odds on returning the self-completion questionnaires by over 100, but is not significant in explaining survey participation at other data collection phases.

The results in Table 4.7 follow our findings from the event history analysis. The maternal language is the closest indicator of people's ethnicity. People with foreign maternal language had reduced likelihood of being contacted and reduced likelihood of co-operation after the health interview. In addition, there is weak evidence that people with foreign maternal language had also reduced likelihood of returning all self-completion questionnaires. However, it is not possible to analyse the survey participation of the language minorities in more detailed grouping in the sequential modelling due to very small group sizes in the sample.

Interviewers are more successful in contacting the elderly people than younger. Nevertheless the elderly people are less likely to participate to the survey and maintaining their full co-operation throughout the data collection phases is more likely to fail than with younger respondents. Females are more often contacted than males, and also more obedient to return the self-completion questionnaires than males. It is more difficult to contact people who live in sheltered homes, oversized dwelling units or whose type of living is unknown in comparison to people belonging to the household population. In addition, contactability is reduced for adult people registered to live at their parental home, single person households and for people receiving income support. The contactability is higher for those in stable socio-economic conditions. Namely people having capital income, living with a partner and/or having children in household improves the probability of being contacted and co-operation. Whilst significant variables describing the affluent economic conditions increase the estimated propensity to co-operate, variables describing income poverty actually reduce the propensity to co-operate. The model suggests that low relative income level and receiving income support reduces the probability of co-operation at many phases while receiving capital income increases the response probability.

In Figure 4.8 and in Table 4.8 we examine the distribution of the predicted probabilities according to the proportion of individuals who fully co-operated at all data collection phases. For sequential modelling, we can observe that the individual sequential models differ in their prediction power. Models predicting the co-operation for symptom interview and medical examination have the largest diversity in the predicted probabilities. In contrary, models predicting the contactability, participation to the health interview and returning self-completion questionnaires are less capable to distinguish between respondents and non-respondents. This can also be observed from the scatter plots of the predicted probabilities across data collection phases in Figure 4.9. For contactability and participation to the health interview the low prediction power of the models can be related to the high contact rate of the survey, multiple underlying processes behind survey non-response on health interview as well as the incapability to control for the topic saliency due to lacking information on peoples health prior the health interview phase. The good explanatory power in models explaining participation to the symptom interview and medical examination is given a significant contribution from the health information available. The occurrence of survey non-response prior to modelling the return of self-completion questionnaires has not simplified the modelling of the non-response at last phase. In contrary, we believe the return of the questionnaire is more related with personality, motivation, self-discipline and cognitive skills than other factors. This type of information is difficult to feed into a quantitative statistical model. The variables indicating social responsibility of men over women, good cognitive skills, further educational qualifications, and couples without children increase the predicted propensity to co-operate. However, the model could still be expected to be improved by some factors that we have been unable to quantify.

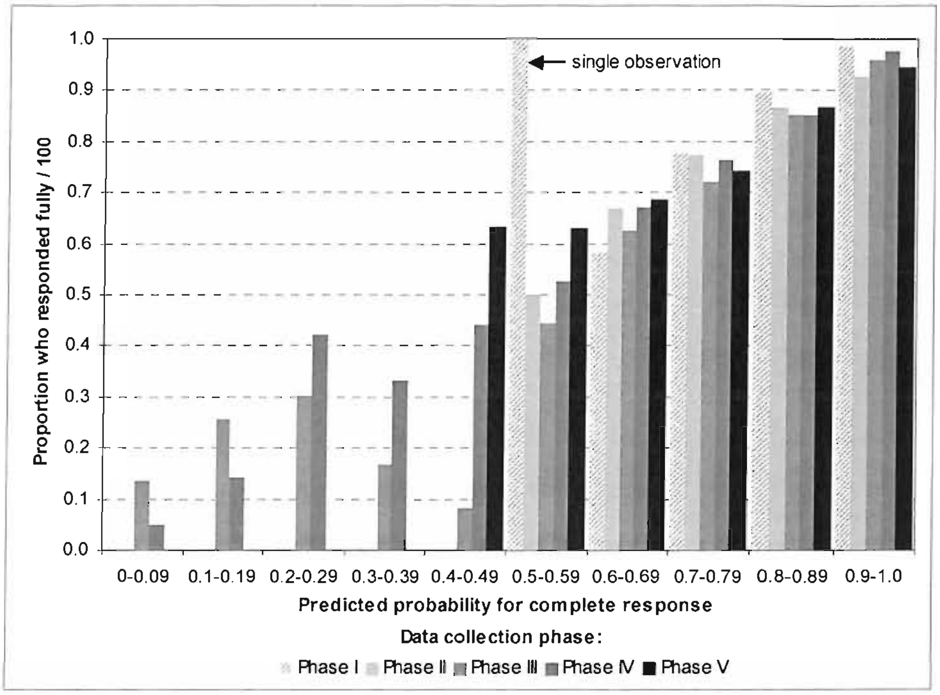
Design-based approach

Two alternative modelling approaches can be implemented depending on the further use of models. The model-based approach represents simple modelling ideology for estimating the response propensities. The model-based approach assumes independency of individuals sampled. This, on the other hand, is often violated in surveys using complex sampling design including features, such as differential inclusion probabilities, stratification and clustering. Methods for analysing complex survey data have been developed by Skinner et al. (1989) and Lehtonen and Pahkinen (1996). These methods permit the inclusion of complex sampling design effects in estimation, as shown in Appendix 4.1.2.

As the Health 2000 survey had a complex survey design, we have compared the approach of model based and design based survey participation models. We focus on the model based estimation as the estimated model and design based response probability models provided almost identical estimates for all models. Model-based approach can be criticised for neglecting the survey design elements, such as the clustering and the differential inclusion probabilities of the sample members. The need of using of design-based approach in response propensity estimation, on the other hand, can be questioned as we will incorporate the estimated models later with methods that account for the survey design in inverse probability weighting. Using the design-based approach can indeed be beneficial in final estimation of survey results. However, the purpose of modelling is to estimate the response propensities of individuals sampled and not the actual survey estimates. Secondly, in our analysis we have found that the design-based models would not have differed by the selection of explanatory variables. In addition, the difference between estimated probabilities in model-based and design-based approach would have been very small. These estimated response propensities will then be used for deriving survey weights that reflect for the inverse of the response propensities, which can then be adjusted with the design weights if necessary.

Figure 4.8 Proportion of individuals sampled who responded at all data collections phases by the product of their predicted probabilities to co-operate fully in the sequential logit models

(a) Across data collection phases



(b) Overall predicted probability

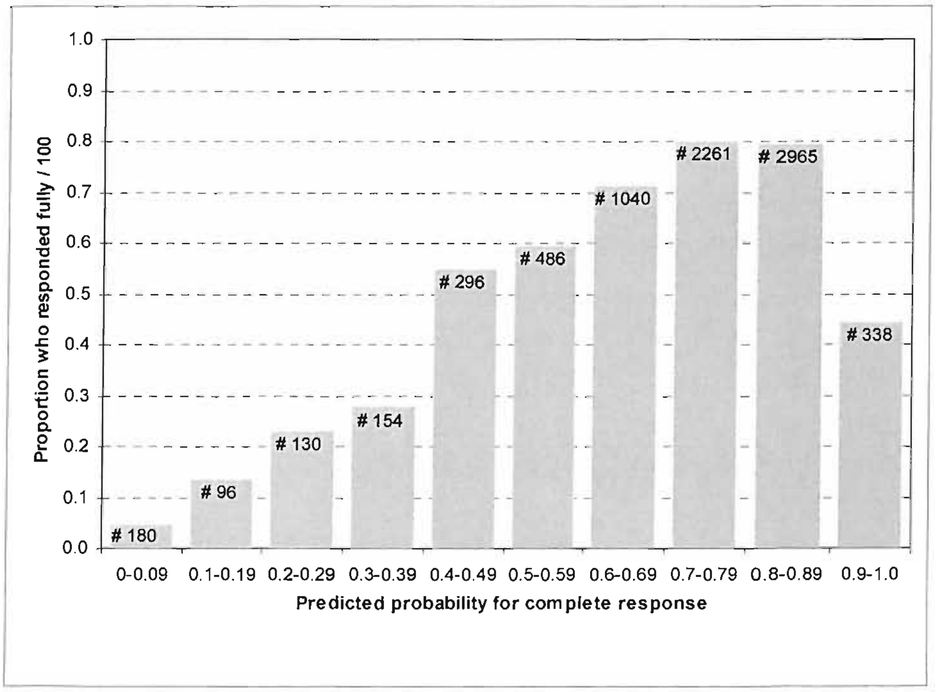


Table 4.7 Parameter estimates for sequential logit model predicting survey participation in multiple data collection phases

	Phase I: Contacting		Phase II: Health Interview		Phase III: Symptom Interview		Phase IV: Health examination		Phase V: Self-completions questionnaires	
Estimated parameters	**Odds Ratio**	**Pr > t-value**	**Odds Ratio**	**Pr > t-value**	**Odds Ratio**	**Pr > t-value**	**Odds Ratio**	**Pr > t-value**	**Odds Ratio**	**Pr > t-value**
Intercept	34.27	0.00	5.04	0.00	2.75	0.00	8.29	0.00	2.53	0.00
Characteristics of the individual:										
Age group:										
- Age 30 - 44 years (vrs 45 - 79)	0.47	0.00	0.84	0.01	1.02	0.87	1.08	0.66	0.96	0.70
- Age 80+ years (vrs 45 - 79)	2.74	0.00	0.99	0.92	0.83	0.14	0.73	0.08	1.12	0.45
Female (vrs male)	2.51	0.00	1.08	0.34	1.14	0.24	0.84	0.20	1.54	0.00
Maternal language:										
- Swedish (vrs Finnish or Sámi)	1.28	0.41	1.04	0.77	1.66	0.01	0.83	0.47	1.01	0.97
- Other (vrs Finnish or Sámi)	0.40	0.00	0.85	0.44	0.36	0.00	0.89	0.78	0.64	0.08
Further education (vrs basic education only)	1.15	0.48	1.26	0.01	1.62	0.00	1.39	0.04	1.62	0.00
Income quintiles of register derived disposable income:										
- 1st quintile or no income (vrs 2nd - 4th quintile)	0.92	0.62	0.93	0.29	0.93	0.48	0.61	0.00	0.77	0.00
- 5th quintile (vrs 2nd - 4th quintile)	0.96	0.80	1.07	0.33	1.11	0.34	1.49	0.01	1.12	0.21
Care benefit for disabled					0.52	0.00	0.46	0.00		
Received income support (vrs none)	0.44	0.00			0.54	0.00			0.59	0.00
Received capital income (vrs none)	1.88	0.01	1.28	0.00						
Savings to additional private pension scheme (vrs none)			1.87	0.00					1.89	0.01
Type of living: shelter homes, over sized DU's or unknown	0.18	0.00			0.48	0.01	0.33	0.00		
Family life cycle:										
- Adult-child in parental home (vrs 'Other type')	0.49	0.05	0.49	0.00	0.57	0.03	1.36	0.49	0.92	0.77
- Single person HH (vrs 'Other type')	0.55	0.00	0.93	0.36	0.84	0.11	0.73	0.04	0.99	0.95
- Couple with no children (vrs 'Other type')	1.92	0.01	1.06	0.53	1.33	0.02	0.90	0.52	1.34	0.01
- Living in HH with children (vrs 'Other type')	2.51	0.00	1.52	0.00	1.29	0.05	1.15	0.47	0.93	0.53
Pensioner (vrs other socio-economic status)							0.50	0.02	0.57	0.00
# of adults 3+ in HH/DU			0.67	0.00						
Type of housing:										
Detached, semi or terraced house (vrs flats)			1.18	0.03						
Type of housing unknown (vrs flats)			0.98	0.86						
Type B: Rental housing[1]							0.30	0.01		

Table 4.7 continues:

Estimated parameters	Phase I: Contacting		Phase II: Health Interview		Phase III: Symptom Interview		Phase IV: Health examination		Phase V: Self-completions questionnaires	
	Odds Ratio	Pr > t-value	Odds Ratio	Pr > t-value	Odds Ratio	Pr > t-value	Odds Ratio	Pr > t-value	Odds Ratio	Pr > t-value
<u>Health interview information of the individual:</u>										
Body mass index (BMI):										
- Information not obtained in health interview					0.07	0.00	0.24	0.00		
- BMI < 18 (vrs 18 <= BMI < 35)					0.85	0.61	0.80	0.57		
- BMI = 35+ (vrs 18 <= BMI < 35)					4.75	0.00	2.54	0.00		
Visited doctor/nurse in last 12 months					0.88	0.03				
What is the current month:										
- 'False (vrs correct)							1.05	0.78	0.67	0.04
- Missing (vrs correct)							0.54	0.00	1.05	0.70
Ability to work:										
- Partly incapable to work							1.86	0.00	1.57	0.00
- Fully incapable to work							0.74	0.08	0.77	0.02
- Missing information							0.44	0.00	0.57	0.00
Showed Sickness Insurance card in HI to interviewer (vrs no)					0.79	0.00				
Difficulties to understand questions and instructions (vrs none)					1.38	0.01	0.65	0.00		
<u>Local area information:</u>										
Crime rate [2]	0.92	0.00	0.98	0.03					0.77	0.01
% Self-employed or entrepreneurs [2]	1.19	0.00								
Children in part-time child care [2, 3]			0.66	0.00					0.69	0.00
Mortality amongst 30 to 64 year olds [2]			6.52	0.00						
Mortality rate [2]					2.15	0.00				
Ratio of institutionalised men over women [2]					1.38	0.00				
Voting ratio of men over women in EU elections [2]									114.81	0.00

[1] Type B: Rental housing constructed with the support by governmental interest rate subsidies

[2] Measured as a deviation from the national estimate per 100 inhabitants

[3] Part-time child care arranged by the local municipality nurseries

Abbreviations:

DU - Dwelling unit; HH - Household; EU - European Union

The models have been estimated using the SAS LOGISTIC-procedure.

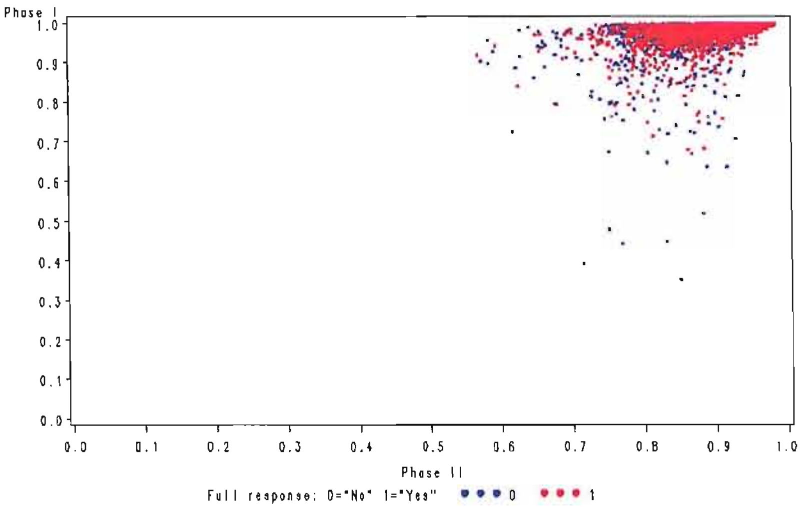
Table 4.8 Distribution of predicted response probabilities in the sequential logit model

Predicted probability	Final Sequential logit % fully co-operated	# of fully co-operated	Mean of sequential response probabilities Stage I	Stage II	Stage III	Stage IV	Stage V
[0 - 0.1]	0.04	8	0.97	0.87	0.30	0.24	0.65
(0.1 - 0.2]	0.14	13	0.96	0.85	0.73	0.48	0.64
(0.2 - 0.3]	0.23	30	0.96	0.85	0.80	0.61	0.68
(0.3 - 0.4]	0.28	43	0.95	0.84	0.87	0.71	0.71
(0.4 - 0.5]	0.55	162	0.96	0.84	0.90	0.83	0.77
(0.5 - 0.6]	0.59	289	0.97	0.85	0.92	0.90	0.81
(0.6 - 0.7]	0.71	740	0.97	0.86	0.94	0.94	0.86
(0.7 - 0.8]	0.80	1812	0.99	0.89	0.96	0.96	0.91
(0.8 - 0.9]	0.79	2352	0.99	0.92	0.97	0.93	0.94
(0.9 - 1.0]	0.44	150	1.00	0.95	0.98	0.81	0.92
Total	0.70	5599	0.98	0.89	0.93	0.90	0.89

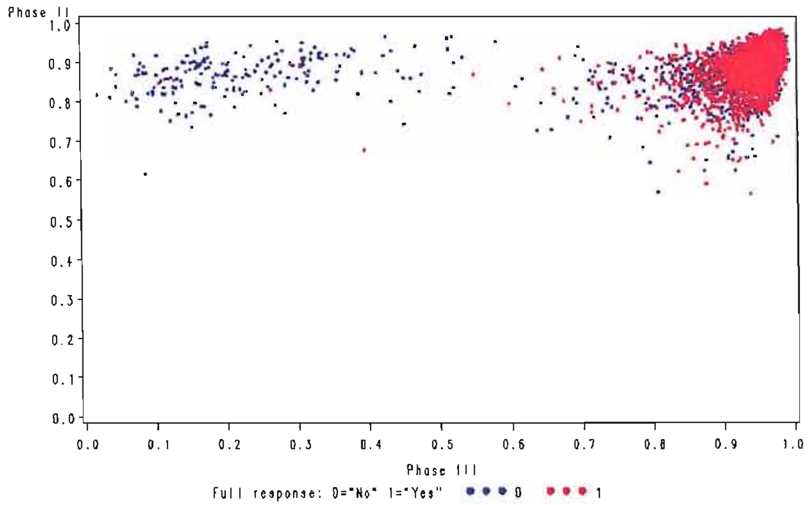
The sequential logit gives us more insight into the survey participation process than any of the previous models. The model estimates are at general level consistent with other models and findings of other non-response research that has been reviewed in Chapter 2. However, the sequential model is the only model of those presented that was able to exploit and demonstrate the importance of the health survey information in the modelling of co-operation at later data collection phases. We will continue assessing the sequential logit models in the next Chapters when examining the interviewer effects on survey participation (Chapter 5), and aiming to improve the non-response adjustment (Chapter 6).

Figure 4.9 Scatter plots of predicted response propensities across data collection phases

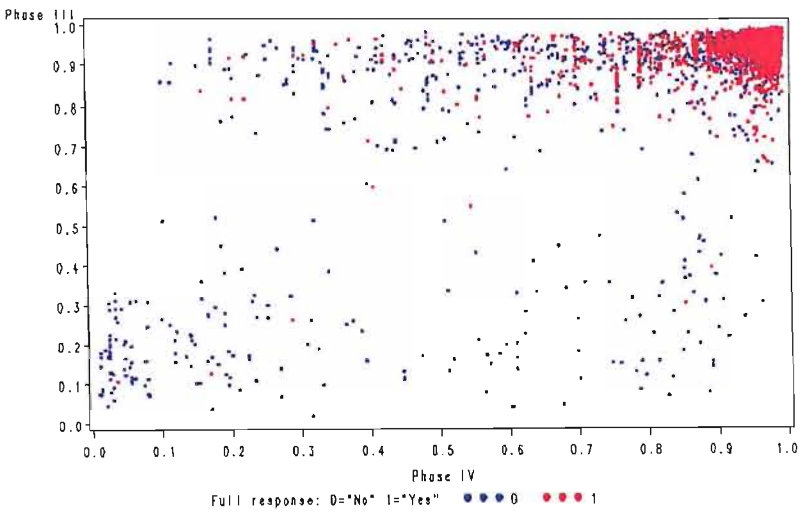
(a) Co-operation at Phase I versus Phase II



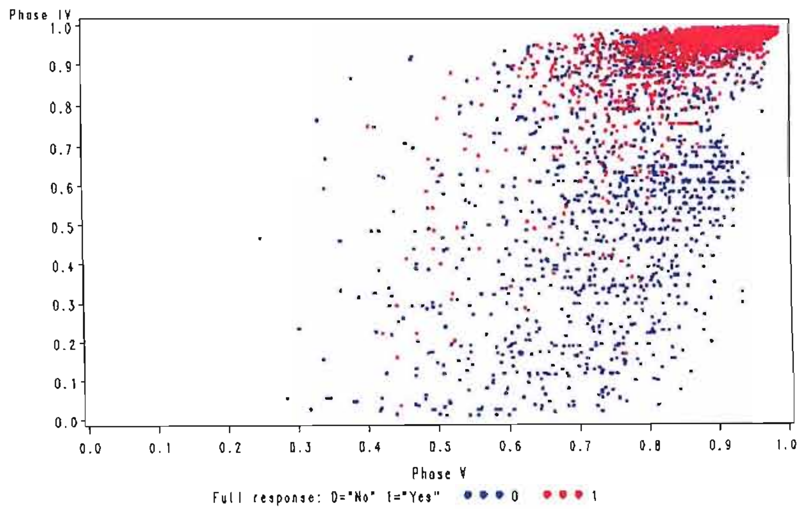
(b) Co-operation at Phase II versus Phase III



(c) Co-operation at Phase III versus Phase IV



(d) Co-operation at Phase IV versus Phase V



4.7. Model comparison and power to predict sampled individuals into respondents and non-respondents

Survey participation behaviour can differ significantly across data collection phases. Therefore the assumption of single binary response variable has been divided into indicators of co-operation at each data collection phase as suggested also by the results from the event history analysis. The results from the event history analysis, discrete-time hazard and cumulative logit models indicate that the expected level of co-operation vary by data collection phases. In addition, the proportionality assumption of the discrete-time hazard and cumulative logit models does not hold and that the parameters of explanatory variables cannot be assumed to be the same at different data collection phases. In comparison to the discrete-time hazard model and the cumulative logit model, the multinomial logit does not condition the probability of the survey outcome on success obtained in the past or future, and it does not distinguish the differences between the past and future events. The implicit difference between the interpretation of the modelling phase in cumulative logit model and discrete-time hazard model is summarised in Table 4.9. The sequential logit uses all the information available for individual level data and in contrast the discrete-time hazard uses the person-event data available at the current data collection phase, for which the success of the survey outcome is been modelled.

The multinomial logit can be seen as a part of the construction of the cumulative logit model, which is more constrained than the multinomial logit. While the multinomial logit allows the β coefficients vary for each x covariate across data collection phases, the cumulative logit fixes the values of β to be the same for each phase t. In contrary to the discrete-time hazard and the sequential logit model both the estimated probability distributions of the multinomial and the cumulative logit models are restricted. The probabilities estimated by the multinomial logit are constrained so that $\sum_{h=1}^{t_{\max}} P(T_i = t) + P(R_{it_{\max}} = 1) = 1$ while the constrained probability of the cumulative logit are $\sum_{h=1}^{t_{\max}} \pi_{ih} = 1$ and $\pi_{it_{\max}} = 1$. Therefore, the cumulative logit can be rewritten containing the estimated probability $P(T_i = t|\mathbf{x}_i)$ of the multinomial logit:

$$\pi_{i[1,t]} = \frac{P(T_i \le t|\mathbf{x}_i)}{P(T_i > t|\mathbf{x}_i)} = \frac{P(T_i = t|\mathbf{x}_i) + P(T_i \le t-1|\mathbf{x}_i)}{P(T_i > t|\mathbf{x}_i)}. \tag{4.25}$$

The nature of the predicted probabilities for full response for individual i are summarised in Table 4.9.

Table 4.9 Summary of logit models used to model the non-response in the Health 2000 data

Model	Event modelled	Key characteristics of the predicted probability	Interpretation of the predicted probability
Discrete-time hazard model	First-time failure at phase T_i conditional on obtaining successfully co-operation at all prior phases	$\lambda_{it} = P\left(T_i = t \middle| T_i \geq t, \mathbf{x}_i\right)$	The conditional probability that individual i experience 1st time wave-non-response at phase t given that acceptable response achieved at all prior phases.
Cumulative logit model	Relative probability that the achieved co-operation ceases at phase T_i or prior to that	$\pi_{t\|1,t\|} = \frac{P(T_i \leq t \| \mathbf{x}_i)}{P(T_i > t \| \mathbf{x}_i)} = \frac{P(T_i = t \| \mathbf{x}_i) + P(T_i < t \| \mathbf{x}_i)}{P(T_i > t \| \mathbf{x}_i)}$	The conditional probability that individual i will co-operate at maximum up to the phase t.
Multinomial logit model	Probability of the failure occurring at phase T_i or probability of the full response	$\pi_{it} = P\left(T_i = t \middle\| \mathbf{x}_i\right)$	The likelihood that individual i co-operates at all subsequent phases until time point t+1. The sum of probabilities of co-operation at each phase t sum up to 1.
Sequential logit model	Co-operation at phase T_i conditional on success at all prior phases	$\pi_{it\|t-1} = P\left(T_i > t \middle\| T_i \geq t, \mathbf{x}_i\right) = 1 - \lambda_{it}$	The product of predicted probabilities for the individual to co-operate at all subsequent phases T conditional that acceptable response achieved at all prior phases.

The discrete-time hazard models, cumulative and multinomial logit models estimate the response and non-response probabilities in a single procedure while the sequential logit models are estimated separately for each phase with the conditionality restriction. The response probabilities of the sequential logit are then derived as a product of the probabilities estimated in the separate models. The estimates of the multinomial logit model illustrate that the impact of one variable can vary greatly at different data collection phases and there can also be major differences in the statistical significance of the parameters across the phases. The results from the sequential logit modelling suggest that completely different set of variables can explain the non-response behaviour in empirical survey data.

Classifying predictions into expected response and expected non-response

The response probability model performances are assessed analysing the prediction power of the models and using the log-likelihood ratio test. Simple measures such as the proportion of correctly classified respondents and non-respondents are compared across the models. Comparing to the simple logit model, all other models have a slightly higher rate for predicting the respondents correctly. The discrete-time hazard model predicts correctly high response probabilities for all respondents in our data, but as emphasised earlier this seemingly outstanding result must be treated with precaution as it is conditional on the response structure. Similarly the sequential logit models are conditional on the knowledge of the response structure. Instead, the unconditional multinomial, cumulative, and simple logit models can be used for estimating the response behaviour independently from the data structure. When comparing the proportion of correctly classified non-respondents only the unconditional multinomial logit model and the multinomial logit conditional on obtaining health interview were capable to improve the prediction rate from that of the simple logit model.

The Hosmer-Lemeshow test provides a comparable goodness-of-fit statistic for assessing the model performance (Hosmer and Lemeshow, 1989). It is derived by ordering the predicted probabilities into deciles and measuring the difference between the observed and model predict counts. The test statistics $\hat{C}$ is obtained by

$$\hat{C} = \sum_{k=1}^{10} \frac{\left(o_k - n_k \bar{\pi}_k\right)}{n_k \bar{\pi}_k \left(1 - \bar{\pi}_k\right)}, \tag{4.26}$$

where n_k is the total number of subject in the kth decile, $k = 1,\ldots,10$, $\bar{\pi}_k$ is the average over the estimated response probabilities in the kth decile and the o_k refers to the number of responses among the c_k covariate patterns in the decile in question (Hosmer and Lemeshow, 1989). The corresponding p-value can be compute from the chi-square distribution.

The problematic feature with all models is that although most of those with high predicted response probabilities are correctly true respondents the models also tend to predict low response probabilities for proportionally large number of respondents. A summary of performance of the logit type of models is provided in Table 4.10. The proportion of correctly estimated respondents out of all true respondents vary between 98 and 64 percentages from all individuals with low predicted response probabilities. On the other hand, this indicates that these individuals are members of sub-populations who are in high risk for becoming non-respondents but that these individuals co-operated with the survey request against the odds.-The goodness of fit comparisons are carried out by analysing the prediction power of the models. The proportion of individuals co-operating fully by their predicted probabilities indicate that the predictive power of the multinomial logit and sequential logit models are reasonable.

In Table 4.10, we also present the likelihood ratio test from the models, comparing the estimated models to the intercept only models for each type of logit model. The likelihood ratio is based on the estimated log-likelihood values for the full model and the intercept only model. Under the null hypothesis that all coefficients of the explanatory variables are zero, minus twice the change in the log-likelihood follows a chi-square distribution with 2 degrees of freedom (Hoshmer and Lemeshow, 1989). According to Powers and Xie (2000), with the assumption of the independence across transition levels, the overall log-likelihood for the model is the sum of the likelihoods from separate models. Thus for sequential logit model, we derive the overall log likelihood statistic as a sum of the dependent sequential logits. In comparison to the simple logit model, the cumulative logit model shows less improvement in the model fit, while the multinomial and sequential logit models seem to have more improvements. However, as the estimated log-likelihoods for the intercept only models for each logit type models differ significantly (except for the multinomial and cumulative logit models), and are possibly greatly affected by the differences in the model assumptions, the likelihood ratio tests cannot be used for further model fit comparisons.

Hosmer and Lemeshow (2000) have recommend to study both the summary measures and individual components of these measures when analysing and comparing the goodness of fit of logistic regression type models. The sensitivity and specificity measures provide measures for assessing how well the models are able to discriminate the respondents and non-respondents in relation to the observed behaviour of individuals sampled. The sensitivity is a proportion between correctly predicted respondents and all respondents. Similarly the specificity is a proportion between correctly classified non-respondents out of the sum of all non-respondents to the survey. Thus the sensitivity provides a measure for the probability of detecting the true respondents and '1–specificity' the probability of detecting false respondents.

The estimated response probabilities of the logistic models are used to classify whether the individuals/units are respondents or non-respondents. One plausible cutpoint used for this classification is the traditional threshold of 0.5. Thus, individuals receiving an estimated response probability less than 0.5 would be grouped into non-respondents and individuals receiving an estimate of 0.5 or above are grouped into respondents based on the definition of the modelled probability. However, depending on the model assumptions and the behaviour of the data, the fixed cutpoint can be very arbitrary, and it can be argued that the cutpoint should vary depending on the estimated model. We use the measures of sensitivity and specificity to assess the optimal choice of the cutpoint for each of the studied models. Thus the sensitivity and specificity rates have been derived for all compared models using the moving cutpoint from 0.01 up to 0.99.

When the sensitivity and specificity are plotted into a same graph, we can observe the optimal cutpoint for the estimated probability to define the respondents and non-respondents from the point in which the curves cross. Thus for the sequential logit, the cutpoint to be used for the conditional response probability the optimal limit is 0.7 rather than 0.5 (Figure 4.10). The suggested cutpoints for the other models vary in more limited range: 0.9 for the discrete-time hazard, and 0.7 for the cumulative, multinomial and simple logit models. The multinomial and cumulative logit models tend to estimate similar response probability distributions regardless the differences in the model definitions.

When the specificity and 1-sensitivity are plotted against each other they form a ROC curve (receiver operating characteristics), which reflects the ability of the model to discriminate true and false respondents at different levels of the reference probability for response. The ROC curves are presented in Figure 4.11. The discrete-time hazard model has seemingly an outstanding capability for the discrimination, but built-in data structure dominates the results from the actual prediction power. The dominance of the data structure on the results of the discrete-time hazard model should be analysed using various type of data. Based on the ROC curve, the simple logit model seems to have a good discrimination power between true and false respondents. The conditional probabilities of the sequential logit model have slightly lower prediction power than the cumulative, multinomial and simple logit models when it comes to predicting correctly the response probability for non-respondents.

The simple logit model performs relatively well in comparison to more complex models. On the other hand, the complexity of the sequential logit model does not reduce its prediction power significantly against the other models. The sequential logit model has the strength in correctly predicting response for the true respondents, and to predict very low response probabilities for a relatively large number of non-respondents. This can be observed clearly from Figure 4.11. However, the model has a lower ability to discriminate the non-respondents, in comparison to simple and multinomial logit.

We will investigate further the model predictions in Chapter 6. Based on the model comparison, we select the simple logit, sequential logit, and the multinomial logit to be assessed further for the purpose of inverse probability weighting. Contrasting with the simple logit, we will study the impact on weighting by allowing the intercept, coefficients and also the explanatory variables to vary across the data collection phases. Thus we will examine whether the complexity of the models adds penalty to the non-response adjustment, or whether the informative nature enables us to better correct for the plausible response bias.

Figure 4.10 Plot of sensitivity and specificity for the conditional response probabilities by estimated survey participation models

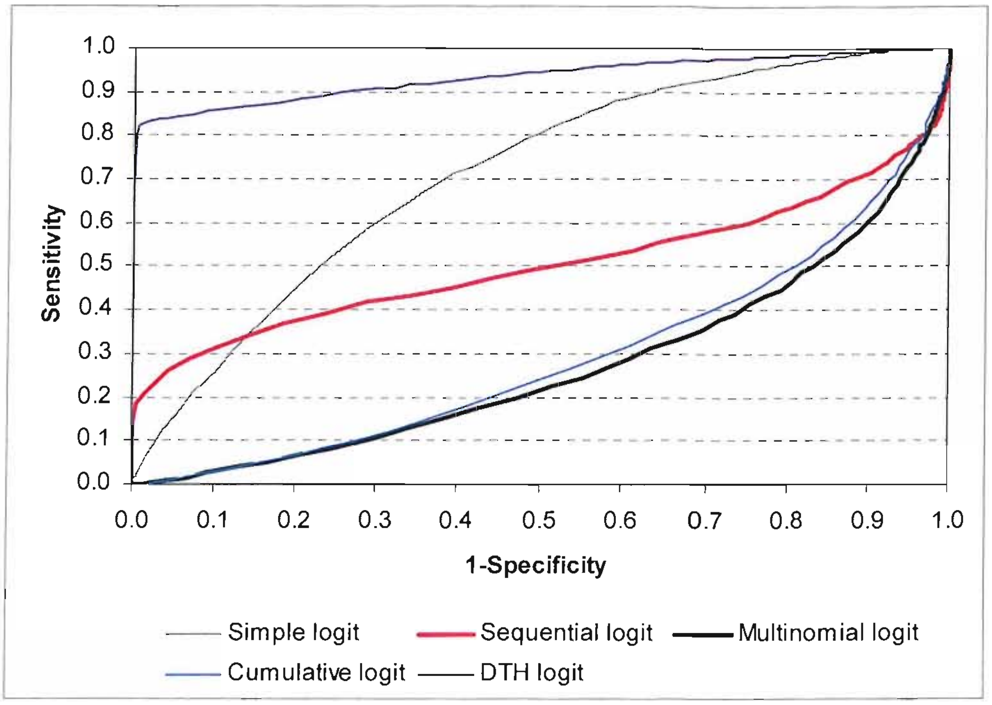


Figure 4.11 ROC Curves by the estimated survey participation models

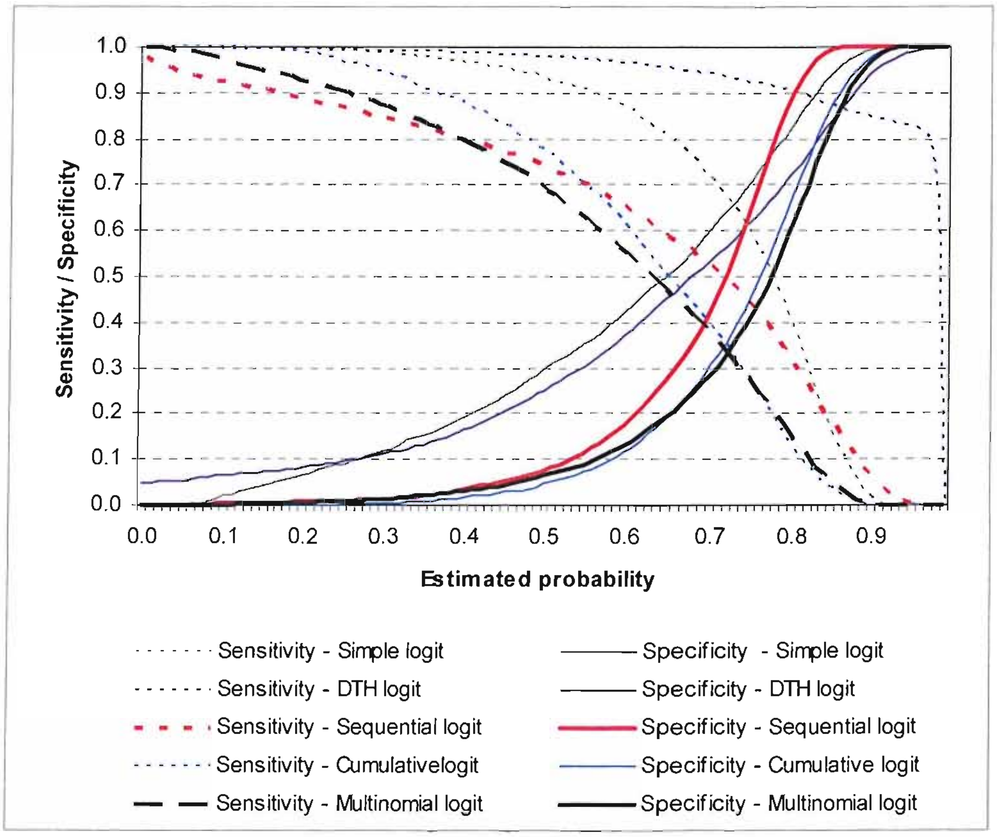


Table 4.10 **Summary of performance of the logit models for the Health 2000 data**

		Cut-off point of sensitivity and specificity curves	% correctly classified respondents	% correctly classified non-respondents	% falsely classified to respondents	% falsely classified to non-respondents	Hosmer-Lemeshow test Ĉ[1]	Likelihood Ratio[2]
Simple Logit Model		0.74	64.6	67.3	32.7	35.4	3.4	1125.9
	(Pr > ChiSq)						(0.9059)	(<0.0001)
Cumulative Logit Model		0.72	66.9	63.3	36.7	33.1		913.8
	(Pr > ChiSq)							(<0.0001)
Multinomial Logit Model		0.73	66.9	65.2	34.8	33.1		1730.3
	(Pr > ChiSq)							(<0.0001)
Sequential logit:								
Stage 1: Contacting		0.98	80.5	67.7	32.3	19.5	7.5	227.7
	(Pr > ChiSq)						(0.4860)	(<0.0001)
Stage 2: Health Interview		0.89	62.9	55.9	44.1	37.1	6.3	181.0
	(Pr > ChiSq)						(0.6124)	(<0.0001)
Stage 3: Symptom Interview		0.92	88.5	54.8	45.2	11.5	9.8	767.3
	(Pr > ChiSq)						(0.2777)	(<0.0001)
Stage 4: Medical measurements		0.94	86.7	70.7	29.3	13.3	4.4	774.5
	(Pr > ChiSq)						(0.8220)	(<0.0001)
Stage 5: Self-completion questionnaires		0.91	70.7	63.7	36.3	29.3	7.5	193.2
	(Pr > ChiSq)						(0.4818)	(<0.0001)
Overall sequential logit		0.73	70.0	50.8	49.2	30.0		2143.7

[1] Observations were divided into 10 groups by their estimated response probability

[2] Likelihood ratio tests compare the model fit of the estimated model with an intercept only model for each logit type model

4.8. Conclusions

In this chapter we have examined the survey participation process with the means of explanatory analysis and statistical modelling. Exploiting vast auxiliary data resources, we have found supporting evidence for the complexity of the phenomena of non-response and that there is no single variable with fixed covariate that would explain the overall survey participation outstandingly better in our models. We have found that decomposing the outcome variable into categorical phases and using more complex models, namely sequential logit and multinomial logit models allowing the explanatory variables vary at different data collection phases, perform better than simple logit model. Our second finding is that while some models with large number of explanatory variables are more informative about the phenomena of survey participation than others. In most cases, the simple nested versions of models actually predict the response probabilities better, than the models characterising in a more multifaceted manner population groups in risk. For this reason, we have kept the inclusion rules for the significance of variables relatively strict in our models.

Our third finding is that the individual and household level characteristics, which reflect socio-economic status, demographics, situational circumstances, community attachment as well as social and political integration, have the strongest explanatory power together with some characteristics of the regional sub-populations, which were of lower significance in the Health 2000 data. More importantly, the health information of the respondents from the health interview is found to be related to respondents' attrition at later data collection phases. We will assess later in Chapter 6 how severe consequences the survey attrition has on the bias of the estimates in various estimation situations.

Using sequential logit models enable to assess the link between survey participation and health information obtained at the initial data collection phase. We also found that the item non-response has a significant impact on individuals' survey participation to later phases. The link between item non-response and survey attrition leading to wave non-response and even to termination of co-operation at later data collection phases raises concerns and needs further analysis in surveys with multiple data collection phases as well as in longitudinal and panel surveys. The finding needs further studying with other data sets. The results indicate that the data collection organisations may benefit from developing more reactive fieldwork operations during the data collection in order to focus in maintaining or winning back the co-operation of those who have item non-response to crucially important survey questions. This also links the survey data quality to survey costs, because in the case of survey attrition, missing data is introduced into the data matrix. At the same time, resources have already been invested to interviewing respondents who break off or who do not provide complete response causing inefficiencies and need for further fieldwork or methodology development. Our results of the predictive power of item non-response on survey non-response at later phases are confounding with the findings of Loosveldt, Pickery and Billiet (2002).

We found significant differences by influential characteristics affecting survey participation across data collection phases. Some people, who were easily contacted, were less willing to participate and some of those willingly co-operating at initial data collection phase were more likely to drop out later. For example, elderly people were relatively easily contacted but less likely to participate and those who did respond to the health interview were still in high risk of dropping out at all subsequent data collection phases. Thus it is important to analyse the non-response not by a single models but diving deeper into the probability of co-operation and into the various ways exploring it. It is important to compare whether the characteristics of non-respondents are consistent over the models or whether they describe similar or various patterns indicating the complexity of the process.

We have found evidence that in the Health 2000 survey, non-response is more likely to occur for some sub-populations that are economically deprived and in the risk of social exclusion. Thus a wider, multi-survey research approach would be needed in studying whether these associations are only related to this survey, or whether they imply that deprived people are in general in higher risk to be excluded from national statistical surveys. In addition, the initial results from the non-response analysis based on the maternal language indicate that there is a need to carry out a large study across surveys to examine the impact of the ethnicity and language on the equal representation of people in national surveys in Finland. Ideally, this future research project would cover a variety of national social and health surveys in order to study also the association of topic salience and survey participation in minority populations.

We found also social capital, projected through family connections to be highly significant in explaining participation independently, and in the conjunction with the foreign maternal language. Single people with foreign maternal language were more likely to fall into non-response than to co-operate fully. Although this population group is relatively small, their proportion may increase significantly in the future due to international migration. If the national surveys fail to capture and measure the social and health experiences of the specific sub-populations with higher risk for social exclusion, for example, they also fail in providing adequate data for policy monitoring to evaluate the social conditions of the population.

While the non-response models based on logistic regression indicate the groups and characteristics increasing the risk of non-response, the nominal predictive power of the models studied in this Chapter and in the non-response literature can still be regarded relatively low in relation to the aims of the research. Our interpretation is that the behavioural characteristics of the respondents that remain to certain extent unquantifiable for the survey methodologists, and the underlying factors may have a stronger impact on response behaviour than auxiliary information. Therefore we will assess the impact of random effects for our data in the following. In addition, the interviewer characteristics may affect their survey fieldwork performance, which will also be studied in Chapter 5.

To conclude our recommendation is that in all multiphase surveys the non-response should be studied according to the main or aggregated phases of the data collection. Both explanatory analyses and statistical modelling should be used and in multi-phase modelling one should explore in addition to simple logit models also few of the ordinal logit models to find out which of them is most applicable to the non-response analysis of the specific survey with the available information structure for the modelling. Further research is also needed in studying the total survey error including the analysis of the impact of all sources of error for assessing the error sources and how their effects on the survey estimates are cumulated or overruled. The most crucial aspect of the non-response and other types of survey errors are whether they cause bias to the survey estimates. Thus the narrow emphasise of the non-response research should move from analysing non-response models into bias analyses of the final survey estimates.

5. Interviewer effects in survey participation

5.1. Introduction

Interviewers may have a significant impact on the quality of survey data, and the level of response can vary across interviewer assignments, which are general concerns in face-to-face interviewing surveys. The literature has studied interviewer effects both in relation to item non-response and unit non-response. In this Chapter, we focus on investigating the latter. We analyse whether the participation of individuals within the same interviewer assignment can depend significantly on the interviewer. The literature of interviewer effects contains partly conflicting and partly confounding results of the association between interviewer performance and their characteristics, such as age and professional experience. Anyhow, previous studies have not been able to detect consistently strong associations of this kind (Hox and de Leeuw, 2002). However, Morton-Williams (1993) emphasised that the social skills of the interviewers are more influential than the attributes of the interviewers. Thereafter, research on interviewer effects has focused increasingly on examining the effects of interviewer perceptions and professional attitudes in conjunction with survey participation. Due to the complexity of the phenomena and the recent development in data collection methods and instruments, the theory of interviewer effects on the survey participation and the overall theoretical framework is still an area under development.

In this Chapter, we aim to study how influential interviewers are in relation to other factors associated with non-response by focusing on surveys collecting data with multiple data collection phases, i.e. multiphase surveys. Previously in Chapter 4, the non-response models assessed the influence of the characteristics of individuals, households, dwellings and geographical sub-populations. In this Chapter, these models are extended to include the impact of the interviewer. The models are therefore modified to take into account not only the attributes of the interviewers but also the hierarchy of the data, i.e. the clustering by the interviewer assignments. Subsequently, we look into complex hierarchy of the data in a situation of cross-classified data defined by overlapping interviewer assignments and local areas.

Lepkowski and Couper (2002) stress the importance of previous survey experience as a predictor of survey co-operation. In relation to this, we aim to study how the initial interviewer-respondent interaction impacts the survey co-operation of individuals at later data collection phases in multiphase surveys. In this assessment, we explore whether survey participation and attrition vary significantly across interviewer assignments and by interviewer related factors. We assess whether interviewers influence survey participation also at later phases in multiphase surveys after the interviewer administration has ceased.

In most multiphase health or social surveys, the data collection begins with contact attempt and face-to-face interviewing followed by further data collection components. The traditional task of the interviewers can be distinguished into tracing the individuals sampled, motivating the contacted to participate in the actual interview, and conducting the actual interview. In multiphase surveys, the interviewers' task is also to encourage the respondent to participate in the subsequent survey components that follow the initial interview. We have found that most non-response studies on interviewer effects focus only on analysing the interviewer effects in relation to contacting the interviewees, achieving co-operation or in relation to an aggregated simple response indicator. Therefore, we extend the conventional analysis of interviewer effects to multiphase surveys. We will explore the significance of the recent survey experiences, beginning from the initial interviewer-respondent interaction, at all subsequent data collection phases. We examine how far in the data collection the initial interviewer approach can influence the survey participation. In addition, we explore whether there are influential interviewer characteristics, attitudes or perceptions that would be significantly related to the non-response also at those data collection phases that are independent from interviewers influence.

Professional attitudes of the interviewers have been found to be associated with fieldwork performance in many studies (e.g. Durbin and Stuart, 1951; Singer, Frankel and Glassman, 1983; Lehtonen, 1996; Hox and de Leeuw, 2002). To illustrate modelling of the interviewer effects we use the Health 2000 data[1] and the related interviewer perception survey. In our empirical study, we report on an interviewer perception survey that was conducted in order to assess to what extent the professional attitudes of interviewers influence the survey participation[2]. This survey has replicated both the set of questions developed by Lehtonen (1996), and some analysed by Couper and Groves (1992). In addition, the questions partly coincide with those analysed by Hox and de Leeuw (2002). Combining the approaches, we aim to explain how well the professional attitudes and characteristics of the interviewers can predict their work performance. Secondly, we compare whether interviewer attitudes influence interviewer performance more than other interviewer characteristics. Thirdly, we are interested to assess how significant the interviewer attitudes are when the response models contain background information also on significantly important variables at the level of individual, household and ecological sub-population.

---

[1] Health 2000 is a health interview and examination survey commissioned by the National Public Health Institute to study the health of the population resident in Finland (see Chapter 3 for more detailed description).

[2] The interviewer perception survey was designed by a team of survey methodologists in Statistics Finland: Nieminen, Laiho, Lehtonen and Vikki. The results have not been published in the series of Health 2000 reports. Ramadan (2001) has studied the data in her Masters Thesis. Laiho (2001) has analysed the results in relation to survey participation. Nieminen (2003) has reported on the interviewer attitudes.

To construct the data for this analysis, we have combined the survey outcome data with the interviewer perception data that was collected after the fieldwork. Both data sets have been explained more in detail earlier in Chapter 3. For using the interviewer survey results correctly, we will address the problem of exploiting the data for further modelling of the non-response of the individuals sampled in the presence of some interviewers not responding to the survey on their professional attitudes. We will also discuss how to adjust for the non-response of the interviewers (in the interviewer perception survey) in order to incorporate their attitudes to the general non-response models of individuals. If the non-responding interviewers would simply be ignored, the non-response models for the individuals sampled to the Health 2000 survey would exclude all those individuals assigned to the non-responding interviewers of the interviewer perception survey, when the interviewer survey information is used in the non-response models.

We begin describing the fieldwork arrangements and evaluating the fieldwork performance of the interviewers by assessing the achieved co-operation within assignments in the Health 2000. Following this, we measure the professional attitudes via interviewer perception survey addressed for those interviewers who collected data for the Health 2000. The key results of the interviewer survey are compared to other similar surveys. We will then focus on detecting interviewer effects from survey participation data. Subsequently, we analyse the impact of interviewer effects on contacting and achieving survey co-operation by assessing the usability of logit models for grouped data, multilevel sequential logistic modelling, and multilevel models with cross-classified data. We will first model the interviewer response rates using solely the interviewer variables as covariates. Later, when modelling the response probabilities in multiple data collection phases in the presence of interviewer effects, we will incorporate the characteristics of respondents and regional sub-populations into the modelling. In the modelling we will also detect the random effects to capture the interviewer level heterogeneity. In order to examine the impact and efficiency of multilevel modelling for survey participation, we will then examine the explained and unexplained heterogeneity of the models.

5.2. Associations between characteristics and perceptions of the interviewers in comparison to their fieldwork performance

In this section, we assess the completion of assignments of the interviewers, who collected the data of the health interviews for the Health 2000 survey and were employed by Statistics Finland. We examine the associations of fieldwork performances in comparison to interviewers' professional attitudes, survey specific perceptions, demographic attributes[3] and length of interviewing experience. In Chapter 4 we found the survey participation associated with demographic and socio-economic factors at the level of individual, household and ecological population. In the following, we initially exclude the area and individual level attributes focusing solely on exploring the impact of interviewer characteristics on their performance. We will then incorporate the rest of the auxiliary information in the framework of multilevel modelling.

5.2.1 Variation of completion rates by interviewer characteristics

When comparing the work performance of all interviewers to those interviewers who responded to the interviewer perception survey, we do not find significant differences in the completion rates. Similarly, there are no significant differences according to the performance and background characteristics between responding and non-responding interviewers as can be seen from the Table 5.1 and the Appendix 5.1. Therefore, we feel confident to incorporate the interviewer perception survey information into the modelling of the interviewer effects. We treat the interviewer-level non-response to the perception survey as informative missing information. Instead of imputing the missing values, all perception variables are coded to contain a separate missing values category. This ensures exploiting all units in the data set without adding assumptions on the interviewer perceptions to those for whom we have not obtained their perceptions. It also enables studying the possible underlying differences in assignment completion between responding and non-responding interviewers.

The completion of the interviewer assignments is associated with some of the background characteristics such as age, educational background, and maternal language, shown in Table 5.1. To test the significance of the association we have carried out one-way ANOVA tests in single variable analysis. The heterogeneity of the completion rates in bivariate analysis is measured by standard deviation. Interviewers aged 60+ were most efficient in tracing and contacting individuals, but less successful in persuading contacted individuals to respond. Though, interviewers aged less than 40 years were least successful both in contacting and in persuading individuals to participate. Interviewer age has more explanatory power than length of their working experience. We can observe a weak indication that interviewers, who are in the early years of their careers, are to some extent less successful than more experienced interviewers. The success of interviewers contacting

[3] The number of male interviewers is proportionally so low in our empirical data, compared to female interviewers, that we cannot compare interviewer performance between female and male interviewers.

and achieving response seems to be inversely associated with their educational level. Interviewers with only basic education level are more successful in achieving response, and especially in contacting people, than interviewers with secondary or higher education. In contrary, the working experience as a professional interviewer does not explain the variation in the survey participation. The education level of the interviewer explains to some extent the differences in the contact rate and the completion rate, but the educational background is insignificant in explaining the variability of the persuasion rate. This may indicate that commitment to interviewing profession is higher amongst those with lower educational background. It can also suggest that in our specific interviewer pool, interviewers with less educational qualifications may have more social skills, helping them in performing successfully in their assignments. However, we cannot test this assumption as the social skills of the interviewers have not been quantified. On the other hand, it is more plausible that the finding indicates that interviewers with further education have been allocated with more difficult assignments. It is also plausible that the finding results from a combination of explanations hypothesised above.

The maternal language of the interviewers is found to explain to some extent the success of the completion and persuasion rate[4]. The results suggest that the interviewers whose language is Swedish had on average lower response and persuasion rates in their assignments. Confoundingly, the response rates of the Swedish-speaking minority have traditionally been lower than for the Finnish-speaking majority in Finland (e.g. Laiho, 1998; Lindqvist, et al. 2001). It is possible that the difference by the interviewing language may also partly be explained to some extent by the regional variation in survey participation, as the geographical distribution of the Swedish-speaking minority is concentrated on the coastal regions and cities in Finland, within which lower response rates were achieved compared to the national average.

The regional differences are found to vary significantly by all outcome rates. The regions indicate roughly the geographic boundaries where interviewers are typically operating. However, because of the survey design, the interviewers were required to operate in wider geographical reach than conventionally. Interviewers operating mainly in the Southern Finland had lowest completion rates while interviewers operating in other areas had significantly higher success rates. While the average completion rate of an assignment was 95 percentages in Northern Finland, it varied between 85 and 87 percentages in Southern Finland, including the capital region.

[4] The interviews were conducted either in Finnish or in Swedish depending on the maternity language and preference of the target person. Unfortunately, we do not have information about the use of interpreters when there were major difficulties in communicating on either of the two languages. Interviewers who participated to the fieldwork have either Finnish or Swedish as their maternal language. Some interviewers are bilingual and conducted interviews on both languages.

Table 5.1 Estimated interviewer success rate by some background characteristics of the interviewer

Interviewer characteristics	All interviewers: # of inter-viewers	Mean of completion rates	Standard Deviation	Anova F-test	p-value	Interviewers participated to the perception survey: # of inter-viewers	Mean of completion rates	Standard Deviation	Anova F-test	p-value
Age group:				4.3	0.01				3.0	0.03
25-39	24	82.8	7.2			21	83.6	7.2		
40-49	43	88.4	6.0			39	88.7	6.0		
50-59	78	88.2	7.1			74	88.2	7.1		
60-66	13	86.8	7.4			11	86.5	7.6		
Gender:				0.7	0.40				0.9	0.36
Female	152	87.4	6.9			139	87.6	6.8		
Male	6	84.9	11.1			6	84.9	11.1		
Education level:				3.5	0.03				3.6	0.03
Basic	92	88.2	6.7			85	88.3	6.6		
Secondary	52	86.9	6.5			46	87.4	6.5		
Tertiary or above	14	83.0	9.6			14	83.0	9.6		
Years of interviewing experience prior this survey:				1.1	0.34				1.1	0.34
0	25	86.4	7.1			21	87.2	6.4		
1	26	86.9	7.5			23	86.8	7.7		
2-9	18	87.3	5.1			17	87.5	5.2		
10-14	28	85.4	9.1			28	85.4	9.1		
15-19	23	89.5	6.6			22	89.8	6.6		
20-24	24	88.2	6.7			21	88.6	6.7		
25+	14	88.7	4.7			13	88.2	4.6		
Main interviewing language of the interviewer:				7.4	0.01				8.1	0.01
Finnish	146	87.8	6.7			133	88.0	6.6		
Swedish	12	82.1	9.6			12	82.1	9.6		
Main regions:				8.4	0.00				7.1	0.00
Larger capital area1)	66	84.7	6.5			60	85.2	6.4		
Other Southern Finland	55	87.3	7.7			53	87.2	7.8		
Eastern Finland	18	90.4	4.3			16	90.9	4.1		
Middle Finland	13	93.3	2.7			11	93.6	2.6		
Northern Finland	6	94.6	3.1			5	94.9	3.3		
All interviewers	158	87.3	7.1			145	87.5	7.0		

[1)] and surrounding municipalities of the capital of Finland i.e. the region of "Uusimaa"

Note: The interviewer success rates are distinguished between contacting individuals sampled and persuading contacted individuals to participate in the Appendix 5.1

5.2.2 Associations of interviewer professional attitudes and fieldwork performance

Previously we have observed variability of interviewer completion rates by their characteristics. We now review the work performance by their professional attitudes and perceptions. We begin with univariate analysis using attitudinal questions for interviewer perception's developed by Couper and Groves (1992) used for assessing the impact of interviewer perception, and the ones designed by Lehtonen (1996) for the general interviewer attitude index. In Appendix 5.2, we have presented distributions of interviewer perceptions by answer categories, and in Appendix 5.3, completion rates of interviewers to selected attitudinal questions. The following aspects can be projected on interviewers' performance in relation to their opinion on facts that increased survey participation. Firstly, professional confidence had a positive association with the success of interviewers' fieldwork. Secondly, recognising positive impacts of the survey incentives appear to be associated with higher completion rates. The thorough health examination appeared to be a relatively effective survey incentive for individuals to participate. Interviewers who recognised the health examination as a positive aspect in survey participation gained completion rates above 87 percentages, in comparison to those who considered the health examination to affect the participation only in small amount, completing only 75 percentages of their assignment.

On reasons for refusals, interviewers were mostly concerned on the impact of the response burden and health of the interviewees. There is weak evidence suggesting that the required home visit in face-to-face interviewing, and the length of the interview lowered the achieved completion rate. More importantly, interviewers concerns on the increased refusals due to the good or bad health condition of the individuals are systematically consistent with reduced completion rates. Successful interviewer performance is associated with determination, persistence but also with flexibility and tailoring. For example, interviewers who indicate that they begin the interview often "before the interviewee has shown any signs of willingness to participate" gain high completion rates of 94 percentages. In comparison, those interviewers who indicate "rarely to begin the interview directly" have a completion rate of 90 percentages, and those who indicated never to begin the interview "without signs of willingness" achieved a completion rate of 87 percentages. On the other hand, flexibility of the interviewer helps to achieve higher completion rates. Those who are strongly supporting or strongly disagreeing the statement of the reluctant respondents, have lower completion rates than others. Also, the completion rate is higher amongst interviewers who strongly agree with respecting the privacy of the respondent than amongst others.

Interviewers who either fully agree or fully disagree on the importance of always emphasising the voluntary nature of the surveys gain higher response rates than those who had no strong opinion. This indicates that a consistent approach can lead to a successful performance. It can also underline that most of these interviewers are operating in areas where a consistent approach is sufficient for gaining higher completion rates without the need for tayloring. At the same time, others may operate within areas where gaining co-operation is more difficult without tayloring.

Results of the interviewers' work motivation contain mixed associations of interviewer perceptions and their fieldwork performance. Also, there are no clear patterns between the work motivation and the conducting or client organisations. Similarly, a mixed pattern can be observed when interviewers are asked about the significance of the survey to the society. This may indicate that two types of interviewers were successful in the fieldwork. Firstly, regardless of a small group size, there is a weak indication of persistent good performance without need for additional motivation. Secondly, a much larger group of interviewers indicated strong positive connection between their work motivation and a special interest on the survey topic as well as with the significance of the survey to the society. The topical motivation and acknowledging the importance of the interviewers work on wider significance, having benefits at national level, were strongly connected with improved completion rates. Interviewers, who consider that their work motivation was significantly improved by these factors, had an average of 89 percent of completion versus 85 percent of those interviewers who indicated that their work motivation increased only slightly by these factors.

The perceived tight schedule in the survey design has a strong positive association with the interviewers' performance[5]. Also, the perceived high workload and the centralised support from Statistics Finland have weak positive associations with fieldwork performance; although not highly significant. This suggests that occasional change of patterns of work and tendencies towards intensive data collection project management can to some extent motivate and enable higher performance amongst interviewers, if they are given at the same time a strong fieldwork support.

Interviewers who participated to the health examination were more successful than on average in gaining co-operation with the contacted individuals. Thus interviewer's own experience of the health examination may have increased their capability to motivate the respondents to co-operate with the survey request. However, in multivariate analysis this association looses its significance. This, on the other hand, may indicate interviewers who participate themselves to the survey components may merely be better performing interviewers, who are also better motivated and familiarise themselves with the survey. Thus the participation to the health examination can also be a latent factor on the work motivation, circumstantial indicator on time use, having the possibilities to make extra professional investment or an indication of interest to the survey topic.

We have found the interviewer performance to vary by significant interviewer attributes and distinct professional attitudes at initial survey steps. We assess these factors in multivariate analysis in the next section. The underlying reasons for interviewers' response may consist of a variety of factors and their behaviours or the nature of their assignment. As the interviewer perception survey was carried after the fieldwork, we cannot make the distinction between whether the positive or negative attitudes affected the fieldwork performance, vice versa, or both.

[5] Collecting the medical data using mobile health clinics imposed a tight fieldwork schedule for the health interviews as the interviews had to be accomplished before the health clinic was due to arrive to the localities.

5.2.3 Completion rates within interviewer assignments across data collection phases

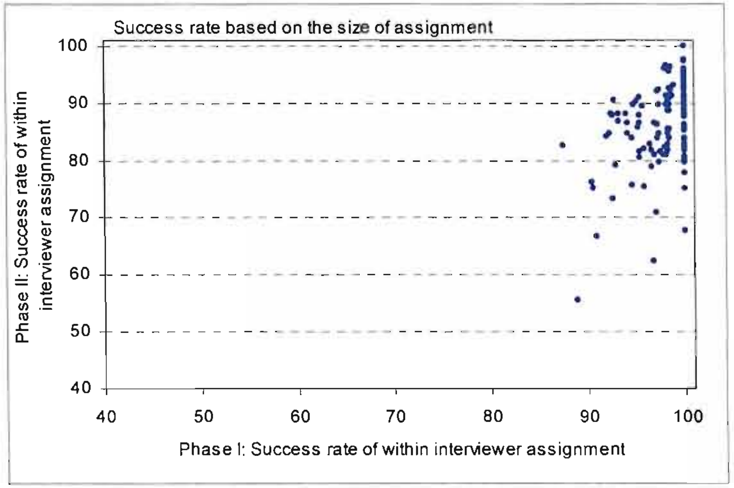
In surveys with multiple data collection phases it is important to assess whether the interviewer performances varies across the phases as interviewers may differ in their contacting or persuasion skills. In addition, in surveys with multiple data collection components, interviewers may differ in their ability to motivate the respondents to fully co-operate at all phases, but the interviewer assignments can differ also in their level of difficulty. Some interviewers are operating in areas with traditionally lower response, and some interviewers may be allocated sampled individuals presumed to be "hard-to-get" based on the prior information available from the sampling frame.

We compare the continuity of the interviewer effect on survey participation across data collection phases assessing the completion rates of interviewer assignments. In Figure 5.1 the associations of the interviewer completion rate are based on the size of the assignment. The interaction of the interviewer completion rates between phases would lay on the diagonal of the scatter plot, if there is no attrition in the following data collection phase. There is only little variation in the contacting rate between interviewer assignments. We cannot observe clear patterns in interviewer performance between contacting and gaining co-operation for the interview as attrition occurred both to interviewers with high and low contacting rates. The patterns of data loss are relatively stable across data collection phases after the first interviewing phase. In assignments with relatively low completion rate to the health interview, the data loss at further phases is not drastically reducing the final obtained data. Large data losses across phases by interviewer assignments are rare. The data loss by assignment is distributed equally between assignments with medium or above medium performance in the previous phase. However, within some assignments there is no data loss in between the transition from one data collection to another. Attrition is occurring at all levels of previously achieved co-operation rates. However, the largest proportions in the reduction of co-operation are occurring in those assignments within which the co-operation has been high at previous phase. There are assignments within which there is no attrition in between two subsequent data collection phases, but a large decrease at the phase following the previous two. This indicates that the survey non-response patterns vary across data collection phases by the interviewer assignments.

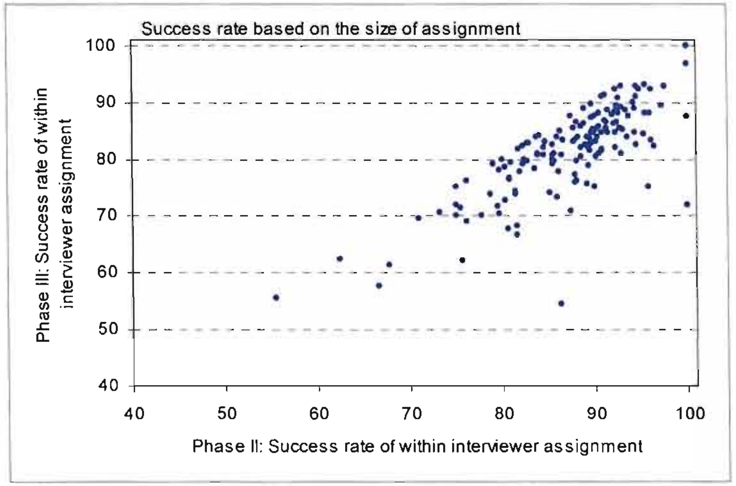
Figure 5.2 demonstrates the achieved response rates at interviewer level across data collection phases. Figure 5.2.a shows a general downward bias trend in survey co-operation across data collection phases. However, when assessing the performance within the interviewer assignments, we can observe more differences visually from the Figure 5.2.b, in which each line represents achieved co-operation rate per interviewer assignment across phases. In some assignments the survey co-operation seems to be high throughout the fieldwork while for some the contactability is initially low, but thereafter the ratio of continuation is high. And for some interviewer assignments the co-operation reduces at all data collection phases.

Figure 5.1 Pair-wise scatter plots of interviewer completion rates across data collection phases based on the original assignment size

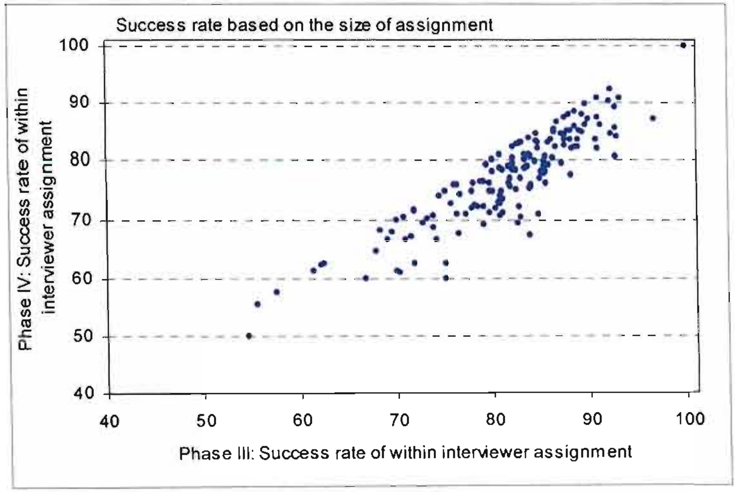
(a) Phase I Contacting versus Phase II Health interview



(b) Phase II Health interview versus Phase III Symptom interview



(c) Phase III Symptom interview versus Phase IV Health examination



(d) Phase IV Health examination versus Phase V Self-Completion

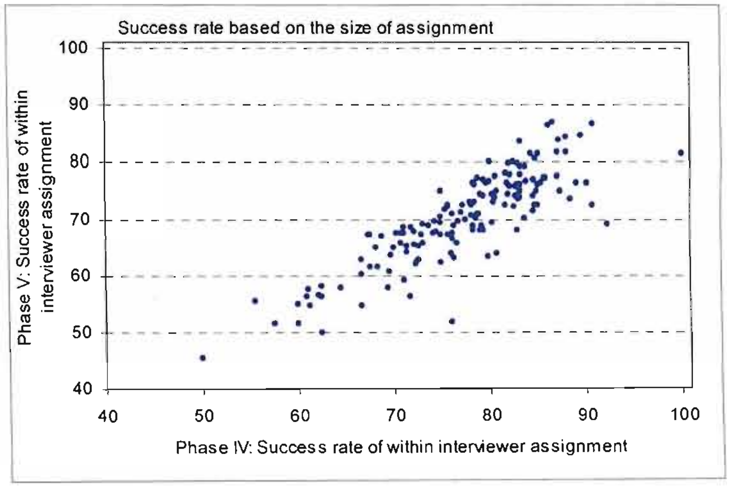
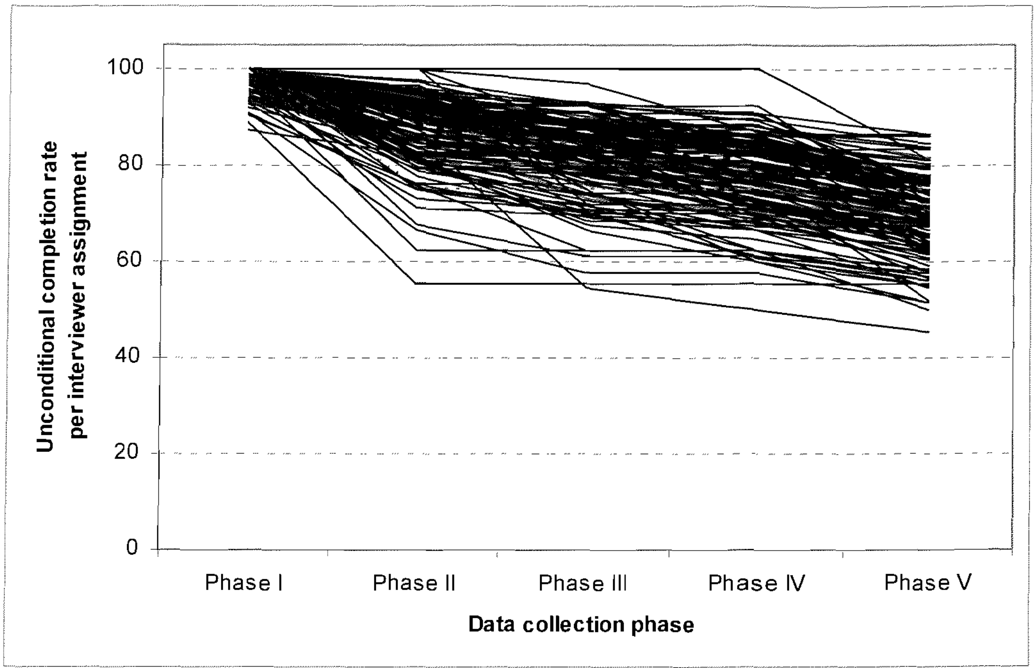
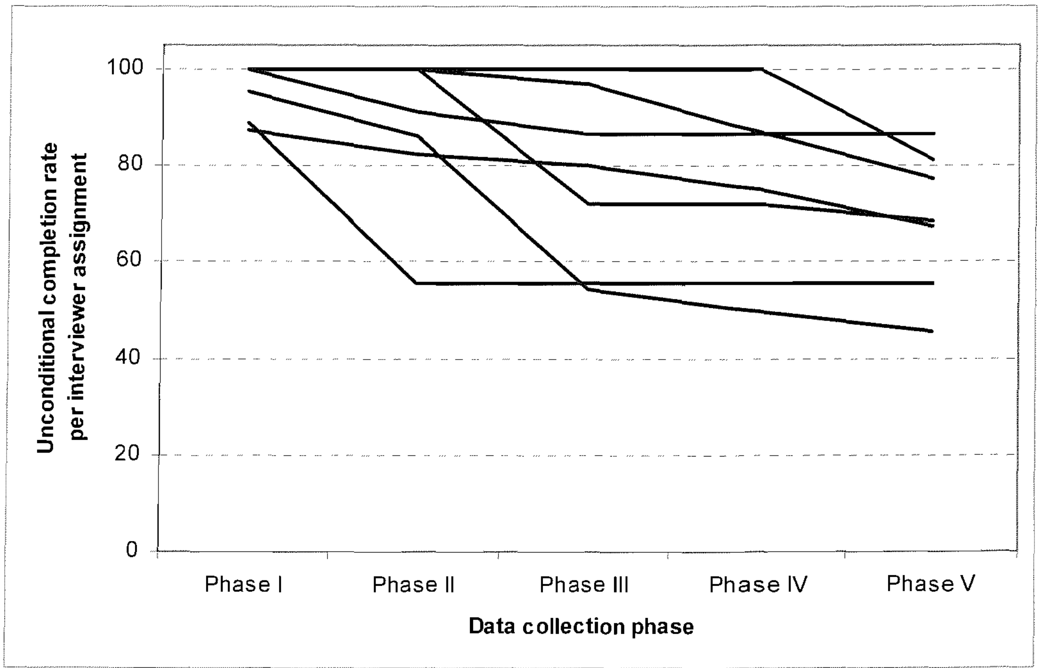


Figure 5.2 Trend of achieved response rates for each interviewer across data collection phases

(a) Continuity of co-operation by assignments



(b) Selected examples of continuity of full co-operation within interviewer assignments



5.2.4 Simple logit model for completion of the assignment at the interviewer level

In this Section, we explore the impact of the interviewer on the individuals' response to the health interview in multivariate context. In order to detect to which extent the interviewer variability explains the differences in the interviewer response rates, the response probabilities are initially modelled using only interviewer level information. If the interviewer level variation is strongly affected by interviewer level variables, it may raise concerns on deploying further adjustments for correcting the survey estimates and needs to assess interviewer effects. A simple logit model is constructed to estimate the probability of the interviewers achieving survey response from the individuals sampled. The model explores the proportion of the successful interviews out of the number n_k of the interviews assigned to interviewer *k*. The model exploits solely the interviewer level information $\mathbf{x}_k$, which consists of the interviewer characteristics, their perceptions and professional attitudes excluding initially the variation at levels of individuals and local areas. Let r_k to denote for the number of successful interviews. Using the interviewer level information we estimate the logit model:

$$r_k \sim Binomial\left(n_k, \pi_k\right), \qquad (5.1)$$
$$\text{logit}\left\{\pi_k\right\} = \boldsymbol{\beta}\mathbf{x}_k .$$

In the model construction, we have explored the importance of all interviewer characteristics and perception data available. For assessing the importance of the survey specific interviewer workload, we transformed the assignment sizes from a count variable into a deviation from the mean assignment size. We then tested the explanatory power whether large positive or negative deviations from the average assignment size affected interviewers' completion rates. In our empirical data there is no significant dependency on the interviewers' performance and their workload. To simplify the interpretation of the modelling and increase the explanatory power of the estimated model, variables measured in years were coded into categorical variables. Also, the other interviewer characteristic, perception and attitudinal variables[6] have been truncated and treated as categorical information in the modelling.

We find that the variables significantly explaining interviewers' completion rate are related to the interviewers' age, educational background as well as interviewers' perception on the impact of individuals' health status in general on survey participation, and to the concerns of data protection. Thus, both interviewer level characteristics and perceptions are significant in multivariate model estimated at interviewer level. The estimation results are shown in Table 5.2. As previously detected in Section 5.2, interviewers with lower education are performing better than interviewers with further education also in multivariate context. This hypothesis will be tested further in the next

[6] The interviewer perception survey questionnaire used mainly categorical answer categories, which could not be transformed unanimously to a linear scale. In data preparation, we merged some relatively close response categories in the interviewer survey questionnaire, as there were no differences in terms of the completion rate.

Section as we will incorporate the individual level information into the model that account for largely to the variation of the difficulty of the assignments. The model suggests that interviewers aged 40 or over are more successful in completion of their assignments than their younger colleagues. Interviewers, who expressed concerns that the bad health condition of sample individuals could have reduced their response propensity, had slightly lower odds ratio in comparison to other interviewers.

Table 5.2 Interviewer level logit model on individuals' response propensity to the health interview

Explanatory variables at interviewer level	Estimate	Standard Error	Pr > Chi Square	Odds Ratio	95% Confidence Limits of Odds Ratio Lower	Upper
Intercept	2.31	0.16	0.00	10.1	7.4	13.7
Interviewer characteristics:						
Aged ≤ 39 (vrs 40-59 years)	-0.14	0.07	0.06	0.9	0.8	1.0
Aged 60+ (vrs 40-59 years)	0.00	0.09	0.96	1.0	0.8	1.2
Maternal language Swedish (vrs Finnish)	-0.26	0.13	0.04	0.8	0.6	1.0
Further education (vrs basic education)	-0.21	0.08	0.01	0.8	0.7	0.9
Interviewer perception survey information:						
Impact of bad health condition or illness of the target person on refusals on general, increased:						
- 'Considerably' (vrs 'Not at all')	-0.28	0.08	0.00	0.8	0.6	0.9
- 'In large amout' (vrs 'Not at all')	0.00	0.07	0.99	1.0	0.9	1.2
- 'In small amount' (vrs 'Not at all')	0.13	0.08	0.10	1.1	1.0	1.3
- 'Missing information' (vrs 'Not at all')	-0.10	0.16	0.55	0.9	0.7	1.2
Concern about inadequote data protection						
- 'Considerably' (vrs 'Not at all')	-0.25	0.21	0.23	0.8	0.5	1.2
- 'In large amout' (vrs 'Not at all')	-0.10	0.12	0.41	0.9	0.7	1.1
- 'In small amount' (vrs 'Not at all')	-0.02	0.08	0.77	1.0	0.8	1.1
- 'Missing information' (vrs 'Not at all')	0.16	0.14	0.24	1.2	0.9	1.5
Interviewer's perception on the Health 2000 survey						
- 'Very positive' (vrs 'Rather positive' or 'Neutral')	0.09	0.07	0.17	1.1	1.0	1.3
- 'Missing information' (vrs 'Rather positive' or 'Neutral')	0.04	0.10	0.70	1.0	0.9	1.3

Technical note: The model has been estimated by SAS Proc Logistic

5.3. Multilevel logit model for survey participation allowing for interviewer level and individual level factors

In this section the individual level data is incorporated with the interviewer information into the analysis of survey participation. We construct a model that relates to interviewer assignments and categorical background factors of individuals linearly to the logit of the probabilities. The analysis is extended to multilevel logit model allowing for the variability across interviewer assignments. We analyse whether the interviewer factors found significant in the previously estimated interviewer performance model ((5.1) and Table 5.2), remain significant when we also control for individual level factors. The responses r_{ik} of individuals *i* within interviewer assignments *k* are modelled simultaneously using logit probabilities of:

$$r_{ik} \sim Binomial\left(1, \pi_{ik}\right)$$

$$\text{logit}\{\pi_{ik}\} = \gamma_0 + \sum_{h=1}^{l} \gamma_h x_{hik} + \sum_{l=r+1}^{p} \varphi_h x_{hk} + u_{0k} \tag{5.2}$$

$$\left[u_{0k}\right] \sim N\left(0, \Omega_u\right)$$

where $h = 1, \ldots, l, \ldots, p$, and l indicates the number of explanatory variables at individual level, and $p - l$ indicates the number of explanatory variables at the interviewer level. The intercept is divided into a fixed intercept γ_0 and into a random intercept u_{0k}, which can be interpreted as a group dependent deviation or correction term for each interviewer assignment *k*. For simplicity, the individual variables are denoted by x_{ik} and the interviewer level variables by x_k, which denotes in this case study for interviewer assignments. The differences of the estimated $\hat{\gamma}$'s provide estimates of the expected differences in logits of individuals' propensity to co-operate within their assignment group. The overall intercept of the logit model is γ_0, and the random intercept adjustments for interviewer assignment groups are denoted by u_{0k}'s respectively. The random intercepts are expected to be normally distributed with 0 mean and variance of Ω_u. In comparison to the previously modelled simple logit model, defined in (5.1), this model incorporates individual level variables with interviewer characteristics and perceptions, allowing also for a random intercept for interviewers[7].

In the Table 5.3, we demonstrate how the individual and interviewer level characteristics affect significantly to the predicted response probabilities. Similarly as in the previous model presented in

[7] There were 158 of interviewers for whom 7946 individuals were allocated out of which 7071 were interviewed on their health by the interviewer.

Table 5.2, both interviewers' characteristics and perceptions are significant explanatory factors. Confounding with the previous model, interviewers aged 40-59 years are more successful in their completion of the assignment than other interviewers. However, the education level and the maternal language of the interviewer found significant in the previous model is not significant in the multi-level model where the individual level attributes are controlled for. This is connected with our hypothesis presented in the last section that the interviewers with higher educational background are allocated more difficult assignments. In the previous model, the education level of the interviewer was thus a latent variable indicating assignment characteristics via interviewer attribute. When the direct controlling of the individuals is introduced, the latent variables loose their explanatory power. Individual characteristics found to be very significant in the model related to the characteristics of non-response risk groups identified previously in Chapter 4 in individual level survey participation models. These risk groups are characterised with young or old age, male, foreign maternal language, single person households, low income level and living in urban areas. The individual level variables thus control for the difficulty of the assignment.

The interviewers' perception variables significant in the presence of the individual level data are the impact of the length of the interview and the adequacy of informing about the survey represent significant interviewer perception variables. Thus, the interviewer perceptions significant in the previous model have also lost significance, and new perception variables become powerful in the presence of the individual level information. Interviewers, who regarded the level of information inadequate, were more successful than other interviewers. This indicates that the perception variable represents underlying information about the interviewer skills. The model suggests that interviewers who are expressing critical views on the survey fieldwork arrangements, or interviewers who are more alert to the impact of informing about the survey to sampled individuals for gaining improved co-operation, may actually be better performing than other interviewers. The model will be used as a base for comparison of the more complex multilevel models to be constructed. In the next Section, we will also examine more in detail the impact of the random intercept in the context of multilevel sequential logit modelling across data collection phases. This will be followed by the modelling cross-classified data, taking into account clustering by the interviewers and regional primary sampling units.

The single variable describing the social environment of the individual is urbanicity. The indicator contrasts cities and towns to other less densely populated or rural areas. Urbanicity reduces the likelihood of co-operation significantly at contacting and persuasion to health interview. In addition it has a weak but non-negligible effect on the participation to the health examination. The significance of other geographic population information will be introduced into multilevel modelling in cross-classified analysis.

Table 5.3. Multilevel logit model on individuals' response propensity to the health interview by interviewer assignments

Explanatory variables at interviewer level	Estimate	Standard Error	Pr > Chi Square	Odds Ratio	95% Confidence Limits of Odds Ratio Lower	Upper
Fixed effects						
Individual level:						
Fixed intercept	2.41	0.10	0.00	11.1	9.2	13.5
Age group						
Aged 30-44 (vrs 45-79)	-0.45	0.09	0.00	0.6	0.5	0.8
Aged 80+ (vrs 45-79)	-0.13	0.12	0.28	0.9	0.7	1.1
Female (vrs male)	0.24	0.07	0.00	1.3	1.1	1.5
Maternal language:						
Swedish (vrs Finnish or Sámi)	-0.37	0.16	0.02	0.7	0.5	0.9
Foreign language (vrs Finnish or Sámi)	-0.67	0.26	0.01	0.5	0.3	0.8
Household type:						
HH with children (vrs no children in HH2+)	0.58	0.10	0.00	1.8	1.5	2.2
Single-person HH (vrs no children in HH2+)	-0.28	0.08	0.00	0.8	0.6	0.9
Disposable income in relation to target population:						
1st income quintile (vrs 2nd-4th)	-0.21	0.09	0.02	0.8	0.7	1.0
5th income quintile (vrs 2nd-4th)	0.29	0.09	0.00	1.3	1.1	1.6
Urban living area (vrs non-urban)	-0.43	0.08	0.00	0.7	0.6	0.8
Interviewer level:						
Interviewer age 26-39	-0.38	0.10	0.00	0.7	0.6	0.8
Interviewer age 60-66	-0.13	0.15	0.39	0.9	0.7	1.2
Interviewer perception on refusals						
Length of the interview:						
- increased considerably or to large extent	-0.06	0.21	0.78	0.9	0.6	1.4
- item missing or no information	0.34	0.25	0.18	1.4	0.9	2.3
Inadequote informing about survey:						
- increased considerably or to large extent	-0.28	0.13	0.03	0.8	0.6	1.0
- item missing or no information	-0.48	0.26	0.07	0.6	0.4	1.0

Random effects	Estimated variance	p-value[1]
Random intercept	0.08	0.03

[1] Based on t-test assuming normality

Technical note: The model has been estimated using 2^{nd} order PQL estimation method in MLwiN

5.4. Modelling survey participation for multiple data collection phases in the presence of interviewer effects

We extend our hypothesis to study whether immediate survey experience at the health interview phase may impact co-operation at later data collection phases such as the participation to the health examination and completion of self-completion questionnaires. Our interest is to assess for how long in the data collection process of multiphase surveys the effects of interviewers can be found in the survey participation behaviour. It is of our interest to analyse whether individuals differ significantly by the level of co-operation and whether the level of co-operation is dependent on the interviewer. We study, whether and how the interviewer effects vary across the survey phases with varying data collection modes. More precisely, we analyse the presence of interviewer effects at the contacting and health interview phase, administrated by the interviewer. We then examine whether the impact of the interviewer are still noticeable with the further phases administrated by the health centre experts. We assess how successfully the individuals co-operated with the further steps of the data collection up to the final data collection phases, returning all self-completion questionnaires. Finally, we model the response probabilities of individuals sampled by taking into account the cross-classification of interviewers and geographical clustering.

5.4.1 Survey participation analysis with random intercepts

We examine the impact of interviewer effects on survey participation across the data collection phases revising the sequential modelling defined in Chapter 4 in (4.23) into the multilevel modelling framework. The successfulness of the co-operation for individuals sampled i are denoted by R_{ikt} for interviewer assignment k in the subsequent data collection phases[8]. The estimated models asses whether the probability to gain fieldwork success at data collection phase t is dependent on the interviewer assignments in addition to explanatory variables:

$$R_{ikt} = \begin{cases} 1 & \text{success} \\ 0 & \text{otherwise} \end{cases}.$$

$$\pi_{ikt} = \Pr\left(R_{ikt} = 1 \middle| R_{ik(t-1)} = 1\right) \tag{5.3}$$

$$\left(R_{ikt} = 1 \middle| R_{ik(t-1)} = 1\right) \sim Binomial\left(1, \pi_{ikt}\right)$$

The intercept of the logit model denoted previously by β_0 is allowed to vary in the multilevel logit model. The intercept is diveded into two parts as $\beta_0 = \gamma_0 + u_{0k}$, conditionally that the random intercept u_{0k} is significant for the interviewer assignments k in the estimated model. Thus we allow a random intercept for interviewer assignments, u_{0k}, to be determined independently in the model for each assignment *k* and we test whether the estimated random intercepts differ significantly from each other. The model for the probability of success in data collection at any phase *t* for individual *i* (assigned for interviewer *k*), π_{ikt}, can be written as follows with the random intercept:

$$\operatorname{logit}\left\{\pi_{ikt}\right\} = \gamma_{0t} + \sum_{h=1}^{l_t} \gamma_{ht} x_{hik} + \sum_{h=l_t+1}^{p_t} \varphi_{ht} x_{hk} + u_{0kt} \tag{5.4}$$

$$u_{0kt} \sim N\left(0, \sigma_{ut}^2\right).$$

In comparison to the previous interviewer effect model (5.2), this model is defined for subsequent data collection phases. The predicted probability of co-operation π_{ikt} at data collection phase *t*, defined in (5.3), is conditional upon the successful co-operation gained at the previous phase $t-1$. The model defines an average intercept γ_{0t} for each data collection phase *t*. Similarly, the random intercepts u_{0kt} estimated for the interviewer *k* are restricted to be normally distributed with expectation on 0 and a variance of $\sigma_{u_0t}^2$. The estimation of the model consists of conditional sequential modelling of separate models. As in (5.2), the denotation of the model (5.4) makes a distinction between individual level and interviewer level explanatory variables. The number of explanatory variables p_t can vary from data collection phase to another. Now l_t specifies the number of individual level variables at phase *t* from all of the p_t explanatory variables.

5.4.2 The impact of individual-interviewer factors on the survey participation

In Table 5.4 we present the empirical results for the predicted survey participation across data collection phases when using the sequential multilevel logit model approach. As previously in Chapter 4, the set of covariates are allowed to vary in each sequential model, although we have constructed the sequential models to contain a fixed set of simple covariates at individual level[9]. Firstly, the variation of their significance across data collection phases in multi-phase survey was studied and, secondly, the impact of the estimation method employed in the modelling[10].

[8] Alternatively, in any survey containing clustering in the survey design, one can define the fieldwork success of individuals *i* within primary sampling units (or regional clusters) *j*.

[9] We examine the stability of some well known survey participation determinants across data collection phases. Age group, sex, family status, educational background, and income quintiles have been regarded as significant factors in the literature of survey participation (see also Chapter 2)

[10] The logit model has been estimated usind two estimation methods. The residual maximum likelihood methods (REML) has been estimated using SAS %GLIMMIX macro, and the second order penalised quasi-likelihood (PQL) has been estimated with MLwiN. Both the estimation methods appear to have generally similar directions with some deviations of the level of some estimates. The estimation methods are described more in detail in the next section 5.4.3.

The interviewer level information does not have a high explanatory power, when rich sources of individual level data are available for the survey participation analysis. Interviewer level factors and covariance parameters are significant in contacting and persuading contacted individuals to participate. However, when the individuals' health information is obtained at health interview and used for the modelling at later phases, the significance of the interviewer level variables and individual level factors decrease. The covariance parameter of the interviewer assignment is still significant in participation to the symptom interview, but reduces significance after that. Thus the findings support the concerns of survey bias in relation to the survey topic and response burden. Factors indicating poor physical health status and low cognitive skills are significantly explaining reduced co-operation at later phases, with the factors on social and economic deprivation. As in the model (4.23) presented in Table 4.7, the factors indicating economic deprivation or prosperity, are very good predictors of survey co-operation at initial data collection phases. Generally, the socio deprivation factors, such as type of living unknown, low education level and living in a single person household, increase the risk of non-cooperation. Generally, factors indicating economically and socially advantaged conditions increase the likelihood to co-operate at all data collection phases, regardless the self-selection occurred in previous phases.

Previously we detected more significant differences in completion rates of interviewer assignments by their characteristics and professional attitudes. However, when the interviewer information data is analysed at the level of individuals, the interviewer level information looses its significance in contrast to the individual level data. Interviewer's age is still found to be related to the contactability of the individuals alike in the previous interviewer level models. More importantly, interviewers' perceptions indicate concerns on self-selection of respondents by their health status, which may increase survey bias already at the first data collection component, the health interview. The impact left with some interviewers indicated that individuals' refused due to their bad health, which is associated with lower estimates on response propensities of individuals within the assignments of these interviewers. Similarly, interviewers, who did not respond to the interviewer perception survey or had not responded to this specific question on the impact of bad health to refusals, were less successful in persuasion of respondents within their given assignment.

Table 5.4 Multi-level sequential logit model on survey participation for individual and interviewer fixed effects with random intercepts on interviewer level[11]

	Phase I: Contacting				Phase II: Health Interview				Phase III: Symptom Interview				Phase IV: Health examination				Phase V: Self-completions questionnaires			
	REML		2nd order PQL		REML		2nd order PQL		REML		2nd order PQL		REML		2nd order PQL		REML		2nd order PQL	
Estimated parameters	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value
Fixed effects:																				
Fixed intercept	265.26	0.00	167.32	0.00	10.41	0.00	10.71	0.00	22.62	0.00	21.43	0.00	34.33	0.00	34.98	0.00	8.86	0.00	11.61	0.00
Individual characteristics:																				
Age group:																				
- Age 30 - 44 years (vrs 45 - 79)	0.59	0.01	0.57	0.01	0.65	0.00	0.65	0.00	0.67	0.01	0.71	0.02	0.79	0.19	0.79	0.23	1.00	0.98	1.00	0.98
- Age 80+ years (vrs 45 - 79)	2.71	0.00	3.00	0.03	0.86	0.21	0.86	0.21	0.63	0.01	0.70	0.04	0.22	0.00	0.22	0.00	1.13	0.54	1.13	0.54
Female (vrs male)	2.36	0.00	2.71	0.00	1.12	0.14	1.12	0.15	1.12	0.35	1.19	0.14	0.84	0.18	0.84	0.22	1.51	0.00	1.51	0.00
Maternal language:																				
- Swedish (vrs Finnish or Sámi)	0.67	0.40	0.71	0.52	0.74	0.08	0.74	0.07	0.79	0.44	0.95	0.87	0.64	0.06	0.65	0.07	0.61	0.01	0.61	0.01
- Other (vrs Finnish or Sámi)	0.25	0.00	0.28	0.00	0.71	0.24	0.70	0.24	0.15	0.00	0.20	0.00	1.54	0.51	1.49	0.57	0.37	0.01	0.37	0.01
Family status:																				
- Family with children (vrs families without children)	2.13	0.00	2.26	0.02	1.51	0.00	1.51	0.00	1.34	0.12	1.05	0.76	1.11	0.60	1.10	0.65	0.88	0.35	0.88	0.36
- Single person household (vrs families without children)	0.44	0.00	0.44	0.00	0.87	0.10	0.86	0.10	0.58	0.00	0.60	0.00	0.71	0.01	0.70	0.02	0.80	0.04	0.80	0.05
Further education (vrs basic education only)	1.02	0.92	1.09	0.70	1.29	0.00	1.29	0.00	1.52	0.00	1.69	0.00	1.53	0.00	1.53	0.01	1.67	0.00	1.67	0.00
Income quintiles of register derived disposable income:																				
- 1st quintile or no income (vrs 2nd - 4rth quintile)	0.70	0.07	0.74	0.25	0.94	0.52	0.94	0.54	0.99	0.93	0.97	0.85	0.54	0.00	0.54	0.00	0.66	0.00	0.66	0.00
- 5th quintile (vrs 2nd - 4rth quintile)	1.33	0.22	0.94	0.84	1.14	0.21	1.14	0.22	1.14	0.45	1.26	0.17	1.42	0.08	1.43	0.10	1.07	0.61	1.07	0.63
Received capital income (vrs none) (household level)	1.65	0.01	1.84	0.02	1.22	0.02	1.22	0.02												
Received income support (vrs none) (household level)	0.33	0.00	0.39	0.00																
Savings to additional private pension scheme (vrs none) (household level)					1.81	0.00	1.82	0.00									1.94	0.01	1.94	0.01
Type of living unknown	0.18	0.00	0.14	0.00					0.48	0.03	0.40	0.01								
Region:																				
Urban (vrs rural or other less densly populated)	0.26	0.00	0.38	0.00	0.74	0.00	0.74	0.00					1.28	0.06	1.26	0.08				

to be continued on the following page

[11] The difference between REML (residual maximum likelihood methods) and 2nd order PQL (second order penalised quasi-likelihood) estimation methods will be explained in the next Section 5.4.3.

Table 5.4 continues

Estimated parameters	Phase I: Contacting				Phase II: Health Interview				Phase III: Symptom Interview				Phase IV: Health examination				Phase V: Self-completions questionnaires			
	REML		2nd order PQL		REML		2nd order PQL		REML		2nd order PQL		REML		2nd order PQL		REML		2nd order PQL	
	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value
Health interview information of the individual:																				
Body mass index (BMI):																				
- Information not obtained in health interview									0.02	0.02	0.02	0.00	0.11	0.00	0.12	0.00				
- BMI < 18 (vrs 18 <= BMI < 35)									0.24	0.00	0.23	0.00	0.32	0.02	0.33	0.03				
- BMI = 35+ (vrs 18 <= BMI < 35)									1.13	0.68	1.43	0.25	1.13	0.65	1.14	0.66				
Interviewer perception on individual's ability to understand speach or instructions:																				
- Apparently mild difficulties (vrs no difficulties)													0.24	0.00	0.24	0.00				
- Clearly observable difficulties (vrs no difficulties)													0.14	0.00	0.14	0.00				
- Barely understands (vrs no difficulties)													0.03	0.00	0.03	0.00				
- Item missing information (vrs no difficulties)													0.01	0.00	0.01	0.00				
Cognitive ability:																				
Which month of the year?																				
- Incorrect (vrs Correct)																	0.44	0.00	0.34	0.00
- Item non-response (vrs Correct)																	1.29	0.06	0.77	0.07
Working capacity:																				
- Partly incapable (vrs Fully capable)																	0.94	0.66	0.94	0.66
- Fully incapable (vrs Fully capable)																	0.40	0.00	0.39	0.00
- Item non-response (vrs Fully capable)																	0.29	0.00	0.29	0.00
Interviewer level infomation:																				
Interviewer characteristics:																				
Age group of the interviewer:																				
- 25 - 39 years (vrs 40 - 59)	0.51	0.04	0.44	0.01	0.80	0.027	0.747	0.016												
- 60+ years (vrs 40 - 59)	7.60	0.02	10.10	0.05	0.78	0.111	0.783	0.122												
Interviewer perception survey information:																				
Impact of bad health condition or illness of the target person																				
- 'In large amout or considerably' (vrs 'Not at all or in small amount')					0.80	0.03	0.80	0.02												
- 'Item missing information' (vrs 'Not at all or in small amount')					0.78	0.11	0.77	0.10												
Interviewer level random effects:	Est. Variance[1]	p-value[2]	Est. Variance[1]	p-value[2]	Est. Variance[1]	p-value[2]	Est. Variance[1]	p-value[2]	Est. Variance[1]	p-value[2]	Est. Variance[1]	p-value[2]	Est. Variance[1]	p-value[2]	Est. Variance[1]	p-value[2]	Est. Variance[1]	p-value[2]	Est. Variance[1]	p-value[2]
Random intercepts	1.29	0.00	0.65	0.01	0.09	0.00	0.08	0.02	0.19	0.01	0.37	0.00	0.13	0.04	0.05	0.48	0.05	0.09	0.04	0.36

[1] Estimated variance of the interviewer level random intercept

[2] p-value based on the t-test derived upon the variance estimates and their estimated standard errors assuming normality

5.4.3 Comparison of estimation methods used for predicting response probabilities with multilevel binomial logistic regression models

The previously estimated sequential models, presented in Table 5.4, used two alternative estimation methods. The residual maximum likelihood methods (REML) are available in mainstream software, but they are generally considered to provide biased estimates. Goldstein et al. (2002) showed the second order penalised quasi-likelihood (PQL) being more accurate than REML. We find that the improvements provided by the second order PQL in comparison to REML are relatively small. When comparing the estimates between methods, their significance and non-significance are relatively stable. The largest departures can be observed in the impact on predicted odds by the age group of the interviewer for contacting and the cognitive capability of the inidividual at the likelihood of returning the self-completion questionnaires.

From analysts' point of view, the crucial feature of an estimation method is the unbiasedness. For example, 1st order penalized quasi-likelihood has been shown, unless corrections are added, to yield biased results for binary outcomes in some circumstances (Breslow and Clayton, 1993; Breslow and Lin, 1995; Rodríguez and Goldman, 1995). The 2nd order PQL estimation method developed by Breslow and Clayton (1993) and implemented by Goldstein (1995) for the MLwiN software, has been considered to improve the estimation like the Markov chain Monte Carlo methods (MCMC) (Hox, 2002; Casella and George, 1992; Gelfand and Smith, 1990). To improve the estimates of the multilevel logit models reducing their bias one can use bootsrap methods (Goldstein, 1998) or the MCMC methods (Browne, 1998).

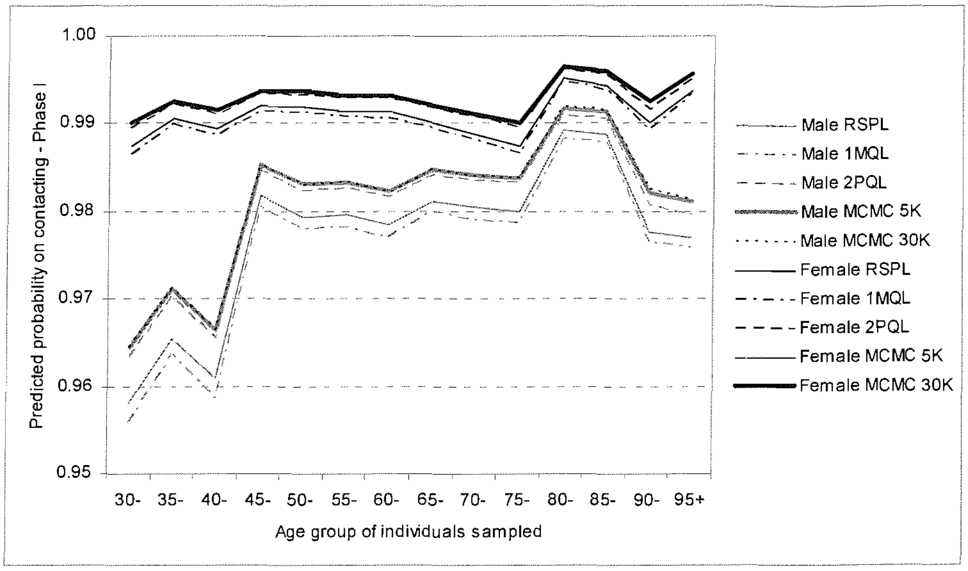
We explore the sensitivity of the model predictions by estimation methods on survey participation of individuals within the interviewer assignments. In Figure 5.3, we compare different estimation methods at the initial two data collection steps for the model (5.4). The mean of the predicted probability to contact and response to the health interview are assessed by sex and five-year age group of the individuals sampled. The differences between 2nd order PQL estimates and the MCMC estimates are minimal. The predictions of these estimation methods deviate largely only for the eldest males, which is a relatively small group. When estimating likelihood of contacting, the RSPL12 and 1st order MQL estimation methods give clearly downward biased estimates. For the probability to respond to health interview, the difference between the estimation methods is smaller. The predicted probabilities for contacting estimated by the REML method are close to those given by the RSPL in Figure 5.3.a. Differences by estimation methods are narrower for health

[12] SAS proc glimmix provides an estimation method of "residual pseudo-likelihood with a subject specific extension" abbreviated by (RSPL) which is a default option for these type of models in SAS proc glimmix. However, modelling the survey participation of Health 2000, the RSPL estimation method converges only for the model on contacting. Thus we have used REML in %glimmix macro and compared the results with 1st order MQL, 2nd order PQL and MCMC estimation in MLwiN.

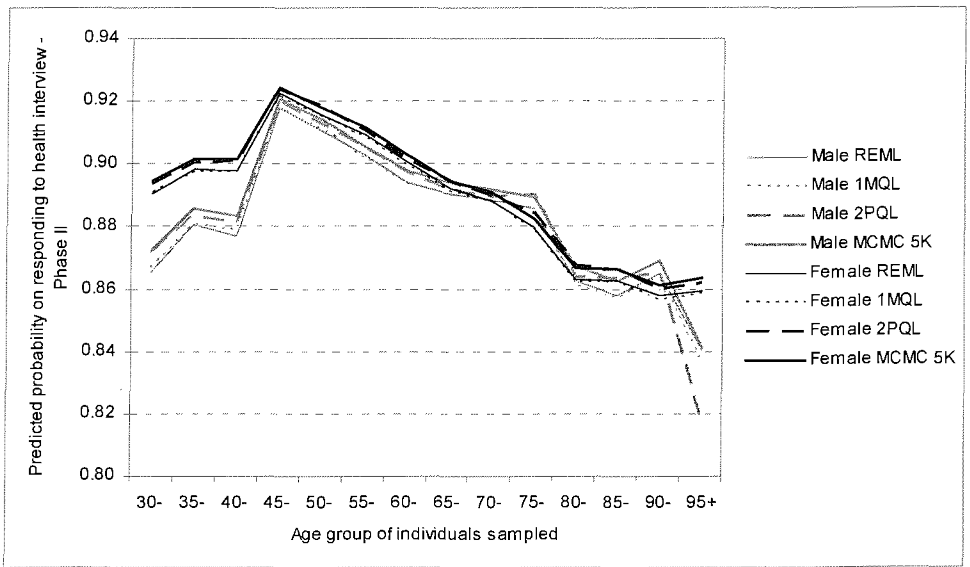
interview phase (Figure 5.3.b) and symptom interview (Figure 5.3.c). Also the model fit of these latter phases is higher than for the model of contactability.

Figure 5.3 Predicted probability of fieldwork success by age, sex and estimation method

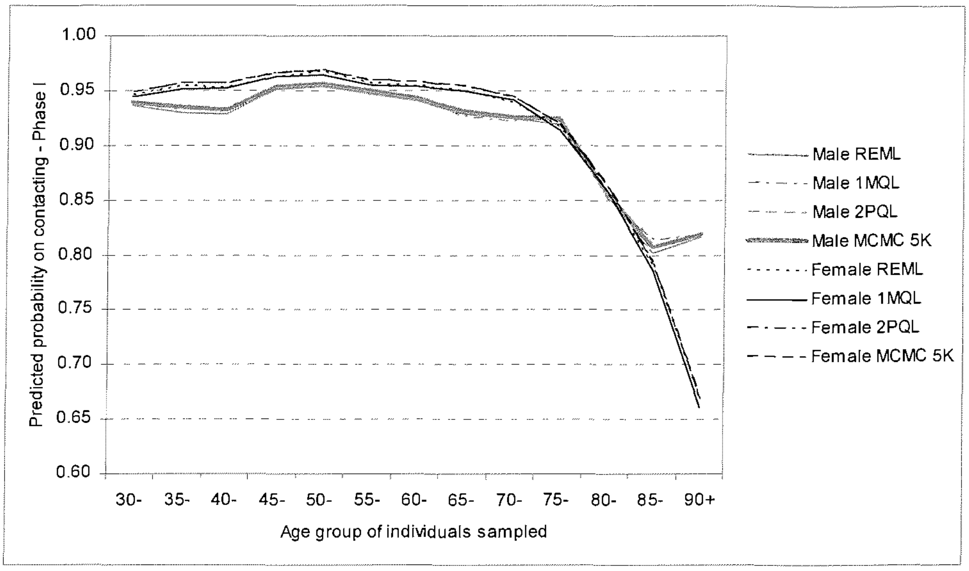
(a) Contacting phase



(b) Health interview phase



(c) Symptom interview phase



5.4.4 Assessing the impact of random intercepts across data collection phases

In previously reported survey non-response studies using the multilevel models, most attention of the interpretation is on the impact of explanatory variables and the interviewer level effects. However, we are intrigued to study further the meaning and impact of the interviewer level random intercept and in later models those of the random slopes. We aim to assess the odds ratios of the random effects in-depth. Deriving the odds ratios of the random effects enable us to interpret, how the random effects affect and vary across the interviewer assignments on the predicted probabilities of the individuals together with the odds ratios of the explanatory variables. Larsen et al. (2000) note that the odds ratios are unobserved random variables for which distributional characteristics can be reported. We will thus examine the distributional aspects of random coefficient using the median odds ratio (MOR), and study the random slopes using the MOR and interval odds ratios (IOR), developed by Larsen et al. (2000).

Traditionally the odds ratio for a variable l, which denotes to the any of the explanatory variables $l = 0,1,\ldots,p$, can be written as:

$$OR_l = \exp(\beta_l). \tag{5.5}$$

The odds ratio for the random intercept at an interviewer level, depends also on the random component, which is by definition normally distributed. Thus, based on the model (5.4) with the random intercept, the odds ratios of intercepts are defined as:

$$OR_{0kl} = \exp(\beta_{l0})\exp(u_{kl0}). \tag{5.6}$$

For a model containing a random slope for a variable l, we can derive the distribution for the odds ratios of the random slope variable. The odds ratio of the random slope at data collection phase t depends on the estimated fixed effect β_{tl} and the interviewer level random slope u_{ktl} for the variable l as follows:

$$OR_{ktl} = \exp\left(\beta_{tl}\right)\exp(u_{ktl}). \tag{5.7}$$

This will be applied later for the model (5.13). Let us note the distribution function of OR_{kt} by

$$F(z)_{tl} = P\left(OR_{ktl} < z\right).$$

This will be the same for all interviewers k, but will depend on t. Under the model in (5.4) we have:

$$F(z)_{tl} = 2\Phi\left(\frac{\log(z)}{\sqrt{2\sigma^2_{u_{tl}}}}\right) - 1,$$

where z represents the percentile points of the cumulative normal distribution. Subsequently, the normal distribution function can be written as:

$$f\left(z\right)_{tl} = \frac{\delta}{\delta z}F\left(z\right)_{tl} = \frac{1}{z}\sqrt{\frac{2}{\sigma^2_{u_{tl}}}}\varphi\left(\frac{\log(z)}{\sqrt{2\sigma^2_{u_{tl}}}}\right). \tag{5.8}$$

MOR is the median of this distribution so it is calculated by solving (Larsen and Merlo, 2005):

$$F(z)_{tl} = 1/2$$

$$2\Phi\left(\frac{\log(z)}{\sqrt{2\sigma^2_{u_{tl}}}}\right) - 1 = \frac{1}{2}$$

$$\left(\frac{\log(z)}{\sqrt{2\sigma^2_{u_{tl}}}}\right) = \Phi^{-1}\left(\frac{3}{4}\right)$$

$$z = \exp\left(\sqrt{2\sigma^2_{u_{tl}}}\,\Phi^{-1}\left(\frac{3}{4}\right)\right).$$

Thus, the MOR for random intercepts at each data collection phase t can be defined as (Larsen et al., 2000):

$$MOR_t = \exp\left[\sqrt{2\sigma^2_{u_{tl}}}\,\Phi^{-1}\left(3/4\right)\right] \tag{5.9}$$

where $\Phi^{-1}(\cdot)$ is the cumulative distribution function for the standard normal distribution.

Distribution of random effects

We wish to examine the impact of the random slope with the fixed effect of the same variable. The distribution of the random effects is by model assumptions allowed to vary within the normal distribution $N(0,\Omega)$. As by definition the $\hat{\pi}_i$ is constrained to be within (0, 1), the log of the odds of survey response, $\text{logit}\{\pi_{ik}\}$, can be of any real value. The estimated values of the fixed intercept, fixed effect and the random effect all contribute to the distribution of the odds of the random effects. Following the MOR definition by Larsen et al. (2000) and Larsen and Merlo (2005), we can generalise the distribution of the odds ratios for the variables containing random effects in addition to the fixed effects as:

$$f(z)_{\text{OR}} = \exp\left(\hat{\beta}_l + \sqrt{2\sigma^2_{u_{kl}}}\,\Phi(z)\right) \tag{5.10}$$

where z refers to the percentile from the normal distribution. Examining the tail areas of the distribution of the odds of the random effects, we can find indication about the magnitude of variation in the fieldwork performance across different type of interviewers, controlling for the individual level random slope variable.

In the lines of the definition of MOR and IOR for the fixed effects in Larsen et al. (2000), we define the MOR and IOR for the random slopes as:

$$MOR_l = \exp\left[\beta_{tl} + \sqrt{2\sigma^2_{u_{tl}}}\,\Phi^{-1}(1/2)\right], \tag{5.11}$$

where the γ_h and u_h denote for the fixed and random effect of the micro level variable as defined in (5.13). The MOR for the random slope variable can also be denoted by IOR_α, where $\alpha = 0.5$ refers to the median. Thus the definition of the MOR can be generalised for the other values of the distribution as follows. The IOR for the impact of the fixed micro level variable h with the random slopes is:

$$IOR_{(\alpha)tl} = \exp\left(\left|\beta_{tl} + \sqrt{2\sigma^2_{u_{tl}}}\,\Phi^{-1}(\alpha)\right|\right). \tag{5.12}$$

$\Phi^{-1}(\cdot)$ is the cumulative distribution function for the standard normal distribution, and the α is a continuous variable between 0 and 1 referring to the percentile or decile groups.

Interpretation of the random intercepts

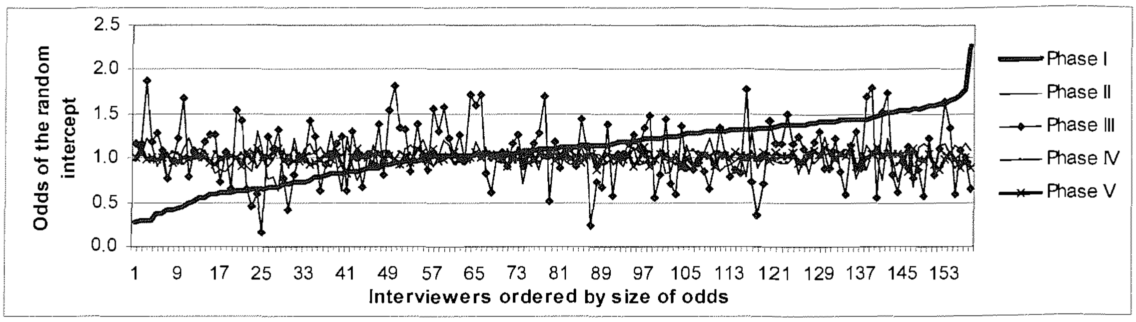
In Table 5.5, we present the MOR for random effects with other distributional statistics that we believe are more informative for assessing the overall effect of random intercepts in multilevel models. It can be observed that the variation by the decile grouping indicates more differences between estimation methods and across data collection phases than the levels of MOR would indicate. The minimum and maximum values for random intercepts are observed from the distribution of the interviewer level random intercepts of the estimated models. The corrected sum of squares indicates the sum of the deviation from the mean for each random intercept. This summary measure has been derived unweighted, as preliminary analysis showed that weights based on the interviewer assignment size did not have a significant impact on the distribution of the interviewer level random effects. In Figure 5.4, we show the distribution of the odds for the interviewer random intercept, in which the interviewer assignments are ordered by the size of the odds of the random co-efficient in the x-axis. In addition, the distributions of the random intercepts across the phases are not similar. For example, the random intercepts for the model of contacting phase has the strongest increase in the upper tail. As the weighted and unweighted odds of the random intercepts fluctuate randomly, this can indicates that there is no significant association between the size of the random intercepts and the size of the interviewer assignments.

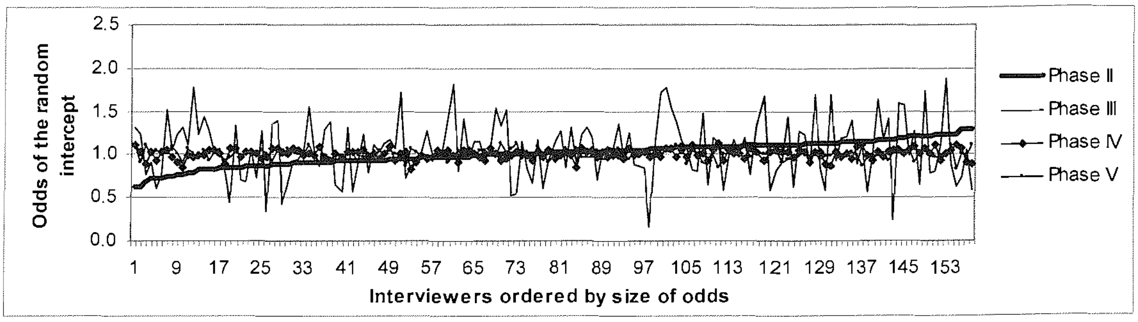
In Figure 5.5 we have compared the random intercepts by REML and second order PQL estimation methods across data collection phases. It can be seen clearly that the variability of the interviewer random intercepts reduces by the data collection phases in our data. Interviewer effects can be observed at the contacting and health interview phases that are administrated by the interviewers. In addition, the random intercepts vary at the subsequent stage, individuals participating to the symptom interview in which the motivation of the interviewer and the intermediate survey experience can have strong impact and can be projected through the interviewer random intercept. However, there is a strong reduction in the interviewer effect after the symptom interview, i.e. in participation to the health examination and returning the self-completion questionnaires. This reduction is even stronger with second order PQL than with REML estimation. Interviewers were not present at these stages, and it seems that the survey experience does not reflect this far in the data collection in Health 2000 data. Thus in the Health 2000 data there is no indication that the success of within the interviewer assignment in the previous data collection phase would significantly impact the success at the following, when controlling for the explanatory variables in the models.

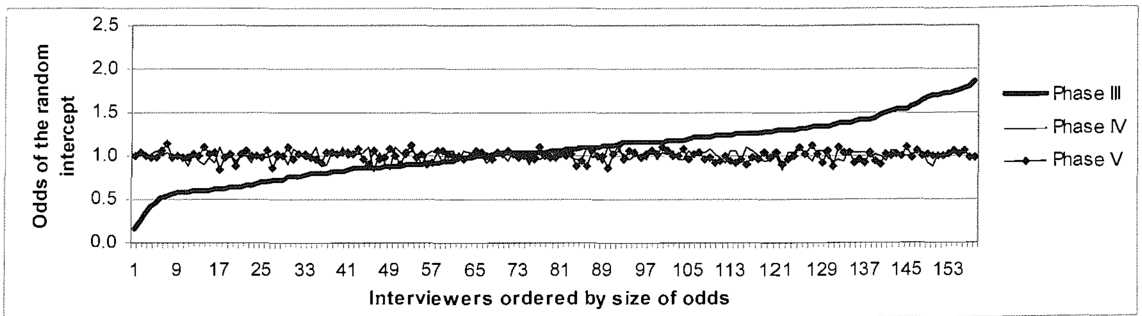
Table 5.5 Distribution of odds of random intercepts in multilevel sequential survey participation models

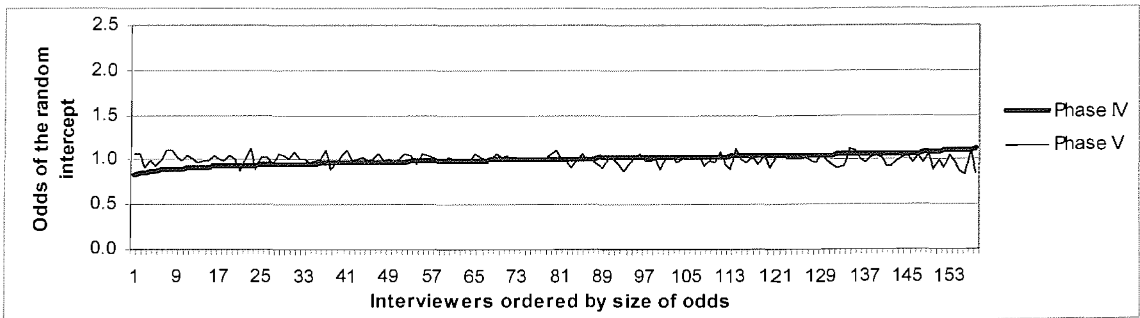
	Phase I: Contacting		Phase II: Health Interview		Phase III: Symptom Interview		Phase IV: Health Examination		Phase V: Self-completion questionnaires	
	REML	2nd order PQL	REML	2nd order PQL	REML	2nd order PQL	REML	2nd order PQL	REML	2nd order PQL
Decile:										
10th %	0.32	0.59	0.81	0.83	0.69	0.61	0.81	0.92	0.88	0.92
20th %	0.52	0.72	0.89	0.90	0.84	0.77	0.88	0.95	0.95	0.96
30th %	0.76	0.92	0.92	0.94	0.90	0.87	0.94	0.97	0.97	0.98
40th %	1.08	1.03	0.98	0.99	1.00	0.98	0.98	0.99	0.99	1.00
50th %; MOR	1.24	1.10	1.03	1.02	1.04	1.06	1.01	1.01	1.01	1.00
60th %	1.41	1.19	1.06	1.05	1.09	1.15	1.06	1.02	1.02	1.02
70th %	1.72	1.30	1.11	1.09	1.15	1.23	1.10	1.04	1.03	1.03
80th %	1.94	1.40	1.14	1.12	1.20	1.33	1.14	1.05	1.07	1.05
90th %	2.20	1.53	1.21	1.17	1.33	1.53	1.19	1.07	1.09	1.07
Min	0.07	0.27	0.58	0.62	0.53	0.15	0.63	0.82	0.78	0.83
Max	3.70	2.27	1.35	1.30	1.53	1.87	1.30	1.13	1.18	1.13
Mean	1.26	1.08	1.01	1.01	1.03	1.06	1.01	1.00	1.00	1.00
CV	56.47	33.37	15.59	13.50	22.28	31.97	14.61	5.81	7.70	5.69
Corrected SS	78.98	20.36	3.92	2.92	6.38	18.12	3.43	0.53	0.94	0.51

Figure 5.4 Ordered odds ratio of the random intercept by interviewer assignments and the data collection phase (2nd order PQL estimation method)









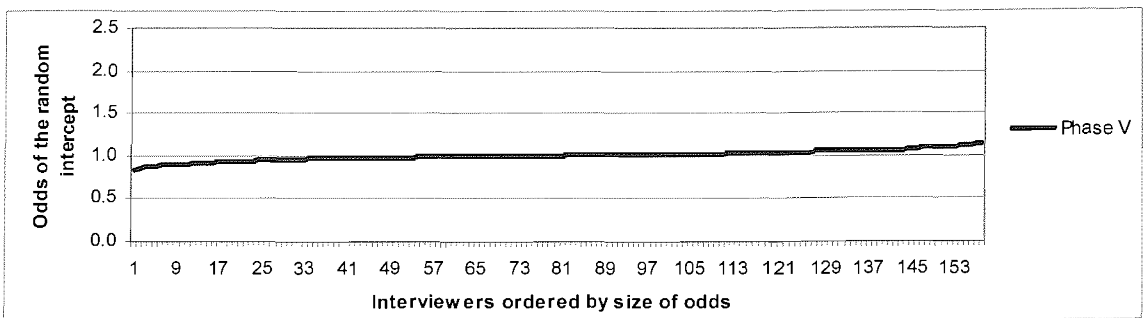
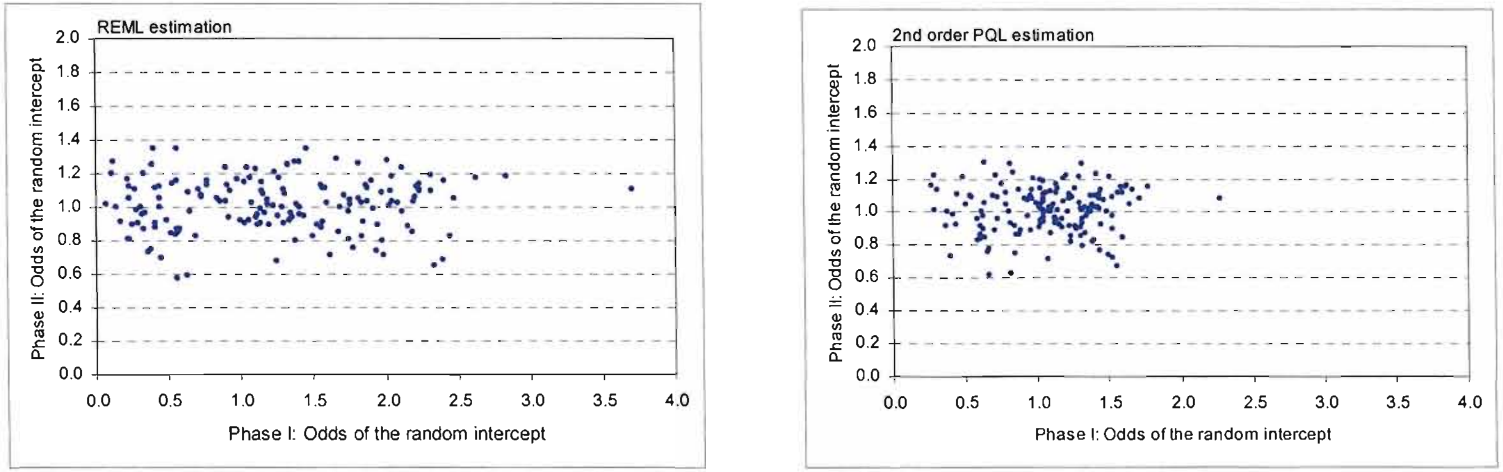
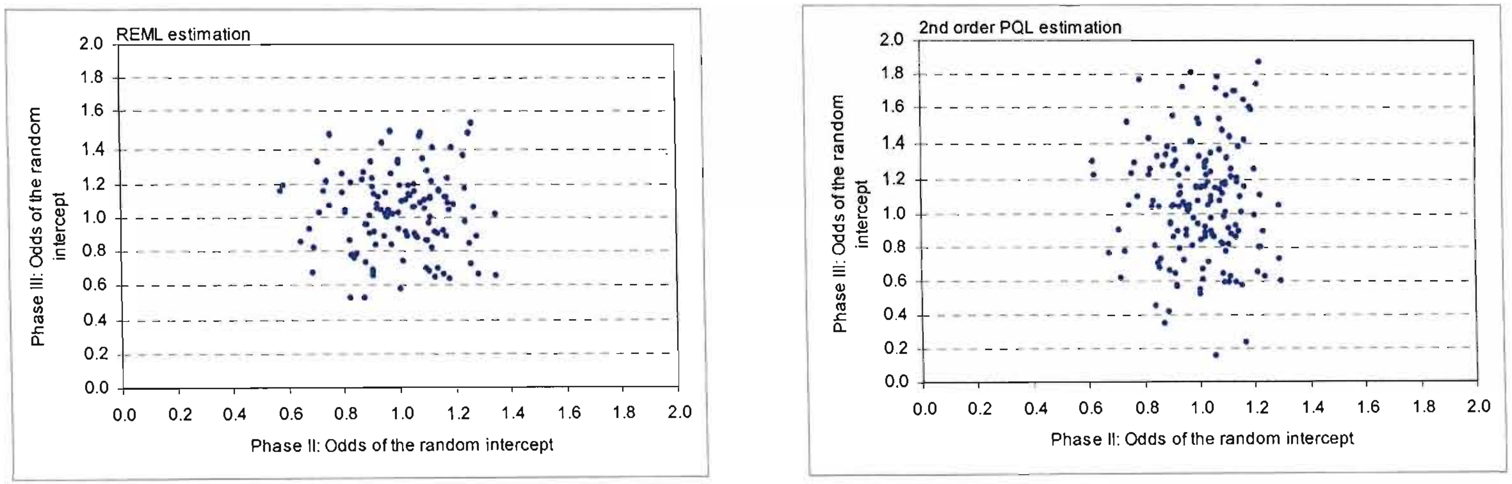


Figure 5.5 Estimated odds ratio of the random intercepts for interviewer assignments across data collection phases by REML and 2nd order PQL estimation methods

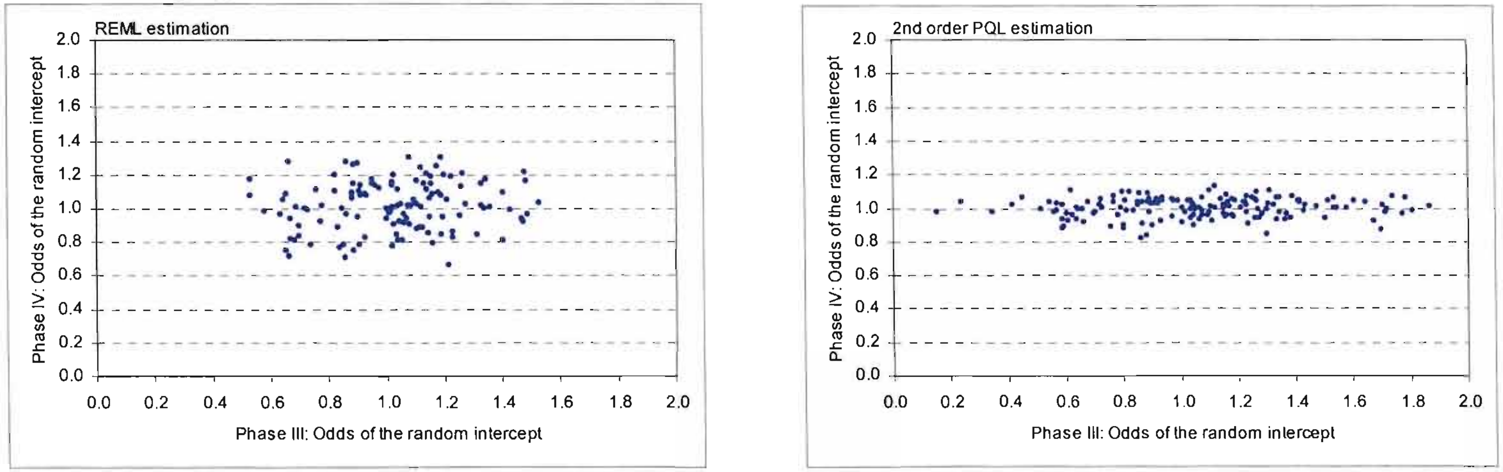
(a) Phase I Contacting versus Phase II Health interview



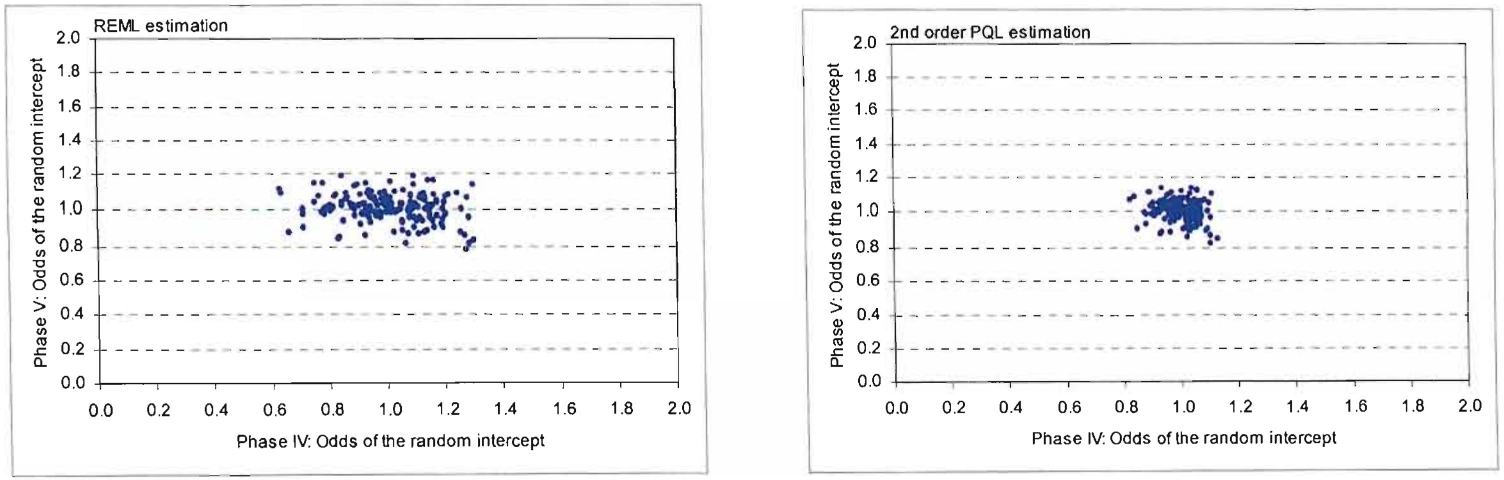
(b) Phase II Health interview versus Phase III Symptom interview



(c) Phase III Symptom interview versus Phase IV Health examination



(d) Phase IV Health examination versus Phase V Self-Completion



Interpretation of the remaining interviewer level heterogeneity in the model

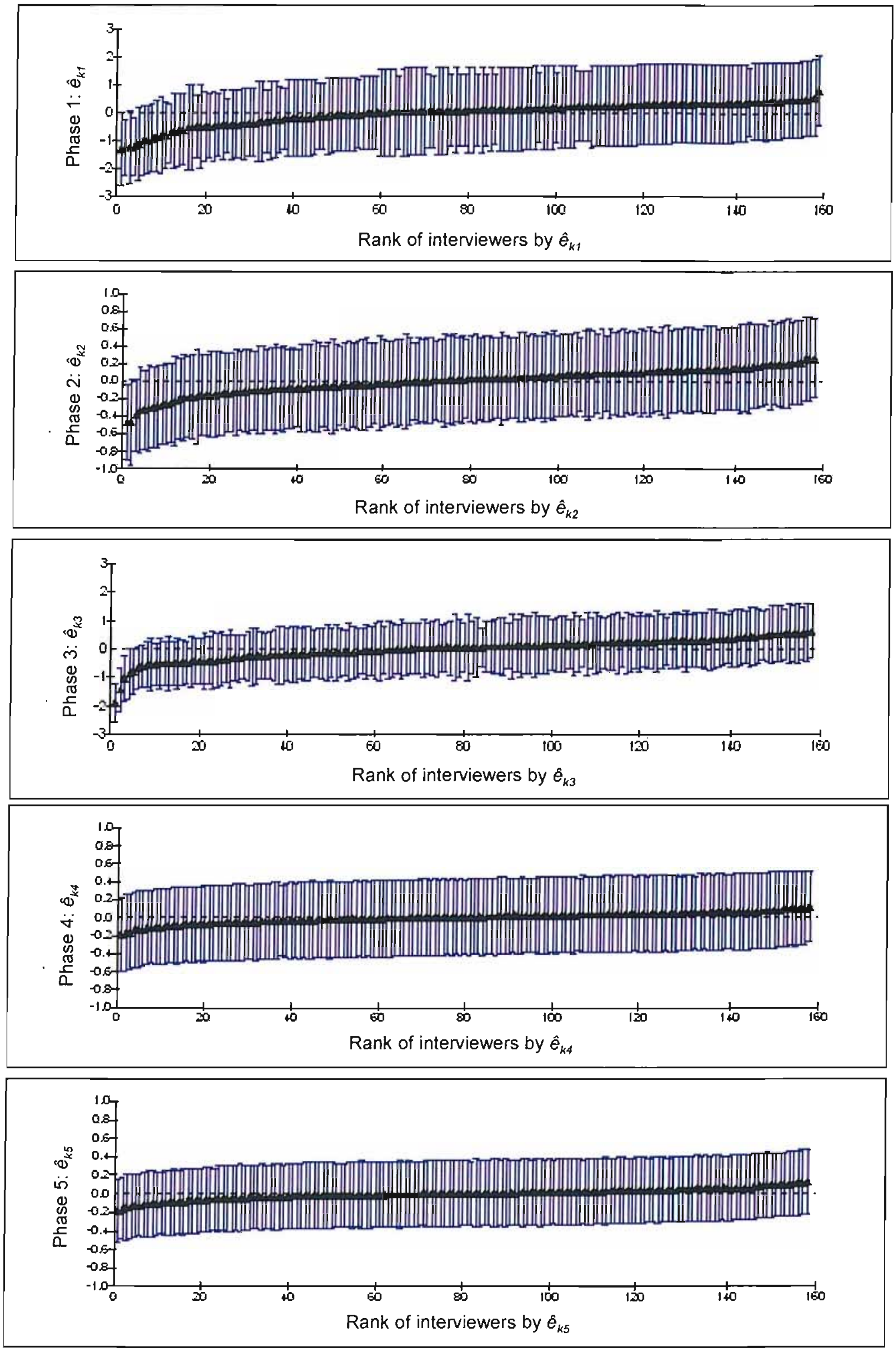
This section aims to analyses the heterogeneity at the interviewer level that is not explained by the sequential multilevel models. We examine the unexplained interviewer heterogeineity i.e. 2nd level residuals. This will allow us also to check whether there are 'outperforming' assignments that have relatively high or low outcome rates, after adjusting for differences in attributes of the individuals sampled and the interviewers. The interviewer-level residual estimates are ranked by their size to examine further the extent of variation across interviewer assignments, presented in Figure 5.6. Comparing the performance of the sequential multilevel logistic models, the Phase I on contactability and the Phase III model on symptom interview seem to have the highest unexplained heterogeneity. In the former there are few assignments that have larger negative residuals.

At the first three data collection phases there are few interviewer assignments having relatively low negative residuals in comparison to the distribution of the others. Using the qualitative data from the interviewer perception survey, we can examine whether these interviewers differ have some generic perceptions and experiences that cannot be quantified but have been communicated via the open ended questions.

Variation of predicted probabilities of survey participation across data collection phases

The sequential modelling is very informative on the development of the survey participation across data collection phases. The distributions of the predicted probabilities vary by explanatory factors, indicating both the risk factors for each stage. There is a significant variation across reference groups, for example, by demographic or socio economic characteristics. Conditional on being contacted, the elderly have the down biased distribution of predicted probabilities in comparison to individuals aged less than 80 years, as can be observed from Figure 5.7. The other risk groups are presented in the Appendix 5.4 with their reference groups. Men, people with foreign maternal language, single people and people with low income all have reduced level of co-operation at all data collection phases, but generally also wider range of predicted response probabilities.

Figure 5.6 Rank of estimated interviewer level residuals by data collection phases

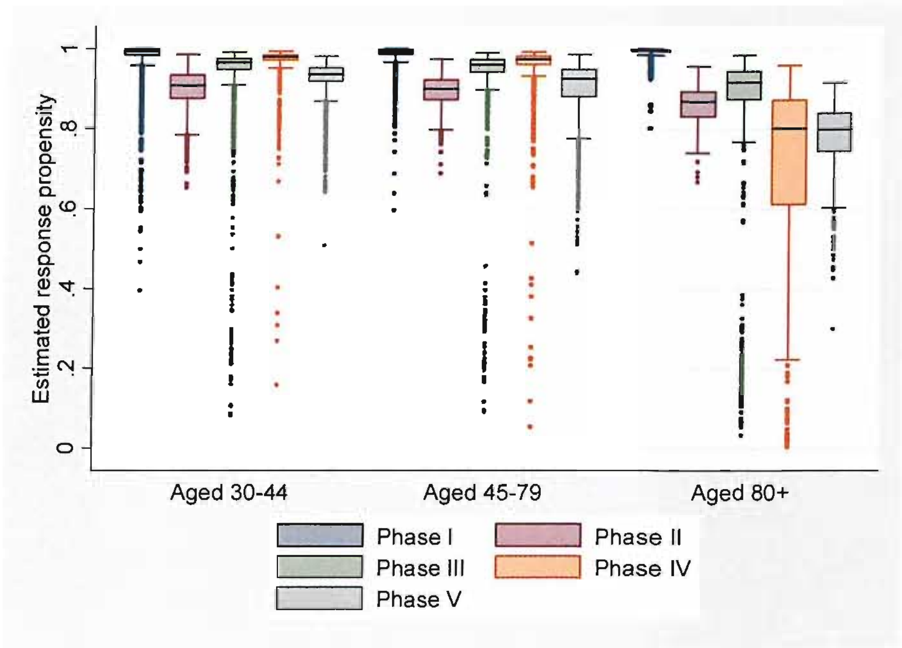


NOTE 1: The vertical bands represent the 95% confidence limit of the estimated residuals for interviewer assignments.

NOTE 2: The residuals are calculated per interviewer assignment and they do not adjust for the number of interviewers specifically, also the overall estimated multilevel model is unweighted. Our preliminary data analysis has shown that weighting by the interviewer assignment size did not improve the prediction power or the performance of the response probability models.

NOTE 3: 2nd order PQL has been used in MLwiN for estimating the sequential multilevel models with random intercepts.

Figure 5.7 Predicted probabilities of sequential co-operation by age group and data collection phases



5.5. Predicting response probabilities in the presence of random effects of interviewer attributes

In this Section, we explore whether the effects of the explanatory variables obtained at individual level have varying effects at interviewer level at any of the data collection phases. In other words, the model in (5.4) is extended by allowing the interviewers' success and ability to contact and gain co-operation to vary across population sub-groups. This extension of the model implies that the predicted probability of success can contain random effects at the interviewer level that depend on the individual level characteristics. These random effects are estimated across the interviewer assignments, with the restriction that they are normally distributed.

The random effects of the interviewer assignments can now be divided into the random intercept and the random slopes that may have varying values of the individual level characteristics. The random intercept is denoted by u_{0kt}, and the random slopes are denoted by $u_{kt}x_{ik}$. The β_h coefficients of the logistic regression model consists of the average coefficient γ_{0t} for the individual level variables x_{hik}, as well as the group level variables x_{hk} with random effects for the group dependent deviations denoted by u_{hkt}. Thus, the coefficient for individual level variable can be written as $\beta_{ht} = \gamma_{hikt} + u_{hkt}$ and the extended model is written as follows:

$$\pi_{ikt} = \Pr\left(R_{ikt} = 1 \middle| R_{ik(t-1)} = 1\right)$$

$$\operatorname{logit}\left\{\pi_{ikt}\right\}=\gamma_{0t}+\sum_{h=1}^{l_t}\gamma_{ht}x_{hik}+\sum_{h=l_t+1}^{p_t}\varphi_{ht}x_{hk}+u_{0kt}+\sum_{h=1}^{q_t}u_{hkt}x_{hik} \tag{5.13}$$

$$\begin{pmatrix} u_{0kt} \\ \vdots \\ u_{qkt} \end{pmatrix} \sim N\left(0,\Omega_u\right)$$

$$\Omega_u=\begin{bmatrix} \sigma^2_{u_0t} & & & \\ \sigma_{(u_0,u_1)t} & \sigma^2_{u_1t} & & \\ \vdots & \vdots & \ddots & \\ \sigma_{(u_0,u_q)t} & \sigma_{(u_1,u_q)t} & \cdots & \sigma^2_{u_qt} \end{bmatrix}$$

where the structure of the model follows closely the model defined in (5.4). p_t indicates the number of all explanatory variables and l_t indicates the number individual level variables; γ_{0t} is an intercept, and u_{0kt} is the group dependent deviation for each interviewer assignment k at data collection phase t. The major difference between the models (5.4) and (5.13) is that the latter contains q_t individual level variables that have both a fixed effect and a random effect varying significantly by the interviewer assignment[13]. If there are significant differences across interviewer assignments in gaining co-operation in sub-population groups these can be defined with the q variables. The variance of the random intercept is denoted by $\sigma^2_{u_0t}$, and the variance of the random slopes by $\left[\sigma^2_{u_1t},\ldots,\sigma^2_{u_qt}\right]$. The covariance of the variance terms are displayed below the diagonal in the variance matrix Ω_u.

In our data analysis we have searched for the possible random slopes regarding all individual level variables that have been significant in the fixed part of the model. The existence of such a factor indicates that the interviewers would have different tendencies in achieving success with individuals with different backgrounds or social conditions, for example. In addition, it can also indicate that the interviewer assignments may differ in their composition unintentionally by some factors that seem to be significant in relation to the survey participation. The model (5.13) can be divided into a fixed part $\gamma_{0t}+\sum_{h=1}^{l_t}\gamma_{ht}x_{hik}+\sum_{h=l_t+1}^{p}\varphi_{ht}x_{hk}$ and a random part $u_{0kt}+\sum_{h=r+1}^{q}u_{hkt}x_{hik}$.

If the individual level information would be scarce or not available it could be useful to test whether survey participation models could be improved by using also the interviewer level information for random effects:

[13] The models could also be defined for regional clustering replacing the k by j, which denotes for regional local areas. This type of models are commonly constructed in the literature to take into account the clustering of individuals by primary sampling units that are often defined by regional clustering (see e.g. Chapter 2).

$$\text{logit}\{\pi_{it}\} = \gamma_{0t} + \sum_{h=1}^{l_t} \gamma_{ht} x_{hik} + \sum_{h=l_t+1}^{p} \varphi_{ht} x_{hk} + u_{0k} + \sum_{h=1}^{q_t} u_{hkt} x_{hik} + \sum_{h=q_t+1}^{g} u_{ht} x_{hk} . \qquad (5.14)$$

We presume that, especially, when rich information resources is available both at individual and at any macro level (such as the interviewer level), the most critical after the model specification is to make full use of the auxiliary data by looking into random effect associations of micro level data at macro level. If the macro level variables have macro level random effects, this would practically only mean that the macro level residual is divided into two parts: normally distributed residuals defined as a function of some macro level variables, and the remaining unexplained residuals. This would mean that part of the random part in the multilevel model could actually be regarded as intermediate, normalised, part that is restricted to follow normal distribution. For model interpretability, random effects defined with the micro level variables are more interesting than the associations of macro level random effects at macro level.

Interpretation of the random intercepts and random slopes for interviewer assignments

We have explored random effects for all covariates used in the sequential logit models for each data collection phase. There was some evidence on differences between interviewer assignments in their work performance at the contacting phase and at the health interview. In addition, we have found some differences in success of motivation the respondents to progress to the symptom interview, which was been held after the health interview at the mobile health examination centres. We did not detect any significant random slopes to be observed in the Health 2000 data for participating to the health examination or returning the self-completion questionnaire. However, using the Health 2000 survey participation data, we found that the second order PQL estimation method was more conservative with detecting random slopes than the REML method. Thus we have compared multilevel sequential logit models with random effects that were estimated with both of these estimation methods, presented in Table 5.6.

In our models, the significant random slopes are detected at micro level. Significant differences were observable by sex of the individuals at the contacting phase, when using the REML estimation. On the other hand, this finding must be treated with specific caution as the model did not converge with the second order PQL estimation method. Instead, the estimation methods give confounding predictions at the phases of health and symptom interview. Interviewers' success in obtaining health interviews varied significantly across individuals belonging to families with children. The model suggests also that interviewers differ on how successfully they were able to direct those with low income to the symptom interview, conditional upon obtaining the health interview.

Table 5.6 Survey participation with interviewer level random effect and random slopes

Estimated parameters	Phase I: Contacting[1] REML estimation Odds Ratio	P-value	Phase II: Health Interview REML Odds Ratio	P-value	Phase II: Health Interview 2nd order PQL Odds Ratio	P-value	Phase III: Symptom Interview REML Odds Ratio	P-value	Phase III: Symptom Interview 2nd order PQL Odds Ratio	P-value
Intercept	220.7	0.00	10.4	0.00	10.8	0.00	20.4	0.00	22.2	0.00
Individual characteristics:										
Age group:										
- Age 30 - 44 years (vrs 45 - 79)	0.6	0.00	0.6	0.00	0.7	0.00	0.7	0.00	0.7	0.02
- Age 80+ years (vrs 45 - 79)	2.8	0.00	0.9	0.19	0.9	0.19	0.7	0.01	0.7	0.03
Female (vrs male)	3.5	0.00	1.1	0.13	1.1	0.14	1.2	0.09	1.2	0.13
Maternal language:										
- Swedish (vrs Finnish or Sámi)	0.7	0.39	0.9	0.49	0.7	0.08	0.9	0.84	1.0	0.90
- Other (vrs Finnish or Sámi)	2.4	0.00	1.1	0.19	0.7	0.24	0.2	0.00	0.2	0.00
Family status:										
- Family with children (vrs families without children)	2.4	0.00	1.6	0.00	1.6	0.00	1.1	0.64	1.1	0.70
- Single person household (vrs families without children)	1.0	0.91	0.9	0.10	0.9	0.11	0.6	0.00	0.6	0.00
Further education (vrs basic education only)	0.7	0.01	1.3	0.00	1.3	0.00	1.7	0.00	1.7	0.00
Income quintiles of register derived disposable income:										
- 1st quintile or no income (vrs 2nd - 4th quintile)	0.7	0.01	0.9	0.49	0.9	0.52	1.0	0.86	1.0	0.88
- 5th quintile (vrs 2nd - 4th quintile)	0.9	0.73	1.1	0.19	1.1	0.24	1.3	0.08	1.3	0.14
Received capital income (vrs none) (household level)	1.9	0.00	1.2	0.01	1.2	0.02				
Received income support (vrs none) (household level)	0.4	0.00								
Private pension payments (household level)			1.8	0.00	1.8	0.00				
Type of living unknown	0.2	0.00					0.4	0.00	0.4	0.01
Region:										
Urban (vrs rural or other less densely populated)	0.3	0.00	0.7	0.00	0.7	0.00				
Health interview information of the individual:										
Body mass index (BMI):										
- Information not obtained in health interview							0.0	0.00	0.0	0.00
- BMI < 18 (vrs 18 <= BMI < 35)							0.2	0.00	0.2	0.00
- BMI = 35+ (vrs 18 <= BMI < 35)							1.4	0.20	1.4	0.27
Interviewer level infomation:										
Interviewer characteristics:										
Age group of the interviewer:										
- 25 - 39 years (vrs 40 - 59)	0.5	0.02	0.7	0.01	0.7	0.01				
- 60+ years (vrs 40 - 59)	9.7	0.01	0.8	0.14	0.8	0.13				
Interviewer perception survey information:										
Impact of bad health condition or illness of the target person on refusals on general, increased:										
- 'In large amout or considerably' (vrs 'Not at all or in small amount')			0.8	0.03	0.8	0.02				
- 'Item missing information' (vrs 'Not at all or in small amount')			0.8	0.11	0.8	0.10				
Interviewer level random effects:										
Random intercept	5.9	0.00	1.1	0.00	1.11	0.01	1.6	0.00	1.6	0.00
Gender of individual	15.0	0.00								
Families with children			1.4	0.02	1.42	0.03				
Income quintiles of register derived disposable income:										
- 1st quintile or no income (vrs 2nd - 4th quintile)							3.0	0.00	2.3	0.02

[1] The multilevel model with random intercept and random slope did not converge with the second order PQL estimation method

Technical note: The model has been estimated using 2[nd] order PQL estimation method in MLwiN

Assessment of the heterogeneity of random intercepts and slopes

Our aim is to apply methods on the heterogeneity measures based on the odds ratios for random intercepts and slopes. The variation of the random effects is explored by the variation of the odds ratio of the random effects. Previously, we have derived the median odds ratio (MOR) and interval odds ratio (IOR) for the random intercepts in **(5.8)**. In this section, we present how the impact of fixed and random effect introduced by the random slope can be assessed jointly.

The heterogeneity measures of odds ratios for the random intercepts and slopes of interviewers from a model assessing the co-operation of individuals within interviewer assignments are presented in Table 5.7. The variation of the random intercepts and random slopes is relatively large. Larsen et al. (2000) recommend using an 80% interval and the median for measuring the impact of the random slopes. Clearly, reporting solely the MOR and IOR, the distributional information given would be limited in informing about the impact and variation of the random slopes. We have already previously emphasized the importance to assess the full distribution as both the median, and the full range of the values are important to assess. The strongest impacts of the random effects lay on the tails of the odds ratios.

Table 5.7. Summary of heterogeneity of interviewer level random effects from single covariate models of survey participation

	Phase I: Contacting[1]		Phase II: Health Interview				Phase III: Symptom Interview			
	REML estimation		REML		2nd order PQL		REML		2nd order PQL	
	Random intercept	Random slope	Random intercept	Random slope	Random intercept	Random slope	Random intercept	Random slope	Random intercept	Random slope
Decile:										
10th %	0.24	0.23	0.76	0.67	0.80	0.69	0.57	0.46	0.62	0.57
20th %	0.42	0.78	0.84	0.79	0.88	0.80	0.72	0.65	0.77	0.76
30th %	0.60	0.84	0.92	0.87	0.94	0.88	0.84	0.81	0.89	0.84
40th %	1.00	0.88	0.97	0.93	0.98	0.94	0.95	0.87	0.98	0.91
50th %	1.29	0.93	1.02	1.01	1.02	1.01	1.10	1.01	1.09	0.99
60th %	1.66	0.99	1.05	1.10	1.06	1.08	1.20	1.16	1.15	1.10
70th %	2.07	1.17	1.14	1.19	1.11	1.17	1.35	1.33	1.28	1.16
80th %	2.30	2.23	1.19	1.29	1.15	1.26	1.53	1.50	1.40	1.34
90th %	2.69	3.55	1.25	1.41	1.20	1.36	1.68	1.88	1.52	1.59
Min	0.04	0.06	0.56	0.45	0.63	0.48	0.10	0.18	0.11	0.29
Max	5.61	9.81	1.53	2.02	1.42	1.91	2.24	9.26	1.98	7.52
Mean	1.42	1.52	1.01	1.04	1.01	1.03	1.10	1.20	1.07	1.12
CV	70.34	101.36	18.64	28.23	15.29	25.88	39.16	78.86	33.09	63.32
Corrected SS	156.58	371.59	5.62	13.57	3.76	11.26	29.36	140.37	19.78	78.85

From Figure 5.8 we can observe the whole variation of the impact of the random slope variables in graphical presentation. In Figure 5.9 we demonstrate that for categorical random slope variables the graphical display is significantly more informative indicating the difference of random slopes between sub-population groups. Figures 5.8 and 5.9 as well as Table 5.7 demonstrate that examining the whole distribution of the random effects is more informative than examining them only via statistical measures not capturing the whole variation. For example, while MOR remains almost constant in the health interview models, there is a larger variation in the coefficient of variation and corrected sum of squares. In Figures 5.10 and 5.11, we compare graphically the unweighted random intercepts and random slopes with the ones weighted by the interviewer assignment size. We find that on overall level the weighted and unweighted model estimates and predicted probabilities do not differ significantly. However, if we look into the random effects we find that the weighted and unweighted random intercepts, for example, may differ even substantially. The weighted random intercepts indicate the volume of the impact of interviewer random intercepts on the survey participation.

Figure 5.8 Distribution of the odds ratio for the normalised random slope of the families with children

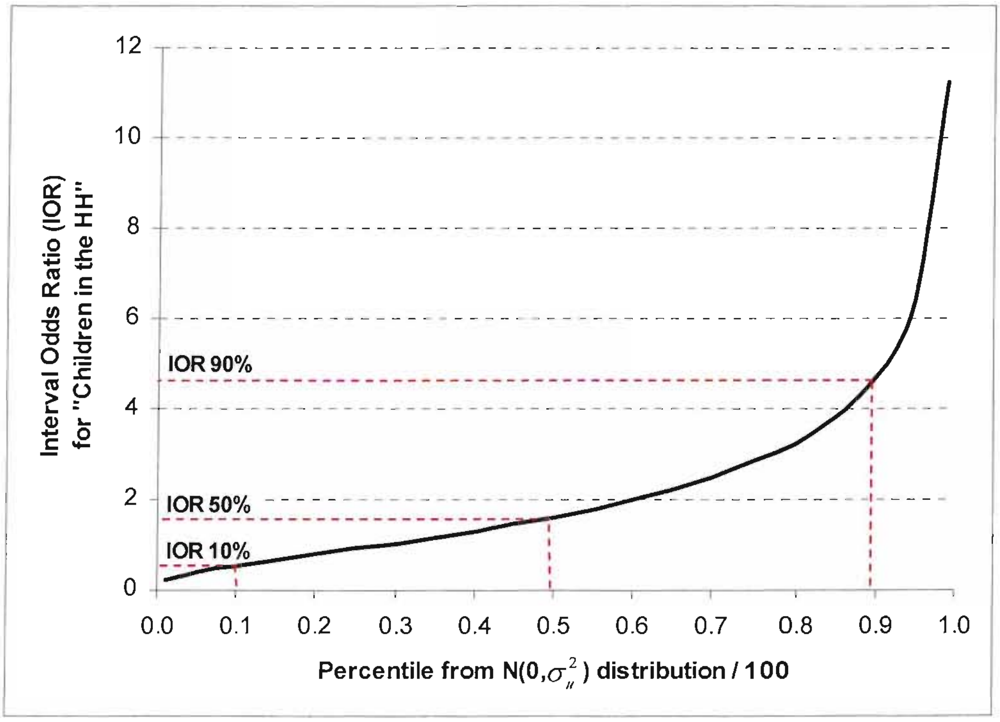


Figure 5.9 Distribution of the odds ratio for the normalised random slope of the lowest income quintile

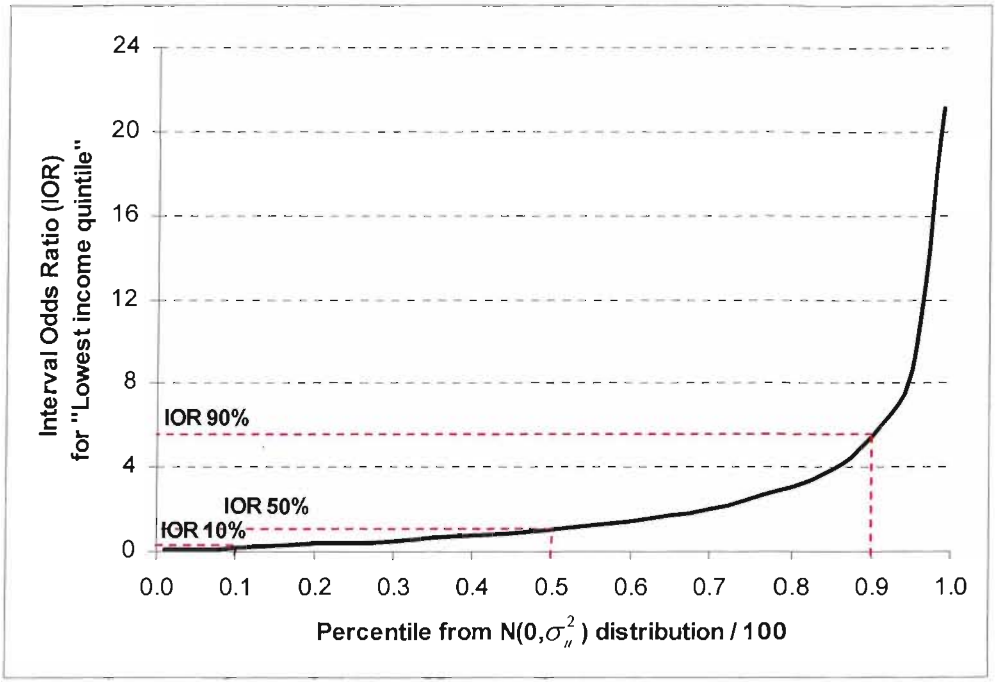


Figure 5.10 Comparison of random effects at health interview (Phase II) and symptom interview (Phase III) by estimation methods

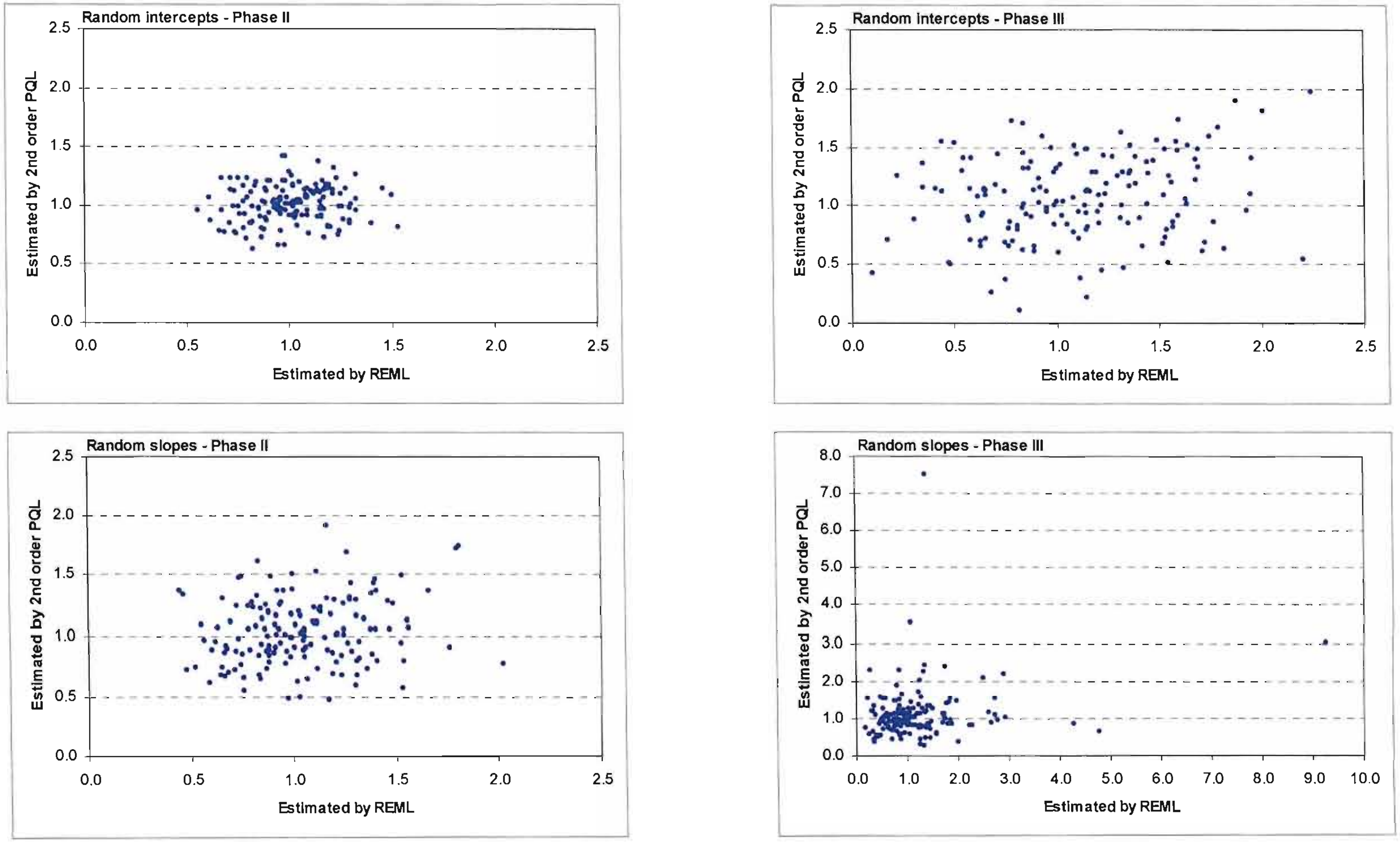
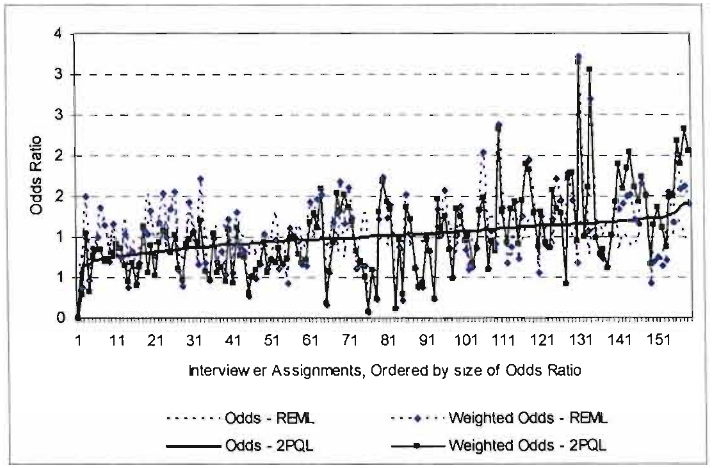
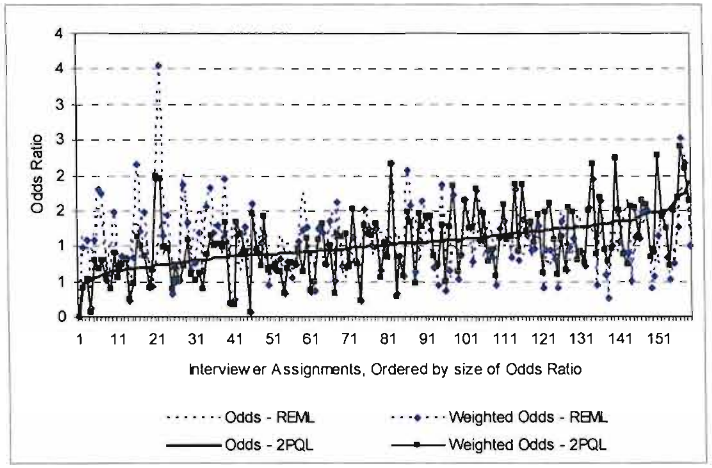


Figure 5.11 Random effects at health interview (Phase II) and symptom interview (Phase III)

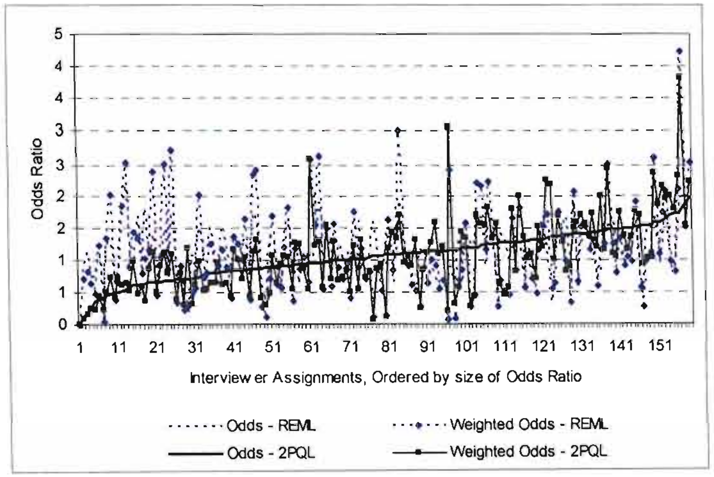
(a) Random intercept for interviewer assignment– Phase II



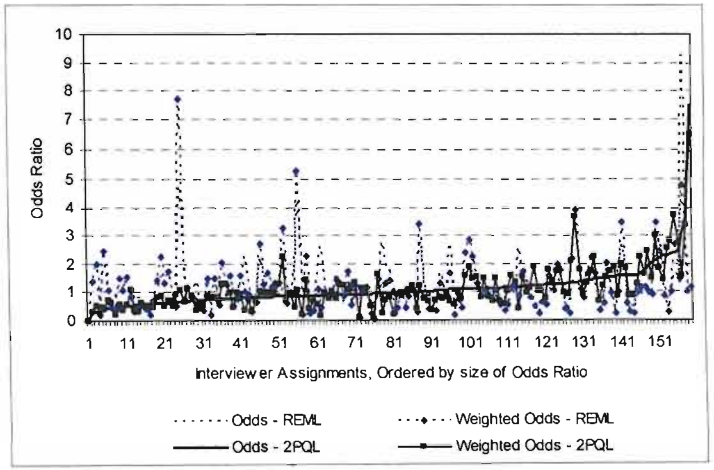
(b) Random slope for having children in the household – Phase II



(c) Random intercept for interviewer assignment– Phase III



(d) Random slope for being in the lowest income quintile – Phase III



5.6. Predicting response probabilities with cross-classified interviewer and local area effects

With the aim of modelling the survey participation correctly we take into account the clustering of the data, identify the significant group levels and study the overlapping hierarchical group levels. However, problems with overlapping hierarchy arise in survey environment, due to regional clustering in the sampling design and interviewer assignments. The interviewers may be working on assignments on two or more primary sampling units while in each primary sampling unit there can be one or more interviewers assigned. Therefore, one cannot see a clear pattern of multi-level hierarchy, in contrary individuals are nested independently and overlapping both within primary sampling units and interviewers. One plausible solution for this problem of unnested random effects is to apply multi-level modelling of survey participation for cross-classified data(e.g. Raudenbush, 1993; Rasbash and Goldstein, 1994; Goldstein, 1995; and Snijders and Bosker, 1999).

In interviewer surveys the interviewers and in clustered sampling design the regional clusters can affect significantly the survey participation. These affects may be controlled for by using multilevel techniques. Regional clusters are geographical small areas and they act as a social environment for all sampled persons and consist of a regional sub-population. There are some separate levels to be considered in the concept of multilevel modelling of survey participation in the Health 2000 such as the interviewer assignments and regional primary sampling units, see Table 5.8[14].

Table 5.8 Typical hierarchy of the data in health surveys

Type of units at levels	Hierarchical levels in the data	Hierarchical levels in survey design	Categories of the levels
Micro-units	Individuals sampled *i*	✓	Level-1 elementary unit
Micro-units	Households		Level-2 elementary units if included in the sampling design (not in Health 2000)
Macro-units	Local areas *j* [1)]	✓	Clusters due to sampling design
Macro-units	Interviewers and/or nurses *k*	✓	Clusters due to fieldwork allocation
Macro-units	Mobile health examination centres		Macro-level unit related to the sampling design and fieldwork arrangements
Macro-units	Major regions [2)]	✓	

[1)] Health centre districts (HCD) were used as local areas i.e. as geographical PSUs
[2)] University hospital districts (UHD) were used as major regions in order to balance the geographical distribution of the PSUs

The hierarchical levels of the Health 2000 data are due to the two-level sampling design using regional clustering, and clustering between interviewers[15]. Therefore the sampling design characterises a two-level hierarchy: (i) local area *j* and (ii) and individual sampled *i*. The individuals are clustered by interviewing assignments in all interviewing surveys. When the sampling design

[14] The Health 2000 is a survey on individuals, and by survey design there is no clustering of individuals within families, households or dwelling units. These cannot be used as level information for multilevel modelling. However, we cannot ignore the possible impact of families, households and/or dwelling unit on individuals survey participation. Thus some explanatory variables have been derived at this level for the modelling purposes.

[15] The data to be analysed consists of 30+ year olds from 80 geographical PSUs, interviewed by 158 interviewers. Due to the needs of the data users, the sampling used the classification of health centre districts in geographical clustering, instead of more conventional division by the municipalities. The number of individuals per local area

contains regional clustering the hierarchy of the data can become very complex, especially if the interviewer assignments and regional PSUs are overlapping. This is also the case in the Health 2000 survey data. In the two-level hierarchy of the data collection, the target persons are clustered by the interviewer assignments k. However, the target persons are nested both within local areas j and interviewers k. A strong overlap appears amongst local areas and interviewer assignments. An example of unnested structure is demonstrated in the Figure 5.12 where some interviewer codes are listed in the left column and some numbers identifying the primary sampling units i.e. local areas are the column headings of other columns. For example, one can observe from the Figure 5.3 that there are at least two interviewers working in the area '677' both of which also work in the area '686'.

Figure 5.12 Cross-classification of the Health 2000 survey data

Interviewers PSUs

Frequency	674	676	677	682	683	686	688
2730	0	0	0	0	0	0	
2756	0	0	0	0	0	0	
2769	0	0	0	0	0	0	
2815	0	0	0	0	0	0	
2828	0	0	0	0	0	0	
2831	0	0	0	0	22	0	
2886	0	0	0	0	0	0	
2932	0	0	0	0	0	0	
3069	0	0	31	0	0	21	
3072	0	0	19	0	0	10	
[illegible]	0	0	0	0	0	0	

The models allow for cross-classification of the interviewers and local areas. This enables the individual level differences to be determined independently of the group levels in the data set. The unnested hierarchical structure can be incorporated to the multilevel model through adding a random effect W_j of local areas j as follows:

$$\operatorname{logit}\{\pi_{ti}\} = \gamma_{0t} + \sum_{h=1}^{r} \gamma_{hk} x_{hik} + u_{0k} + \sum_{h=1}^{p} u_{hk} x_{hik} + w_{j(i,k)} \tag{5.15}$$

where the random effect of the local areas is defined as follows using group indicator variable b for crossed effects indication the cross-classification in the data:

$$W_{j(i,k)} = \sum_{j=1}^{J} W_j b_{jik} \tag{5.16}$$

and

varied from 48 to 884 individuals sampled. At the same time, the size of the interviewer assignment varied from 4 to 129 interviewees.

$$b_{jik} = \begin{cases} 1 \text{ if individual } i \text{ sampled in health district centre } j \text{ is assigned to an interviewer } k \\ 0 \text{ if the individual } i \text{ is assigned to another interviewer} \end{cases}$$

The structure of the variance of the cross-classified multilevel model becomes more complex than in (5.13). There are more than one macro levels which contain overlapping clustering. Thus, in our empirical case, the variance components include also a third dimension $\sigma^2_{(u_{qk}, u_{(q+1)j})t}$.

The multilevel models for cross-classified data enable to account properly for the hierarchical structure of the data. The data has an overlap between interviewer assignments and regional clustering. Following, the survey response behaviour for the interviewer attempt, *R*, can be constructed in a similar way as above for contactability. Using the sequential approach defined earlier in Chapter 4, the model for the probability of interviewers gaining co-operation from the assigned individuals at each subsequent phase is defined similarly. The covariate vector $\mathbf{x}$ may consist of significant covariates from different levels of analysis such as characteristics of the individuals themselves or their dwelling units, characteristics of the ecological sub-population at the level of local area and characteristics of the interviewer.

The impact of the interviewer on motivating the respondent to co-operate at later data collection phases

In Table 5.9, we present the estimation results from survey response modelling with crossed effects by interviewer assignment and the regional primary sampling unit. Factors affecting the survey participation have been studied with detected differences across data collection phases, which are based on the dependence of interviewer, region or their crossed-effects. The model diagnostics of unconditional means suggest that both interviewers and ecological sub-populations differ in terms of contacting the target persons and gaining their co-operation in survey participation. As with sequential logit model in Chapter 4, factors characterising socioeconomic wellbeing or deprivation of the individuals are most influential variables. When both the interviewer and regional variables are included, this affects to the model composition as some variables loose their explanatory power. For example, none of the regional factors found significant in Chapter 4, were influential in the crossed effect multilevel models.

Previously we have modelled survey participation with individual, household, dwelling and regional information (in Chapter 4) as well as excluding the regional information replaced by the interviewer level information (previously in Chapter 5). In Table 5.9, we demonstrate that the success in contacting or gaining target persons' co-operation is dependent not only on the characteristics of the target person, but also both on some characteristics of the interviewer and the interviewer attitudes towards the survey and the survey fieldwork arrangements. However, their impact varies hugely across the data collection phases. Comparison of the models of contacting target persons and gaining their co-operation in survey participation show that the simple demographic characteristics of individuals sampled are most significant in both models. These characteristics are namely age, gender and family status of the individual, and an indicator whether the sampled individual belongs to a register derived family. The crossed-effects models re-emphasise the importance of urbanicity of the local area on individuals' response behaviour and indicate a weaker explanatory power for dwelling level information. In addition, the fusion of both regional and interviewer level information causes changes in model dynamics. In addition, the complex structure of random effects reduces the explanatory power from some variables that previously seemed to have significant explanatory power. As in sequential logit modelling, survey variables obtained in the health interview describing physical or cognitive abilities are strongest indicators for the continuation of co-operation at later phases in cross-classified modelling.

Table 5.9 Estimated sequential multilevel models for survey participation allowing for the cross-classification by data collection phases[16]

Estimated parameters from cross classified multilevel models	Phase I: Contacting		Phase II: Health Interview		Phase III: Symptom Interview		Phase IV: Health examination		Phase V: Self-completion questionnaires	
	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value
Fixed effects:										
Fixed intercept	193.98	0.00	10.93	0.00	25.68	0.00	36.34	0.00	11.81	0.00
Individual characteristics:										
Age group:										
- Age 30 - 44 years (vrs 45 - 79)	0.57	0.01	0.64	0.00	0.69	0.01	0.80	0.25	0.99	0.95
- Age 80+ years (vrs 45 - 79)	3.27	0.01	0.86	0.20	0.67	0.02	0.22	0.00	1.12	0.57
Female (vrs male)	3.37	0.00	1.12	0.14	1.18	0.15	0.84	0.21	1.51	0.00
Maternal language:										
- Swedish (vrs Finnish or Sâmi)	0.77	0.61	0.73	0.08	0.88	0.68	0.65	0.07	0.62	0.01
- Other (vrs Finnish or Sâmi)	0.26	0.00	0.71	0.28	0.19	0.00	1.78	0.45	0.38	0.01
Family status:										
- Family with children (vrs families without children)	2.33	0.01	1.64	0.00	1.07	0.69	1.10	0.65	0.88	0.34
- Single person household (vrs families without children)	0.43	0.00	0.87	0.12	0.59	0.00	0.70	0.02	0.80	0.05
Further education (vrs basic education only)	1.11	0.63	1.30	0.00	1.72	0.00	1.53	0.00	1.67	0.00
Income quintiles of register derived disposable income:										
- 1st quintile or no income (vrs 2nd - 4rth quintile)	0.73	0.21	0.94	0.54	0.99	0.96	0.54	0.00	0.66	0.00
- 5th quintile (vrs 2nd - 4rth quintile)	0.94	0.83	1.14	0.21	1.29	0.12	1.44	0.00	1.07	0.62
Received capital income (vrs none) (household level)	1.87	0.01	1.22	0.02						
Received income support (vrs none) (household level)	0.39	0.00								
Savings to additional private pension scheme (vrs none) (household level)			1.89	0.00					1.99	0.01
Type of living unknown	0.13	0.00			0.41	0.01				
Region:										
Urban (vrs rural or other less densly populated)	0.36	0.00	0.73	0.00						

to be continued on the following page

16 Markov Chain Monte Carlo (MCMC) estimation has been used. The models have been estimated by 30000 simulations using MLwiN software.

Estimated parameters	Phase I: Contacting Odds Ratio	Phase I: Contacting p-value	Phase II: Health Interview Odds Ratio	Phase II: Health Interview p-value	Phase III: Symptom Interview Odds Ratio	Phase III: Symptom Interview p-value	Phase IV: Health examination Odds Ratio	Phase IV: Health examination p-value	Phase V: Self-completion questionnaires Odds Ratio	Phase V: Self-completion questionnaires p-value
Health interview information of the individual:										
Body mass index (BMI):										
- Information not obtained in health interview					0.01	0.00	0.10	0.00		
- BMI < 18 (vrs 18 <= BMI < 35)					0.25	0.00	0.34	0.00		
- BMI = 35+ (vrs 18 <= BMI < 35)					1.48	0.20	1.16	0.00		
Interviewer perception on individual's ability to understand speach or instructions:										
- Apparently mild difficulties (vrs no difficulties)							0.24	0.00		
- Clearly observable difficulties (vrs no difficulties)							0.14	0.00		
- Barely understands (vrs no difficulties)							0.02	0.00		
- Item missing information (vrs no difficulties)							0.01	0.00		
Cognitive ability:										
Which month of the year?										
- Incorrect (vrs Correct)									0.34	0.00
- Item non-response (vrs Correct)									0.78	0.08
Working capacity:										
- Partly incapable (vrs Fully capable)									0.94	0.69
- Fully incapable (vrs Fully capable)									0.39	0.00
- Item non-response (vrs Fully capable)									0.29	0.00
Interviewer level infomation:										
Interviewer characteristics:										
Age group of the interviewer:										
- 25 - 39 years (vrs 40 - 59)	0.44	0.00	0.74	0.02						
- 60+ years (vrs 40 - 59)	17.38	0.03	0.80	0.17						
Interviewer perception survey information:										
Impact of bad health condition or illness of the target person on refusals on general, increased (Q37i.):										
- 'In large amout or considerably' (vrs 'Not at all or in small amount')			0.80	0.03						
- 'Item missing information' (vrs 'Not at all or in small amount')			0.77	0.10						
Random effects:	**Estimated variance**	**p-value[1)]**	**Estimated variance**	**p-value[1)]**	**Estimated variance**	**p-value[1)]**	**Estimated variance**	**p-value[1)]**	**Estimated variance**	**p-value[1)]**
Random intercept for interviewer assignment effects	0.20	0.45	0.08	0.03	0.03	0.64	0.05	0.43	0.02	0.30
Random intercept for PSU effects	0.69	0.14	0.03	0.11	0.69	0.00	0.05	0.31	0.04	0.30
Random slope	0.79	0.15	0.43	0.02	1.21	0.01				
Random slope variable	*Female*		*Children in HH*		*1st Income quintile*		-		-	
Covariance between random intercept of crossed effects and random slope	-0.17	0.86	-0.11	0.03	-0.57	0.01	-	-	-	-

[1)] p-value based on the t-test derived upon the variance estimates and their estimated standard errors assuming normality

5.7. Conclusions

We have analysed the interviewer performance from contacting to achieving co-operation at health interview and to motivating the co-operation at later data collection phases. This analyse situation was designed to gain more knowledge on the mechanisms of attrition in surveys with multiple data collection and high response burden. Studying the survey participation of the Finnish Health 2000 survey, we have detected interviewer level heterogeneity. Interviewers have varying performance even when the variability is controlled for significant characteristics at individual, interviewer and local area level to avoid the ecological fallacy. As the response propensity of individuals appears to depend from fieldwork arrangements, there is a need for further research on suitable adjustment methods in response probability modelling. Although, our data contained clustering both by the interviewer assignments and regional primary sampling units, the multilevel crossed-effect modelling provided less improvement than the approach with random slopes. This indicates that the explanatory variables capture more efficiently the differences of the regional primary sampling units, while the differences in interviewer performance are more difficult to quantify with direct or latent variables.

We have explored how far in data collection the interviewer effects can be detected in data collection. Our finding is that there are significant interviewer level random effects at contacting, health interview and at following symptom interview. The symptom interviews took place together with the medical examinations and were conducted by medical staff i.e. by other than the professional interviewers whose performance is under investigation. This means that there are observable differences across interviewer assignments also after the health interview on how successfully the respondents progressed to the symptom interview, after the interviewer administration has ceased. Controlling for the individual and interviewer level variables, we have found that interviewers differ in contacting males and gaining co-operation form households with children. In addition, the data suggests that there is systematic difference across interviewer assignments on how people with low income are persuaded to participate to the symptom interview. The interpretation of the random effect on sex of the individual is possibly linked to the prior knowledge of the gender of the individual sampled, and varying expectations of the interviewer on the success affecting the level of efforts invested in contacting in relation to the required efforts. The interviewer level variation in achieving co-operation from households with children is affected by tailoring skills, level of flexibility and patience on finding suitable timing for interview.

Although, the estimated models were allowed to adjust for the variability across the interviewer assignments, the overall model fit was not improved significantly. In fact, in most of the sequential models, there was a reduction in the model fit by log likelihood tests. However, this is partly due to the impact of increasing parameters in the model when the random effects are allowed for. Thus suitable testing methods for assessing should be developed further.

Assessing the informative nature of the random effects was also studied in this chapter. The summary measures such as median odds ratio (MOR) and interval odds ratio (IOR) were applied to examine the random intercepts and slopes. Although these measures are useful for assessing the overall impact of random effects, we found them uninformative in studying the variation of the impact across the interviewer assignments. For exploring the differences across interviewer assignments, we recommend studying the entire distribution of odds of the random effects as well as studying the deciles, the lower and upper tails for extreme values. We also recommend graphical assessment of the random effects as a routine procedure for assessing estimated multilevel models and for examining their interpretation. This may enable survey organisations to better monitor the overall fieldwork operations and implement the necessary operations to improve the survey quality.

The assessment of interviewer effects generally implies that the variation in the interviewer work performance is examined. However, there is an unexplained factor whether the interviewer level variation may also arise from differences in the allocation of the interviewer assignments. In other words, the differences can develop from varying interviewer skills but also from unequal allocation of difficult assignments. In this case, there was no metadata available on the allocation rules of individuals into interviewer assignments. Thus, we can only aim to control for the association of the influential interviewer characteristics and difficulty of the assignment by incorporating individual-level data into the analysis. Applying and reporting systematic rules for constructing interviewer assignment allocations, compares with the necessity of implementing logical editing rules in comparison to subjective manual editing. Exploiting this type of metadata could be used for detecting weaknesses in the survey process, leading possibly for improved data quality. Therefore, further development of applications in this area, are called for.

One obvious challenge is to measure the interviewer attitudes and experiences in a usable manner for this kind of analyses. We have focused on using a quantitative approach to collect and analyse the interviewer perception. The obtained data has been incorporated with information on survey participation to test the significance of the various factors plausibly affecting the survey response. However, one of our findings is that using quantitative interviewer perception surveys, it can be very difficult to capture the underlying causes of the interviewers' attitudinal responses. Thus we recommend that in the next large scale surveys, further analysis of interviewer perceptions would be examined by combining the approaches of quantitative and qualitative analysis. We believe that the combined use of qualitative and quantitative approach can enrich in-depth understanding and correct interpretation of the interviewer attitudes in relation to their work performance. Similarly, we recommend focus group analysis for studying the perceptions of people who are sampled, for example, to pilot studies on their immediate reactions towards the survey and their attitudes towards participation. This information could enrich our knowledge of survey response, and take us further on our learning curve of the survey participation in surveys with multiple data collection phases and high response burden.

Furthermore, it will be a challenge to interpret the survey estimation models and examining the true contribution of interviewer effects. We have gained further understanding in survey participation, and more emphasis needs to be addressed for training the interviewers to encounter even more challenging situations they face in the fieldwork. Assessing the qualitative data from the interviewer perception survey, we found out that few interviewers had faced situations in which there was a problem with gate keepers, outside the household, in a survey on individuals. Adult children had told their elderly parents of not to participate to the survey. With the ageing population, problems such as this need to be addressed more efficiently in future surveys, and efforts to must be invested for developing the advance letters, for example, to become more proactive in terms of refusals. To conclude, we have detected various sources of interviewer effects. This imposes a challenge for the task of adjusting properly for non-response in a situation where it is not independent from the interviewer assignment allocation.

6. Weighting Adjustment for Non-response

6.1. Introduction

In this Chapter we consider whether the survey participation models developed in the Chapter 4 and 5, can be used to improve weighting adjustment. Conventional non-response weighting adjustment methods include techniques such as weighting classes, calibration, post-stratification or raking ratio. The auxiliary information is commonly used in the form of population margins, totals or weighting cells. Auxiliary information generally consists of demographic, socio-economic or geographic factors. As the causes of non-response can be due to complex set of factors or their interactions, limited information used in the conventional weighting methods may not withhold correct adjustment. Groves and Couper (1995) argue that the theoretical bases in post-survey adjustment should lie in the socio-psychological theories that specify the human behaviour affecting survey participation. At present, the use of survey participation modelling in the adjustment of survey estimation is more of an idea and an area of research than a generally accepted practise in the statistical offices. These models can ideally have a meaningful interpretation, given that the models connect closely with the chosen theoretical framework.

In this Chapter, we exploit extensively auxiliary information from register data sources to assess the performance of alternative weighting methods for our survey data. Previously in Chapter 4 and 5, we have detected that survey participation of Health 2000 data has been significantly dependent on auxiliary information on the individuals and interviewers. In this Chapter, we use further these previously estimated response probabilities for estimation in the presence of non-response. The use of estimated response propensity models will be explored in weighting adjustment. In particular, this involves methods based on overall response probabilities for each sampling unit and the use of variants of the Horvitz-Thompson estimator (Särndal et al., 1992). The performance of simple Horvitz-Thompson type estimator based on the estimated response probabilities is compared to other weighting methods such as the calibration together with the sample selection models adjusting for the complex survey design. To emphasise, the weights will depend on the modelling assumptions and the selection of the study variables. In the empirical assessment of this Chapter, we link highly correlated proxy estimates of the survey variables from the health registers indicating whether respondents, non-respondents and population members have been diagnosed with long-term illnesses. We compare various weighting methods by estimating the prevalence of long-term illnesses and distribution of socio-economic conditions. We then benchmark the conventional methods against estimates derived by weighting methods based on our response behaviour models.

The use of auxiliary information for non-response adjustment has been studied in the past (e.g. Bethlehem, 2002), but the efficiency testing of the different sets of auxiliary information has been lacking quantified indicators. Särndal and Lundström (2005) have hypothesised that the use of auxiliary information improves the estimator more than the mathematical methods. We will investigate whether this holds in our data. First, we conduct a sensitivity analysis by assessing how the selection of explanatory variables in the response probability models affects the survey estimates. Secondly, we examine the bias of survey estimates given by the alternative methods, data collection phases, and non-response reduction efforts.

For assessing bias of the estimates, we use information that can be measured similarly both for the sample and for the total population as emphasised by Bethlehem and Kersten (1985). We exploit taxation, social security and health register information available that contain highly correlated variables with the survey variables and with the propensity to respond. The definitions of health register variables are based on the condition whether the individuals have been diagnosed with specified long-term illness and are entitled for reimbursement on medical expenses caused by the treatment of the disease. We simulate estimation of the prevalence of diabetes mellitus, chronic cardiac insufficiency, connective tissue diseases, chronic asthma and similar chronic obstructive pulmonary diseases, chronic hypertension, and chronic coronary heart disease. The estimate for the population totals by socio-demographic factors for the purpose of the ratio estimator can also be derived from taxation and social security registers and linked to the survey data.

We begin by an introduction of the conventional weighting methods and presentation of the Horvitz-Thompson type inverse response probability weighting. Afterwards, we apply response marginality weighting in simple definition of non-response and the inverse probability weighting methods to our empirical data. The weighting methods are then assessed using theoretical comparison and quantitative indicators for weight comparison developed recently by Kalton and Florence-Cervantes (2003) and Särndal and Lundström (2005). In addition, we present an indicator measuring the change in the impact of the interviewer due to non-response adjustment. Subsequently, analysis of error is studied in order to estimate the relative bias of survey estimates. In the analysis of error, we study the bias correction of the non-response correction efforts. Secondly, we examine the impact on survey estimates of the level of co-operation achieved at various data collection phases. We conclude by discussion on the performance of various weighting methods and compare their strengths and weaknesses.

6.2. Response marginality weighting in simple definition of non-response

The main purpose of weighting is to adjust the responding sample for unit-nonresponse. Weights are derived exploiting the known information for the target population, non-respondents and respondents. The choice of a weighting method is affected by restrictions imposed by sampling design, response structure and the needs of specific estimators. Survey weights are often developed in series of phases to initially compensate for the differential inclusion probabilities and non-coverage errors, and subsequently compensating for the non-response and non-sampling fluctuations from known population values (Brick and Kalton, 1996).

We construct sampling weights, also called design weights, based on information both from the sample S and the target population U. In addition to the differential selection probabilities, the design weights can be adjusted to take into account the differences between the original sample and the obtained responding sample in regional clusters. For comparison, we use an alternative base weight that relates to the underlying perceived probability of individuals to respond to surveys. The estimated response probabilities $\hat{\pi}_i$ for individual i are conditioned on probabilities that the individual is both sampled π_{si} and responds $\hat{\pi}_{r|si}$ so that

$$\hat{\pi}_i = \hat{\pi}_{r|si}\pi_{si} = P\left(i \subset S \cap R_i = 1 \middle| \mathbf{X}_i\right), \tag{6.1}$$

where R_i is a dummy indicator for the response of the individual i defined in (2.1). Assuming simple random sampling, the inclusion probabilities π_{si} would equal to unity for all sample members. When using sampling from a register or list of individuals with known characteristics, the inclusion probabilities π_{si} can be derived while the unknown response probabilities $\pi_{r|si}$ must be estimated. Generally, the inclusion probabilities form the basis for the sample weights. The use of design weights as base weights is often motivated by differential inclusion probabilities in surveys with complex designs. Design weights reflect to varying sampling fractions, when the size of the target population and the size of the population within each stratum are known, or expected to be known with reasonable accuracy. Further, it is assumed that the inclusion probabilities of each sampled individual depend only on the individual characteristics or their stratum or cluster, but not on the inclusion probabilities of other sampled individuals. In addition, Oh and Scheuren (1983) stress that two assumptions are made for probability samples first each individual in the population can be sampled with the probability of greater than zero (regardless of their characteristics or their expected values on the survey variables) and secondly all individuals are assumed to have a probability above zero to respond to the survey. In Appendix 6.1, we explain the construction of design weights for the Health 2000 survey data used in our analysis[1].

[1] The Health 2000 survey weights and their construction has previously been reported by Djerf and Laiho (2004)

6.2.1 Weighting classes

Weighting classes method represents weighting based on response rates and assumption that the respondents can be unambiguously divided in to groups. These adjustment cells are determined by auxiliary information available for all sampled individuals. The method strictly defines that this allocation of respondents is both exclusive, and that every respondent is allocated solely to one adjustment cell. Weighting class method makes a strong assumption that the response probabilities of all sampled individuals are equal to the response rate within their adjustment cells (Kalton and Maligalig, 1991). The estimated response probability conditional on being a member of the sample $\hat{\pi}_{r|si}$ is then defined as:

$$\hat{\pi}_{r|si} = \sum_{h=1}^{H} \frac{m_h}{n_h} \tag{6.2}$$

where m_h denotes for the respondents and n_h for sampled individuals in weighting class h. Thus the final survey weight w_i can be written as:

$$w_i = \frac{n_h}{m_h} d_i = \frac{n_h N}{\hat{\pi}_{ih} m_h N_e} \tag{6.3}$$

where d_i denotes for the design weight and $\hat{\pi}_{ih}$ is defined as:

$$\hat{\pi}_{ih} = \hat{\pi}_{h|r} \pi_{si} = P\left(i \subset S \cap i \subset h \middle| \mathbf{X_i}\right), \tag{6.4}$$

in which $\hat{\pi}_{h|r}$ represents the response rate in the adjustment cell h. The definition of the $\hat{\pi}_{ih}$ resembles the general definition of $\hat{\pi}_i$ given in (6.1), the difference being in the allocation between the weighting classes h.

Weighting classes are based on the available information for the respondents and non-respondents, the appropriate grouping in relation to response behaviour as well as the anticipated optimal bias reduction. The key criteria for the success of using the weighting classes method is the identification of homogenous weighting classes that are reasonably justified, and that the non-response is unrelated to the survey variables within the weighting classes. A potential disadvantage of weighting classes methods is that it can lead to large variablity in the distribution of the weighting adjustments, there by inflating the variance of the survey estimates (Kalton and Flores-Cervantes, 2003).

6.2.2 Calibration using auxiliary population information

Calibration estimators have been proposed by Deville and Särndal (1992) referring to earlier work by Lemel (1976) and Deville (1988) on post-survey correction. Calibration is one method of re-weighting aiming to correct for the effect of non-response on the final attained sample. Secondly, the objective is to generalise the final data to represent the target population. Calibrated weights are readjusted by a specified distance function, which aims to meet the marginal conditions set for the estimators and retain the calibrated weights as close as possible to the original weights (Deville and Särndal, 1992)[2].

We denote the estimated survey variable by y and the individual level multivariate auxiliary information by a vector of $\mathbf{x}_i = (x_{i1},\ldots,x_{iP})'$. We can observe both y_i and $\mathbf{x}_i$ for all survey respondents. For estimating the population total $Y = \sum_{i=1}^{U} y_i$ in the target population U, calibration adjusts the starting weights d_i, which are generally design weights based on the responding individuals. The starting weights are adjusted to correct the weighted population distributions and totals to follow the known auxiliary information $\mathbf{X} = \sum_{i=1}^{U} \mathbf{x}_i$. The adjustment is restricted with a distance measure, which is to minimize the difference between the adjusted and initial design weights. When choosing auxiliary variables used in calibration, one needs to judge the accuracy and reliability of the information, while assuring that the basic assumptions of the calibration method are met. The calibration equations are written as:

$$\sum_{i=1}^{m} w_i \mathbf{x}_i = \mathbf{X}, \tag{6.5}$$

where m refers to the responding individuals for whom we observe the value of y_i. Deville and Särndal, (1992) show that the calibrated estimator:

$$\hat{Y}_C = \sum_{i=1}^{m} w_i^* y_i \tag{6.6}$$

follows closely the Horvitz-Thompson estimator, defined as

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_{si}} \tag{6.7}$$

[2] The CALMAR macro developed by Sautory (1993) and CLAN97 (Anderson and Nordberg, 1998) can be used to minimise the function that measures the distance between the calibrated weighting coefficient and the sample weight. The former will be used in this analysis.

where π_{si} is the inclusion probability for the individual i in sample S. The distance measure is a quantity based on the calibration equations and a reminiscent of the chi-square:

$$\sum_{i=1}^{m}\left(w_i-d_i\right)^2/d_iq_i=\sum_{i=1}^{m}d_i\left(w_i/d_i-1\right)^2/q_i\,. \tag{6.8}$$

The difference should be minimised given the obtained sample. Deville and Särndal, (1992) show how its minimization leads to the calibrated weight:

$$w_i=d_i\left(1+q_ix_i'\lambda\right). \tag{6.9}$$

Calibration can be useful when auxiliary information is available both for all respondents and population margins. The method has been adopted as the standard way of constructing weights in sample surveys at Statistics Finland (see Statistics Finland, 2002). Also the Health 2000 survey weights are based on calibration (Djerf, Laiho and Härkänen, 2004)[3]. For this reason, we will later compare the response propensity weighting to the calibration.

6.2.3 Post-stratification

Post-stratification is a simple form of calibration. Post-stratification forces the marginal survey distributions and the size of the weighting classes to follow the equivalent population distributions by assigning identical adjustment weights to all elements in the same weighting class. The method can be extended in the situation of several marginal distributions and population totals into calibration method. The sample is weighted according to the known population distribution information of the frame; in individual-based surveys generally according to demographic characteristics such as age, gender and area of residence (Deville, Särndal and Sautory, 1993). Post-stratification gives a population weighted survey estimator, which exploits the availability of population level information such as the size of the target population at each weighting class. Assuming SRS, the response probability weight for weighting class h is defined as

$$w_h=\frac{N_h}{N}\frac{n}{n_h}. \tag{6.10}$$

However, the information on population totals is limited to the knowledge obtained from administrative records or registers i.e. to the registered population. Therefore, the actual response probability weight used for post-stratification is more precisely expressed as:

$$w_h=\frac{N_{h_e}}{N_e}\frac{n}{n_h}, \tag{6.11}$$

[3] Previously derived Health 2000 survey weights are based on calibrating the non-response adjusted design weights by age, language and regional distributions. These weights have been used in reporting the survey results. They are also included into the data files released for researchers.

where e denotes to the population found eligible representing estimates obtained from auxiliary sources (Bethlehem, 2002).

6.3. Inverse response probability weighting for sequential non-response

The previously presented weighting methods ignore the differential response probabilities that were estimated in Chapters 4 and 5 using logistic regression type modelling. However, we have previously demonstrated that the health of the individuals affected their co-operation at later data collection phases in Health 2000 survey. In addition, the survey co-operation was found strongly associated with socio-economic wellbeing or deprivation of individuals. In this Section we present weighting methods that also account for variability in estimated response probabilities and later in this Chapter we assess the usability of the underlying weighting structure of the previously estimated models. In simplest case, individuals are assigned with weight of the inverse of their estimated response probability. The inverse probability weighting is expected to balance the obtained sample. People having lower response propensity have a higher weight in comparison to those who have higher probability to respond. However, depending on the distribution of the estimated probabilities, this method can lead into large variation of weights. So in the construction of response probability estimates, one must balance between informative weighting, convergence, and reduced variability with the objective of unbiased and accurate estimators. A sequence of non-response adjustments may be employed in panel and multiphase surveys, where non-response may occur at each successive phase of data collection (e.g. Clayton, et al. 1998; Kalton and Flores-Cervantes, 2003; and Iannacchione, 2003).

The study variable Y is a dichotomous variable indicating the prevalence of the health condition in the population:

$$Y_i = \begin{cases} 1 & \text{if the individual } i \text{ has the specific health condition} \\ 0 & \text{otherwise} \end{cases} \tag{6.12}$$

The association between the study variables and response probabilities are unknown. The response probabilities are estimated to using informative auxiliary information matrix $\mathbf{X}_i$. Our basic assumption for the inverse probability weighting is that co-operation of sampled individuals with the survey request is independent from each other.

6.3.1 Explicit response model weighting and inverse probability weights - An extension of a naïve Horvitz-Thompson estimator

Horvitz-Thompson (H-T) estimator represents a simple approach of using inverse probabilities in weighting expansion. The estimator was originally used for the principle of "π expansion" to estimate the population total for survey variable (Särndal, Swensson, and Wrettman, 1992 referring to Horvitz and Thompson, 1952). The inclusion probabilities π_{si} for individual i can by definition be assumed positive for all i in the target population. As the H-T estimator defined in (6.7) is design unbiased, it assumes that information is obtainable for all members of the target population (Djerf, 2001).

The Horvitz-Thompson estimator has been used traditionally to predict population totals or ratios using the design weights i.e. inclusion probabilities of the sampled individuals. We study these estimators by assessing the performance of H-T estimators built solely the predicted response probabilities $\hat{\pi}_{r|si}$ and alternatively on the product of the inclusion probability and the response probability i.e. $\hat{\pi}_{r|si}\pi_{si}$. Following the (6.7), the Horvitz-Thompson based simple ratio estimator for the prevalence of a property Y in the population is presented accordingly:

$$\hat{\bar{Y}}_{HT} = \frac{1}{N_e}\sum_{i=1}^{m}\frac{y_i}{\hat{\pi}_{r|si}}. \tag{6.13}$$

The alternative Horvitz-Thompson type of estimators are weighted both for differential inclusion and response probabilities:

$$\hat{Y}_{HT} = \sum_{i=1}^{m} w_i y_i = \sum_{i=1}^{m}\frac{1}{\pi_{si}\hat{\pi}_{r|si}} y_i \tag{6.14}$$

where w_i is the survey weight for the individual i, π_{si} is the inclusion probability and $\hat{\pi}_{r|si}$ is the estimated response probability for sampled individuals.

The unbiased estimation of the population mean $\bar{Y}$ would generally require full response of all sampled individuals, which is a rare situation in sampling surveys. Then the unbiased estimate of the mean of explained variable would be:

$$\hat{\bar{Y}} = \frac{1}{N_e}\sum_{i=1}^{R} d_i y_i \tag{6.15}$$

where d_i is the design weight. For estimating the population mean in the presence of non-response, we aim to adjust for the non-response to correct for the impact of not observing the values of Y on the non-respondents. Ekholm and Laaksonen (1991) point out that for estimating the true mean in the target population we should concentrate on the following equation of the population total:

$$\hat{Y} = \hat{N}\hat{\bar{Y}} \tag{6.16}$$

which enables us to focus estimating the total of $\hat{Y}$ as $\hat{N}$ is obtained as a special case.

In our data structure, it is plausible to estimate the $\pi_{r|si}$ as a function of different settings of sample indicator D, response indicator R, and auxiliary information X using logistic regression based models across data collection phases t. The $\pi_{r|si}$'s are estimated for all $i \subset S$ using the available explanatory variables, linked directly to the individuals, ecological sub-populations or interviewers. Subsequently, we use the inverse probability $\left(\hat{\pi}_{r|si}\right)^{-1}$ for deriving the weighted survey estimates for the prevalence of some health condition indicator Y:

$$\hat{\bar{Y}} = \frac{1}{m}\sum_{i=1}^{R}\frac{1}{\hat{\pi}_{r|si}}D_i R_i y_i \,. \tag{6.17}$$

Alternatively, the non-response adjustment can depend on the product of design weights and the estimated unknown response propensity $\hat{\pi}_{r|si}$:

$$\hat{\bar{Y}} = \frac{1}{N_e}\sum_{i=1}^{R}\frac{1}{\pi_{si}}\frac{1}{\hat{\pi}_{r|si}}y_i \,. \tag{6.18}$$

Weights based on logistic regression models are restricted to be above 1 by definition. We scale the weights so that the final survey weights weight the responding sample to the level of eligible target population, i.e. $\sum_{i=1}^{R} w_i = N_e$. Thus the scaled final survey weight w_i can be written as:

$$w_i = \frac{1}{\left(\pi_{si}\hat{\pi}_{r|si}\right)^{*}}, \tag{6.19}$$

where the * denotes for the scaled product of weights. In simple random sampling or when ignoring the survey design, the final survey weight for population level would be defined simply accordingly:

$$w_i = \frac{N_e}{\hat{\pi}^{*}_{r|si}} \,. \tag{6.20}$$

If the logistic regression model contains continuous explanatory variables then the response probabilities are estimated at individual level, otherwise the categorical variables form a set of groups resembling the weighting class method or post-stratification, presented in Section 6.2.1 and 6.2.3. In contrary to the raking, post-stratification and weighting classes methods, logistic regression base weighting can take into account features of hierarchical data structures, complex latent dependencies and longitudinal elements of the data collection. An early application of logistic regression for weighting purposes has been provided by Ekholm and Laaksonen (1991)[4]. However, using only categorical variables their method resembles post-stratification.

[4] They presented a method that employs the auxiliary information available for both respondents and non-respondents as they model the response propensity by logistic regression. The explanatory factors contain information from household structure, urbanism, region and indicator of capital income.

6.3.2 Response probability weight adjustment across multiple data collection phases

As in Chapter 4 and 5, the conditional probability for individual i at data collection phase t is defined with the binary response indicator R_{ti}:

$$\pi_{it|t\text{-}1} = P(R_{it} = 1 \left| \prod_{h=1}^{t-1} R_{ih} = 1, \mathbf{X}_i \right.) . \tag{6.21}$$

For conditional response probability in sequential logistic regression, we define further that after the initial contacting phase the response probability is estimated solely for sub-group responding successfully at the previous data collection phase $t-1$ i.e. for the sub-group that may still co-operate fully with the survey request:

$$0 \le P\left(R_{ti} = 1 \middle| D, Y, R_{(t-1)i}\right) \le 1 .$$

The corresponding estimator of population mean $\bar{Y}$ is

$$\hat{\bar{Y}} = \frac{1}{m} \sum_{i=1}^{N} \frac{1}{\hat{\pi}_{rt|si}} D_i R_i Y_i ,$$

where m denotes for the number of responding units and $\hat{\pi}_{rt|si} = P\left(R_{ti} = 1 \middle| D, Y, R_{(t-1)i}\right)$.

6.3.3 Calibration with estimated response probabilities

Alternative formulation of the calibration has been presented by Lundström and Särndal (2001) in which the initial weight consists of inverse of the product of design weight and estimated response propensity. This approach of calibration is closely connected with our approach modelling the propensity of survey response and the use the inverse of the response probabilities as an element of weighting. The calibration equations can be re-written in a following way (Särndal and Lundström, 2005):

$$\hat{Y} = \sum_{i=1}^{m} \frac{1}{\pi_{r|si}} d_i y_i ,$$

where $\pi_{r|si}$ is the unknown response propensity of individual i, estimated by $\hat{\pi}_{r|si}$. Särndal and Lundström (2005) presume that this $\hat{Y}$ estimator is biased in comparison to calibration estimator in (6.6).Thus they suggest a bias reduction using auxiliary population total information denoted by $\sum_{i=1}^{U} x_i^*$ which is incorporated into the estimator in a following manner:

$$\hat{Y} = \sum_{i=1}^{m} \frac{g_{i\hat{\theta}}}{\hat{\pi}_{r|si}} d_i y_i \,, \tag{6.22}$$

where $g_{i\hat{\pi}} = 1 + \lambda_i' x_i^*$ and $\lambda_i' = \left(\sum_{i=1}^{U} x_i^* - \sum_{i=1}^{m} \frac{d_i}{\hat{\pi}_{r|si}} x_i^* \right)' \left(\sum_{i=1}^{m} \frac{d_i}{\hat{\pi}_{r|si}} x_i^* x_i^{*'} \right)^{-1}$.

6.4. Comparison of weighting methods

In this Section we review indicators developed for comparing weighting methods. We also present an extended indicator for the purpose of interviewing surveys. The main purpose of weighting is to reduce the bias of the survey estimates. The survey weights or their re-adjustment may however result also in increased variablity of the weights. However, ideally the accuracy of the estimates is not reduced by the efforts to improve the unbiasedness. Our aim is to find a weighting method with informative non-response bias correction. Thus, we compare inverse response probability weighting methods with unconventional and alternative calibration approaches. We present calibration with alternative schemes of auxiliary information using standard demographic variables, variables indicating socio-economic conditions and income inequalities. For inverse probability weighting, we use the individual level response probabilities estimated in Chapter 4 and 5.

Traditionally, the weighting distribution has been assessed by its variation. High variation in weights is regarded as negative feature as the increased standard errors of the survey estimates reduce the accuracy of the predictions. Kalton and Flores-Cervantes (2003) stress that measuring the variance inflating factor F is useful for comparison of weighting methods:

$$F = 1 + CV(w_i)^2 \,,$$

where $CV(w_i)$ is the coefficient of variation of the survey weights w_i, where $CV(w_i)^2$ indicates the variation of the weights. The variance inflating factors for comparing the weighting of our empirical data has been presented in Table 6.1.

A recent development for weighting assessment has been presented by Särndal and Lundström (2005). They argue that one can improve the estimators more with efficient use of auxiliary information rather than with the choice of weighting method. Using the Health 2000 survey and register data we study whether the impact of the method is less substantial than the informativeness of the auxiliary information. Previously, Lundström and Särndal (2001) have emphasized three principles for the selection of an auxiliary vector which have been extended for the comparison of weighting methods in this Chapter. Firstly, the auxiliary vector should sufficiently explain the variation of the response probabilities. Secondly, the auxiliary vector should have good explanatory power on the variation of the main study variable. And thirdly, it should identify the most important domains.

Särndal and Lundström (2005) have proposed an indicator to measure how well the auxiliary vector explains the response influence i.e. the inverse of the response probability, and an alternative indicator to measure how well the auxiliary vector explains the target variable. According to Särndal and Lundström (2005) the crucial is to construct an auxiliary vector that gives a large value of the first indicator, since this will reduce the non-response bias for all estimates. The indicator measures the relative variation within the weight correction term for the estimated response propensity as:

$$IND_1 = \sum_{i=1}^{R} d_i \left(w_i - \bar{w}\right)^2 / \sum_{i=1}^{R} d_i \qquad (6.23)$$

where w_i is an estimated value for the unknown inverse of the response propensity $\hat{\pi}_{r|si}$. Särndal and Lundström (2005) demonstrate how the indicator is applied for any one-way classification by some grouping variable. In interviewing survey with complex survey designs, this indicator can be derived in addition to other classification, both for the regional clustering j as well as for the allocation of the sample into interviewer assignments k. The indicator is then extended to make a comparison how much the bias of the estimation is corrected by taking into account the interviewer level as follows:

$$IND_{1k} = \frac{1}{\sum_{i=1}^{r} d_i} \left(\sum_{k=1}^{K} \frac{\left(\sum_{i=1}^{n_k} d_i\right)^2}{\sum_{i=1}^{r_k} d_i} - \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{\sum_{i=1}^{r} d_i} \right). \qquad (6.24)$$

Similarly, the indicator can be calculated for regional clusters. This will allow us to make the choice whether the clustering due to survey design features on the data collection have more impact than the sampling design features. We also suggest that the indicator based on inclusion probabilities would be extended by replacing the sample weights by inverse response probability weights:

$$IND_{1k} = \frac{1}{\sum_{i=1}^{r} w_i} \left(\sum_{k=1}^{K} \frac{\left(\sum_{i=1}^{n_k} w_i\right)^2}{\sum_{i=1}^{r_k} w_i} - \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{r} w_i} \right), \text{ and}$$

$$IND_{1k} = \frac{1}{\sum_{i=1}^{r} d_i w_i} \left(\sum_{k=1}^{K} \frac{\left(\sum_{i=1}^{n_k} d_i w_i\right)^2}{\sum_{i=1}^{r_k} d_i w_i} - \frac{\left(\sum_{i=1}^{n} d_i w_i\right)^2}{\sum_{i=1}^{r} d_i w_i} \right).$$

The second indicator proposed by Särndal and Lundström (2005) measures the capacity of the auxiliary vector used in weighting to explain the specific study variable:

$$IND_2 = 1 - \frac{\sum_{i=1}^{r} d_i w_i (y_i - \hat{y}_i)^2}{\sum_{i=1}^{r} d_i w_i (y_i - \bar{y}_r)^2}. \tag{6.25}$$

We propose an extension of this for the interviewer surveys to assess the within and across interviewer assignment deviations in relation to the bias assessment:

$$IND_{2k} = 1 - \frac{\sum_{k=1}^{K} \sum_{i=1}^{r_k} d_i w_i (y_{ik} - \hat{y}_{ik})^2 + \sum_{k=1}^{K} (\hat{y}_k - \hat{y})^2}{\sum_{k=1}^{K} \sum_{i=1}^{r_k} d_i w_i (y_{ik} - \bar{y}_{r_k})^2 + \sum_{k=1}^{K} (\bar{y}_k - \bar{y}_r)^2}. \tag{6.26}$$

Särndal and Lundström (2005) state that large values of the indicator 1 indicate improved bias correction of the weighting method. If the auxiliary vector under assessment results into a large values for Indicators 1 and 2, the bias for the estimate based on the particular target variable is reduced even further. However, they also note that the indicator 1 can be manipulated too large if the group size of the auxiliary variable projects too many groups of small size. The implementation of the second indicator would anyhow require applying prediction models with the individual level risk factors for each long term disease. The information on risk factors is limited to socio-demographic factors, missing the health related information crucial to separate population at risk for a specific disease. Thus the second indicator suggested by Särndal and Lundström (2005) will be applied in a follow-up survey focusing more in detail on the specific health conditions and their estimation in the population. Instead, the weighting methods are reviewed comparing the variance inflating factor, indicator 1 and a measure for change in the underlying interviewer effect to be introduced in the following.

Introducing an indicator on the impact of the interviewer in the obtained sample

As a final indicator to assess weighting methods, we present a measure for assessing the change in the underlying impact of the interviewers due to non-response and weighting. This approach provides one step towards assessing the total survey error. We assume that in interviewer surveys interviewers can contribute to the survey error not only by their varying response rates but also by interviewer related measurement errors. The interviewer error and achieved completion rate can be anticipated in advance and balanced in the assignment allocation. However, it can be more complex to anticipate the effect of the non-response adjustment methods. When non-response has been adjusted using weighting, the impact of interviewer level measurement errors in survey results are likely to change from the original allocation. Therefore, we introduce an additional indicator for weighting assessment to measure the relative change in interviewer impact due to weighting in terms of their proportional share of the issued and obtained sample.

The survey non-response may vary across the interviewer assignments, as was observed in Health 2000 data in Chapter 5. In addition, it is reasonable to assume that the interviewer measurement errors vary across interviewer assignments. Thus the non-response cause imbalance to the obtained sample, not only due to sampling errors, but also due to changes in the relative power of interviewer effects related to the survey design features. Let us note the interviewer weight by:

$$w_{ik} = \frac{n_k}{\sum_{k=1}^{K} n_k} = \frac{n_k}{n}, \qquad (6.27)$$

which is assumed contain an optimal allocation between the interviewers that would minimise the effect of the interviewer measurement error on survey results. This allocation is affected by the survey non-response, so that in the presence of non-response the impact of the interviewer has changed from w_{ik} into:

$$w_{ik}^{*} = \frac{r_k}{n} \qquad (6.28)$$

or when assuming that the allocation is based on estimated interviewer achieved response rate, which is closely predicted:

$$w_{ik}^{*} = \frac{r_k}{m}. \qquad (6.29)$$

Further re-weighting may change the values of interviewer weight, denoted by w_{ik}^{**} and defined as, when the sum of weights is scaled to the level of the sample:

$$w_{ik}^{**} = \frac{w_i}{n}\frac{r_k}{m} \qquad (6.30)$$

or when taking into account the survey design

$$w_{ik}^{**} = \frac{d_i w_i}{n}\frac{r_k}{m}. \qquad (6.31)$$

The indicator aims to give an indication on the imbalance the survey non-response may cause to the interviewer related measurement errors in the obtained data.:

$$IND_{k'} = \sum_{k=1}^{K}\sum_{i=1}^{r_k}\left|w_{ik}^{*} - w_{ik}^{**}\right|. \qquad (6.32)$$

The next phase in survey estimation in minimising interviewer effects would be to minimise the difference between distributions of (6.27) or (6.28) and the re-weighted weights w_{ik}^{**}.

Comparison of weights

The indicators presented in the previous section are applied to various weighting methods, presented in Table 6.1. The variance inflating factor (Kalton and Flores-Cervantes, 2003) indicates the lowest variation of weights for the sequential multilevel logit model based inverse response propensity weights allowing for interviewer effects. Also the inverse response propensity weights based on single-level sequential logit models have a relatively small variation in comparison to other methods. Calibrated weights tend to have double the variance inflating factor than the related inverse probability weights. We can observe from the Table 6.1, that the design weight adjustment in any inverse probability adjustment yields only slightly higher values for the indicator, pointing towards minor improvement in bias reduction.

According to Särndal and Lundström (2005), the grouping variables for indicator 1 enable us to detect which grouping would reduce the bias of the estimators. In the table, simple design weights are compared to design weights with non-response correction, inverse probability weights (IPW) from selected modelling, and IPW with design weight adjustment. We have derived the indicators based on grouping variables that are related to the survey and sampling design as well as auxiliary information, which importance is assessed for improving the weighting. Taking into account interviewer assignments is suggested to correct more bias than accounting for local areas. The variable that would mostly reduce the bias is the age of the individual. In the design weights and IPW models age of the individual has been incorporate at more aggregated level (See Chapter 3 and 4). For design weights, this is due to the sampling design which contained over-sampling of those aged 80 or over with double inclusion probability. For modelling, the categorical grouping of age was chosen by the distinctively varying co-operation levels by age groups. Thus, the result suggests that age of the individual should be accounted for in further adjustment of weights, i.e. in the calibration. In addition, the results favour that the socio-economic grouping should be considered as another calibration variable. The geographical major areas are less important in terms of the bias reduction.

The indicator for interviewer impact shows that significantly the largest deviation in the impact can be observed in calibration by demographic variables. The smallest impact is observed for weighting method based on the IPW multinomial logit model without design weight adjustment. Of the calibrated weights, schemes using socio-economic or income inequality variables have improved performance in comparison to exploiting solely demographic information. All in all based on results in Table 6.1, the IPW sequential multilevel modelling allowing for the interviewer effects seems to be the most promising weighting method. However, in a multi-purpose survey there is typically no single optimal non-response weighting adjustment (Oh and Scheuren, 1983). Therefore, sensitivity analyses are needed to compare the inference and impact of alternative response mechanism specifications.

Table 6.1 Comparison of weighting methods by indicators (multiplied by 1000) for health interviewed

		Indicator for response influence by auxiliary variables					Interviewer impact	
Weighting method	**Variance inflating factor**	**Interviewer assignments (157 interviewers)**	**Local area (80 districts)**	**Major area (5 regions)**	**Socio-economic status (5 groups)**	**Age (10 year)**	**Change from original assignment allocation**	**Change from interviewer level response rate**
Design weights	17.0	22.1	3.9	1.8	35.9	90.1	101.4	65.9
Design weights with non-response correction at stratum level	75.4	32.1	21.2	1.9	67.1	152.5	110.5	89.2
Inverse response probability weighting (IRPW):								
- simple logit model	348.3	61.7	54.0	4.5	146.9	364.1	92.7	69.4
- simple logit model and design weight	346.6	63.3	54.4	4.3	148.9	370.7	104.6	84.5
- sequential logit	38.3	19.6	14.7	1.8	13.8	32.2	101.6	36.3
- sequential logit and design weight	43.6	20.2	15.6	1.8	14.1	32.8	126.3	88.1
- multinomial logit	424.1	59.3	52.5	4.5	153.8	393.2	88.8	68.4
- multinomial logit and design weight	424.9	61.5	52.8	4.4	156.1	400.8	102.3	80.9
- sequential multilevel logit model (2nd order PQL estimation)	32.8	25.2	19.0	2.4	27.6	74.8	107.7	34.1
- sequential multilevel logit model (2nd order PQL estimation) and design weight	37.8	25.9	19.9	2.4	28.1	76.1	122.7	76.3
- sequential multilevel logit model (MCMC estimation)	32.6	25.8	19.1	2.4	27.8	75.4	108.4	34.2
- sequential multilevel logit model (MCMC estimation) and design weight	37.7	26.4	20.0	2.4	28.3	76.7	122.3	76.1
Calibration of weights (for health interviewed)								
Demographic variables:								
- IRPW of sequential logit	88.4	9.5	7.8	0.5	6.8	15.7	296.8	283.0
- IRPW of multinomial logit (MCMC)	85.6	11.7	9.4	0.6	12.8	36.0	298.9	278.3
- design weight	77.7	12.3	9.0	0.5	15.1	42.0	285.7	271.0
- design weight * IRPW of sequential logit	88.7	9.5	7.8	0.5	6.7	15.6	296.0	281.7
- design weight * IRPW of multinomial logit (MCMC)	85.8	11.8	9.4	0.6	12.8	36.0	298.3	277.3
Socio-economic variables:								
- IRPW of sequential logit	69.4	9.2	6.9	0.4	7.1	16.3	101.4	40.5
- IRPW of multinomial logit (MCMC)	66.0	11.5	8.4	0.4	13.2	36.5	107.7	34.2
- design weight with demographic variables	64.8	11.3	7.2	0.3	13.2	33.9	107.1	61.0
- design weight with demographic variables * IRPW of sequential logit	73.5	9.6	7.4	0.4	7.3	16.8	127.0	89.1
- design weight with demographic variables * IRPW of multinomial logit (MCMC)	69.4	11.9	9.0	0.5	13.6	37.5	123.0	76.3
Income inequality variables:								
- IRPW of sequential logit	76.5	9.1	6.9	0.3	8.0	17.6	113.9	57.1
- IRPW of multinomial logit (MCMC)	73.5	11.5	8.5	0.4	14.0	36.7	115.8	47.4
- design weight with demographic variables	74.2	11.5	7.5	0.3	13.0	29.4	109.1	61.5
- design weight with demographic variables * IRPW of sequential logit	80.7	9.5	7.4	0.3	8.2	18.2	142.1	102.1
- design weight with demographic variables * IRPW of multinomial logit (MCMC)	77.2	11.7	9.4	0.6	12.8	36.0	135.5	88.9

6.5. Estimation of population totals and prevalence of long-term illnesses under survey non-response

In cross sectional health surveys one is generally interested on prevalence of illnesses in the population. Furthermore, the health inequalities between population domains are of specific concern in many studies. To reflect these aspects, we have linked directly a set of health register variables to the sample which are closely connected with some survey variables. We focus on relatively general long term diseases in the Finnish adult population that can lead to reduced physical capacity and life quality, and without proper treatment and medication they can be fatal. The motivation for the choice of variables is that they are both closely connected with survey variables, and give an indication on the use of social insurance and medical costs for specified diseases. As we can observe the diseases for respondents, non-respondents and the target population, we can at the same time study the initial non-response error, as well as the impact of sampling error, survey attrition, and non-response reduction efforts.

To simulate the non-response bias, we estimated the prevalence of severe long-term disease conditions from the Health 2000 survey sample members using directly linked data from health registers. We focus on assessing the stability of simulated survey estimates on diabetes mellitus, chronic cardiac insufficiency, connective tissue diseases, chronic asthma and similar chronic obstructive pulmonary diseases, chronic hypertension, and chronic coronary heart disease[5]. More precisely, the estimates indicate the number of people in the population diagnosed with the disease and having an entitlement for reimbursement on medical expenses due to the treatment of the disease. In addition, we assess health related estimates linked directly from the taxation register to the sampled individuals. These estimates contain information on socio-economic class of the individuals sampled, experience of unemployment spells, whether they received social benefits on disability for work (restricted to people aged 30-64), and entitlement for other social benefits. All variables are obtained for the target population as well as for responding and non-responding individuals.

We compare IPW weighting methods with complex survey design element to simple model based response propensity estimates and to calibration. For Horvitz-Thompson type estimation, we use both the estimated response probabilities from single level non-response models estimated in Chapter 4, and the multilevel models with interviewer effect, estimated in Chapter 5. The inverse probability weights are based on simple, multinomial, and sequential logit models as well as on sequential logit models with interviewer effects using 2nd order PQL and MCMC estimation. In addition, we use three alternative calibration schemes. The schemes are based on the use of auxiliary information in a set of variables characterised as demographic, socio-economic and income inequality variables. The five weights to be calibrated within each auxiliary information type are design weight, inverse probabilities from single level sequential logit models, multilevel sequential logit models (with MCMC estimation)[6], and their design weight adjusted products. In addition, we have derived all weights described above for the aggregated data collection phases to allow for assessing the impact of non-response and attrition to survey estimates. We examine how severely the survey attrition across data collection phases affects the survey estimates.

Oh and Scheuren (1983) have suggested to produce weighting coefficients for different data collection phases. This approach seems plausible due to the multiple data collection phases in the Health 2000 survey and varying non-response patterns. In Table 6.2, the estimation results are presented for chronic hypertension across data collection phases, comparing alternative weighting methods. In Appendix 6.5-6.9 similar estimates are given for other long term diseases. The setting allows us to compare the true population value with the estimates given by Horvitz-Thompson type weight estimation ignoring the sample design (Table 6.2.a) to more complex weighting that combines the estimated response probabilities and design weights (Table 6.2.b), and to calibration of weights with and without estimated response propensities (Table 6.2.c). In addition, the variables used in the calibration can influence the final estimates through weighting. We can observe from the tables that all weighted estimates for long term diseases studied tend to be biased, but the bias is varying to some extent by age and sex. Generally the estimates are mostly biased upwards. However, the estimates for chronic hypertension are biased downwards for elderly men. The variation of the estimates by the weighting method is minor for all age and sex groups, in comparison to the deviation of the level of the estimates from the true population values. Similarly, we can observe minor differences between the estimates based on sub-sets of co-operating sample across data collection phases. The prevalence estimates of chronic hypertension increases slightly for men and decreases more dramatically for women across data collection phases.

The variation of the estimates by weighting methods is relatively minor also for other health estimates except for the chronic cardiac insufficiency, shown in Appendix 6.6. For chronic cardiac insufficiency, the inverse probability weights arising from the simple and multinomial logit models add significantly to the bias of the estimates. Similar tendencies of upwards biased estimates, but with significantly smaller bias, can be observed for the prevalence estimates of connective tissue diseases, rheumatoid arthritis and comparable diseases based on inverse probability weights from simple and multinomial logit model, shown in Appendix 6.7. Comparing the prevalence estimates for all health estimates by ignoring or adjusting for the design weights, we cannot argue that the design weight adjustment would correct or even improve significantly the precision of the estimates.

---

[5] The information source is the health register maintained by The Social Insurance Institution.

[6] Preliminary analysis has shown that the weights based on the MCMC estimation perform slightly better than the ones based on 2nd order PQL method. Thus we present only the MCMC based weights representing the multilevel modelling of response propensities in the wider comparison of weights.

Table 6.2 Simulation of the ratio estimator for prevalence of chronic hypertension[7] by weighting methods for sub-sets of survey participation in the Health 2000 Survey[8]

(a) Inverse probability weighting

Prevalence estimates of chronic hypertension, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**1.26**	**10.40**	**27.49**	**26.01**	**12.82**	**1.25**	**9.26**	**29.45**	**36.23**	**15.30**
Unweighted sample estimates										
Sampled	3.31	12.69	30.61	19.12	15.07	1.65	13.44	33.54	39.02	19.95
Health interviewed	3.47	12.43	30.44	18.92	15.14	1.59	13.58	33.74	38.57	19.73
Symptom interviewed	3.42	12.62	31.20	19.23	15.42	1.27	13.47	33.80	39.66	19.31
Medical measurements	3.51	12.46	31.05	19.35	15.23	1.16	13.55	33.14	36.79	18.07
Full response	3.68	13.03	30.77	19.59	15.38	1.07	13.39	31.89	36.24	17.09
Weighted by inverse of response probabilities from simple logit model										
Sampled	3.05	12.85	28.34	16.67	14.61	1.78	15.45	33.21	32.85	22.57
Health interviewed	3.39	12.02	30.68	17.69	15.30	1.58	13.49	34.38	38.66	23.63
Symptom interviewed	3.37	12.40	31.51	18.63	15.73	1.31	13.39	34.54	41.79	23.20
Medical measurements	3.49	12.20	31.64	17.02	15.37	1.13	13.49	33.90	36.56	20.06
Self-completion questionnaires	3.68	12.87	31.80	18.02	15.66	1.05	13.27	32.76	37.94	19.19
Weighted by inverse of response probabilities from sequential logit model										
Sampled	3.38	12.83	30.68	19.76	15.01	1.67	13.45	32.79	39.86	18.12
Health interviewed	3.49	12.53	30.78	18.87	15.08	1.57	13.61	33.23	39.66	17.97
Symptom interviewed	3.43	12.57	31.23	19.82	15.26	1.20	13.44	33.08	39.13	17.62
Medical measurements	3.51	12.41	31.00	20.14	15.09	1.19	13.54	32.46	38.24	17.11
Self-completion questionnaires	3.66	12.99	30.63	20.30	15.29	1.09	13.47	31.25	36.46	16.24
Weighted by inverse of response probabilities from multinomial logit model										
Sampled	3.04	11.99	30.21	18.41	14.82	1.67	13.27	34.10	38.31	23.88
Health interviewed	3.36	12.02	30.55	18.53	15.35	1.63	13.46	34.42	37.87	23.61
Symptom interviewed	3.33	12.37	31.34	19.51	15.71	1.36	13.35	34.54	40.04	22.94
Medical measurements	3.46	12.21	31.45	17.23	15.31	1.14	13.43	33.91	35.68	20.02
Self-completion questionnaires	3.65	12.88	31.64	18.16	15.60	1.05	13.20	32.72	37.43	19.15
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL)										
Sampled	3.39	12.80	30.86	19.31	15.00	1.69	13.41	33.18	39.88	18.71
Health interviewed	3.58	12.48	30.65	18.99	15.01	1.61	13.55	33.39	39.39	18.28
Symptom interviewed	3.46	12.51	31.18	19.64	15.16	1.25	13.38	33.28	39.56	17.79
Medical measurements	3.55	12.31	30.95	19.32	14.94	1.19	13.47	32.52	37.88	17.04
Self-completion questionnaires	3.70	12.87	30.53	19.76	15.12	1.11	13.40	31.28	36.44	16.18
Weighted by inverse of response probabilities from multilevel logit model (MCMC)										
Sampled	3.42	12.79	30.82	19.37	14.99	1.69	13.40	33.19	39.82	18.70
Health interviewed	3.60	12.48	30.59	19.05	15.00	1.61	13.55	33.40	39.30	18.28
Symptom interviewed	3.49	12.51	31.14	19.69	15.15	1.25	13.38	33.30	39.51	17.78
Medical measurements	3.58	12.32	30.92	19.37	14.93	1.19	13.47	32.54	37.85	17.04
Self-completion questionnaires	3.73	12.87	30.50	19.81	15.12	1.11	13.39	31.29	36.42	16.18

[7] Disease which entitled the patient to receive reimbursement of medicine costs under the Higher Special Refund Category; Disease code 205.

[8] Source: Individual level data linkage from health registers of National Pension Institute for the sample and target population of the Health 2000 survey. The population totals have been derived separately according to the definitions of the target population from the same register.

(b) Inverse probability weighting with design weight adjustment

Prevalence estimates of chronic hypertension, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**1.26**	**10.40**	**27.49**	**26.01**	**12.82**	**1.25**	**9.26**	**29.45**	**36.23**	**15.30**
Weighted by design-weights										
Sampled	3.28	12.64	30.38	19.14	14.96	1.59	13.33	33.02	38.62	19.70
Health interviewed	3.45	12.36	30.23	18.84	15.02	1.55	13.47	33.15	38.17	19.46
Symptom interviewed	3.43	12.53	30.99	19.06	15.28	1.24	13.35	33.18	39.38	19.05
Medical measurements	3.52	12.35	30.84	19.20	15.10	1.15	13.42	32.59	36.72	17.86
Self-completion questionnaires	3.71	12.91	30.52	19.31	15.25	1.09	13.25	31.35	36.61	16.93
Weighted by design-weights with non-response adjustment										
Sampled	3.22	12.70	30.40	19.02	15.17	1.59	13.31	32.83	38.40	21.92
Health interviewed	3.38	12.40	30.24	18.62	15.20	1.55	13.43	32.95	37.92	21.52
Symptom interviewed	3.37	12.55	31.02	18.74	15.43	1.24	13.31	32.98	39.07	21.00
Medical measurements	3.47	12.37	30.86	18.85	15.23	1.16	13.38	32.40	36.48	19.20
Self-completion questionnaires	3.67	12.94	30.47	18.88	15.35	1.10	13.23	31.22	36.49	18.17
Weighted by inverse of response probabilities from simple logit model and design weight										
Sampled	3.10	12.02	30.40	17.51	14.86	1.69	13.39	34.55	39.37	24.16
Health interviewed	3.42	12.09	30.81	17.58	15.38	1.64	13.61	35.01	39.21	23.97
Symptom interviewed	3.37	12.50	31.67	18.68	15.85	1.36	13.51	35.21	42.29	23.54
Medical measurements	3.49	12.32	31.85	17.04	15.50	1.15	13.62	34.47	36.54	20.25
Self-completion questionnaires	3.66	12.99	32.09	18.19	15.79	1.04	13.40	33.32	37.43	19.29
Weighted by inverse of response probabilities from sequential logit model and design weight										
Sampled	3.40	12.86	30.96	19.71	15.13	1.73	13.57	33.34	40.13	18.35
Health interviewed	3.50	12.61	31.02	18.99	15.21	1.61	13.74	33.85	39.94	18.23
Symptom interviewed	3.42	12.67	31.48	19.93	15.40	1.22	13.58	33.72	39.36	17.89
Medical measurements	3.50	12.52	31.24	20.26	15.23	1.20	13.68	33.06	38.32	17.34
Self-completion questionnaires	3.63	13.11	30.93	20.52	15.43	1.08	13.62	31.83	36.05	16.44
Weighted by inverse of response probabilities from multinomial logit model and design weight										
Sampled	3.07	12.05	30.35	18.17	14.90	1.75	13.37	34.63	38.77	24.17
Health interviewed	3.39	12.10	30.66	18.35	15.43	1.70	13.58	35.05	38.38	23.94
Symptom interviewed	3.33	12.48	31.49	19.55	15.83	1.42	13.47	35.22	40.54	23.27
Medical measurements	3.45	12.33	31.62	17.35	15.45	1.15	13.57	34.49	35.72	20.21
Self-completion questionnaires	3.62	13.00	31.90	18.44	15.73	1.04	13.33	33.29	36.97	19.26
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL) and design weight										
Sampled	3.42	12.82	31.13	19.18	15.11	1.76	13.52	33.72	40.36	18.96
Health interviewed	3.59	12.55	30.90	19.02	15.14	1.66	13.67	34.03	39.84	18.55
Symptom interviewed	3.45	12.60	31.45	19.68	15.30	1.28	13.51	33.94	39.93	18.06
Medical measurements	3.54	12.43	31.21	19.39	15.08	1.20	13.61	33.12	37.97	17.27
Self-completion questionnaires	3.67	12.99	30.83	19.92	15.26	1.10	13.55	31.86	36.05	16.38
Weighted by inverse of response probabilities from multilevel logit model (MCMC) and design weight										
Sampled	3.45	12.82	31.09	19.24	15.11	1.76	13.51	33.73	40.28	18.95
Health interviewed	3.62	12.55	30.84	19.09	15.13	1.65	13.67	34.03	39.73	18.55
Symptom interviewed	3.48	12.61	31.40	19.71	15.29	1.27	13.51	33.96	39.87	18.06
Medical measurements	3.57	12.43	31.17	19.43	15.08	1.20	13.61	33.13	37.94	17.26
Self-completion questionnaires	3.70	13.00	30.80	19.96	15.26	1.09	13.54	31.87	36.03	16.37

(c) Calibration with alternative auxiliary information structure for the health interviewed

Prevalence estimates of chronic hypertension, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**1.26**	**10.42**	**27.52**	**26.04**	**12.84**	**1.26**	**9.27**	**29.48**	**36.28**	**15.32**
Unweighted sample estimates										
Sampled	3.31	12.69	30.61	19.12	15.07	1.65	13.44	33.54	39.02	19.95
Prevalence estimates of chronic hypertension, % - weighted by calibrated weights based on:										
Design weights										
Health interviewed:										
Demographic variables	3.69	12.70	30.90	21.43	15.21	1.96	14.28	35.19	42.21	19.71
Socio-economic variables	3.54	12.43	32.03	17.32	15.19	1.71	13.98	34.59	37.91	19.04
Income inequality variables	3.59	12.35	30.15	20.27	14.80	1.55	13.79	33.97	38.07	18.76
Design weights and sequential logit response probability										
Health interviewed:										
Demographic variables	3.64	12.81	31.46	20.24	15.72	1.93	14.12	34.29	41.64	18.94
Socio-economic variables	3.80	12.66	33.20	18.45	16.19	1.72	13.76	35.30	40.16	19.57
Income inequality variables	3.86	13.20	30.49	17.22	15.69	1.55	14.41	31.57	33.59	18.49
Design weights and multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	3.84	12.74	31.22	20.10	15.61	1.98	14.02	34.54	41.15	19.25
Socio-economic variables	3.90	12.51	32.83	17.62	15.79	1.79	13.70	35.38	39.96	19.65
Income inequality variables	3.97	13.21	30.23	16.94	15.60	1.62	14.45	31.72	33.52	18.77
Sequential logit response probability										
Health interviewed:										
Demographic variables	3.65	12.80	31.41	20.27	15.70	1.93	14.12	34.24	41.69	18.93
Socio-economic variables	3.77	12.62	32.95	18.37	16.09	1.67	13.67	34.68	39.84	19.34
Income inequality variables	3.85	13.14	30.32	17.17	15.57	1.52	14.27	31.12	33.41	18.27
Multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	3.85	12.73	31.17	20.14	15.59	1.98	14.03	34.48	41.18	19.24
Socio-economic variables	3.86	12.47	32.58	17.64	15.69	1.73	13.61	34.75	39.49	19.41
Income inequality variables	3.96	13.15	30.04	17.03	15.49	1.58	14.31	31.27	33.23	18.54

As the measurement error can be excluded in register based simulation, the estimation results suggest that the analysis data is contaminated by systematic error, arising either from the use of a combination of multiple register sources or from the sampling. Comparing with the inverse probability weighting, calibration of weights does not provide improvement for the biased estimators regardless the alternative auxiliary information sets. However, if we examine estimates for background factors on health inequalities and health related social benefits based on the taxation register, the estimates are less biased. Further assessment of the sampling design would require simulated samples and replicated analysis of this type. This assessment should also contain alternative sampling designs that aim to account for the health inequalities of the population.

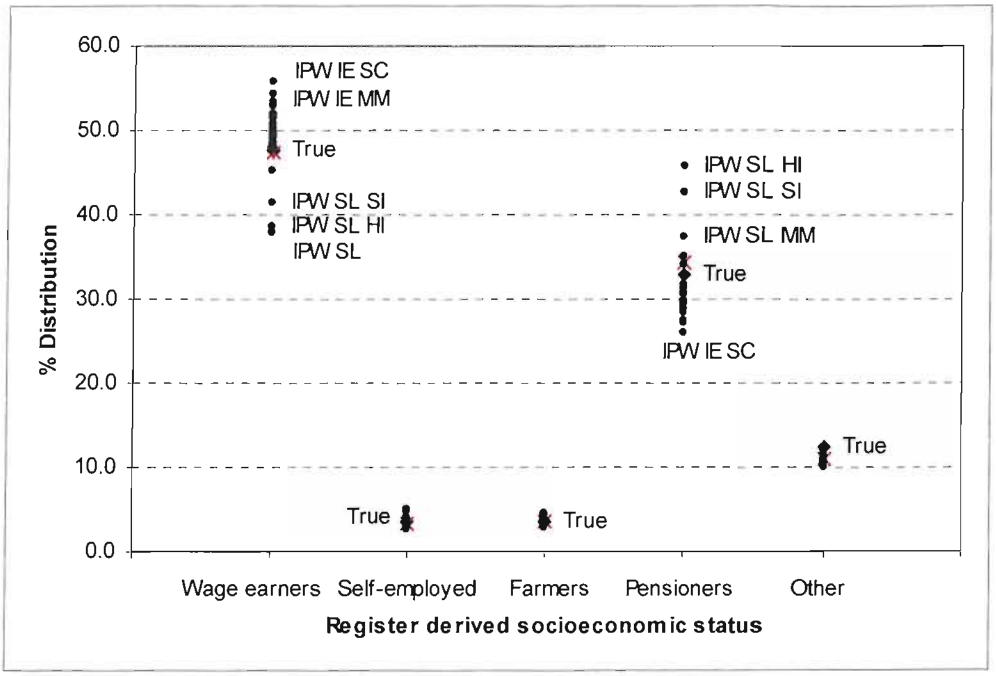
In Figure 6.1, we show the distribution of the socio-economic status by weighting methods. The sample and true population values are very close to each other. Extreme estimates are given by the simple logit based inverse probability weighting, and by sequential logit allowing for interviewer effects. The Figure 6.1 demonstrates that although the sample and true population values are very close to each other, weighted distributions can still have large deviations. For the categorical groups of socio-economic status, all weighting methods give precise estimates for some groups, while the proportional sizes of other groups vary considerably. Weighting methods adjusting for interviewer effects, for example, over estimate the proportion of wage earners and under estimate the proportion of pensioners.

In Figure 6.2, we examine the performance of the weighting methods for the health estimates comparing the range of inverse probability weighted as well as calibration weighted estimates to the true population value. We observe from Figure 6.2 that although the inverse probability weighted estimates are biased upwards for most sex and age groups, the calibration of weights does not reduce the bias. Most strikingly, it can be observed from the Figure 6.2 that the bias of the estimates is varying by survey estimates. This is worrying in terms of estimating those health conditions that do not have population totals available for controlling the reliability of the survey estimates. The finding also underlines the need for further assessment of the total survey error.

The size of the bias varies slightly by selected long term diseases with reimbursement entitlement on medical expenses as well as by age and sex groups. For example, the estimates of people receiving medical reimbursement for chronic cardiac insufficiency are precise for men and women aged 30 to 79, but with positive bias for elderly men and negative bias for elderly women. On the other hand, for connective tissue diseases, the positive bias is significant for elderly men and across all age groups of women. The estimates of chronic asthma behave similarly, but the magnitude of bias is lower. The prevalence of chronic hypertension is slightly over estimated from the sample for all other age and sex groups except for elderly men for whom the prevalence is underestimated by all weighting methods. Thus we can conclude that the simulated prevalence estimates show that survey estimates are contaminated with different types of errors. For finding the reason for survey bias and its source of variation, we have examined the estimates also by regions as it was possible to link to the register information by local areas.

In Figure 6.3.a, we show the unweighted estimates for long term diseases across geographical primary sampling units (PSU), which represent the 15 largest cities (PSUs 6 to 21) and the remaining university hospital districts (PSUs 1 to 5) of the country. It can be observed that the variation of the bias by diseases is large, especially within the 15 largest towns. Largest variation in the bias can be observed for the chronic hypertension and chronic coronary heart disease. In every region some health estimates are unbiased. However, the diseases with precise estimates vary across the PSUs. In Chapter 3 it has been described that also the sample size varies across the PSUs. From Figure 6.3.b we can observe that also the design weighted estimates vary largely. Our conclusion is that the possibility of sampling error cannot be excluded. It is possible that the sample sizes within clusters have not been adequate in relation to the within region health inequalities. Also, the ecological population may differ across and within clusters by significant latent characteristics that remain unquantifiable. However, we must also consider that the simulated variables contain an element of assuming that people apply for reimbursement they are entitled to. Tendency for this type of civil activity can also have regional variation. Therefore for further analysis we have included two health related variables from taxation register. In Section 6.6, we examine the bias of survey estimates on people receiving unemployment pension or care support with the long term disease estimates.

Figure 6.1 Estimated distribution of register derived socio-economic class by alternative weighting methods



Source: Rough socio economic status has been derived from the taxation register information for each individual sampled and to the entire target population noted by "True".

Abbreviations used in the Figure 6.1 for weighting methods:

IPW IE SC – Inverse probability weighting, multilevel sequential logit allowing for interviewer effects (2nd order PQL estimation) for those returning self-completions

IPW IE MM - Inverse probability weighting, multilevel sequential logit allowing for interviewer effects (2nd order PQL estimation) for those participating to medical measurements

IPW SL SI - Inverse probability weighting, simple logit for those interviewed on their symptoms

IPW SL HI - Inverse probability weighting, simple logit for those interviewed on their health

IPW SL - Inverse probability weighting, simple logit for sampled individuals

IPW SL MM - Inverse probability weighting, simple logit for those participating to medical measurements

Figure 6.2 True prevalence, inverse probability weighted ratio estimators and calibration estimates of health interviewed for prevalence of long-term diseases

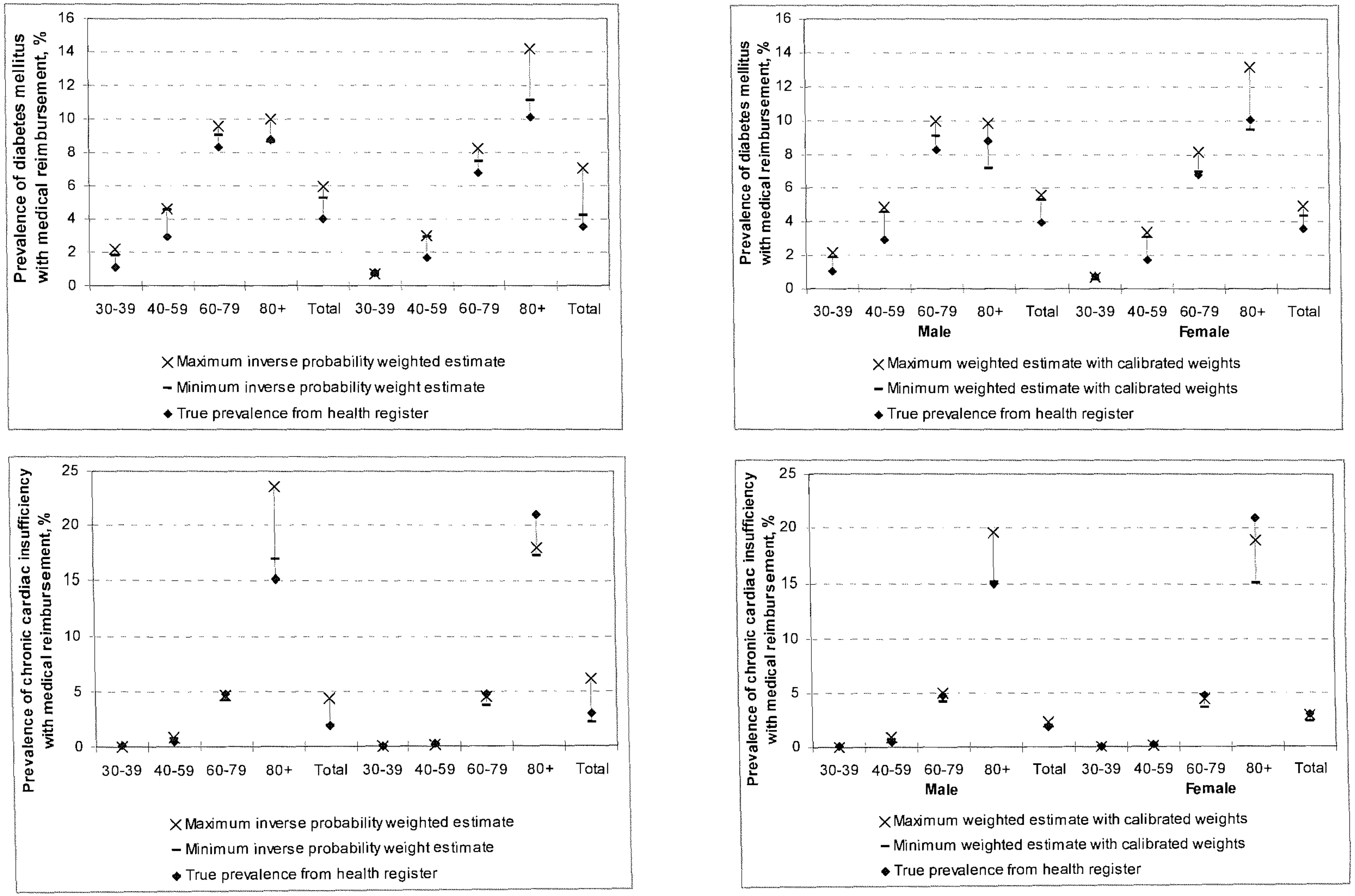
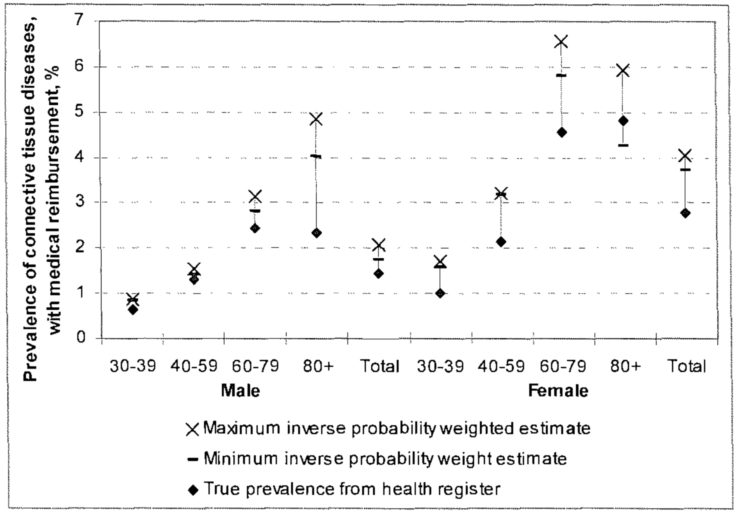
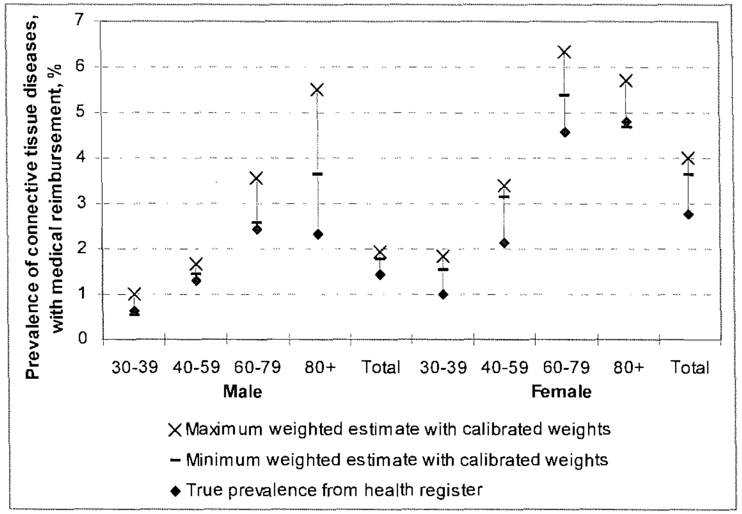
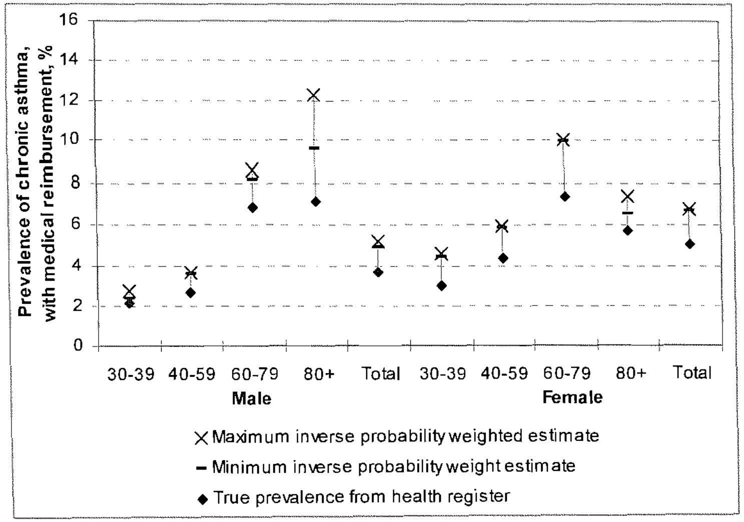


Figure 6.2 continues







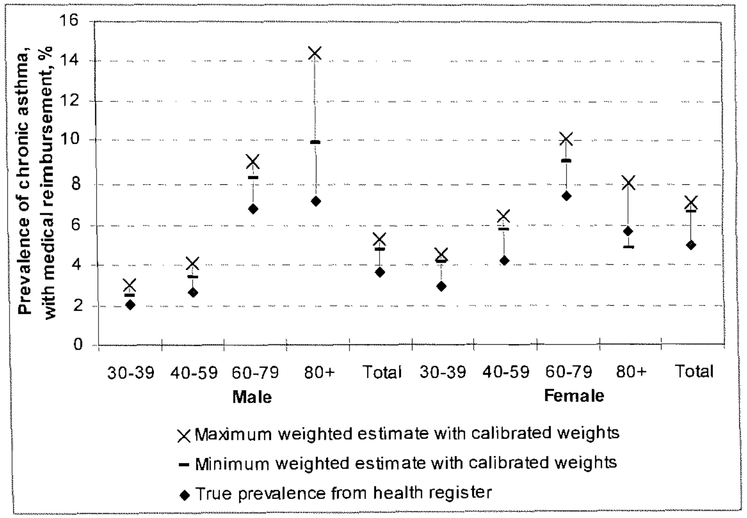


Figure 6.2 continues

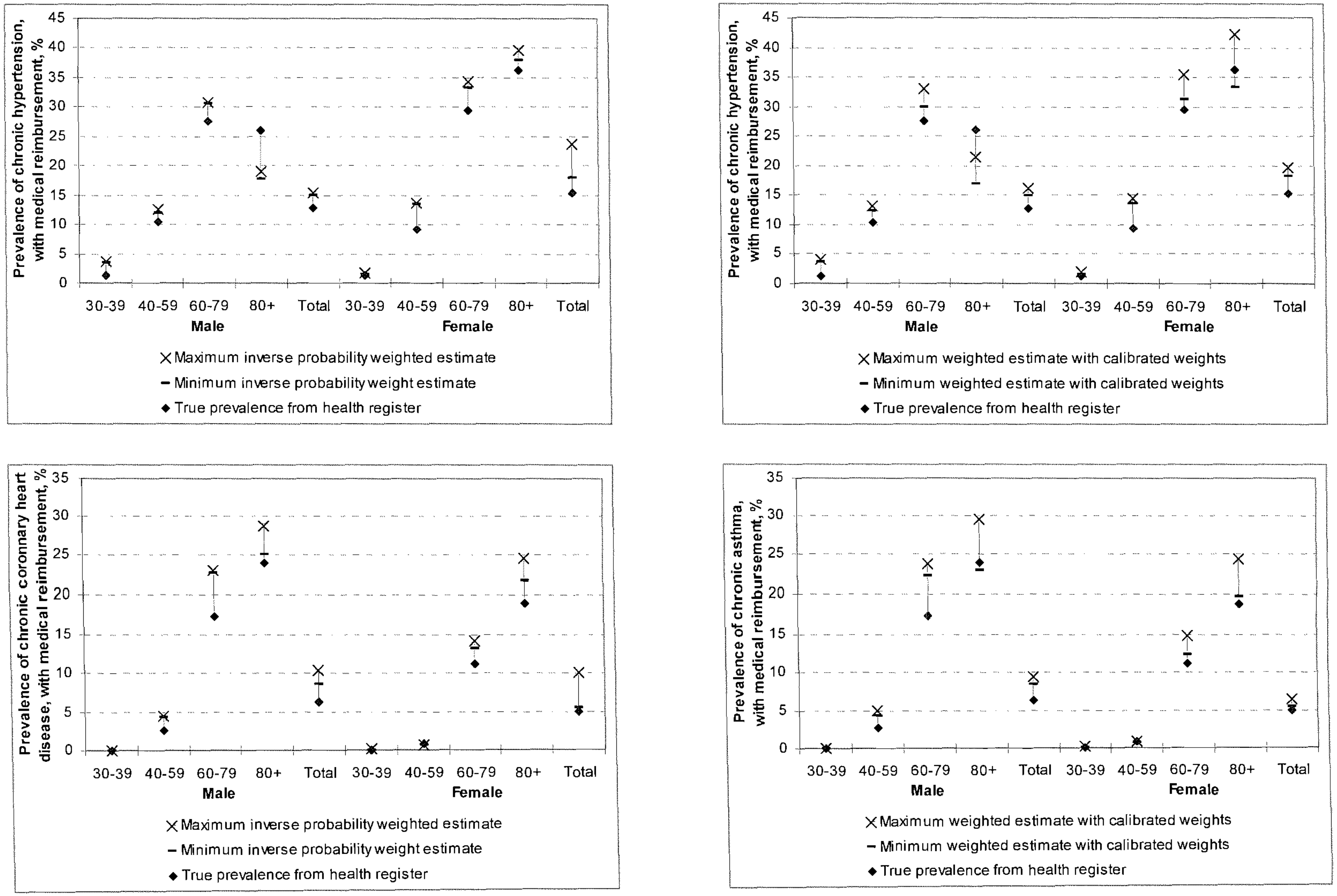
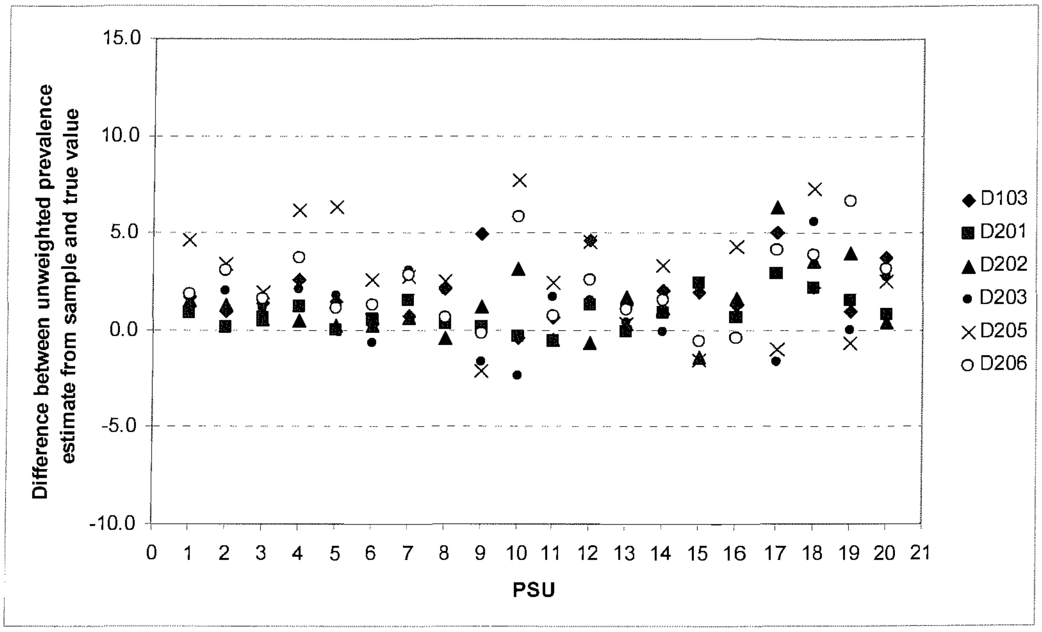
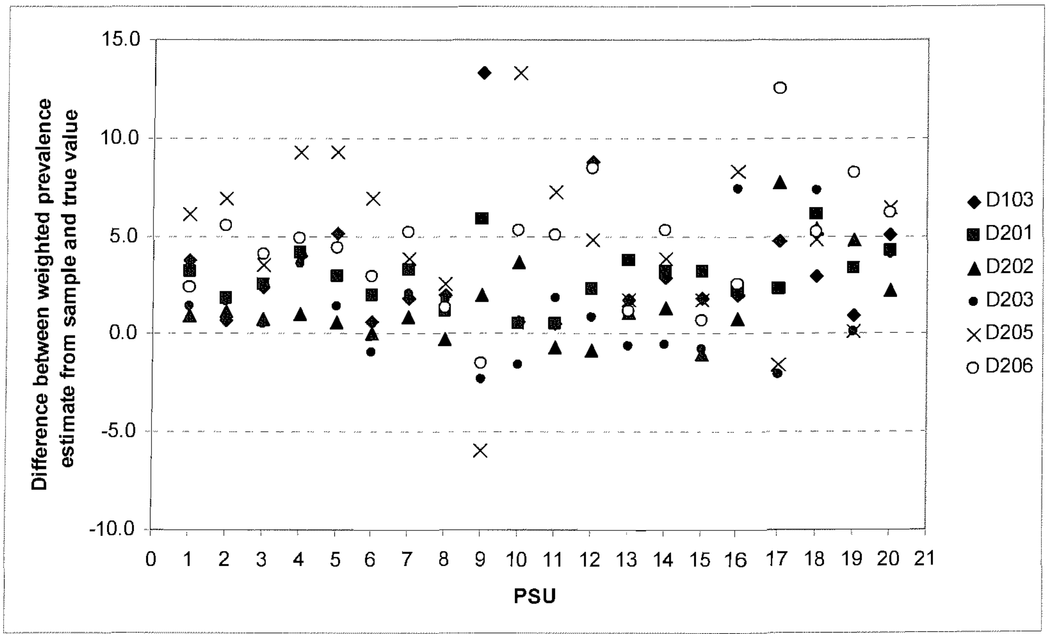


Figure 6.3 Unweighted and design weighted prevalence from the true prevalence of long term diseases[1)] by regional primary sampling units

(a) unweighted



(b) weighted by design weight



1) Codes for long term diseases:
D103 Diabetes mellitus
D201 Chronic cardiac insufficiency
D202 Connective tissue diseases, rheumatoid arthritis and comparable diseases
D203 Chronic asthma and similar chronic obstructive pulmonary diseases
D205 Chronic hypertension
D206 Chronic coronary heart disease

6.6. Analysis of error

In this Section we will assess the bias of the estimates and compare the weighting methods and the auxiliary information vectors. We examine the bias correction of non-response reduction efforts and examine the impact of survey attrition across data collection phases. We aim to minimize the bias of the estimates by using efficient and informative auxiliary information in the final estimate $\hat{Y}$. Särndal and Lundström (2005) stress that standard adjustment variables such as age, sex and region may be inadequate for this purpose and they encourage seeking more informative variables. The auxiliary information we have used for estimating $\hat{\pi}_i$ is described more in detail in Chapters 3, 4 and 5 for IPW methods, and in the Appendix 6.2 – 6.4 for calibration.

The prevalence estimator of a long-.term illness can be postulated using the notation for the population mean $\bar{Y}$ which is defined as

$$\bar{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k \, ,$$

and it is estimated by the weighted sample mean if weights are scaled to the population level

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{n} w_i y_i$$

or more realistically in the presence of non-response

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{r} w_i^* y_i \, ,$$

where w_i is the estimate of the response probability from the logistic regression suggested by Little (1986), and the w_i^* is the scaled weight of the respondents.

The weight of the prevalence estimator of long-term disease can be defined as a product of the sample weight (inverse of the inclusion probability) and the inverse of the estimate for the unknown response probability $\pi_{r|si}$ of individual i as:

$$\hat{\bar{Y}} = \frac{1}{N}\sum_{i=1}^{m} \frac{1}{\pi_{si}} \frac{1}{\hat{\pi}_{r|si}} y_i \, ,$$

where it is assumed that the response probabilities are independent of the realised sample s (Lundström and Särndal, 2001).

As we can use the auxiliary information from the registers we can compare the simulated survey prevalence estimates to the true values. Thus the difference i.e. contrast c between respondents r and non-respondents nr on the survey prevalence estimate $\bar{y}$ can be formulated as:

$$c = \bar{y}_r - \bar{y}_{nr},$$

and

$$E(\bar{y}_r) = \bar{Y}_r.$$

If $\bar{Y}_r$ differs significantly from $\bar{Y}$ then $\bar{y}_r$ is biased estimator of $\bar{Y}$ and the bias $B(\bar{y}_r)$ is equal to

$$B(\bar{y}_r) = E(\bar{y}_r - \bar{Y}) = \bar{Y}_r - \bar{Y}. \tag{6.33}$$

Impact of the non-response reduction efforts on the precision of the estimates

In this Section we examine the impact of non-response reduction efforts on correcting the survey prevalence estimates. We use the simulated health information from registers comparing the distributions of initial and converted respondents. Direct data linkage from registers and deriving totals for the exact target population allows us to simulate the impact of non-response reduction on health related variables. We use a measure of relative impact of the absolute values of survey bias contrasting the bias of the estimate of health interviewed to attained sample by original fieldwork of interviewers and further non-response reduction:

$$IND_B = \frac{|\bar{y}_{r_1} - \bar{y}|}{|\bar{y}_{r_1+r_2} - \bar{y}|} = \frac{|c_{r_1}|}{|c_{r_1+r_2}|}, \tag{6.34}$$

where r_1 indicates respondents to health interview and r_2 respondents to non-response reduction efforts. The improvements of the non-response reduction can be seen if the value of the relative indicator is above unity. Similarly, increased bias is detected by values below the unity.

We can observe from the Figure 6.4 that the relative impact of non-response reduction varies slightly across eight simulated survey estimates[9] and five alternative weighting methods. The three weighting methods based on Horvitz Thompson type estimation exploit design weight, inverse probability from sequential logit model, and multilevel sequential logit model with interviewer effects. In comparison, two calibration methods both use income inequality factors as auxiliary information contrasting the starting weights of design weight to inverse probability from multilevel sequential logit model with interviewer effects. We review the impact of non-response reduction by sex and age groups due to their importance as background factors in health analysis. The observable impact of non-response reduction also varies by sex and age groups for different variables. For chronic hypertension all estimates are improved by non-response reduction for younger females and males, while for connective tissue diseases all weighting methods produce improved estimates for middle aged men. In contrary, all weighting methods, except the calibration of design weight, tend to give more biased diabetes prevalence estimates for elderly men. Also the

[9] We examine the simulated survey estimates of diabetes mellitus, chronic hypertension, chronic coronary heart disease, chronic cardiac insufficiency, connective tissue disease, chronic asthma, care support due to physical or mental impediments, and receiving unemployment pension (for the working age population).

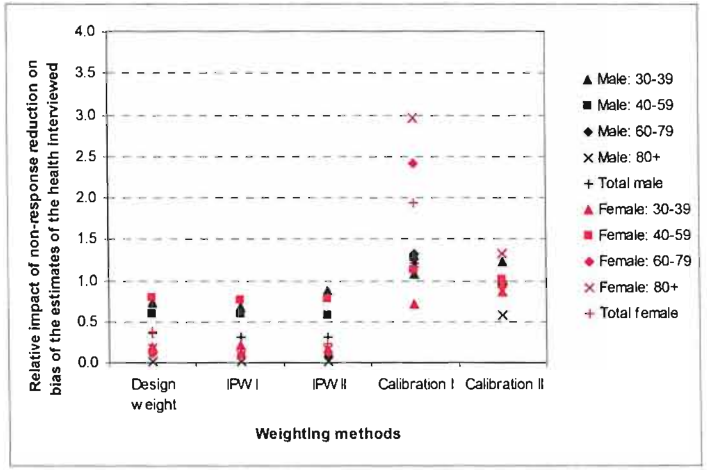
proportion of elderly men receiving care support has an increasing bias, if estimation is based also on those people who did not participate until the non-response reduction efforts.

The non-response reduction efforts provide an improvement for the prevalence estimate of diabetes, when using calibration based on design weights with income inequality variables for all sex and age groups except for young females. Also the chronic coronary heart disease estimates have slightly improved estimates by all weighting methods, with the exception for the elderly men and women when using design weight based calibration. The inverse probability weighting with interviewer effects or calibration with income inequality variables seem to be slightly more efficient than other weighting methods. However, the non-response reduction does not provide improvement for chronic hypertension estimates among the elderly and the middle aged. In contrary, the total estimates by sex groups have an increasing bias when estimates are based also on those who were converted in non-response reduction efforts. The estimates of the chronic cardiac insufficiency are presented at log-scale in Figure 6.4.d. The results are dominated by the effect of the precise or exact estimate after non-response reduction for total of males when using calibration with design weight. For connective tissue diseases, there is a similar impact observed for the middle aged men when calibrating the inverse probability allowing for the interviewer effects.

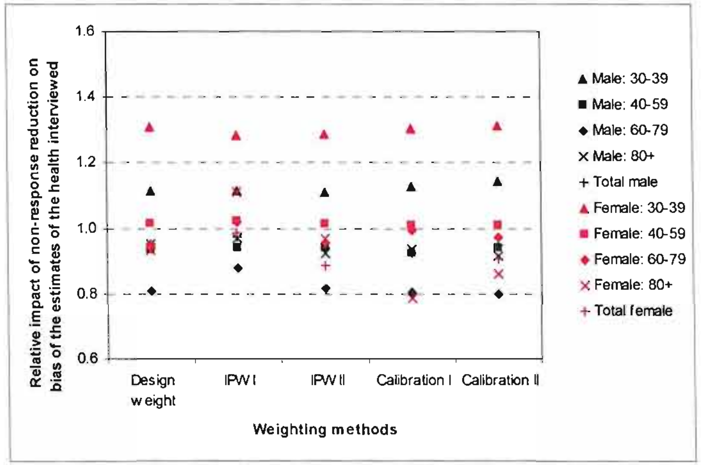
Comparing the performance of the weighting methods, we find that the calibration of design weights can result both with the highest bias correction and with the highest increase in bias of the estimates, depending on the estimated variable. The impact of non-response reduction on the bias of the estimates is seemingly low for estimates weighted by the inverse probability arising from the sequential logit models. The most stable weighting methods for the sub-data containing both the health interviewed and converted non-respondents are inverse probability and calibration based on multilevel sequential logit with interviewer effects. To conclude, we find that the impact of the non-response reduction on survey estimates varies largely by sex and age group, and the variable to be analysed. There are no obvious gains observed from the non-response reduction for all groups. As the non-response reduction has also shown to increase the bias and the achieved information structure in real data can be contaminated with measurement and context effects, we do not recommend attempts to merge this part of the data to the analysis with the originally obtained data. We emphasize that for most survey variables in the real data, the consequences cannot be controlled or simulated for, due to the lack of available population level control variables.

Figure 6.4 Relative impact of non-response reduction on the bias of estimated proportion of people receiving reimbursement of medical expenses[1)]

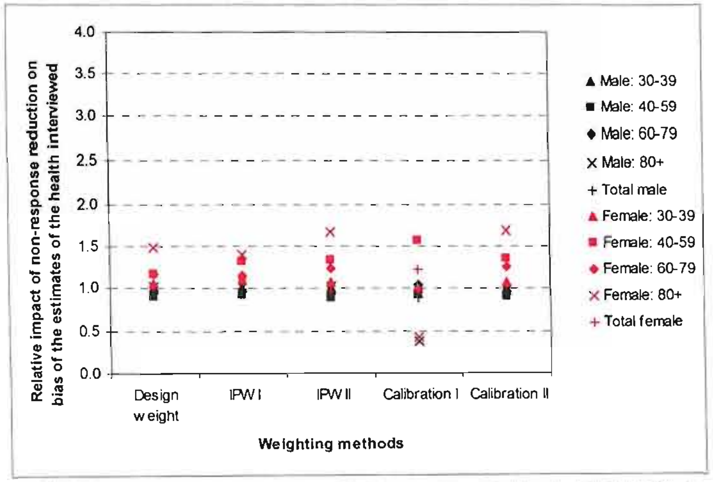
(a) Diabetes mellitus



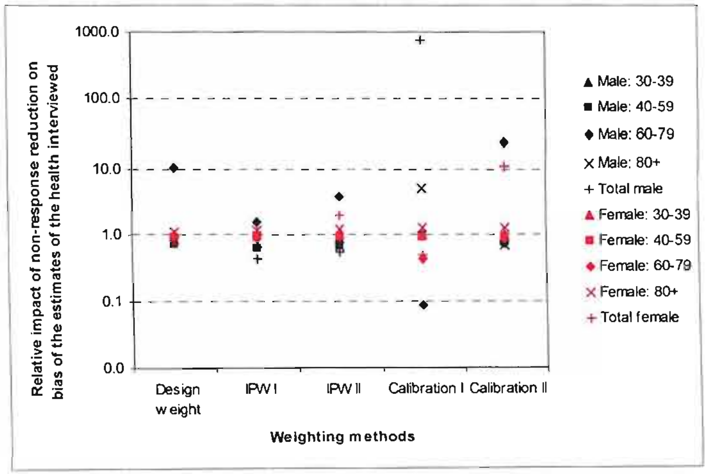
(b) Chronic hypertension



(c) Chronic coronary heart disease

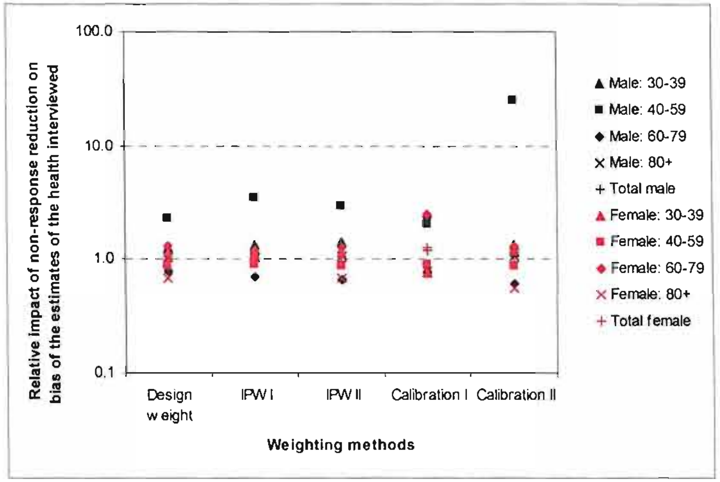


(d) Chronic cardiac insufficiency

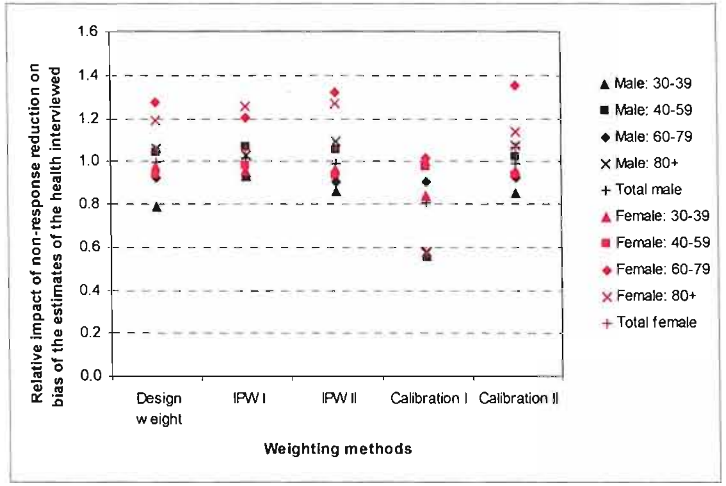


2) IPW I: Inverse response propensity from sequential logit; IPW II with interviewer effects using MCMC; Calibration I with design weights; Calibration II with inverse probability weights with interviewer effect

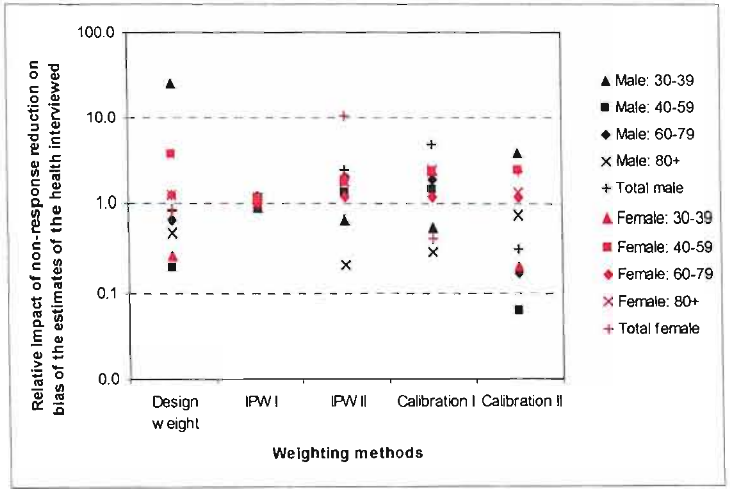
(e) Connective tissue diseases



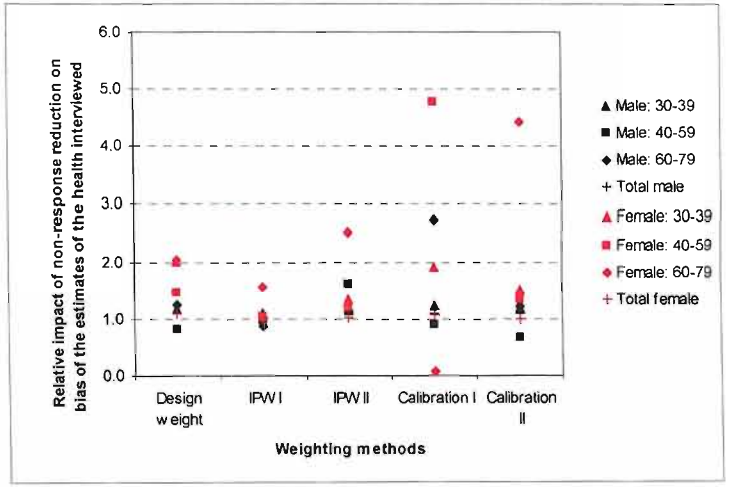
(f) Chronic asthma



(g) Care support due to physical or mental impediments [2)]



(h) receiving unemployment pension [2) 3)]



[2)] Source: Individual level data linkage from the register of The Social Insurance Institution for the sample, and target population totals derived from the taxation register file

[3)] Proportion of working aged people receiving unemployment pension due to long term unemployment and incapability to work

Analysis of error across data collection phase

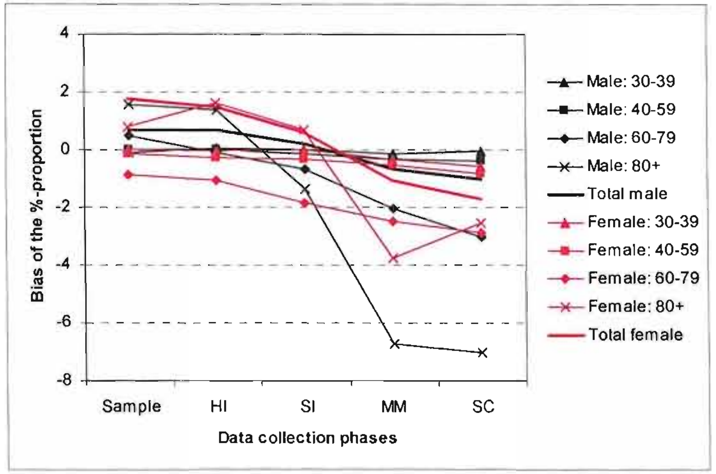
We return back to the research setting of the multiphase survey with aggregated data collection phases implemented previously in this Thesis in Chapter 4 and 5. We examine the development of bias of the simulated survey estimates across the phases. We use the health estimates of the previous Sections and compare the performance of four weighting methods. Horvitz Thompson type estimation based on design weights, inverse probability from sequential logit and multilevel sequential logit allowing for interviewer effects are compared to calibration of the weight from the latter. Bias of the proportional prevalence estimates is calculated simply by (6.33). Graphical presentation of the survey bias is given in Figures 6.5, 6.6 and 6.7, and in the Appendices 6.10 to 6.14 for the same variables analysed in the previous section.

We can observe from the Figure 6.5 that all weighting methods have increasing problem with bias of the care support estimate. There is a severe problem with under estimation of the elderly receiving care support. This suggests that elderly people co-operating with the survey request fully have less physical and mental impediments than their counterparts who are unable to co-operate fully with the survey request. The survey design described earlier in Chapter 3, had the element of over sampling the elderly people aged 80 or over. However, our finding suggests that over sampling a sub-population with increasing health problems can increase their absolute number in the final obtained data and thus increase arbitrarily the accuracy indicators of the estimates. However, over sampling does not reduce the actual bias of the estimates due to self-selection of individuals, which is more of a severe problem.

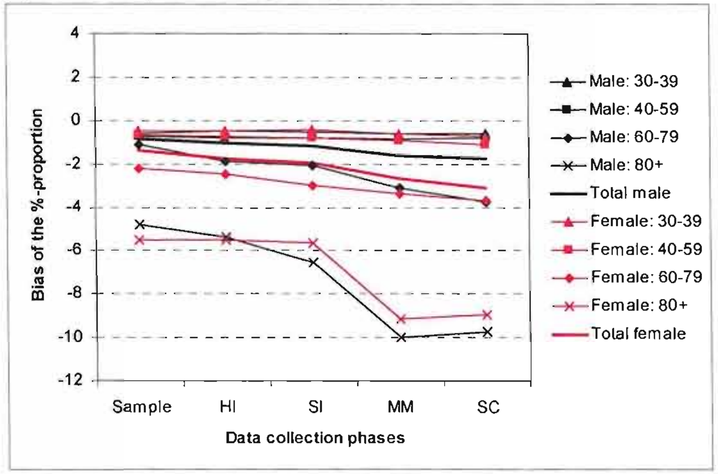
In Figure 6.6, we examine the simulated estimate of chronic cardiac insufficiency. The sample has an outstanding over-representation of elderly men with this long term disease while other sex and age groups have almost unbiased estimates at the sample level. The impact of the survey attrition, however, reduces the positive bias of the elderly men towards the estimates of the other sex and age groups. Thus, the initial bias is almost overruled by the non-response error of the elderly men. Completely opposite development can be observed from the Figure 6.7 examining the estimates of connective tissue diseases. The survey attrition is inversely related with the diagnosis of the disease for elderly men. Similar phenomena can be observed from Appendix 6.14 with the prevalence estimates of chronic asthma. The prevalence estimates increase significantly across the data collection phases. Thus we presume that some long term diseases do not reduce the survey co-operation, and that the effect can be strongly the opposite than generally presumed. This has also been supported by findings of Duhart et al. (2001) that the co-operation to the health surveys can also be positively associated with bad health of the elderly as they may have more interest to participate when having more health-related incidents to report. However, when the diseases of the individuals are severely limiting their capabilities to participate, we can observe the opposite effect leaving more healthy people co-operating across data collection with increasing response burden.

Figure 6.5 Bias of estimated proportion of people receiving care support across data collection phases

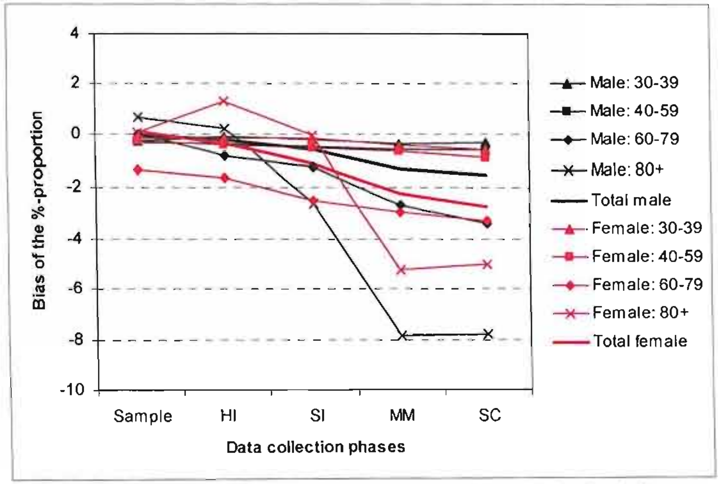
(a) Design weight



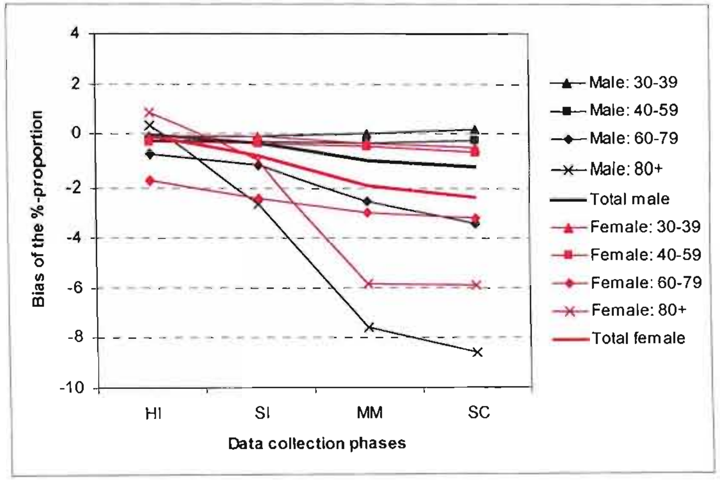
(b) Inverse probability weight (IPW)



(c) Inverse probability weight with interviewer effects (MCMC)



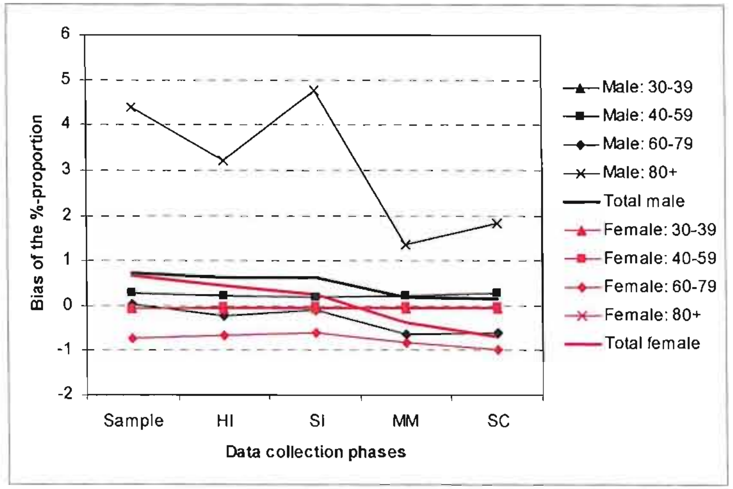
(d) Calibration of IPW with interviewer effects



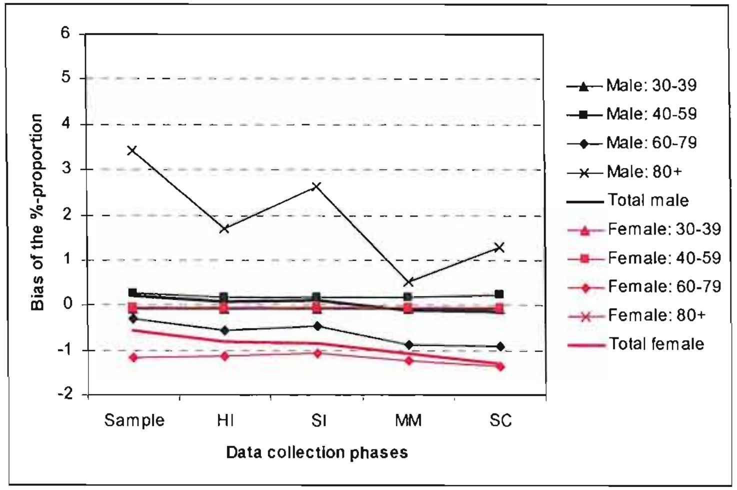
Source: Individual level data linkage from the register of The Social Insurance Institution for the sample, and target population totals derived from the taxation register file

Figure 6.6 Bias of estimated proportion of people receiving medical reimbursement on chronic cardiac insufficiency

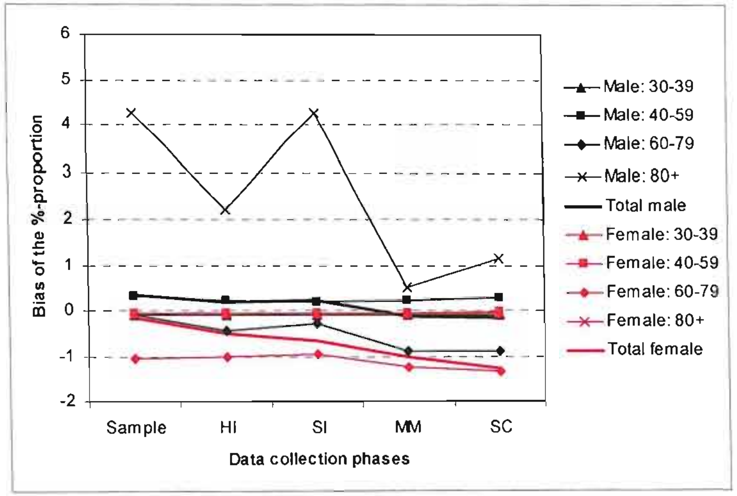
(a) Design weight



(b) Inverse probability weight (IPW)



(c) Inverse probability weight with interviewer effects (MCMC)



(d) Calibration of IPW with interviewer effects

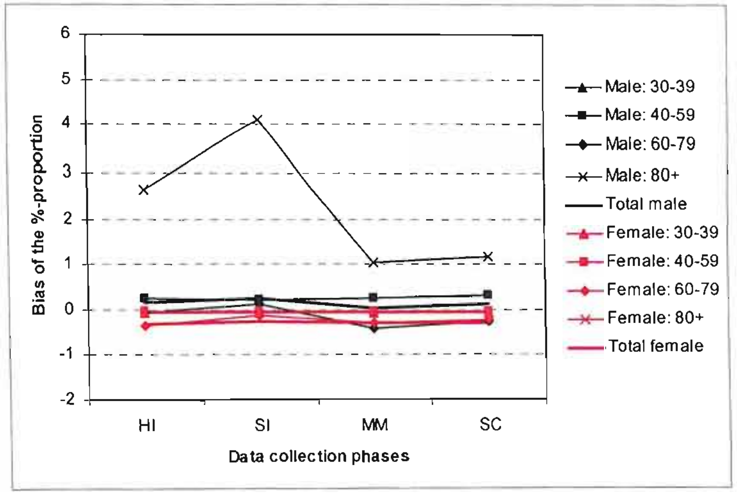
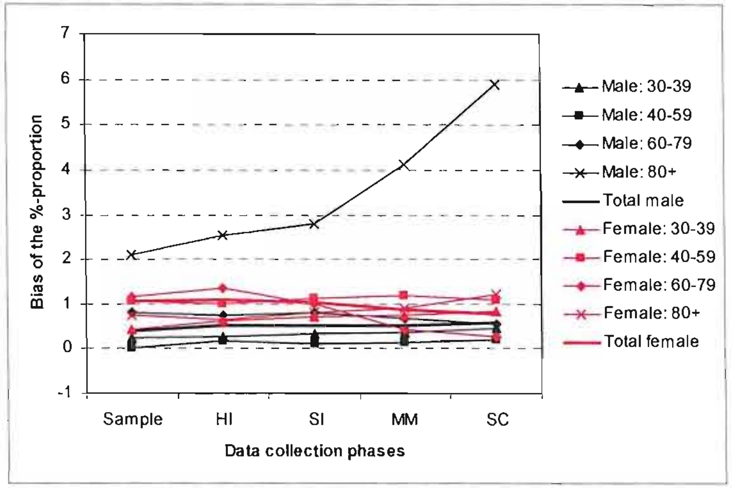
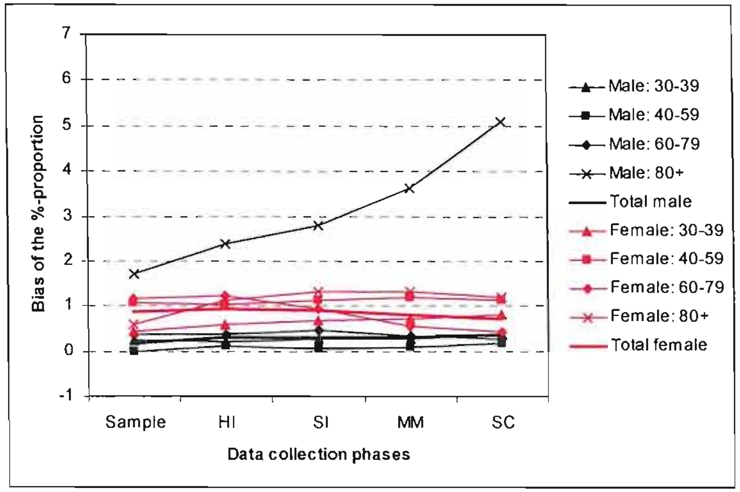


Figure 6.7 Bias of estimated proportion of people receiving medical reimbursement on connective tissue diseases

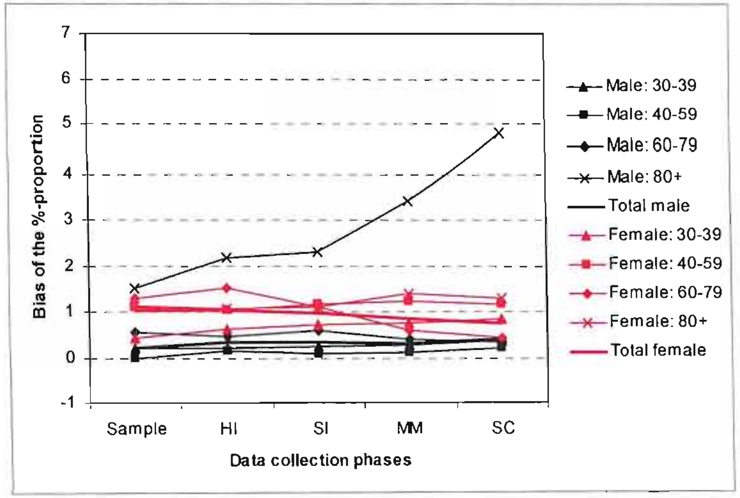
(a) Design weight



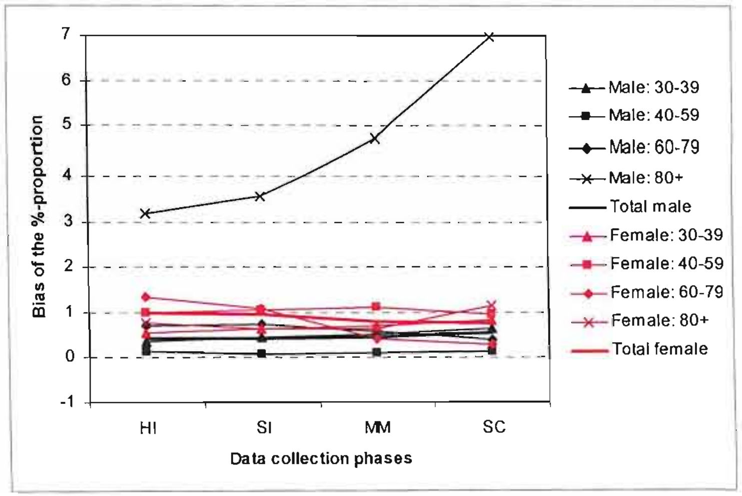
(b) Inverse probability weight (IPW)



(c) Inverse probability weight with interviewer effects (MCMC)



(d) Calibration of IPW with interviewer effects



6.7. Conclusions

In this Chapter we have examined the weighting of the Finnish Health 2000 survey. We have used health related register data with direct linkage to simulate the survey estimation with the achieved fieldwork performance on non-response, survey attrition and non-response reduction efforts. The weighting methods contain variants of Horvitz-Thompson estimator and calibrated weights. The quantitative indicators developed for assessing the weights reveal that the weighting methods differ in the bias reduction, interviewer impact, and survey estimates. In our empirical study, the design weighted adjustments did not provide improvement for estimation. Examining the bias of the survey estimates, we find that in this multi-purpose survey the direction and magnitude of the bias can vary largely by the simulated analysis variables.

The impact of survey attrition affects in various patterns the estimates based on the remaining co-operating sub-sample. This is found to be partly dependent on the population age and sex groups, but more importantly on the impact the estimated health condition may have on the individuals' ability to participate in surveys, which is confounding with the finding of Duhart et al. (2001). Estimates of the chronic cardiac insufficiency or people receiving help support are decreasing drastically with the survey attrition, while the estimates of chronic asthma are increasing across data collection phases. This means that people less capable to participate and with severe health problems fall into non-response in the presence of high response burden affecting the survey results in a health survey. At the same time people with less limiting long term illnesses maintain their co-operation, leading to high increase in the prevalence estimates of some diseases. Thus one needs to examine, how in a real survey situation the auxiliary health information can be exploited for correctly adjusting for the non-response.

Combining the information from late respondents, i.e. from those individuals not co-operating until the non-response reduction efforts, has diverse effects on the bias of the survey estimates. For some analysis variables the bias is decreased while for some the bias can be increased. The results indicate that the bias correction depends largely both on the type of disease and the weighting method. When late respondents from the non-response reduction efforts are included in the estimation of diabetes mellitus, the calibration with design weight provides bias reduction. However for chronic asthma, the calibration with design effect has the poorest performance in terms of the bias, when including the late respondents. Overall, the most stable performance is given by the inverse probability weighting based on sequential logit models. All in all, there are no obvious gains observed from the non-response reduction for all groups and survey estimates. As the non-response reduction may also increase the bias and the achieved information structure in real data can be contaminated with measurement and context effects, we do not recommend attempts to merge late respondents to the analysis with the originally obtained data.

Our results show that the weighting methods based on the design effect, do not add for the correction of the estimates, regardless the complex survey design. In addition, the calibration of weights does not provide improved estimates for all analysis variables. Instead, we have found that, in our empirical study, weighting methods dealing with the human behaviour affecting survey participation both by the respondents and the interviewers, yield generally better survey estimates than traditional survey weights or calibration of design weights. This result is based on the precision of the health related estimates, and to the variation of the weights. Thus our finding supports the argument made by Groves and Couper (1995) on the need to develop post-survey adjustment using the socio-psychological theories that specify the human behaviour affecting survey participation. More importantly, we have shown that it is not only the characteristics of the respondent and their ecological population that need to be taken into account in weighting adjustment, but also the characteristics and professional attitudes of the interviewer. Traditionally the interviewer effects have been dealt with in standard errors.

In future analysis of the survey data, it would be recommendable to examine the suitability of weighting separately for each analysis situation, as has been hypothesized previously also by Oh and Scheuren, 1983. We have conducted initial analysis on the estimation of the prevalence of diabetes mellitus and the chronic cardiac insufficiency controlling for their joint risk factor, the prevalence of chronic blood pressure. Using calibration, we controlled the analysis also for age and sex of the individuals. However, the bias reduction gained was not significant. Thus, further research would be needed to develop estimation exploiting efficiently health register information and other auxiliary information that quantify the risk factors of long term illnesses. In addition, we recommend further assessment of the sampling design, containing simulated samples and replicated analysis of this type using register derived data. This assessment should also contain alternative sampling designs that aim to account for the health inequalities of the population.

7. Conclusions and discussion

In this thesis we have reviewed the survey participation literature from the point of view of multiphase surveys with mixed data-collection components. In Chapter 2, we have linked the theory of response effects (Sudman and Bradburn, 1974), the conceptual theory of survey participation (Groves and Couper, 1996) and the route map of social exclusion (Atkinson, 1998), to extend the survey non-response framework for studying the associations between social exclusion and non-response. This theoretical framework contains elements of conditions of individuals, social environment and the society. These elements consist of various dimensions that affect the connectedness to the society, co-operation and trust to government and societal institutions, leading to underlying propensity to respond to social surveys.

In the empirical part of this thesis, we have linked the Finnish Health 2000 survey with rich sources of quantitative auxiliary information. In addition, we have conducted an interviewer perception survey, and incorporated the results into survey participation analysis. In Chapters 3 and 4, we have modelled survey participation behaviour of individuals in the presence of high response burden. We have found that the survey participation behaviour is greatly affected by individuals' socio-economic circumstances, social capital, and social connectedness. People with affluent circumstances are more co-operative than people with any of the risk factors of social exclusion. We have demonstrated that a single model oversimplifies the survey participation in a survey with multiple data collection phases. Instead, sequential logit modelling allows more flexibility and better use of cumulating data resources for non-response modelling. In addition, the profiles and risk groups of survey non-respondents vary by data collection phases due to differential data collection modes and increasing response burden. Confounding with Loosveldt, Pickery and Billiet (2002), the occurrence of item-nonresponse at initial interview has been found to predict non-response at later phases.

In Chapter 5, we have found that the interviewer effects in face-to-face interviewing survey are influential at contacting and interviewing phases. Interviewers may also affect the success of further data collection components, which by survey design are independent from the presence of the interviewers. We have found significant differences across interviewer assignments on how successfully the respondents of the health interview progressed to the next phase of data collection, administrated by other than the interviewers. Our finding is that interviewers differ systematically in contacting males, and gaining co-operation form households with children, even when controlling for individual and interviewer level variables. In addition, the data suggests that there is systematic difference across interviewer assignments on how successfully interviewers motivate people with low income to participate to the next phase of data collection after the health interview.

In Chapter 6, we have found that the simulated health estimates can be improved, if we account the survey non-response propensity weighting also for the characteristics and perceptions of the interviewers. This finding is remarkable as it shows that the interviewer effect can contaminate the obtained survey information not only at individual level, but also at the level population distributions for the survey estimates. We have also found survey attrition causing instability for the estimates based on the remaining co-operating sub-samples. In addition, we have observed that the non-response reduction may also increase the bias of the estimates. The results indicate that the bias correction depend largely both on the health variables to be estimated and the weighting method implemented. In real survey data, the achieved information structure of the late respondents can also contaminate the data with measurement and context effects. Thus, based on these results we do not recommend merging late respondents to the analysis with the originally obtained data.

Are socially excluded also excluded from social surveys?

The results from our empirical analysis point out that factors increasing the survey participation are projections of economic and social wellbeing, whilst factors reducing the participation are associated with poor living conditions, lower social and economic status. Subsequently, the survey participation can be related to factors associated with social exclusion. We have found some evidence that social benefits aiming to increase the activity of people, who are in the risk of social exclusion, may be associated with increased survey participation. In contrary, benefits providing support without intension to improve the economical independency of the individuals in the long term decrease the survey participation. For example, rehabilitation programs supporting people to actively participate to the society and help their entry or return to labour market, have been found to be associated with increased social participation and connectedness to the society in the form of responding to surveys. At the same time, people with disabilities receiving long-term support and people receiving benefits such as income support have increased risk not to co-operate fully with survey request.

The significance of risk factors of social exclusion vary across data collection phases. Thus the population at risk both for social exclusion and non-response or survey attrition is not a homogenous group that could be characterised by few quantified variables or factors. Instead, we can identify different sub-populations at risk for non-contact, non-cooperation and drop-out in the presence of high response burden. Although some population groups have a very high risk of not to co-operate fully, they may still have relatively minor weight in overall models, if they represent only a small minority of the target population. The importance of the small sub-populations should not anyhow be neglected especially when studying phenomena or evaluating social policies in which they are of any potential interest. More importantly their impact in the population may also change over the time, if their relative proportion in the population would alter. Therefore, it is important to recognise all groups, which have high risk to be excluded from surveys due to self-selection.

The survey participation of minority groups should be investigated in large assessment, across national surveys, to gain reliable results with the possibility to examine the possible impact of the survey topic. In Health 2000 data, people who did not belong to the household population and who had been unemployed for many weeks in 2000 had a high risk of dropping out. Similarly, single men with foreign maternal language had very low propensity to participate. Immigrants may encounter prevailing social and economic problems in adapting to the Finnish society, for example, in facing negative attitudes, prejudice and in finding employment corresponding to their educational background. Thus it would be important to study across surveys whether their survey participation is at the same level or below the Finnish speaking majority and Swedish speaking minority. In addition, there can be large variation in survey participation depending on the integration of the minority groups, across the immigration generations and across the age groups of members of minority groups. For studying the associations between the social integration and survey participation, we would need to have a large data set to allow more detailed analysis.

Does the quantified data represent the perceptions of the individual?

The analysis variables and classifications derived for our analysis have been constructed with the aim to reflect objectively the heterogeneity in the population, recognising the breakdown of the conventional norms. In Chapter 2, we have noted that the variable transformations, classifications and the relationship of quantified and qualified information have not been discussed in great detail in the previous survey participation studies. The typology of the conceptual models should, however, be developed further to enable the non-response models to better account for the influential factors relevant for the data loss. In addition, the psychological components referred to in the conceptual models, have proven difficult to be quantified and direct auxiliary information is negligible in this area for the purpose of large scale surveys. This is, however, rarely discussed in studies applying this type of frameworks.

Quantifying characteristics from external sources and higher hierarchical levels may possibly violate the interpretation of the behaviour of individuals. Using external data and logical classification rules in the analysis of behavioural models, we can also falsely group the individual into a group they would not identify themselves with. Risks of categorising individuals against their perceptions are high in classifications such as socio-economic group, family type, or ethnicity. For example, according to the ILO[1] definitions, a student in part-time employment should be coded as employed, although their identity and social behaviour may resemble more closely to that of the student population. The dilemma of classifications in the context of modelling behaviour of individuals and the interpretability of the explanatory information, have been given too little attention in the survey non-response analysis. Thus it is the responsibility of the researcher to reconstruct variables and classifications that are, as far as possible, both objective and reflecting reality rather than the conventional norms.

[1] ILO refers to International Labour Organisation

Improving survey requires changes in the survey practise

Our findings indicate that the sampling design of the Health 2000 survey has overlooked the needs arising from the increasing polarisation in the Finnish society. Instead, the sampling design has reflected the traditions in designing surveys in Finland and aimed to respond to the needs of studying health inequalities by age, sex and regional differences. However, if one would need reliable data for studying health inequalities across other sub-population groups, the sampling design should target to obtain adequate representation of these sub-population groups, not only over-sampling the elderly population. For example, we have found strong implications on the reduction in survey co-operation amongst people in the risk of social exclusion. Due to the self-selection, these sub-groups have a diminishing representation in the final obtained data. At the same time people who have economically or socially affluent circumstances are proportionally over represented in the data.

We have found that the impact of survey attrition affects in various patterns the simulated health estimates in our study, based on the remaining co-operating sub-sample. This is partly dependent on the age and sex groups, but more importantly on the individuals' ability to participate in surveys due to their health condition. This means that people less capable to participate and with severe health problems fall into non-response in the presence of high response burden affecting the health survey results. At the same time people with less limiting long term illnesses maintain their co-operation, leading to high increase in the prevalence estimates of some diseases. Thus one should examine, how in a real survey situation the auxiliary health information can be exploited for correctly adjusting for the non-response. Actually, little can be achieved with simple over-sampling of sub-population groups. Over-sampling of the elderly, for example, does not solve the problem of bias as it only seems to increase the number of healthy elderly in the final obtained survey data.

Opening of the labour market to the people living in the new member states of the European Union is likely to increase the immigration, adding more challenges for survey organisations in the future. Thus, one may anticipate a turning point when survey practise must increase the survey languages to contain also other languages than Finnish and Swedish. In addition to overcome language barriers, training specialised interpreters for data collection should become a normal practise in the future. In Health 2000 surveys, interpreters were used in some cases, but not systematically.

The sampling design of the Health 2000 survey has not sufficiently exploited the auxiliary data sources available, for example, information from the health registers. In addition, the regional clustering used in sampling has possibly introduced problems due to small sample sizes within clusters, which has increased the potential sampling error in the total sample. The health information linked from the health register showed upward bias for most health estimates by age and sex. Non-response adjustment methods did not remove the bias entirely. In future surveys one needs to define the sampling design using auxiliary health register information to accommodate both the differences within and across regional clusters. Also, the sampling design should exploit error calculations based on register derived health indicators, not solely the geo-demographic factors of the population. We recommend further assessment of the sampling design, containing simulated samples and replicated analysis of this type using register derived data. This assessment should also contain alternative sampling designs that aim to account for the health inequalities of the population and unbiased health estimators in small population domains.

Further research is also needed in extending the use of survey process data in computer assisted interviewing, towards ad hoc analyses during the data collection. Empirical studies should test, whether non-response in risk groups could be reduced by interactive feedback systems to interviewers. Prevention of non-response could also benefit from analysing response propensities of the remaining interviewees, and respondents with item-nonresponse in multiphase surveys. The use of process data can potentially improve the understanding of performance differences across interviewer assignments. For example, in the qualitative information obtained from the interviewer perceptions survey, we found that interviewers, with lowest interviewer level random effects in survey participation models, had indicated experiencing problems in survey fieldwork operations. Being able to separate objectively and reliably the impact of these problems from speculations of interviewer skills, would be crucially important in a sensitive area such as evaluation of work performance.

Future research and development is needed to improve further use of data

We have shown that rich auxiliary information has been available for survey non-response analysis and how detailed information improves the explanatory power of the models. The information identified crucially important on survey non-response analysis in this thesis should be examined for the purpose of re-designing the current sampling designs. We recommend similar survey participation assessments using micro level auxiliary information in other social and health surveys in Finland and in other European countries. This would be especially important in surveys measuring the wellbeing of people in relation to their health, social or economic wellbeing, as non-ignorable non-response may increase the bias of survey estimates. Secondly, further research would be needed on non-response and social exclusion possibly including meta-analysis with multiple data sets and qualitative data on the perceived social exclusion. Thirdly, we call for developing the survey designs to better reflect the changing conditions and diversification in the society for identifying and taking into account the risk groups of survey non-response.

We conclude with our statement expressed in Chapter 4: *If the national surveys fail to capture and measure the social and health experiences of the specific sub-populations with increased risk for social exclusion, they will also fail in providing adequate data for policy monitoring to evaluate the social conditions of the population; especially amongst those who belong to the target groups of social welfare policies.*

APPENDIX 2.1. Comparison of Non-Response in European National Health Surveys 1998 - 2002

(based on the table presented in Räty et al., 2003; slightly edited version)

Survey name in English	Year	Country	Type of survey	Frequency	Instutionalised included	Age restriction	Type of sample	Non-response rate, %
Microcensus	1999	Austria	HIS	Irregular	No	No restriction	HH	18 (I)
Health Interview Survey	2001	Belgium	HIS	4-yearly	Yes	No restriction	HH	38 (HH)
Health and Morbidity in Denmark	2000	Denmark	HIS	6-7 yearly	Yes	16+	I	26 (I)
Health Behaviour Survey among the Elderly Population	2001	Finland	HIS	Irregular	Yes	65-84	I	81 (I)
Health Behaviour Survey among the Adult Population	2001	Finland	HIS	Yearly	Yes	15-64	I	70 (I)
Health Interview and Examination Survey 2000	2000	Finland	HIS/HES	Irregular	Yes	18+	I	11 (I)
Survey on health behaviour	2000	Finland	Other	Yearly	Yes	15-64	I	30 (I)
The National Finrisk Study	2002	Finland	HIS/HES	5-yearly	No	Unknown	I	35 (I)
Health and Social Protection Survey	2002	France	HIS	2-yearly	No	No restriction	I	
The INSEE Survey on Handicaps, Disabilities and Dependency	2001	France	Disability survey	10-yearly	Yes	No restriction	I	22 (I)
Health Barometer	1999	France	HIS	3 yearly	No	12-75	HH	25 (HH)/ 6 (I)
Health and Social Protection Survey	1998	France	Other	2 yearly	No	No restriction	I	34 (HH)
Handicaps, Disabilities and Dependency Survey	1999	France	Disability survey	Unknown	Yes	No restriction	I	31 (I)
German National Health Examination and Interview Survey	1998	Germany	HIS/HES	6-7 yearly	No	18-79	I	39 (I)
Questions on Health	1999	Germany	HIS	4 yearly	Yes	No restriction	HH	9 (I)
National Greek Survey: Psychosocial Factors and Health	1998	Greece	HIS	5-yearly	No	12 to 64	I	20 (I)
Health and Living Conditions in Iceland	89/99	Iceland	HIS	Irregular	No	18-75	I	31 (I)
Survey of Lifestyle, Attitudes and Nutrition (SLÁN)	1998	Ireland	HSCS	4 yearly	Yes	18+	I	38 (I)
Survey of Lifestyle, Attitudes and Nutrition (SLAN) Health Behaviour in School-aged Children (HBSC)	2002	Ireland	HSCS	4-yearly	No	18+ (SLAN); 10-17 (HBSC)	I	37 (I)
Health Conditions and the Use of Health Services	99/00	Italy	HIS	4 yearly	No	No restriction	HH	10 (HH)/ 10 (I)
National Health Interview Survey	98-99	Portugal	HIS	Irregular	No	No restriction	HH	20 (HH)
Impairments, Disabilities and Health Status Survey	1999	Spain	Disability survey	Irregular	No	No restriction	HH	
Spanish National Health Survey	2001	Spain				Unknown		
Swiss Health Survey 2002	2002	Switzerland	HIS	5 yearly	No	15+	HH	
Continuous Quality of Life Survey	2001	The Netherlands	HIS/HES	Continuous	No	No restriction	I	45 (I)
Continuous Quality of Life Survey	1998	The Netherlands	HIS/HES	Continuous	No	No restriction	I	38 (I)
Patient Survey - Second Dutch National Survey of General Practice	2001	The Netherlands	HIS	Irregular	No	No restriction	I	35 (I)
The Scottish Health Survey	1998	United Kingdom	HIS/HES	Irregular	No	2-74	HH	23 (HH)
The Health Survey for England	2000	United Kingdom	HIS/HES	Continuous	Yes		HH	25 (HH)
The Health in Wales Survey	1996	United Kingdom	HIS	Irregular	No	18-79	HH	78 (HH) / 67 (I)
Welsh Health Survey	1998	United Kingdom	HSCS	Irregular	No	18+	I	59 (I)

HIS: Health Interview Survey
HES: Health Examination Survey
HSCS: Health Self-Completion Survey

HH: Households
I: Individuals

APPENDIX 3.1 Data collection phases in Health 2000 survey

(Aromaa and Koskinen 2002)

At respondent's home (conducted by Statistics Finland):

- Health interview
- Self-completion questionnaire 1

At health examination centre (conducted by National Public health Institute):

1 Registering

- Information, informed consent, symptom interview
- Handing over the self-completion questionnaire 2 and the urine sample container

2 Anthropometrical measurements

- height, circumference, ECG, and blood pressure

3 Measurements

- spirometry, bioimpedance, and heel bone density

4 Laboratory

- drawing blood samples (100 ml), handling of samples

5 Dental health

- clinical oral examination, orthopantomography

Snack and filling the self-completion questionnaire 2

6 Functional capacity tests

- physical and cognitive capacity, vision and hearing

7 Clinical examination

8 Mental health interview

9 Final interview

- checking that all examinations and questionnaires have been completed
- handing over the self-completion questionnaire 3 and the dietary questionnaire
- information about the previous and possible further examinations

[In total approximately 3 hours and 15 minutes]

At home:

- (Health examination for those not attending the health examination proper at the health centre)
- Filling in questionnaire 3 and dietary questionnaire

University hospitals or research centres:

- More detailed studies for sub-samples

From registers:

- Register data

APPENDIX 3.2 Contents of the Interview Perception Survey

1. Background information

- Educational background,
- Region,
- Interviewing language,
- Telephone interviewing,
- Experience from previous health surveys

2. Experience of the Health 2000 Survey

2.1 Training and materials

- Interviewer training,
- Advance letter,
- Questionnaire,
- Classifications,
- Difficult questions,
- Sensitive questions,
- Agreement form,
- Interviews in institutions,
- Functionality of materials and systems

2.2 Achieving co-operation, and survey non-response

- Refusal letter,
- Refusal conversion,
- Factors affecting participation,
- Factors affecting refusals,
- Own professional behaviour and attitudes,

2.3 Support during the fieldwork

- Adequacy of support during the fieldwork

2.4 Attitudes towards Health 2000 Survey

- Work motivation,
- Post-survey image/survey experience,
- Participation to the medical examination

Questionnaire in BLAISE
56 questions in total

APPENDIX 3.3 Auxiliary information on the ecological population used for survey participation analysis of the Health 2000 survey

Theoretical context	Variable	Data source
Population composition:		
	Ratio of males to females in total population	PIS & GIS
	Ratio of males to females in 15-24 population	PIS & GIS
	Ratio of males to females 65+ population	PIS & GIS
	Dependency ratio	PIS & GIS
	Population density	PIS & GIS
Mortality		
	Mortality rate	PIS & GIS
	Mortality rate for infants (0 to 1 years)	PIS & GIS
	Mortality rate for 0-17	PIS & GIS
	Mortality rate for 18-29	PIS & GIS
	Mortality rate for 30-64	PIS & GIS
Composition of dwelling units:		
	Average size of dwelling units	PIS & GIS & RD
	% of pensioners	PIS & TR
	% single-parents	PIS & GIS & RD
	Ratio of male single parents over female	PIS & GIS & RD
	% young female single-parents (15-24 year olds)	PIS & GIS & RD
	% young single-parents (15-24 year olds)	PIS & GIS & RD
	Ratio of young male single parents over young female (15-24 year olds)	PIS & GIS & RD
	% young female single-parents (15-24 year olds)	PIS & GIS & RD
	% young single-parents (15-24 year olds)	PIS & GIS & RD
	Ratio of young male single parents over young female (15-24 year olds)	PIS & GIS & RD
	Adult males living in parental home	PIS & GIS & RD
	Adult females living in parental home	PIS & GIS & RD
Divorce:		
	% of divorced	PIS & GIS
	% of divorced females	PIS & GIS
	% of divorced males	PIS & GIS
	Ratio of divorced males over females	PIS & GIS
	% of divorced young people	PIS & GIS
	% of divorced young females	PIS & GIS
	% of divorced young males	PIS & GIS
	Ratio of divorced young males over females	PIS & GIS
People living in institutions:		
	% Institutionalised of the population	PIS & GIS
	Ratio of institutionalised men over women	PIS & GIS
	% Institutionalised 30-64 males	PIS & GIS
	% Institutionalised 30-64 females	PIS & GIS
	Ratio of institutionalised 30-64 men over women	PIS & GIS
Educational background:		
	% with no further education	CED & PIS & GIS
	% with further education degree	CED & PIS & GIS
	% with university degree	CED & PIS & GIS
Socio-economic grouping:		
	% self-employed	TR & PIS & GIS
	% salaried employee / white collar	TR & PIS & GIS
	% worker / blue collar	TR & PIS & GIS
	% employment status unknown	TR & PIS & GIS

Unemployment:	
Rate of economically active population	ER & PIS & GIS
Unemployment rate	ER & PIS & GIS
Long-term unemployment rate	ER & PIS & GIS
% unemployment amongst young (15-24 year olds)	ER & PIS & GIS
% unemployment amongst young men (15-24 year olds)	ER & PIS & GIS
% unemployment amongst young women (15-24 year olds)	ER & PIS & GIS
% Long-term unemployment amongst young (15-24 year olds)	ER & PIS & GIS
% Long-term unemployment amongst young men (15-24 year olds)	ER & PIS & GIS
% Long-term unemployment amongst young female (15-24 year olds)	ER & PIS & GIS
Ratio of long-term unemployed young male over female (15-24 year olds)	ER & PIS & GIS
% unemployment amongst 50+	ER & PIS & GIS
% unemployment amongst 50+ men	ER & PIS & GIS
% unemployment amongst 50+ women	ER & PIS & GIS
% Long-term unemployment amongst 50+ men	ER & PIS & GIS
% Long-term unemployment amongst 50+ female	ER & PIS & GIS
Ratio of Long-term unemployed 50+ male over female	ER & PIS & GIS
Housing conditions:	
% of homeless people	GIS & PIS
% of people living in shelter/care homes	GIS & PIS
% of people living in institutions	GIS & PIS
% of people with unknown type of living	GIS & PIS
% of people living in dwellings where all under 15	GIS & PIS
% of dwelling units (DU) living in cramped dwellings	RD& GIS & PIS
% of DUs with children out of those in cramped dwellings	RD& GIS & PIS
% of DUs of foreign origin in cramped dwellings	RD& GIS & PIS
% of residential buildings with electricity	RD
% of residential buildings with plumbing	RD
% of residential buildings with running water	RD
% of residential buildings with hot water	RD
Income poverty:	
% of DUs received income support out of DUs	SII & GIS & PIS
% persons received income support per 1000 inhabitants	SII & GIS & PIS
% of young DUs received income support (15-24 year olds)	SII & GIS & PIS
% of mature DUs received income support (50+ year olds)	SII & GIS & PIS
% of single men receiving income support	SII & GIS & PIS
% of single women receiving income support	SII & GIS & PIS
% of single parent women receiving income support	SII & GIS & PIS
Participation to society:	
EU-election (with relatively low poll):	
% voting in EU elections in 1999	EIS & GIS
% male voting in EU elections in 1999	EIS & GIS
% female voting in EU elections in 1999	EIS & GIS
Ratio male to female voting in EU elections in 1999	EIS & GIS
% pre-election votes / election day votes	EIS & GIS
% male pre-election votes / election day votes	EIS & GIS
% female pre-election votes / election day votes	EIS & GIS
Ratio male to female pre-voting in EU elections in 1999	EIS & GIS
% deserted votes of all votes given in EU elections in 1999	EIS & GIS
Local elections (with average poll and local impact):	
% voting in local elections in 1999	EIS & GIS
% male voting in local elections in 1999	EIS & GIS
% female voting in local elections in 1999	EIS & GIS
Ratio male to female voting in local elections in 1999	EIS & GIS
% pre-election votes / election day votes	EIS & GIS
% male pre-election votes / election day votes	EIS & GIS
% female pre-election votes / election day votes	EIS & GIS
Ratio male to female pre-voting in local elections in 1999	EIS & GIS
% deserted votes of all votes given in local elections in 1999	EIS & GIS
% of female candidates in local elections in 2000	EIS & GIS
% of female candidates elected in local elections in 2000	EIS & GIS

Use of cultural, recreation and social services, and their public support:	
Pupils in preparatory studies	RA
Students in evening schools in Autumn term	RA
Education hours given in evening schools, hours per year	RA
Loans from public libraries, per year	RA
Visits to public libraries, per year	RA
Fire and rescue function alarms, rescue alarms per year	RA
Sport clubs and teams received allowance from municipality, number of clubs and teams	RA
Youth clubs and societies received allowance from municipality, number of clubs and societies	RA
Cultural clubs and societies received allowance from municipality, number of clubs and societies	RA
Constructed and maintained green areas, hectares	RA
Flats in council housing	RA
Use of public social and health services:	
Children in full time day care	RA
Children in part time day care	RA
Children in full time day care in private day care in families	RA
Children in full time day care in private day care in families	RA
Children in day care supported by the local municipality	RA
Children in organised play activities, on average during the day of activity	RA
Families in family affairs conciliation, per year	RA
Visits to child health centres, per year	RA
Elderly people in care homes	RA
Attendance days in care homes of the elderly	RA
Mentally deficient or handicapped in care homes	RA
People received help in home care	RA
People received help services	RA
Elderly people in full time care in care homes	RA
Elderly people in part time care	RA
Elderly people in full time care in care homes or in part time care	RA
People used services for the handicapped or services of the personal assistant	RA
Reported crime:	
Crime rate per 100 inhabitants	SSCO & PIS & GIS
Murders, violent deaths, physical violence and sexual crimes per 100 inhabitants	SSCO & PIS & GIS
Offences against property per 100 inhabitants	SSCO & PIS & GIS
Sexual crimes per 100 inhabitants	SSCO & PIS & GIS
Traffic offences per 100 inhabitants	SSCO & PIS & GIS
Offences against alcohol laws per 100 inhabitants	SSCO & PIS & GIS
Other offences per 100 inhabitants	SSCO & PIS & GIS

List of abbreviations in alphabetical order:

CED Register of Completed Education and Degrees

DU Dwelling Unit

EIS Election Information System of the Ministry of Justice

ER Employment Register of the Ministry of Labour

GIS Geographical Information System (here namely the geographical point coordinates of the GIS system have been used to map the information from registers into the level of local areas)

HH Household

ISCO Statistical System on Criminal Offences

PIS Population Information System

RA Regional Accounts and Records from Municipalities

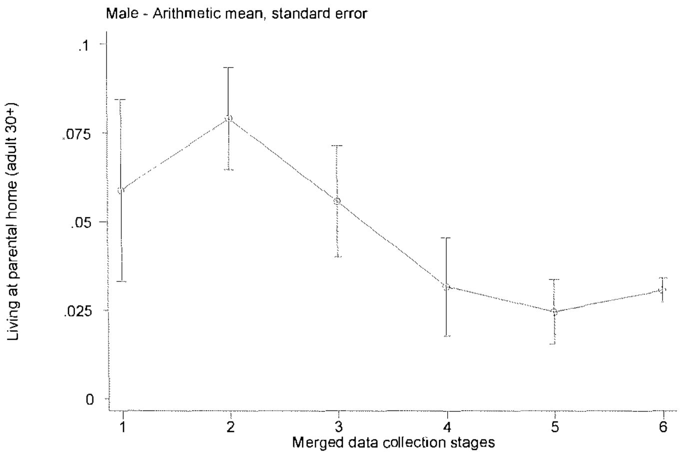
RD Register of Dwellings

SII Social Insurance Institution Information System

TR Register maintained by the Tax Administration

APPENDIX 3.4 Arithmetic mean and standard error of 1st occurrence of non-response by demographic characteristics

Figure 3.4.1 Proportion of adults living at parental home of 1st time non-respondents by sex and data collection phase



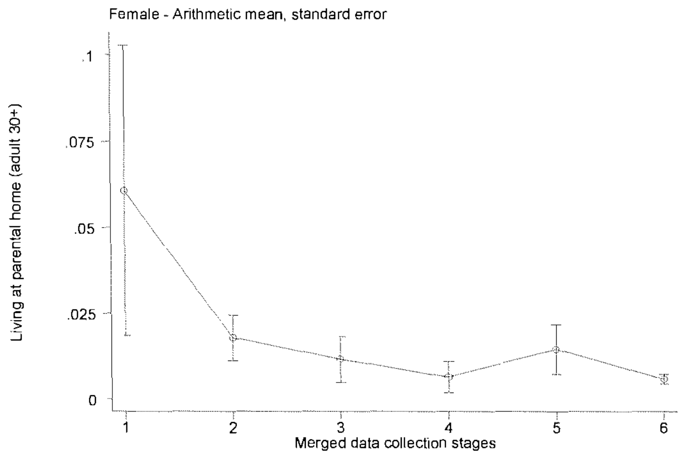
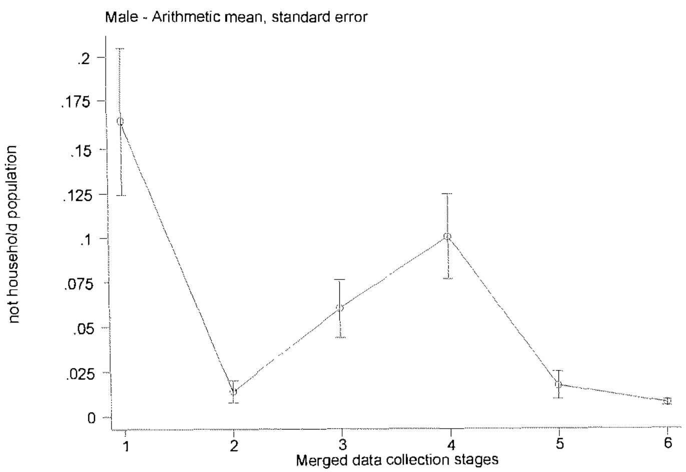


Figure 3.4.2 Proportion of people not belonging to the household population of 1st time non-respondents by sex and data collection phase



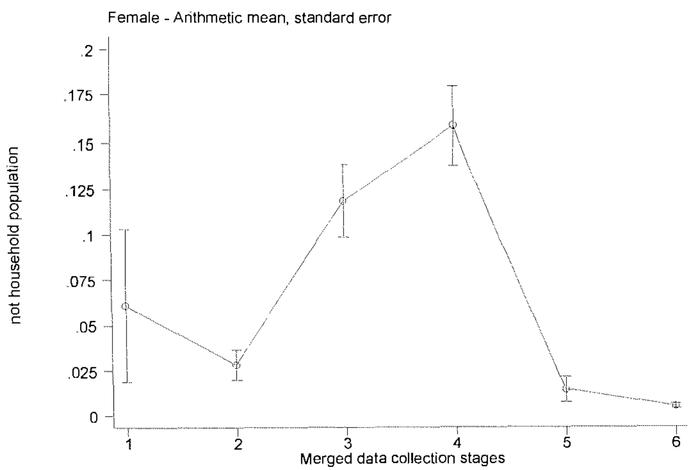
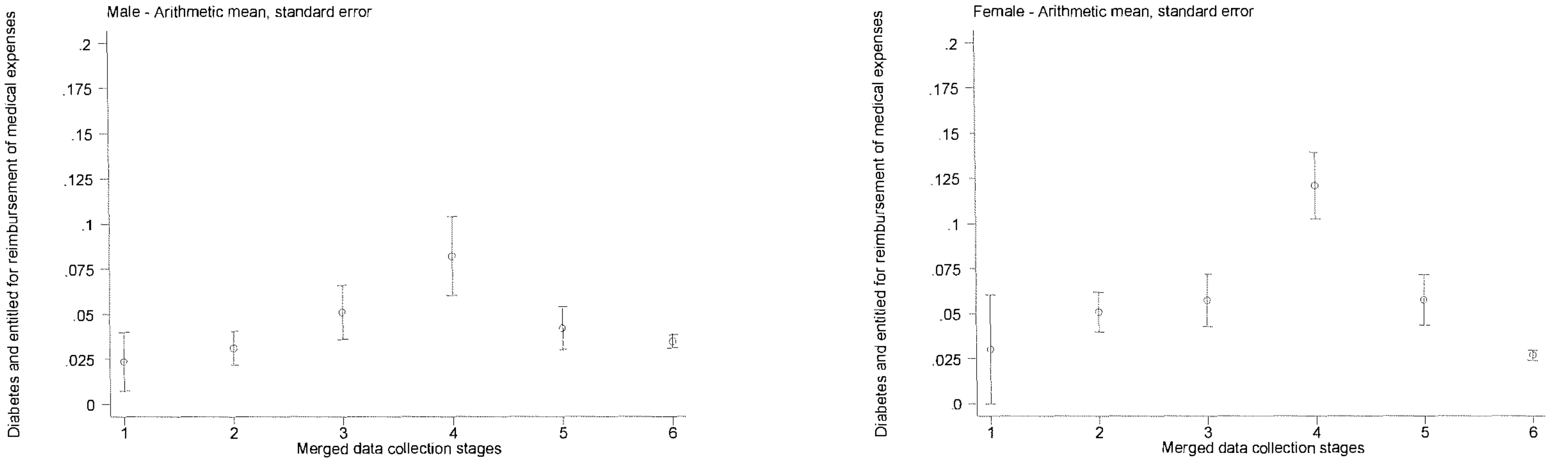


Figure 3.4.3 **Proportion of people diagnosed with diabetes and entitled for reimbursement of medical expenses of 1st time non-respondents by sex and data collection phase**



NOTE: phase 6 refers to those participated at an acceptable level to all phases

APPENDIX 3.5 Characteristics of sampled individuals or the local area by sex and data collection phase

Figure 3.5.1 Age of the youngest child in the dwelling unit

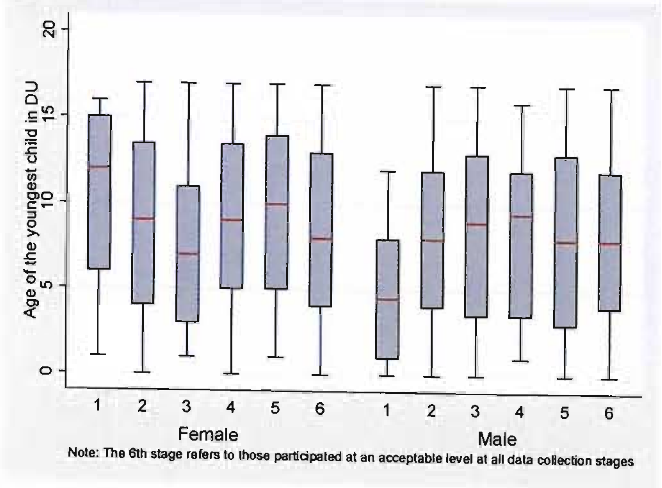


Figure 3.5.2 Size of dwelling unit

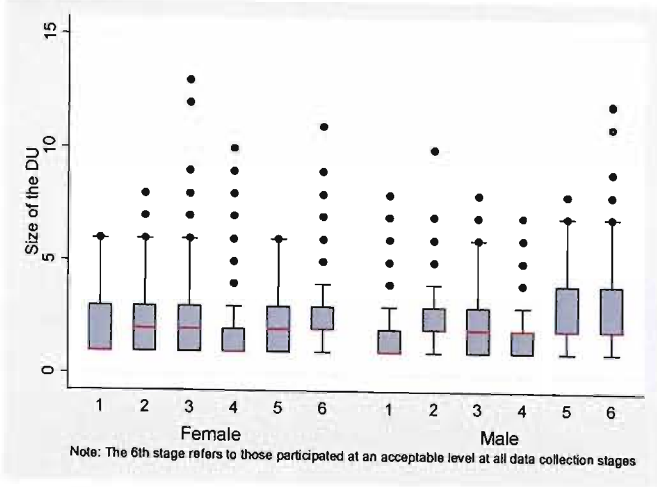


Figure 3.5.3 Decile of register derived disposable income

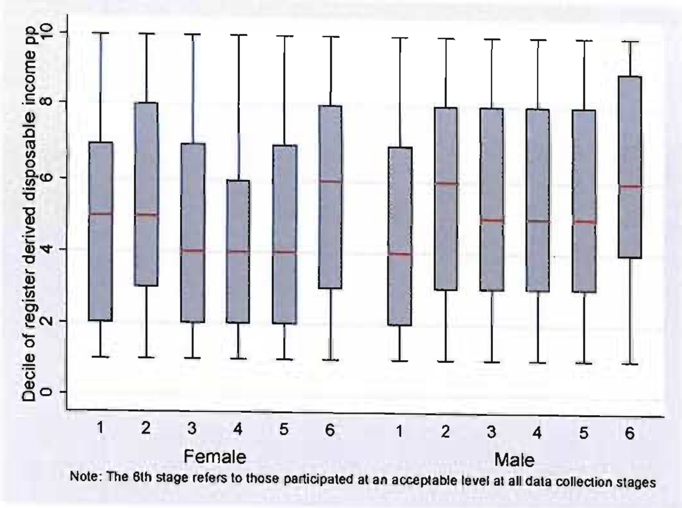


Figure 3.5.4 Dependency ratio in HCD

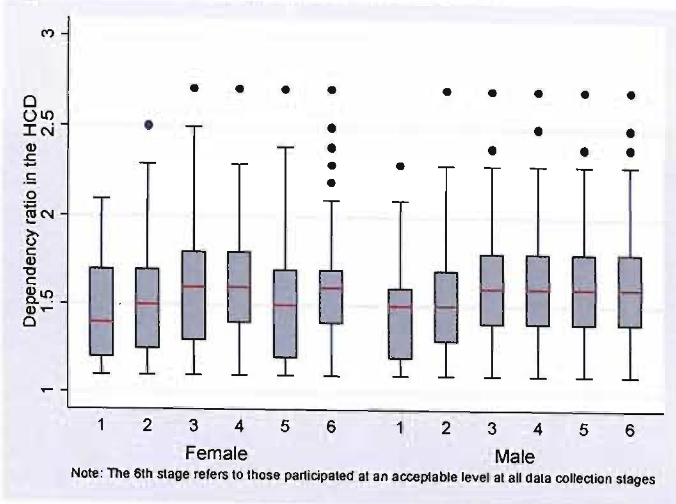


Figure 3.5.5 Crime rate in the local area

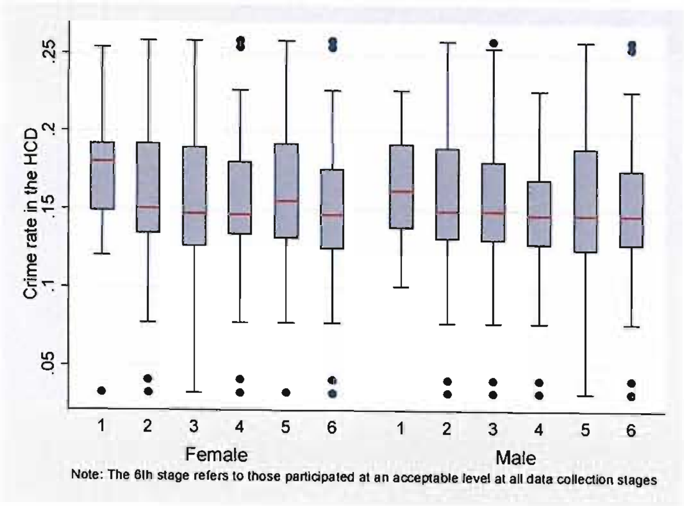


Figure 3.5.6 Unemployment rate in the local area

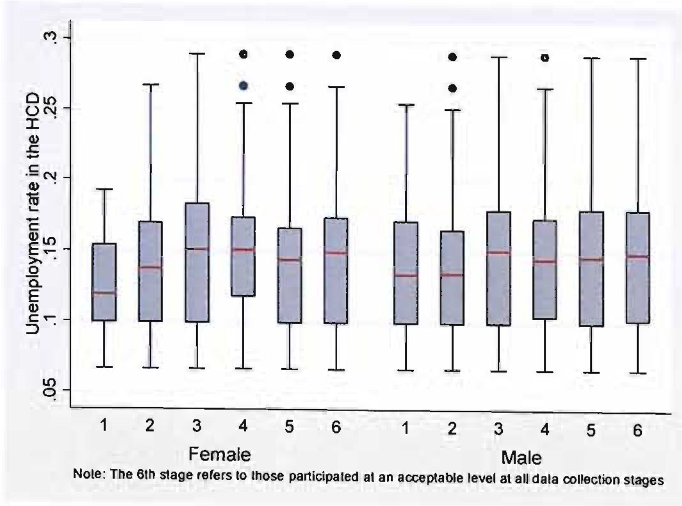


Figure 3.5.7 Proportion self-employed in the local area

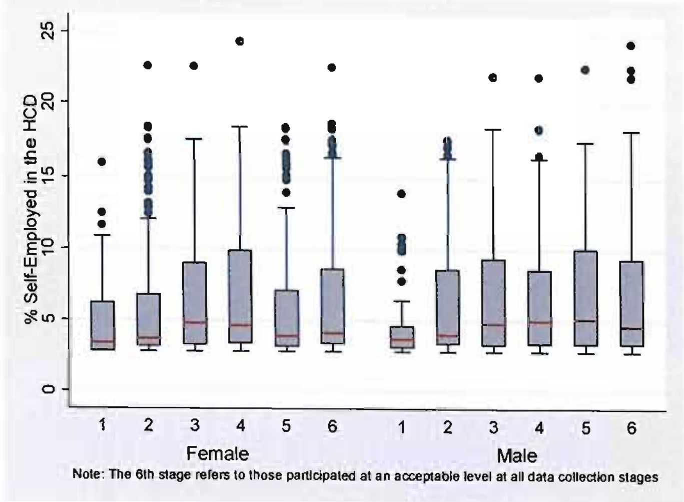
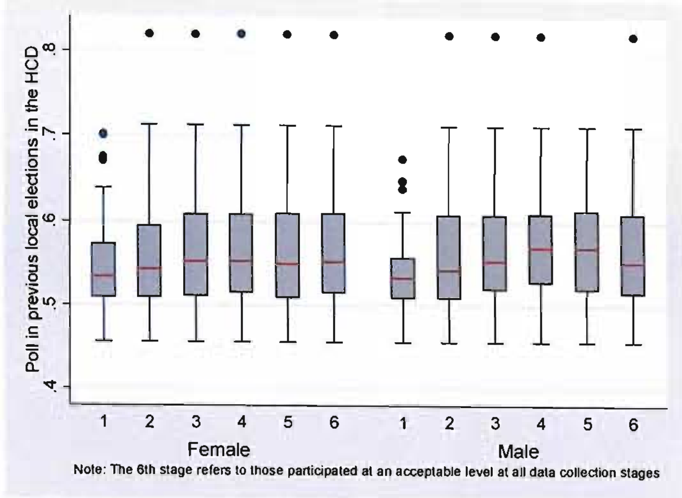


Figure 3.5.8 Poll in previous local elections in the local area



APPENDIX 4.1 Design-based response probability modelling

4.1.1 Design-based survival and hazard functions

The design based estimation of survival and hazard functions take into account the weighting by the inverse of the differential inclusion probabilities θ_i of individuals i. The design weights form estimates that aim to generalise the survey estimates from the sample level to the level of the target population. Alternatively, they can be scaled so that the sampling distributions are balanced to weight up to the size of the net sample only. We denote the scaled weights by

$$w_i^* = \frac{1}{\theta_i^*} = \frac{1}{\theta_i} \frac{n^e}{\hat{N}} \tag{4.1.1}$$

where the θ_i^* refers to the scaled weights, the $\hat{N}$ to the estimated number of eligible individuals in the target population and n^e to the net sample size consisting on of eligible individuals. (The derivation of the inclusion probabilities is explained more in detail in Chapter 6.) Here the scaled weights are used, because they enable to compare the unweighted and weighted estimates in the survival and hazard functions as well as in the response probability models.

The probability of maintaining co-operation until t is defined as earlier in (4.1):

$$S(t) = \mathrm{P}(T_i \geq t) = P\left(T_i > t-1\right).$$

In design-based approach this is estimated by $\hat{S}(t)^*$, which is the weighted survival function for the individuals sampled representing the target population:

$$\hat{S}(t)^* = \prod_{h=1}^{t-1} \left(\frac{n_h^* - d_h^*}{n_h^*} \right) = \prod_{h=1}^{t-1} \left(\frac{\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i d_{ih}}{\sum_{i=1}^{n} w_i} \right) \tag{4.1.2}$$

where d_{ih} is a binary indicator for individual i at data collection phase h is a drop-out:

$$d_{ih} = \begin{cases} 0 \text{ if individual } i \text{ co-operates at phase } h \\ 1 \text{ if individual } i \text{ fail to co-operate at phase } h \end{cases}$$

$h = 1, \ldots, t-1$ and the weights w are defined in (4.27). The standard error of the survival function is modified from the one by presented by Greenwood (1926) into the one following:

$$\hat{s}_t^* = \hat{S}(t)^* \sqrt{\sum_{h=1}^{t-1} \left(\frac{\sum_{i=1}^{n} w_i d_{ih}}{\sum_{i=1}^{n} w_i \left(\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i d_{ih} \right)} \right)} \tag{4.1.3}$$

where $t = 1, \ldots, t_{\max}$ (in our analysis the $t_{\max} = 5$) and T_i takes possible values of $1, \ldots, t_{\max}$.

The discrete-time hazard function $\lambda(t)$ describes the probability:

$$\lambda(t) = P(T_i = t | T_i \geq t) = \frac{P(T_i = t)}{P(T_i \geq t)} = \frac{P(T_i = t)}{S(t)},$$

and in a design based approach it is estimated by the hazard function accordingly by:

$$\hat{\lambda}(t)^* = \frac{\sum_{i=1}^{n} w_i d_{ih} / \sum_{i=1}^{n} w_i}{\left(1 - \left(\sum_{i=1}^{n} w_i d_{ih} / \sum_{i=1}^{n} w_i \right) / 2 \right)} \tag{4.1.4}$$

and we estimate its standard error by

$$\hat{s}_{\hat{\lambda}_h}^* = \hat{\lambda}_h^* \sqrt{\frac{1 - (\lambda_h^* / 2)^2}{\sum_{i=1}^{n} w_i d_{ih}}}. \tag{4.1.5}$$

The weighted estimates of survival and hazard functions have been presented with the unweighted functions in the tables 4.2 and 4.3.

4.1.2 Design-based estimation of the response and non-response probabilities

The objective is to estimate the response probability for each eligible individual sampled using the background information available taking into account the complex sampling design features. The estimated probability is defined in logit model as previously in Chapter 2:

$$\text{logit}(\pi) = \ln\left(\frac{\pi}{1 - \pi} \right)$$

and subsequently the weighted response probability for the whole sample is derived using the response indicator and scaled design weights w_i^*

$$\hat{\pi} = \frac{\sum_{i=1}^{n} w_i^* r_i}{\sum_{i=1}^{n} w_i^*}$$

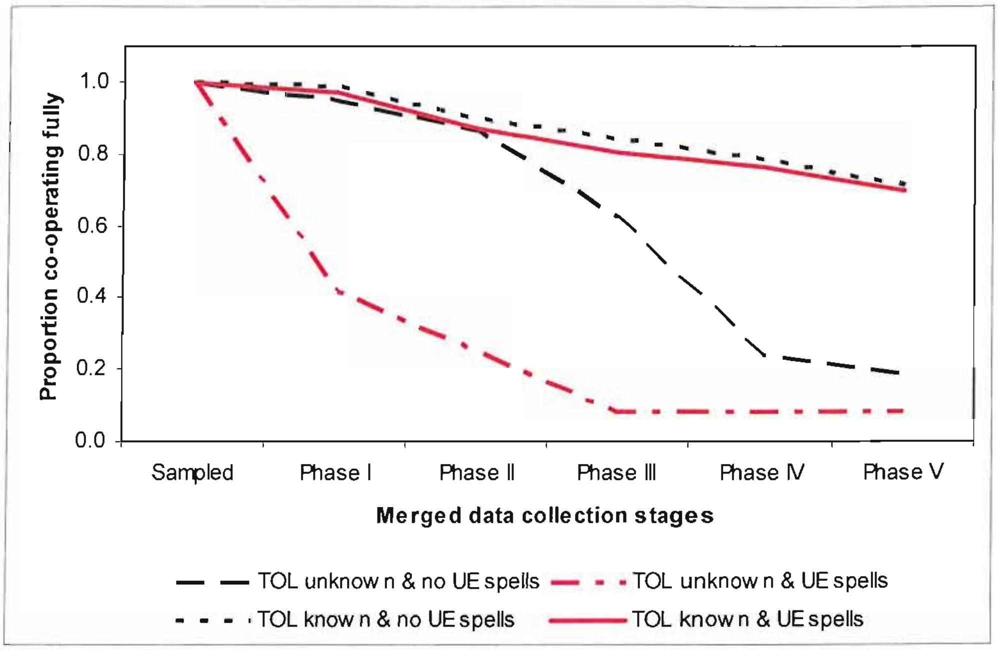
And the individual response probabilities are estimated by the function

$$g(\mathbf{x}_i) = \ln\left[\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)}\right] = w_i\beta_0 + w_i\beta_1 x_{1i} + \ldots + w_i\beta_P x_{Pi} = w_i \sum_{p=1}^{P} \beta_p x_{pi} \ .$$

In the estimation procedure of the design-based mode the survey design features are defined so that the clustering and stratification are taken into account in the variance estimation which affects the standard errors and the test statistics. This is explained more in detail in the SUDAAN manual provided by Research Triangle Institute (2004, Chapter 10). The design based estimates for response or non-response probabilities are derived accordingly when using weighted discrete-time hazard model, weighted sequential, cumulative or multinomial logit models.

APPENDIX 4.2. Survival plots by characteristics of sampled individuals

Figure 4.2.1 Survival plot by type of living and experience of unemployment



Note: TOL refers to type of living
UE refers to unemployment

Figure 4.2.2 Survival plot by family status

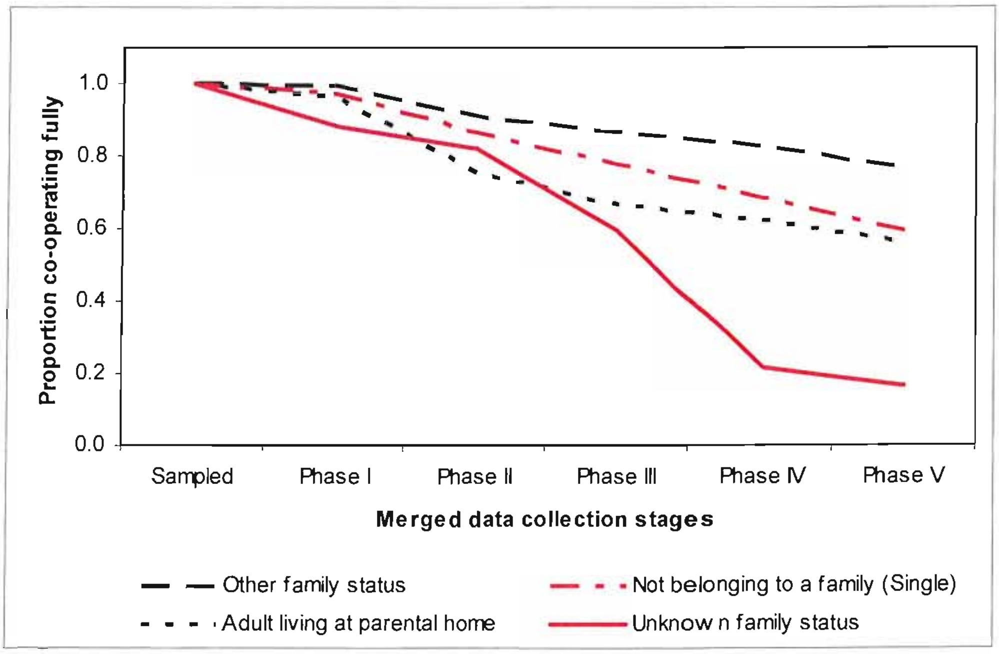


Figure 4.2.3 Survival plot by family status and maternal language

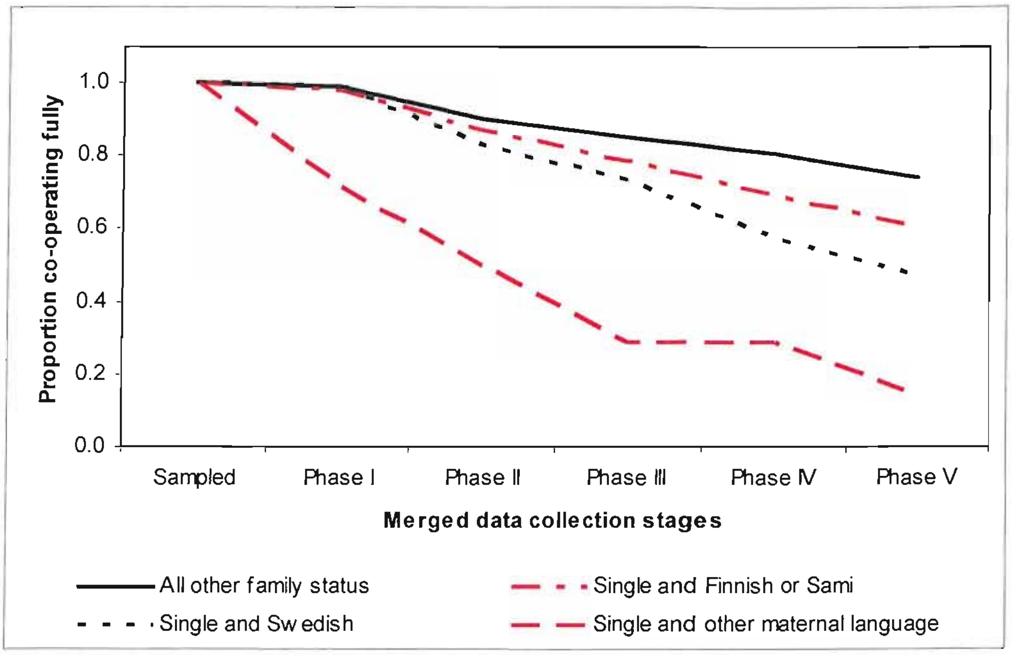


Figure 4.2.4 Survival plot by educational background and maternal language

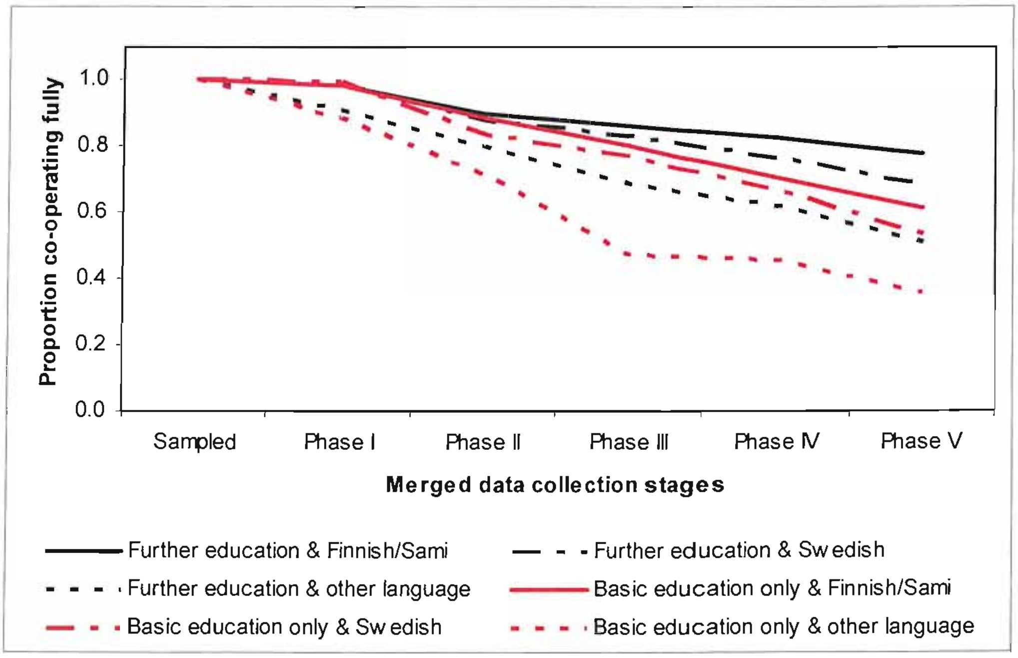


Figure 4.2.5 Survival plot by self-assessed health for those responded to the question in health interview

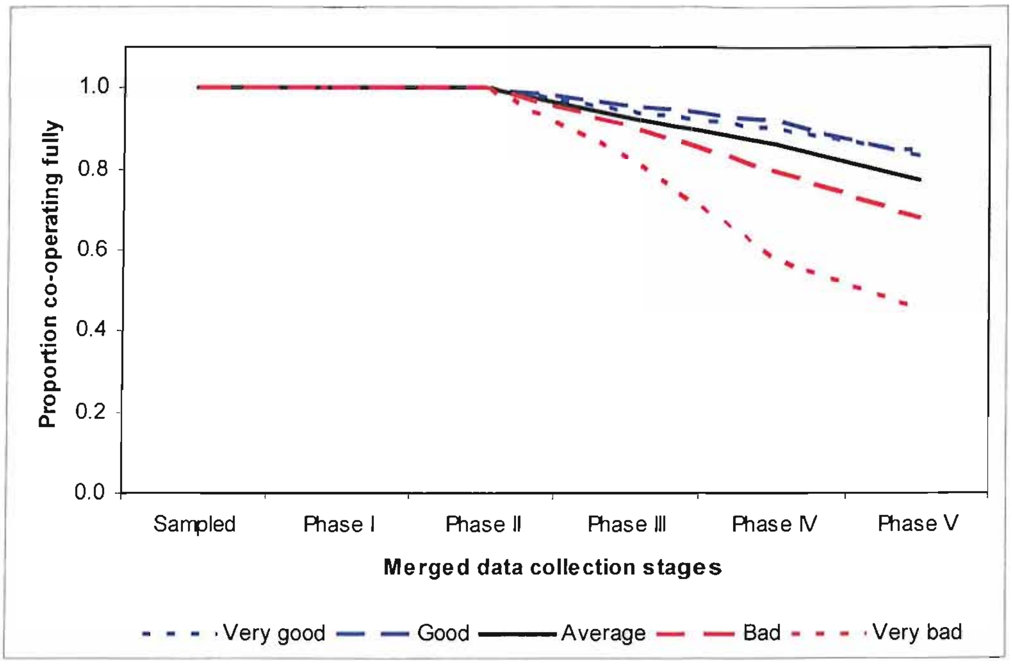


Figure 4.2.6 Survival plot by physical mobility for those responded to the question in health interview

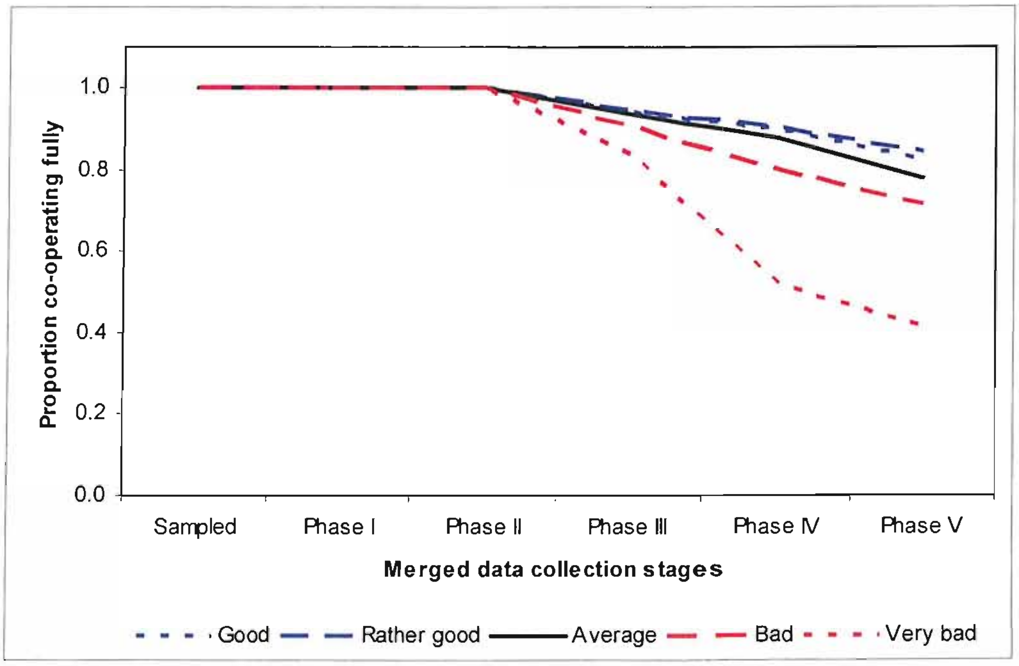


Figure 4.2.7 Survival plot by whether entitled to reimbursement on medical expenses on diabetes

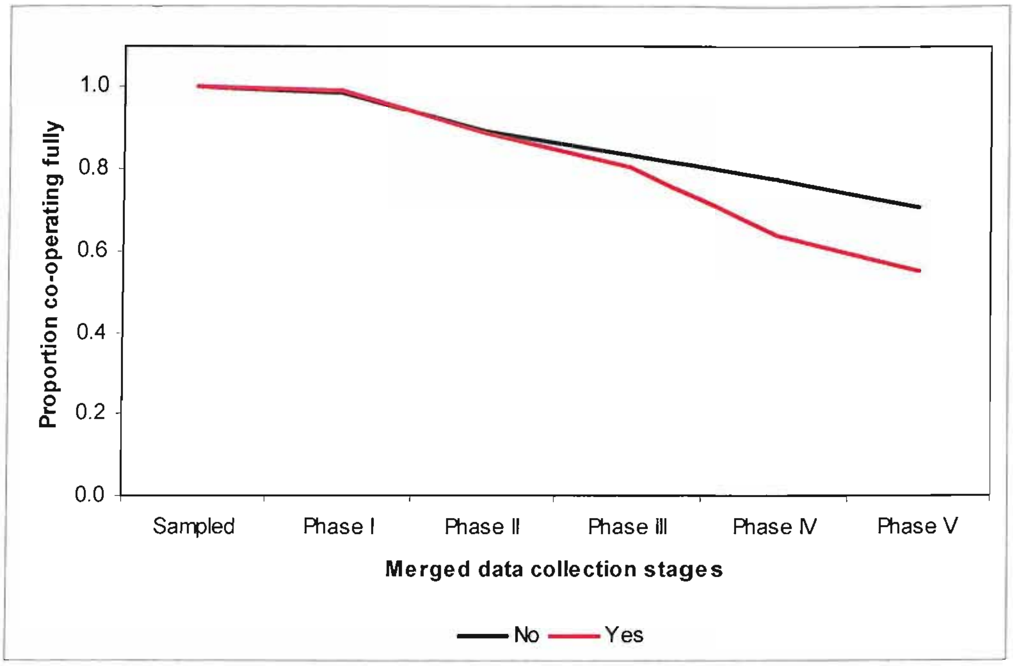
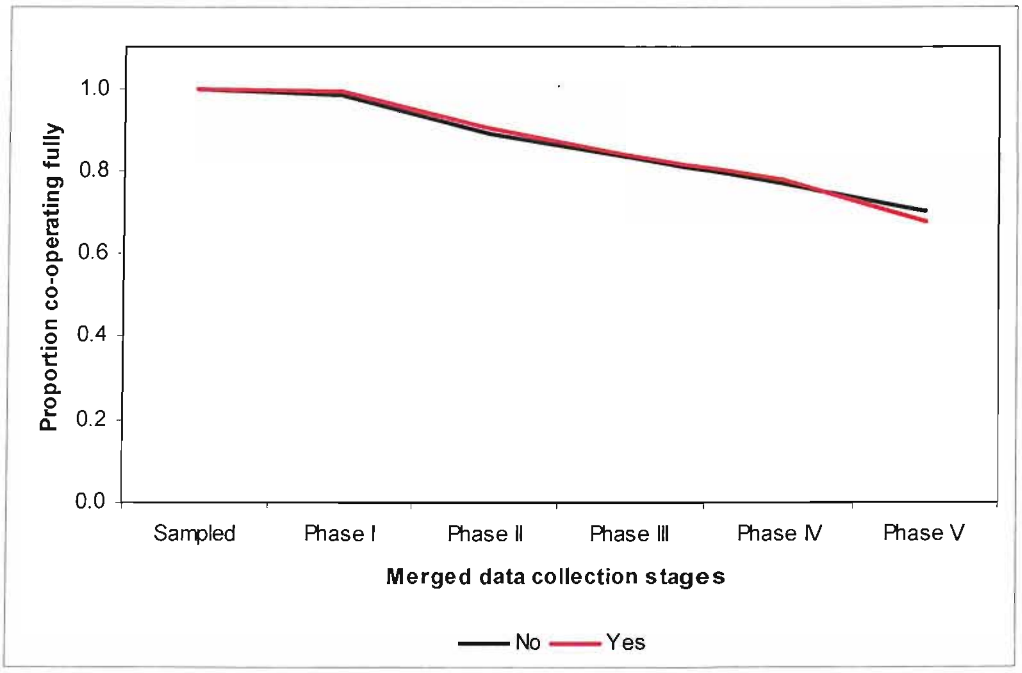


Figure 4.2.8 Survival plot by whether entitled to reimbursement on medical expenses on asthma



APPENDIX 4.3. Hazard functions by characteristics of sampled individuals

Figure 4.3.1 The hazard function by member of household population and experience of unemployment

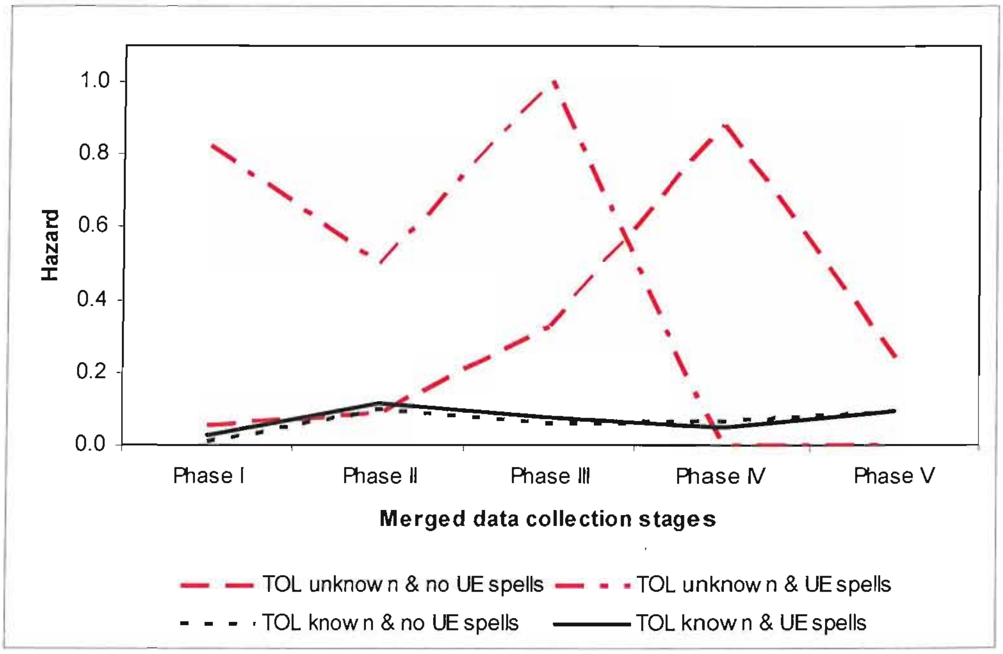


Figure 4.3.2 The hazard function by family status

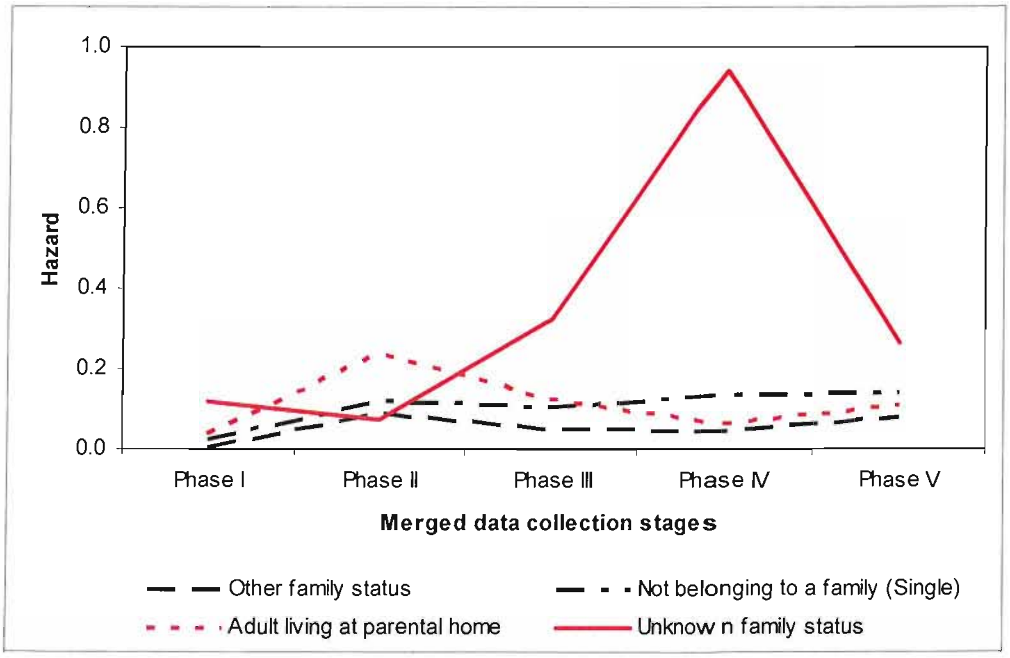


Figure 4.3.3 The hazard function by family status and native language

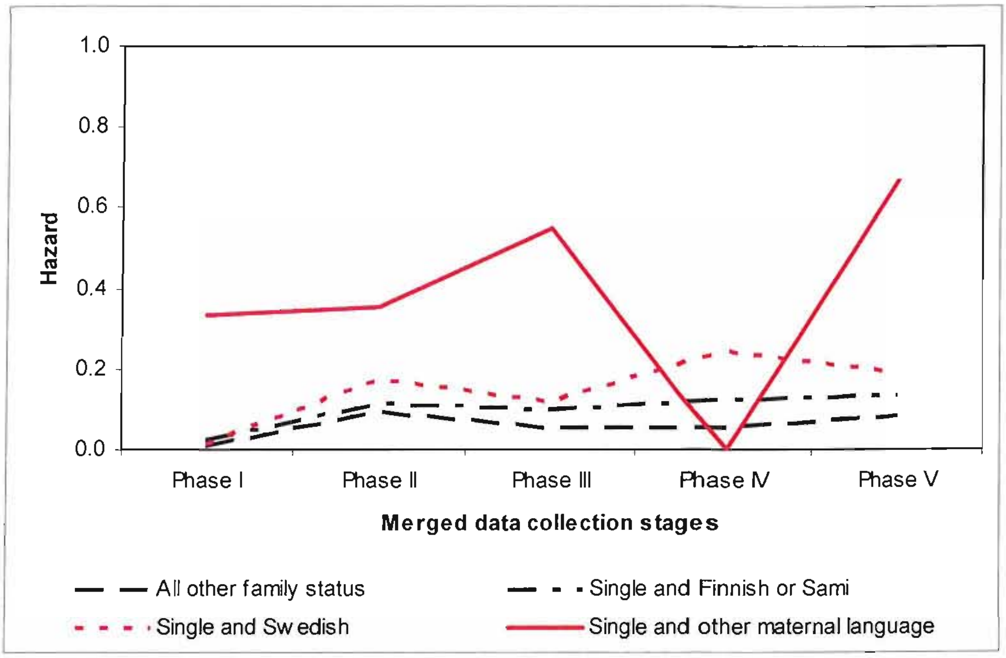


Figure 4.3.4 The hazard function by educational background and native language

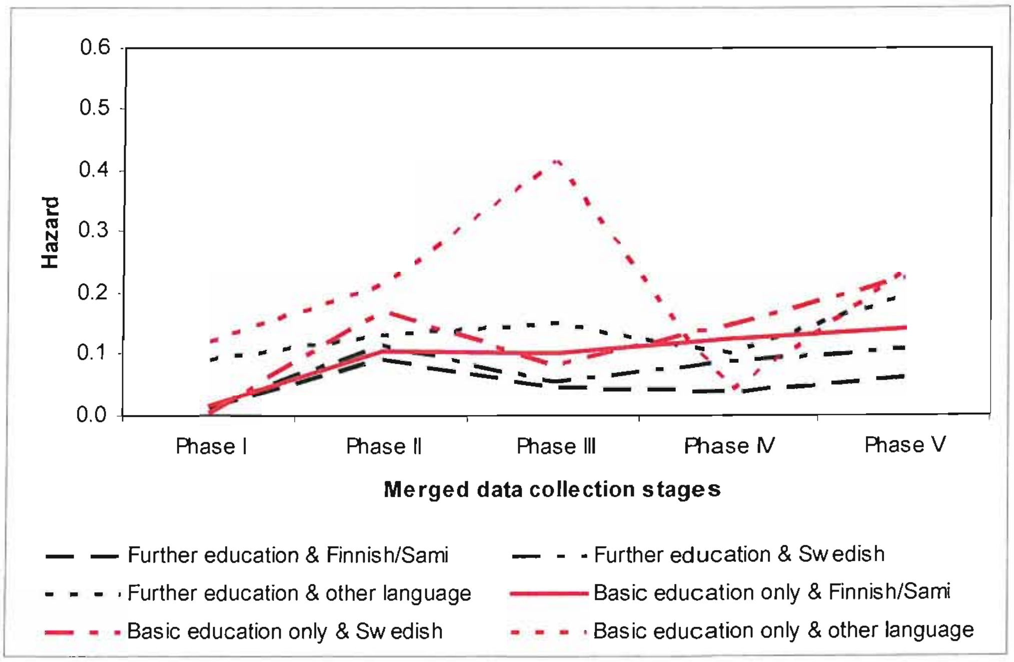


Figure 4.3.5 The hazard function by self-assessed health

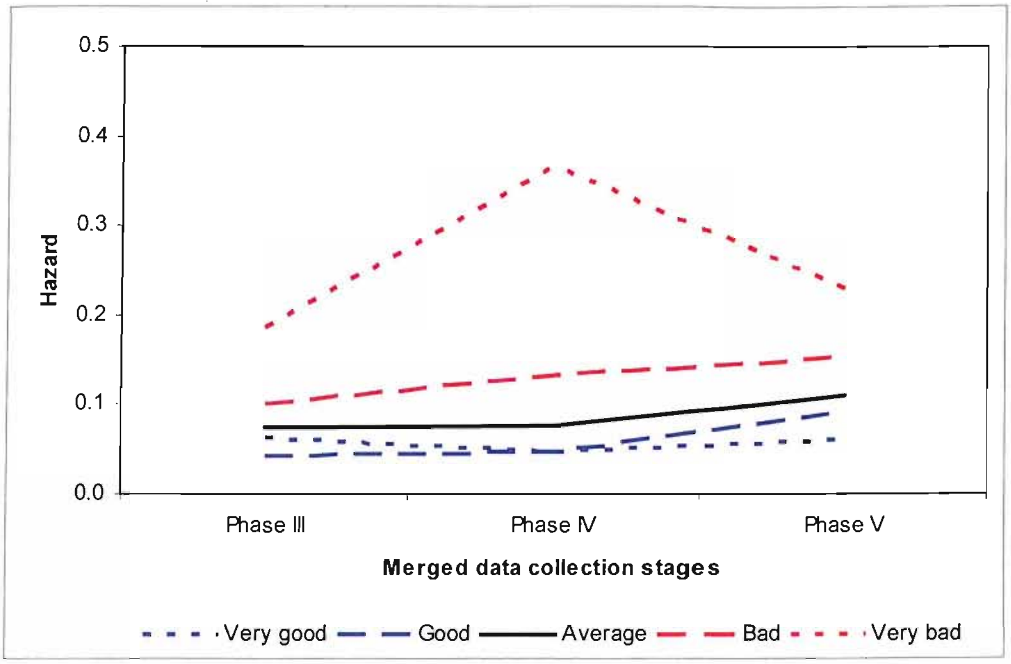


Figure 4.3.6 The hazard function by physical mobility

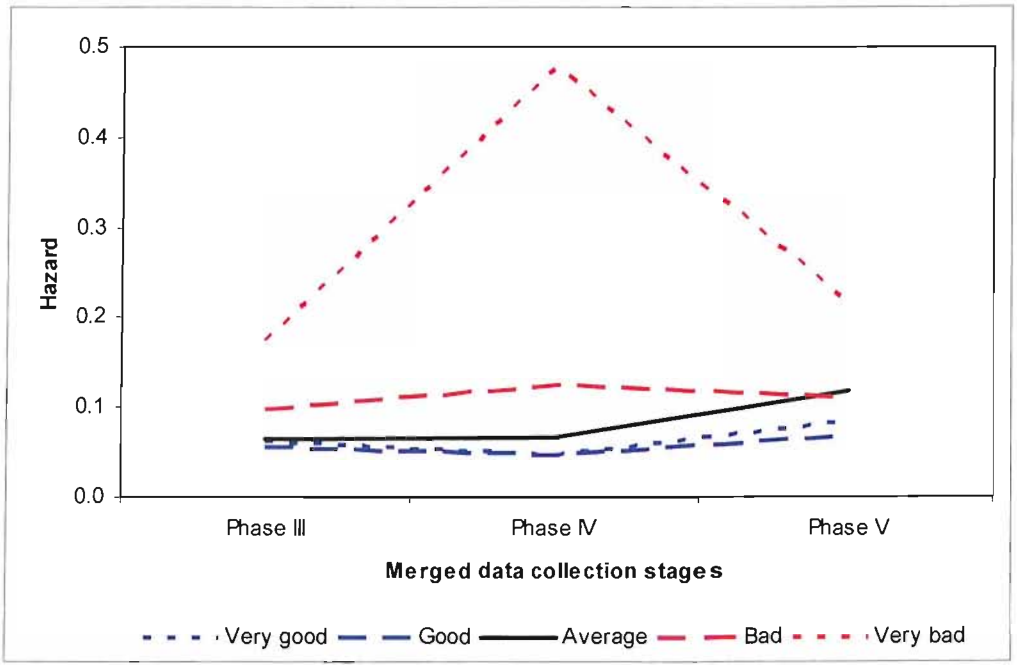


Figure 4.3.7 The hazard function by right for medical expenses' reimbursement based on diabetes

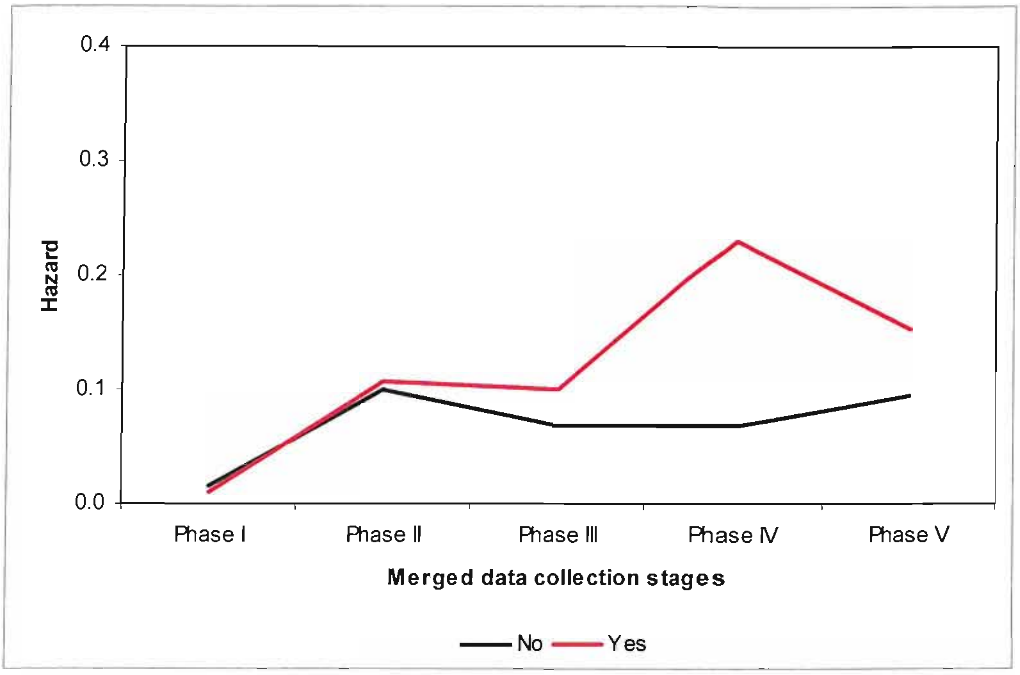
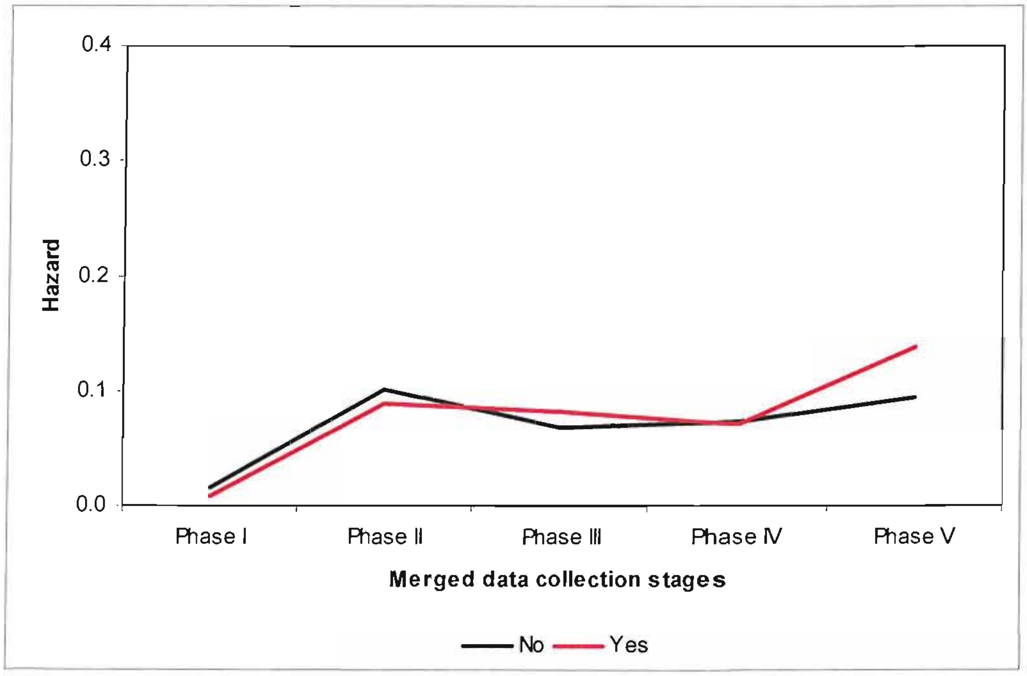


Figure 4.3.8 The hazard function by right for medical expenses' reimbursement based on asthma



APPENDIX 5.1 Interviewer performance by interviewer background characteristics

Table 5.1.1 Estimated interviewer contact rate by some background characteristics of the interviewer

Interviewer characteristics	All interviewers # of inter-viewers	Mean of completion rates	Standard Deviation	Anova F-test	p-value	Interviewers participated to the perception survey # of inter-viewers	Mean of completion rates	Standard Deviation	Anova F-test	p-value
Age group				3.1	0.03				3.6	0.02
25-39	24	97.1	2.7			21	96.8	2.7		
40-49	43	98.3	2.6			39	98.2	2.7		
50-59	78	98.2	2.9			74	98.3	2.7		
60-66	13	99.9	0.5			11	99.8	0.5		
Gender				5.0	0.03				5.1	0.03
Female	152	98.3	2.5			139	98.3	2.5		
Male	6	95.8	5.3			6	95.8	5.3		
Education level				5.2	0.01				5.6	0.00
Basic	92	98.5	2.4			85	98.5	2.2		
Secondary	52	98.2	2.5			46	98.1	2.6		
Tertiary or above	14	96.0	4.4			14	96.0	4.4		
Years of interviewing experience prior the survey				1.9	0.11				2.5	0.03
0	25	98.4	2.2			21	98.2	2.3		
1	26	98.3	2.5			23	98.1	2.6		
2-9	18	98.7	1.9			17	98.6	2.0		
10-14	28	96.9	3.1			28	96.9	3.1		
15-19	23	98.6	2.8			22	98.7	2.8		
20-24	24	98.6	2.7			21	99.2	1.6		
25+	14	97.9	3.5			13	97.8	3.6		
Main interviewing language of the interviewer				1.2	0.27				1.2	0.27
Finnish	146	98.2	2.6			133	98.2	2.5		
Swedish	12	97.3	4.1			12	97.3	4.1		
Main regions				3.8	0.01				3.1	0.02
Larger capital area1)	66	97.5	2.8			60	97.5	2.6		
Other Southern Finland	55	98.1	3.2			53	98.0	3.2		
Eastern Finland	18	99.2	1.0			16	99.2	1.0		
Middle Finland	13	99.9	0.4			11	99.9	0.5		
Northern Finland	6	99.7	0.7			5	99.6	0.8		
All interviewers	158	98.2	2.7			145	98.2	2.7		

[1)] and surrounding municipalities i.e. region of "Uusimaa"

Table 5.1.2 Estimated interviewer persuasion rate by some background characteristics of the interviewer

Interviewer characteristics	All interviewers: # of inter-viewers	All interviewers: Mean of completion rates	All interviewers: Standard Deviation	All interviewers: Anova F-test	All interviewers: p-value	Interviewers participated to the perception survey: # of inter-viewers	Interviewers participated to the perception survey: Mean of completion rates	Interviewers participated to the perception survey: Standard Deviation	Interviewers participated to the perception survey: Student's t-test	Interviewers participated to the perception survey: p-value
Age group				4.1	0.01				2.7	0.05
25-39	24	85.3	6.5			21	86.3	6.1		
40-49	43	90.0	5.6			39	90.3	5.6		
50-59	78	89.8	6.3			74	89.7	6.4		
60-66	13	86.9	7.3			11	86.6	7.5		
Gender				0.0	0.89				0.1	0.82
Female	152	88.9	6.3			139	89.2	6.2		
Male	6	88.6	9.3			6	88.6	9.3		
Education level				1.7	0.19				1.6	0.21
Basic	92	89.6	6.3			85	89.6	6.2		
Secondary	52	88.5	5.7			46	89.0	5.5		
Tertiary or above	14	86.4	9.1			14	86.4	9.1		
Years of interviewing experience prior beginning of the fieldw				0.9	0.50				0.7	0.65
0	25	87.8	6.9			21	88.8	6.0		
1	26	88.3	6.5			23	88.5	6.6		
2-9	18	88.4	5.4			17	88.8	5.4		
10-14	28	88.0	7.9			28	88.0	7.9		
15-19	23	90.7	5.8			22	90.9	5.8		
20-24	24	89.5	6.6			21	89.3	6.7		
25+	14	90.6	4.0			13	90.3	4.0		
Main interviewing language of the interviewer				7.2	0.01				8.3	0.00
Finnish	146	89.3	6.1			133	89.6	6.0		
Swedish	12	84.2	8.2			12	84.2	8.2		
Main regions				5.6	0.00				5.0	0.00
Larger capital area1)	66	86.9	6.4			60	87.3	6.1		
Other Southern Finland	55	88.9	6.9			53	88.9	6.9		
Eastern Finland	18	91.1	4.1			16	91.7	3.9		
Middle Finland	13	93.5	2.6			11	93.7	2.5		
Northern Finland	6	94.9	3.4			5	95.3	3.6		
All interviewers	158	88.9	6.4			145	89.1	6.3		

[1] and surrounding municipalities i.e. region of "Uusimaa"

APPENDIX 5.2 Interviewer perception survey questions used in the analyses[1]

36. How much to your opinion the following facts influenced to the survey participation of the interviewees:

		Consid- erably	In large amount	In small amount	Not at all	Consid- erably, %	In large amount, %	In small amount, %	Not at all, %
a.	Survey topic on health?	114	31	0	0	78.6	21.4	0.0	0.0
b.	Conducted by the National Public Health Institute?	31	72	32	4	22.3	51.8	23.0	2.9
c.	Statistics Finland collected the data?	22	62	41	10	16.3	45.9	30.4	7.4
d.	Your own professional experience?	40	66	29	2	29.2	48.2	21.2	1.5
e.	Letter for refusals?	0	8	73	57	0.0	5.8	52.9	41.3
f.	The motivation reasoning memo provided by the NPHI?	7	42	76	17	4.9	29.6	53.5	12.0
g.	Health examination?	117	26	2	0	80.7	17.9	1.4	0.0
h.	Good health condition of the target person?	2	33	85	17	1.5	24.1	62.0	12.4
i.	Bad health condition of the target person / illness?	13	71	52	5	9.2	50.4	36.9	3.6
j.	High age of the interviewee?	9	46	64	23	6.3	32.4	45.1	16.2
k.	Providing blood-pressure measurement tools for home use?	3	17	57	65	2.1	12.0	40.1	45.8
l.	Re-allocation of refusal cases to another interviewer?	1	22	58	42	0.8	17.9	47.2	34.2
m.	Publicity the survey gained in the media?	48	62	31	3	33.3	43.1	21.5	2.1

37. How much to your opinion the following facts influenced <u>to the refusal</u> of the interviewees:

		Consid- erably	In large amount	In small amount	Not at all	Consid- erably, %	In large amount, %	In small amount, %	Not at all, %
a.	Survey topic on health?	10	30	69	28	7.3	21.9	50.4	20.4
b.	Conducted by the National Public Health Institute?	0	5	51	83	0.0	3.6	36.7	59.7
c.	Statistics Finland collected the data?	0	0	47	90	0.0	0.0	34.3	65.7
d.	Data collection required home visit?	5	27	71	35	3.6	19.6	51.5	25.4
e.	Sensitive questions?	6	34	58	41	4.3	24.5	41.7	29.5
f.	The length of the interview?	14	41	58	28	9.9	29.1	41.1	19.9
g.	Health examination?	5	35	68	29	3.7	25.6	49.6	21.2
h.	Good health condition of the target person?	18	48	50	21	13.1	35.0	36.5	15.3
i.	Bad health condition of the target person / illness?	23	50	42	24	16.6	36.0	30.2	17.3
j.	High age of the interviewee?	22	38	52	28	15.7	27.1	37.1	20.0
k.	Considerations on inadequate data protection?	2	12	60	60	1.5	9.0	44.8	44.8
l.	Insufficient information at the beginning of the survey?	5	12	57	64	3.6	8.7	41.3	46.4

[1] Translated from Finnish into English and answer categories shown are those recoded for the purpose of this survey

Appendix 5.2 continues

39. Please consider your own behaviour during the contact of the interviewees at the data collection in Health 2000. How often did you:

	Allways	Often	Rarely	Never	%	Often,%	Rarely,%	Never,%
a. Mention something positive about interviewees home or themselves? [1]	33	57	42	10	23.1	39.9	29.4	7.0
b. Refer that most people are happy to participate to the survey?	21	80	34	10	14.5	55.2	23.5	6.9
c. Inform how the survey results benefit the respondents themselves?	94	45	5	1	64.8	31.0	3.5	0.7
d. Mention that the interview must be carried out by a certain date?	13	27	77	28	9.0	18.6	53.1	19.3
e. Begin the interview before the interviewee has shown any signs of willingness to participate?	1	2	23	112	0.7	1.5	16.7	81.2

[1] One interviewer had respondent '8' indicating other than the given options to the question 39A.

40. To what extent do the following claims relate to your opinion on interviewers professional role?

	Fully agree	Agree to some extent	No opinion	Disagree to some extent	Disagree fully	Fully agree, %	Agree to some extent, %	No opinion, %	Disagree to some extent, %	Disagree fully, %
a. Reluctant respondent should always be persuaded to participate	19	56	9	43	18	13.1	38.6	6.2	29.7	12.4
b. With enough efforts even the most reluctant respondent can be converted to respond	5	26	14	58	42	3.5	17.9	9.7	40.0	29.0
c. An interviewer should respoect the privacy of the respondent	115	26	2	1	0	79.9	18.1	1.4	0.7	0.0
d. If refusal is reluctant, refusal should be accepted	8	31	11	79	16	5.5	21.4	7.6	54.5	11.0
e. Voluntariness of participation always be emphasised	22	25	10	60	28	15.2	17.2	6.9	41.4	19.3

Appendix 5.2 continues

How much the following affected to your work motivation regarding the Health 2000 survey:

	Increased significantly	Increased some extent	No effect	Decreased some extent	Decreased significantly	Increased significantly, %	Increased some extent, %	No effect, %	Decreased some extent, %	Decreased significantly, %
Interviewer training?	65	68	11	1	0	44.8	46.9	7.6	0.7	0.0
Centralised support from Statistics Finland?	16	59	68	1	1	11.0	40.7	46.9	0.7	0.7
Deedback from interviewees?	87	46	11	0	0	60.4	31.9	7.6	0.0	0.0
Survey topic that interested me?	86	52	7	0	0	59.3	35.9	4.8	0.0	0.0
The significance of the survey to the society?	87	54	4	0	0	60.0	37.2	2.8	0.0	0.0
Publicity the survey gained in the media?	53	65	26	1	0	36.6	44.8	17.9	0.7	0.0
Regional meetings of the interviewers?	27	64	52	0	0	18.9	44.8	36.4	0.0	0.0
Other contacts with the interviewers?	63	58	24	0	0	43.5	40.0	16.6	0.0	0.0
Feedback regarding the progress of the fieldwork?	16	53	71	2	0	11.3	37.3	50.0	1.4	0.0
Contacts to the National Public Health Institute?	5	34	94	7	3	3.5	23.8	65.7	4.9	2.1
High work load in autumn 2000?	29	23	88	4	0	20.1	16.0	61.1	2.8	0.0
Unexceptional regional clustering of the interviews?	14	33	92	4	2	9.7	22.8	63.5	2.8	1.4
Possibility to participate myself to the health examination?	28	39	75	1	1	19.4	27.1	52.1	0.7	0.7
Tight schedule of the fieldwork period?	15	36	84	9	1	10.3	24.8	57.9	6.2	0.7

3. What was your perception on:

	Very positive	Rather positive	Neutral	Rather negative	Very negative	Very positive, %	Rather positive, %	Neutral, %	Rather negative, %	Very negative, %
the interviewee's predominant attitudes towards the survey?										
Their attitude was generally:	109	35	1	0	0	75.2	24.1	0.7	0.0	0.0
on the Health 2000 survey as a whole?	84	55	4	1	1	57.9	37.9	2.8	0.7	0.7

Did you participate yourself to the health examination organised by the National Public Health Institute?

	Yes	No	Yes, %	No, %
	102	43	70.3	29.7

APPENDIX 5.3 Estimated interviewer success rate by their perceptions

(a) Opinion of the survey participation of the interviewees

How much to interviewer's opinion the following facts influenced to the survey participation of the interviewees:	Completion of the interviewer assignment, % Consid-erably	In large amount	In small amount	Not at all	Anova F-test	p-value
Survey topic on health?	87.4	88.1	-	-	0.2	0.62
Conducted by the NPHI[1]?	87.9	87.7	86.7	87.5	0.2	0.90
Statistics Finland collected the data?	89.0	88.7	85.7	84.4	2.5	0.07
Your own professional experience?	89.3	88.2	84.7	87.0	2.8	0.04
Letter for refusals?	-	86.5	87.1	87.8	0.2	0.81
The motivation reasoning memo provided by the NPHI[1]?	88.3	88.2	87.5	85.2	0.8	0.50
Health examination?	87.7	87.9	75.3	-	3.2	0.04
Good health condition of the target person?	88.2	87.2	87.3	89.1	0.3	0.80
Bad health condition of the target person / illness?	87.3	87.0	87.7	92.4	0.9	0.44
High age of the interviewee?	85.4	87.6	87.7	87.4	0.3	0.84
Providing blood-pressure measurement tools for home use?	89.0	88.5	88.2	87.2	0.4	0.77
Re-allocation of refusal cases to another interviewer?	..	89.8	87.2	87.2	1.4	0.23
Publicity the survey gained in the media?	87.9	87.8	86.0	92.8	1.1	0.35

1) National Public Health Institute

\- no observation

.. only single observation

Appendix 5.3 continues

(b) Opinion of the refusal of the interviewees

How much to interviewer's opinion the following facts influenced to the refusal of the interviewees:	Completion of the interviewer assignment, % Consid-erably	In large amount	In small amount	Not at all	Anova F-test	p-value
Survey topic on health?	87.4	87.6	86.7	89.0	0.7	0.53
Conducted by the NPHI[1]?	-	86.4	87.1	87.7	0.2	0.84
Statistics Finland collected the data?	-	-	86.3	88.1	2.2	0.14
Data collection required home visit?	82.5	86.0	87.3	89.6	2.5	0.07
Sensitive questions?	88.8	87.4	87.0	87.9	0.2	0.87
The length of the interview?	85.5	87.3	86.8	90.5	2.4	0.07
Health examination?	85.5	86.3	87.4	89.5	1.3	0.29
Good health condition of the target person?	84.7	87.1	87.5	91.7	3.6	0.02
Bad health condition of the target person / illness?	83.6	87.1	88.8	90.0	4.1	0.01
High age of the interviewee?	84.6	86.7	88.3	89.2	2.3	0.08
Considerations on inadequate data protection?	83.6	86.1	86.5	88.8	1.5	0.22
Insufficient information at the beginning of the survey?	89.9	86.5	86.3	88.5	1.3	0.28

1) National Public Health Institute

- no observation

.. only single observation

Appendix 5.3 continues

(c) Interviewer behaviour in interviewer-respondent interaction

Interviewer behaviour during the contact of the interviewees at the data collection in Health 2000. How often did:	Completion of the interviewer assignment, % Always	Often	Rarely	Never	Anova F-test	p-value
Mention something positive about interviewees home or themselves?	87.8	88.1	87.2	85.4	0.4	0.83
Refer that most people are happy to participate to the survey?	83.4	88.7	87.9	85.3	3.8	0.01
Inform how the survey results benefit the respondents themselves?	87.0	88.2	91.2	¨	0.9	0.46
Mention that the interview must be carried out by a certain date?	88.7	88.2	87.2	87.3	0.3	0.84
Begin the interview before the interviewee has shown any signs of willingness to participate?	¨	93.5	90.1	87.0	1.8	0.15

- no observation

¨ only single observation

(d) Professional attitudes of interviewers

To what extent do the following claims relate to your opinion on interviewers professional role?	Completion of the interviewer assignment, % Fully agree	Agree to some extent	No opinion	Disagree to some extent	Disagree fully	Anova F-test	p-value
Reluctant respondent should always be persuaded to participate	85.56	88.81	87.11	88.50	83.50	2.7	0.04
With enough efforts even the most reluctant respondent can be converted to respond	87.88	88.77	88.09	87.55	86.50	0.5	0.77
An interviewer should respect the privacy of the respondent	87.90	86.16	80.78	¨	-	1.2	0.31
If refusal is reluctant, refusal should be accepted	89.48	86.92	85.77	87.77	87.75	0.4	0.81
Voluntariness of participation always be emphasised	86.47	86.38	82.78	88.35	89.31	2.2	0.08

- no observation

¨ only single observation

Appendix 5.3 continues

(e) Interviewer work motivation

How much the following affected to work motivation regarding the Health 2000 survey:	Completion rate of the interviewer assignment, % Increased significantly	Increased to some extent	No effect	Decreased to some extent	Decreased significantly	Anova F-test	p-value
Interviewer training	88.17	86.97	86.88	..	-	0.4	0.74
Centralised support from Statistics Finland	89.81	88.31	86.26	..	..	2.4	0.05
Feedback from interviewees	88.36	86.47	87.17	-	-	1.2	0.31
Survey topic that interested me	88.78	85.34	88.38	-	-	4.1	0.02
The significance of the survey to the society	88.70	85.47	89.86	-	-	3.9	0.02
Publicity the survey gained in the media	88.49	87.91	84.46	..	-	2.2	0.09
Regional meetings of the interviewers	86.95	88.26	86.84	-	-	0.7	0.51
Other contacts with the interviewers	87.79	88.23	85.13	-	-	1.8	0.18
Feedback regarding the progress of the fieldwork	87.40	87.25	87.69	86.86	-	0.0	0.99
Contacts to the NPHI[1]	83.54	90.06	86.72	89.29	84.83	2.1	0.09
High work load in autumn 2000	90.06	88.99	86.79	84.99	-	2.4	0.07
Unexceptional regional clustering of the interviews	89.90	89.65	86.28	85.39	97.44	3.1	0.02
Possibility to participate to the health examination	86.82	86.37	88.36	..	..	0.7	0.58
Tight schedule of the fieldwork period	89.10	88.89	87.42	84.03	..	7.4	0.00

1) National Public Health Institute

- no observation

.. only single observation

Appendix 5.3 continues

(f) Interviewer perception of the survey

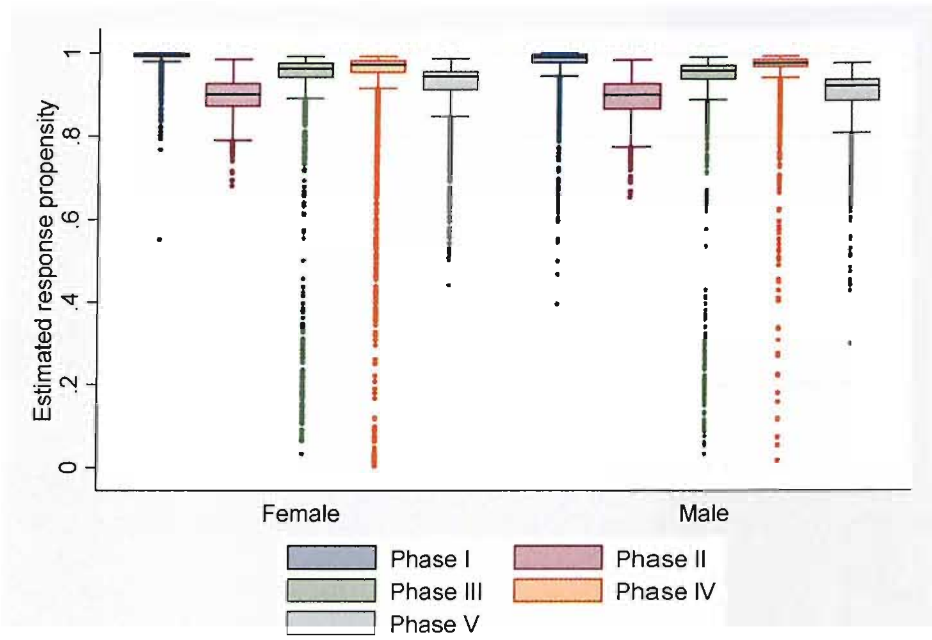
Perception:	Completion rate of the interviewer assignment, % Very positive	Rather positive	Neutral	Rather negative	Very negative	Anova F-test	p-value
Interviewee's predominant attitudes towards the survey	88.52	84.98	¨	-	-	8.1	0.00
Interviewers' own perception on the Health 2000 survey as a whole	89.32	85.07	85.39	¨	¨	4.0	0.00
Interviewers' own experiences on the health examination	88.75	87.82	88.04	90.25	88.43	0.3	0.87
	Yes	**No**				**Anova F-test**	**p-value**
Interviewer participeted to the health examination	88.32	85.65				4.5	0.04

- no observation

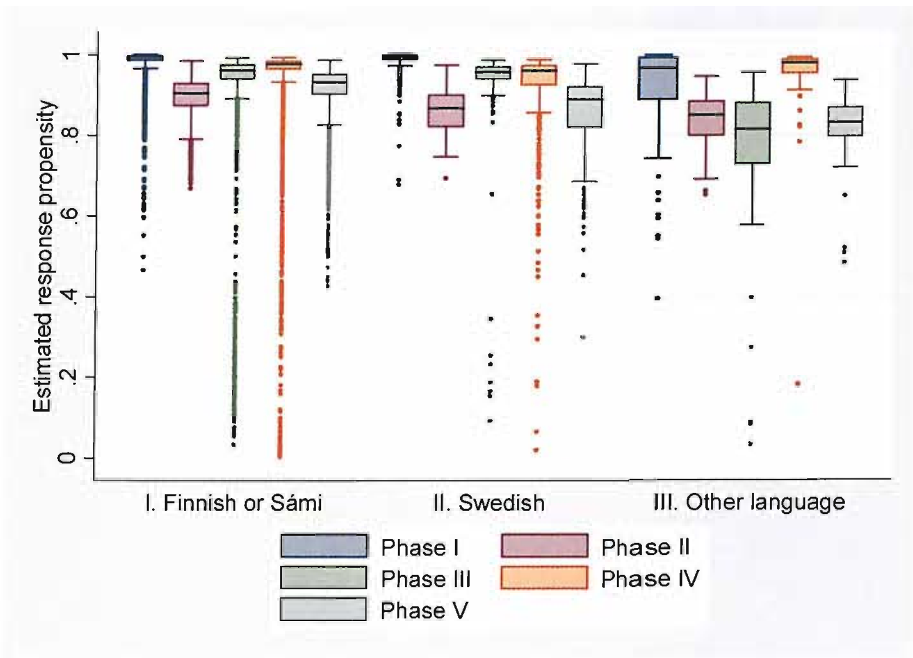
¨ only single observation

APPENDIX 5.4 Predicted probabilities of sequential co-operation by population subgroups and data collection phases

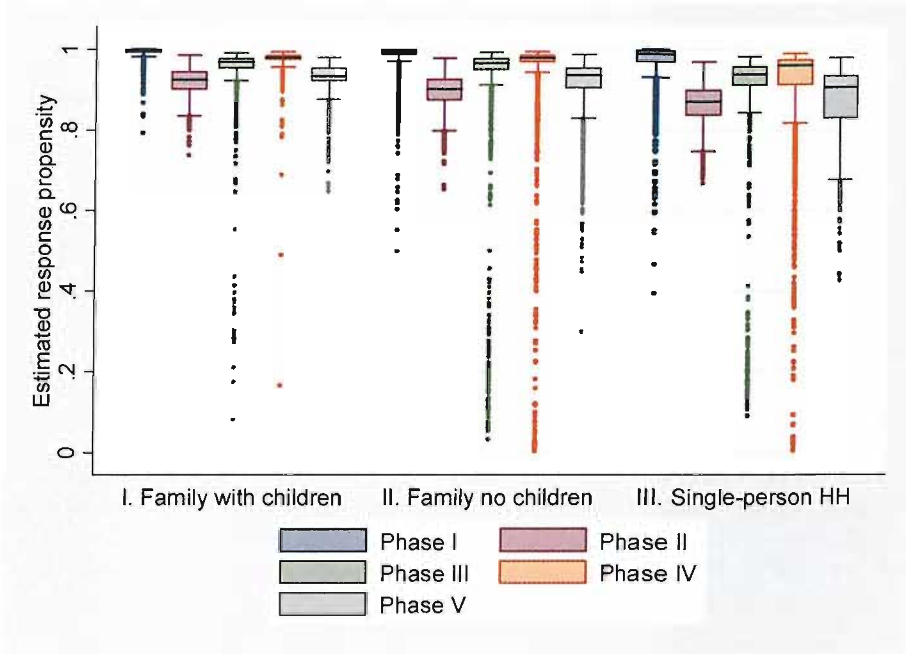
(a) by sex of the individual



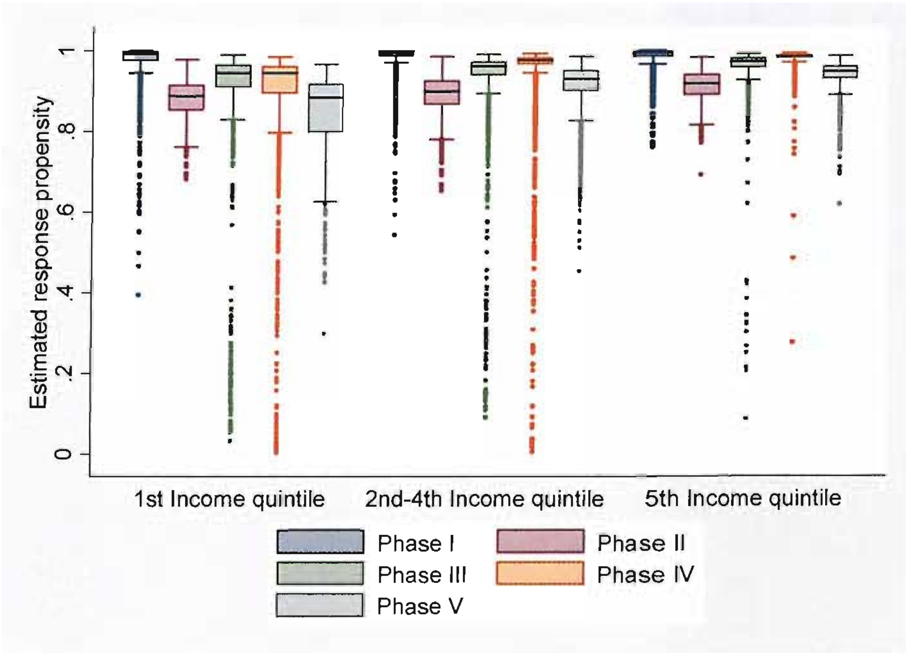
(b) by language group of the individual



(c) by family background of the individual



(d) by income quintile of register derived disposable income of the individual



APPENDIX 6.1 Description of the sampling weights for the Health 2000 data

In the Finnish Health Examination Survey 2000 there were $j = 80$ health centre districts (PSUs) sampled in the clustered sampling design which were allocated according to the five university hospital districts (UHD) so that each district have $m_j = 80/5 = 16$ PSUs sampled.

The sample size n_j for each PSU was defined based on relative allocation i.e.:

$$n_j = n(N_j / N) \text{ and } \sum_{j=1}^{J} n_j = n,$$

where N refers to the size of the total target population in the sampling frame and N_j to the size of the target population within the jth PSU and n to the pre-defined sampled size. However, equal allocation was used to define the sample size n_j in the 15 largest towns. The total target population is thus divided along the subsets $N = N_1 + N_2$, where the N_1 refers to the subset of the 15 largest towns and the N_2 to the remaining PSUs. Thus the inclusion probability for the individuals at PSU j (health centre district) was defined as:

$$\theta_j = \begin{cases} \dfrac{n_j}{N}\dfrac{N}{N_1} & \text{if } j \subset N_1 \\ \dfrac{m_l n_j}{N_2} & \text{if } j \subset N_2 \end{cases},$$

where m_l refers to the number of PSUs sampled within the UHD l and n_j refers to the sampling size within each PSU. The subscript 2 refers to the second phase of sampling for the derivation of which the 15 largest towns were excluded. The inverse of these inclusion probabilities define the sample weights.

At the 2nd phase of the sampling the PSUs were ordered by gender and age. People aged 80+ were sampled with double inclusion probabilities in relation to other age groups. Subsequently people aged 80 or over were sampled with double inclusion probability than the other within the clusters; the sampling design has been described in detail in Chapter 3. The initial inclusion probabilities for each individual are thus defined as

$$\theta_i = \begin{cases} \theta_j & \text{if } 30 \leq \text{age} < 80 \\ 2\theta_j & \text{if age} \geq 80 \end{cases}$$

, and the initial weight d is defined as

$$d_i = \frac{1}{\theta_i} \text{ and } \sum_{i=1}^{n} d_i = N.$$

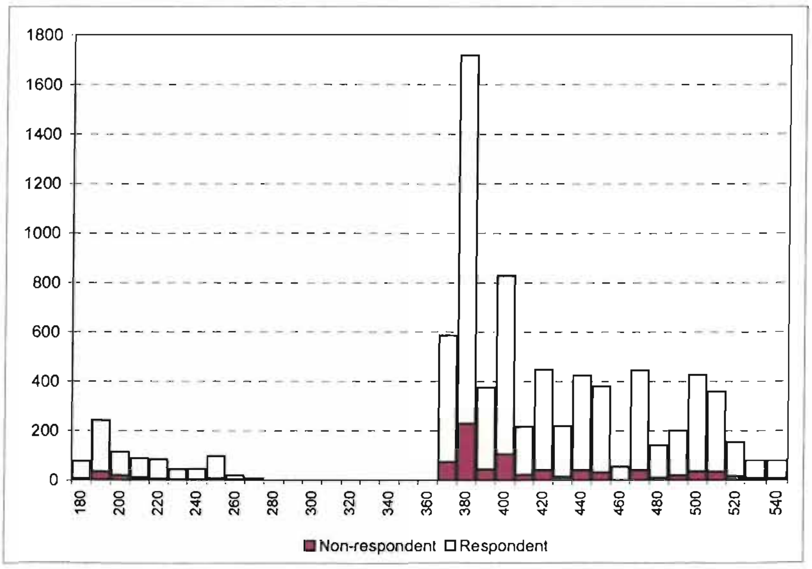
Subsequently the design weights d_i can be scaled in the obtained sample due to over-coverage detected so that the estimates can be generalised to the level of the target population within the sampling frame:

$$d_i^* = \frac{1}{\theta_i^*} = \frac{1}{\theta_i}\frac{N_e}{N},$$

where the N_e refers to the subset of the eligible individuals in the sampling frame that is not necessarily known in advance but about which our knowledge is increasing during the fieldwork period.

The purpose of weighting in sampling surveys is to adjust to sampling and non-response errors in relation to the frame and the target population in order to obtain proper inference of survey estimates and reduce the bias of the estimates. Final survey weights allow raising the sample level estimates to the level of the frame population. In most cases the weight construction is based on the inverse of the original inclusion probabilities i.e. sample weights, which are dependent on the sample design. However, these design weights are not sufficient for most practical situations because after sampling and data collection there may be total errors due to frame errors, sampling errors, non-response and/or measurement errors. Therefore the sample weights have to be adjusted on the basis of model assumptions and that phase of weighting is called re-weighting as the original weights are adjusted utilising auxiliary information.

Figure 6.1.1 Distribution of sample weights by the response status



APPENDIX 6.2 Calibration estimates for sampling weight on demographic variables

SAS CALMAR macro, health interviewed with design weight

Iteration criteria
Criteria after iteration 1: 1.33527
Criteria after iteration 2: 0.00010
Criteria after iteration 3: 0.00000

Comparison between the sample distributions (with the initial sampling weights) and the actual distribution of values in the target population

Variable	Value of variable	Distribution Sample	Distribution Population	Percentage Sample	Percentage Population
MATERNAL LAUNGUAGE					
	FINNISH OR SÁMI 1	3029178.28	3029011.18	93.07	93.07
	SWEDISH 2	190973.25	172520.88	5.87	5.30
	OTHER 3	34529.47	53148.94	1.06	1.63
REGIONAL GROUPS	1	335655.78	485715.57	10.31	14.92
	2	326610.12	438952.32	10.04	13.49
	3	510073.66	757592.10	15.67	23.28
	4	411666.44	559053.30	12.65	17.18
	5	350738.83	436911.63	10.78	13.42
	6	1319936.17	576456.08	40.56	17.71
AGE AND SEX GROUP	MALE 30-39	323470.39	366380.00	.	.
	MALE 40-49	352918.30	392888.00	.	.
	MALE 50-59	343782.02	367322.00	.	.
	MALE 60-69	214713.56	225395.00	.	.
	MALE 70-79	131977.66	146129.00	.	.
	MALE 80+	85925.72	47788.00	.	.
	FEMALE 30-39	353104.38	351929.00	.	.
	FEMALE 40-49	394816.34	383007.00	.	.
	FEMALE 50-59	353310.23	365789.00	.	.
	FEMALE 60-69	251893.08	254092.00	.	.
	FEMALE 70-79	207405.65	226153.00	.	.
	FEMALE 80+	241363.68	127809.00	.	.

APPENDIX 6.3 Calibration estimates for sampling weight on socio-economic variables

SAS CALMAR macro, health interviewed with design weight

Iteration criteria
Criteria after iteration 1: 0.80729
Criteria after iteration 2: 0.00047
Criteria after iteration 3: 0.00000

Comparison between the sample distributions (with the initial sampling weights) and the actual distribution of values in the target population

Variable	Value of variable	Distribution Sample	Distribution Population	Percentage Sample	Percentage Population
MATERNAL LAUNGUAGE					
	FINNISH OR SÁMI 1	3029178.28	3029011.18	93.07	93.07
	SWEDISH 2	190973.25	172520.88	5.87	5.30
	OTHER 3	34529.47	53148.94	1.06	1.63
SES	WAGE EARNER	1573507.52	1548902.69	48.35	47.59
	SELF-EMPLOYED	112356.94	117493.98	3.45	3.61
	FARMER	119675.15	115215.71	3.68	3.54
	PENSIONER	1106573.47	1066884.43	34.00	32.78
	OTHER	342567.92	406184.19	10.53	12.48
INCOME SUPOORT	YES	143382.65	169243.41	4.41	5.20
	NO	3111298.35	3085437.59	95.59	94.80
CAPITAL INCOME	YES	1009058.09	996583.32	31.00	30.62
	NO	2245622.91	2258097.68	69.00	69.38
AGE AND SEX GROUP	MALE 30-39	323470.39	366380.00	.	.
	MALE 40-49	352918.30	392888.00	.	.
	MALE 50-59	343782.02	367322.00	.	.
	MALE 60-69	214713.56	225395.00	.	.
	MALE 70-79	131977.66	146129.00	.	.
	MALE 80+	85925.72	47788.00	.	.
	FEMALE 30-39	353104.38	351929.00	.	.
	FEMALE 40-49	394816.34	383007.00	.	.
	FEMALE 50-59	353310.23	365789.00	.	.
	FEMALE 60-69	251893.08	254092.00	.	.
	FEMALE 70-79	207405.65	226153.00	.	.
	FEMALE 80+	241363.68	127809.00	.	.

NOTE: SES denotes for socio-economic status; a register derived grouping developed for the purpose of the Income distribution survey is used here.

APPENDIX 6.4 Calibration estimates for sampling weight on income inequality variables

SAS CALMAR macro, health interviewed with design weight

Iteration criteria
Criteria after iteration 1: 1.44709
Criteria after iteration 2: 0.00010
Criteria after iteration 3: 0.00000

Comparison between the sample distributions (with the initial sampling weights) and the actual distribution of values in the target population

Variable	Value of variable	Distribution Sample	Distribution Population	Percentage Sample	Percentage Population
MATERNAL LAUNGUAGE					
	FINNISH OR SÁMI 1	3029178.28	3029011.18	93.07	93.07
	SWEDISH 2	190973.25	172520.88	5.87	5.30
	OTHER 3	34529.47	53148.94	1.06	1.63
INCOME QUINTILE	0	57885.58	27013.85	1.78	0.83
	I	539247.73	645533.43	16.57	19.83
	II	633942.02	645533.43	19.48	19.83
	III	623926.76	645533.43	19.17	19.83
	IV	652937.70	645533.43	20.06	19.83
	V	746741.21	645533.43	22.94	19.83
CAPITAL INCOME	YES	1009058.09	996583.32	31.00	30.62
	NO	2245622.91	2258097.68	69.00	69.38
RECEIVED INCOME SUPPORT	YES	143382.65	169243.41	4.41	5.20
	NO	3111298.35	3085437.59	95.59	94.80
RECEIVED UNEMPLOYMENT BENEFIT					
	YES	430433.19	464442.98	13.23	14.27
	NO	2824247.81	2790238.02	86.77	85.73
AGE AND SEX GROUP	MALE 30-39	323470.39	366380.00	.	.
	MALE 40-49	352918.30	392888.00	.	.
	MALE 50-59	343782.02	367322.00	.	.
	MALE 60-69	214713.56	225395.00	.	.
	MALE 70-79	131977.66	146129.00	.	.
	MALE 80+	85925.72	47788.00	.	.
	FEMALE 30-39	353104.38	351929.00	.	.
	FEMALE 40-49	394816.34	383007.00	.	.
	FEMALE 50-59	353310.23	365789.00	.	.
	FEMALE 60-69	251893.08	254092.00	.	.
	FEMALE 70-79	207405.65	226153.00	.	.
	FEMALE 80+	241363.68	127809.00	.	.

NOTE: Income quintile is based on register derived incomes, including earnings and social benefits that can be retrieved from register sources. The zero group denotes for those who do not have any earnings based on the registered information.

APPENDIX 6.7 Distribution of the ratio estimator for prevalence of diabetes mellitus[2] by weighting methods

(a) Inverse probability weighting

Prevalence estimates of diabetes mellitus, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**1.06**	**2.93**	**8.28**	**8.77**	**3.95**	**0.77**	**1.68**	**6.78**	**10.04**	**3.55**
Unweighted sample estimates										
Sampled	1.84	4.91	9.72	7.84	5.50	0.71	3.06	7.74	13.01	5.22
Health interviewed	1.88	4.59	9.48	8.65	5.40	0.66	2.99	7.70	12.21	4.97
Symptom interviewed	1.71	4.44	9.83	8.97	5.38	0.70	3.03	7.67	12.98	4.91
Medical measurements	1.76	4.42	9.66	8.06	5.24	0.58	2.91	7.26	11.43	4.28
Full response	1.93	4.56	9.20	8.25	5.21	0.61	2.86	6.89	11.47	4.01
Weighted by inverse of response probabilities from simple logit model										
Sampled	1.92	4.90	9.69	8.93	5.79	0.72	3.04	8.03	15.20	7.46
Health interviewed	2.12	4.56	9.47	9.78	5.86	0.68	3.01	8.19	14.18	7.05
Symptom interviewed	1.97	4.41	10.05	8.54	5.68	0.73	3.05	8.02	15.68	6.87
Medical measurements	2.04	4.29	9.96	7.19	5.40	0.53	2.92	7.58	12.79	5.22
Self-completion questionnaires	2.24	4.53	9.48	8.06	5.44	0.57	2.88	7.30	13.43	4.99
Weighted by inverse of response probabilities from sequential logit model										
Sampled	1.81	4.95	9.27	7.65	5.32	0.67	3.06	7.54	12.87	4.51
Health interviewed	1.78	4.64	9.06	8.94	5.21	0.60	2.96	7.46	11.06	4.19
Symptom interviewed	1.63	4.54	9.29	9.40	5.20	0.63	2.98	7.40	11.70	4.20
Medical measurements	1.67	4.52	9.06	8.96	5.10	0.60	2.88	7.03	10.77	3.89
Self-completion questionnaires	1.83	4.61	8.65	8.81	5.06	0.63	2.84	6.56	10.58	3.64
Weighted by inverse of response probabilities from multinomial logit model										
Sampled	1.84	4.86	9.78	9.19	5.79	0.76	3.05	7.97	14.70	7.39
Health interviewed	2.03	4.60	9.56	10.01	5.90	0.73	3.02	8.12	13.80	7.03
Symptom interviewed	1.86	4.44	10.15	8.19	5.64	0.78	3.07	8.00	14.66	6.71
Medical measurements	1.93	4.35	10.05	7.13	5.41	0.53	2.93	7.57	12.34	5.20
Self-completion questionnaires	2.12	4.59	9.56	8.02	5.44	0.57	2.88	7.27	12.99	4.95
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL)										
Sampled	1.81	4.14	9.80	28.29	5.85	0.62	2.16	7.19	47.12	7.57
Health interviewed	1.97	4.62	9.01	8.61	5.22	0.65	2.97	7.64	11.82	4.41
Symptom interviewed	1.75	4.46	9.25	9.26	5.16	0.68	2.99	7.50	12.97	4.37
Medical measurements	1.79	4.42	9.00	8.52	5.03	0.61	2.88	7.13	11.36	3.94
Self-completion questionnaires	1.97	4.51	8.58	8.35	4.99	0.64	2.85	6.71	11.54	3.72
Weighted by inverse of response probabilities from multilevel logit model (MCMC)										
Sampled	1.92	4.96	9.49	7.27	5.38	0.76	3.04	7.88	13.44	4.84
Health interviewed	1.99	4.62	8.99	8.62	5.21	0.65	2.97	7.63	11.64	4.40
Symptom interviewed	1.77	4.45	9.24	9.25	5.16	0.68	2.99	7.50	12.85	4.36
Medical measurements	1.81	4.42	8.98	8.51	5.03	0.61	2.88	7.13	11.35	3.94
Self-completion questionnaires	1.99	4.50	8.57	8.32	4.99	0.64	2.84	6.72	11.55	3.72

[2] Disease which entitled the patient to receive reimbursement of medicine costs under the Higher Special Refund Category; Disease code 103.

(b) With design weight adjustment

Prevalence estimates of diabetes mellitus, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**1.06**	**2.93**	**8.28**	**8.77**	**3.95**	**0.77**	**1.68**	**6.78**	**10.04**	**3.55**
Weighted by design-weights										
Sampled	1.82	4.96	9.66	7.88	5.49	0.66	3.08	7.59	12.70	5.14
Health interviewed	1.84	4.65	9.44	8.67	5.40	0.62	3.02	7.51	11.89	4.88
Symptom interviewed	1.70	4.49	9.75	9.17	5.38	0.66	3.05	7.48	12.70	4.83
Medical measurements	1.75	4.47	9.61	8.50	5.27	0.55	2.93	7.10	11.25	4.23
Self-completion questionnaires	1.92	4.60	9.15	8.66	5.23	0.58	2.87	6.73	11.30	3.97
Weighted by design-weights with non-response adjustment										
Sampled	1.77	5.01	9.52	7.83	5.59	0.66	3.07	7.60	12.64	6.05
Health interviewed	1.79	4.69	9.29	8.62	5.55	0.63	3.00	7.52	11.82	5.68
Symptom interviewed	1.66	4.53	9.59	9.17	5.54	0.66	3.03	7.49	12.69	5.64
Medical measurements	1.71	4.51	9.44	8.51	5.37	0.56	2.91	7.14	11.30	4.78
Self-completion questionnaires	1.88	4.64	8.97	8.61	5.32	0.59	2.85	6.77	11.33	4.46
Weighted by inverse of response probabilities from simple logit model and design weight										
Sampled	1.94	4.86	9.76	9.32	5.86	0.78	3.02	8.19	15.65	7.62
Health interviewed	2.16	4.50	9.50	10.20	5.91	0.73	2.97	8.39	14.69	7.23
Symptom interviewed	1.97	4.36	10.11	8.51	5.68	0.78	3.02	8.21	16.13	7.03
Medical measurements	2.04	4.26	9.99	6.74	5.37	0.55	2.90	7.73	12.92	5.27
Self-completion questionnaires	2.23	4.50	9.52	7.63	5.40	0.60	2.87	7.44	13.50	5.02
Weighted by inverse of response probabilities from sequential logit model and design weight										
Sampled	1.82	4.90	9.38	7.38	5.32	0.72	3.04	7.68	13.13	4.57
Health interviewed	1.81	4.59	9.13	8.64	5.20	0.64	2.93	7.65	11.32	4.26
Symptom interviewed	1.63	4.50	9.39	9.03	5.20	0.67	2.95	7.60	11.97	4.27
Medical measurements	1.67	4.48	9.13	8.47	5.09	0.64	2.86	7.22	11.02	3.95
Self-completion questionnaires	1.83	4.57	8.71	8.36	5.05	0.67	2.83	6.74	10.79	3.69
Weighted by inverse of response probabilities from multinomial logit model and design weight										
Sampled	1.85	4.85	9.70	9.06	5.77	0.76	3.06	7.96	14.80	7.41
Health interviewed	2.07	4.54	9.58	10.52	5.96	0.79	2.99	8.31	14.31	7.21
Symptom interviewed	1.87	4.39	10.20	8.16	5.64	0.84	3.04	8.19	15.10	6.85
Medical measurements	1.94	4.32	10.08	6.69	5.38	0.56	2.90	7.72	12.51	5.25
Self-completion questionnaires	2.12	4.56	9.60	7.60	5.41	0.60	2.87	7.40	13.10	4.98
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL) and design weight										
Sampled	1.93	4.98	9.30	7.32	5.36	0.71	3.07	7.70	13.37	4.79
Health interviewed	2.01	4.56	9.07	8.48	5.21	0.69	2.94	7.84	12.11	4.49
Symptom interviewed	1.76	4.41	9.35	8.98	5.16	0.73	2.96	7.71	13.26	4.45
Medical measurements	1.80	4.38	9.05	8.06	5.01	0.64	2.86	7.32	11.60	4.00
Self-completion questionnaires	1.97	4.47	8.63	7.94	4.98	0.67	2.83	6.89	11.76	3.77
Weighted by inverse of response probabilities from multilevel logit model (MCMC) and design weight										
Sampled	1.94	4.98	9.28	7.33	5.37	0.71	3.07	7.69	13.24	4.77
Health interviewed	2.03	4.56	9.05	8.50	5.21	0.69	2.93	7.83	11.92	4.47
Symptom interviewed	1.77	4.40	9.33	8.97	5.15	0.73	2.96	7.71	13.14	4.44
Medical measurements	1.82	4.37	9.04	8.05	5.01	0.64	2.85	7.33	11.59	4.00
Self-completion questionnaires	1.99	4.46	8.62	7.91	4.97	0.67	2.83	6.89	11.77	3.77

(c) Calibration with alternative auxiliary information structure for the health interviewed

Prevalence estimates of diabetes mellitus, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**1.07**	**2.94**	**8.30**	**8.80**	**3.97**	**0.78**	**1.69**	**6.79**	**10.05**	**3.56**
Unweighted sample estimates										
Sampled	1.84	4.91	9.72	7.84	5.50	0.71	3.06	7.74	13.01	5.22
Prevalence estimates of diabetes mellitus, % - weighted by calibrated weights based on:										
Design weights										
Health interviewed:										
Demographic variables	2.05	4.51	9.99	9.83	5.41	0.62	3.14	7.96	13.16	4.72
Socio-economic variables	1.95	4.78	9.78	9.19	5.45	0.67	3.35	7.87	12.94	4.78
Income inequality variables	1.84	4.65	9.68	7.72	5.29	0.62	3.18	7.65	12.09	4.58
Design weights and sequential logit response probability										
Health interviewed:										
Demographic variables	1.87	4.57	9.74	9.51	5.45	0.56	3.07	7.54	11.76	4.36
Socio-economic variables	1.93	4.83	9.51	8.84	5.56	0.60	3.16	7.90	11.84	4.70
Income inequality variables	1.87	4.77	9.23	7.44	5.38	0.60	3.25	7.04	9.65	4.39
Design weights and multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	2.17	4.54	9.66	9.07	5.46	0.61	3.08	7.74	12.54	4.60
Socio-economic variables	2.15	4.75	9.33	8.96	5.49	0.68	3.15	8.09	12.53	4.89
Income inequality variables	2.10	4.78	9.14	7.17	5.39	0.66	3.25	7.21	10.08	4.57
Sequential logit response probability										
Health interviewed:										
Demographic variables	1.88	4.55	9.73	9.55	5.44	0.54	3.07	7.53	11.72	4.35
Socio-economic variables	1.90	4.86	9.44	9.11	5.56	0.58	3.20	7.77	11.56	4.65
Income inequality variables	1.84	4.83	9.19	7.65	5.39	0.57	3.29	6.92	9.46	4.34
Multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	2.17	4.52	9.64	9.09	5.45	0.60	3.08	7.74	12.51	4.59
Socio-economic variables	2.10	4.79	9.28	9.05	5.48	0.65	3.19	7.94	12.23	4.84
Income inequality variables	2.06	4.85	9.11	7.29	5.41	0.62	3.30	7.07	9.87	4.52

APPENDIX 6.8 Distribution of the ratio estimator for prevalence of chronic cardiac insufficiency[3] by weighting methods

(a) Inverse probability weighting

Prevalence estimates of chronic cardiac insufficiency, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**0.08**	**0.49**	**4.77**	**14.98**	**1.87**	**0.06**	**0.18**	**4.84**	**20.94**	**3.02**
Unweighted sample estimates										
Sampled	0.00	0.80	4.80	20.10	2.64	0.00	0.11	4.09	18.37	3.70
Health interviewed	0.00	0.73	4.54	18.92	2.56	0.00	0.12	4.15	17.83	3.49
Symptom interviewed	0.00	0.71	4.70	20.51	2.57	0.00	0.13	4.21	18.51	3.27
Medical measurements	0.00	0.72	4.16	16.94	2.10	0.00	0.13	3.97	18.57	2.64
Full response	0.00	0.80	4.18	17.53	2.07	0.00	0.14	3.83	18.35	2.33
Weighted by inverse of response probabilities from simple logit model										
Sampled	0.00	0.84	4.94	24.03	4.29	0.00	0.11	4.41	17.36	6.23
Health interviewed	0.00	0.83	4.68	22.48	4.30	0.00	0.12	4.51	17.25	6.05
Symptom interviewed	0.00	0.81	4.98	25.22	4.24	0.00	0.13	4.65	19.10	5.74
Medical measurements	0.00	0.84	4.38	18.76	2.96	0.00	0.13	4.39	18.33	4.07
Self-completion questionnaires	0.00	0.94	4.41	16.82	2.66	0.00	0.14	4.28	17.99	3.60
Weighted by inverse of response probabilities from sequential logit model										
Sampled	0.00	0.77	4.45	18.40	2.07	0.00	0.11	3.67	18.27	2.46
Health interviewed	0.00	0.67	4.19	16.69	1.94	0.00	0.12	3.69	17.10	2.21
Symptom interviewed	0.00	0.65	4.31	17.61	1.97	0.00	0.13	3.75	17.38	2.18
Medical measurements	0.00	0.67	3.88	15.51	1.74	0.00	0.13	3.61	17.79	1.95
Self-completion questionnaires	0.00	0.73	3.84	16.27	1.72	0.00	0.14	3.48	17.24	1.71
Weighted by inverse of response probabilities from multinomial logit model										
Sampled	0.00	0.84	4.98	25.02	4.38	0.00	0.12	4.45	17.60	6.44
Health interviewed	0.00	0.83	4.69	23.51	4.40	0.00	0.13	4.56	17.36	6.24
Symptom interviewed	0.00	0.80	4.99	26.29	4.28	0.00	0.14	4.69	19.17	5.87
Medical measurements	0.00	0.83	4.38	19.41	2.95	0.00	0.14	4.42	18.61	4.20
Self-completion questionnaires	0.00	0.93	4.41	17.08	2.62	0.00	0.15	4.30	18.46	3.72
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL)										
Sampled	0.00	0.80	4.67	19.14	2.22	0.00	0.11	3.78	18.72	2.86
Health interviewed	0.00	0.71	4.31	17.10	2.05	0.00	0.12	3.82	17.74	2.51
Symptom interviewed	0.00	0.68	4.50	19.15	2.09	0.00	0.12	3.88	18.09	2.34
Medical measurements	0.00	0.70	3.90	15.40	1.73	0.00	0.13	3.61	18.45	1.99
Self-completion questionnaires	0.00	0.77	3.87	16.01	1.72	0.00	0.13	3.49	17.85	1.75
Weighted by inverse of response probabilities from multilevel logit model (MCMC)										
Sampled	0.00	0.81	4.67	19.24	2.23	0.00	0.11	3.77	18.76	2.87
Health interviewed	0.00	0.72	4.31	17.20	2.06	0.00	0.12	3.81	17.80	2.52
Symptom interviewed	0.00	0.69	4.49	19.23	2.09	0.00	0.12	3.87	18.15	2.35
Medical measurements	0.00	0.70	3.89	15.50	1.73	0.00	0.13	3.61	18.50	1.99
Self-completion questionnaires	0.00	0.77	3.87	16.12	1.72	0.00	0.13	3.49	17.91	1.75

[3] Disease which entitled the patient to receive reimbursement of medicine costs under the Higher Special Refund Category; Disease code 201.

(b) With design weight adjustment

Prevalence estimates of chronic cardiac insufficiency, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**0.08**	**0.49**	**4.77**	**14.98**	**1.87**	**0.06**	**0.18**	4.84	**20.94**	**3.02**
Weighted by design-weights										
Sampled	0.00	0.78	4.80	19.37	2.58	0.00	0.12	4.10	18.13	3.68
Health interviewed	0.00	0.71	4.52	18.18	2.49	0.00	0.13	4.16	17.55	3.46
Symptom interviewed	0.00	0.68	4.66	19.75	2.51	0.00	0.13	4.22	18.28	3.25
Medical measurements	0.00	0.70	4.14	16.35	2.06	0.00	0.13	4.00	18.23	2.64
Self-completion questionnaires	0.00	0.77	4.17	16.82	2.03	0.00	0.14	3.85	17.95	2.32
Weighted by design-weights with non-response adjustment										
Sampled	0.00	0.78	4.80	19.22	3.45	0.00	0.11	4.10	18.08	5.45
Health interviewed	0.00	0.71	4.50	18.04	3.34	0.00	0.12	4.15	17.54	5.12
Symptom interviewed	0.00	0.69	4.63	19.60	3.36	0.00	0.13	4.23	18.26	4.80
Medical measurements	0.00	0.70	4.13	16.32	2.66	0.00	0.13	4.02	18.23	3.85
Self-completion questionnaires	0.00	0.78	4.17	16.74	2.57	0.00	0.14	3.89	17.86	3.36
Weighted by inverse of response probabilities from simple logit model and design weight										
Sampled	0.00	0.86	4.95	24.84	4.43	0.00	0.11	4.39	17.50	6.26
Health interviewed	0.00	0.86	4.69	23.33	4.44	0.00	0.12	4.50	17.43	6.09
Symptom interviewed	0.00	0.83	5.01	25.95	4.35	0.00	0.13	4.64	19.18	5.75
Medical measurements	0.00	0.86	4.41	19.49	3.05	0.00	0.13	4.35	18.63	4.08
Self-completion questionnaires	0.00	0.97	4.41	17.50	2.72	0.00	0.14	4.25	18.30	3.60
Weighted by inverse of response probabilities from sequential logit model and design weight										
Sampled	0.00	0.79	4.47	19.00	2.12	0.00	0.11	3.65	18.44	2.46
Health interviewed	0.00	0.70	4.22	17.39	1.98	0.00	0.12	3.67	17.34	2.22
Symptom interviewed	0.00	0.68	4.35	18.33	2.02	0.00	0.13	3.74	17.66	2.19
Medical measurements	0.00	0.69	3.91	16.17	1.78	0.00	0.13	3.59	18.15	1.95
Self-completion questionnaires	0.00	0.76	3.86	17.07	1.76	0.00	0.13	3.45	17.63	1.71
Weighted by inverse of response probabilities from multinomial logit model and design weight										
Sampled	0.00	0.86	4.99	25.81	4.52	0.00	0.12	4.44	17.73	6.46
Health interviewed	0.00	0.85	4.70	24.35	4.55	0.00	0.13	4.55	17.54	6.28
Symptom interviewed	0.00	0.82	5.03	26.95	4.38	0.00	0.13	4.68	19.28	5.88
Medical measurements	0.00	0.85	4.40	20.18	3.04	0.00	0.14	4.38	18.93	4.21
Self-completion questionnaires	0.00	0.96	4.41	17.76	2.68	0.00	0.15	4.27	18.79	3.72
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL) and design weight										
Sampled	0.00	0.82	4.69	19.84	2.27	0.00	0.11	3.77	18.93	2.87
Health interviewed	0.00	0.74	4.35	17.81	2.11	0.00	0.12	3.81	18.00	2.52
Symptom interviewed	0.00	0.71	4.55	19.93	2.14	0.00	0.12	3.86	18.36	2.36
Medical measurements	0.00	0.73	3.92	16.07	1.77	0.00	0.12	3.59	18.84	1.99
Self-completion questionnaires	0.00	0.79	3.88	16.79	1.76	0.00	0.13	3.47	18.25	1.75
Weighted by inverse of response probabilities from multilevel logit model (MCMC) and design weight										
Sampled	0.00	0.83	4.68	19.95	2.28	0.00	0.11	3.76	18.98	2.88
Health interviewed	0.00	0.75	4.34	17.92	2.11	0.00	0.12	3.80	18.07	2.53
Symptom interviewed	0.00	0.71	4.54	20.02	2.14	0.00	0.12	3.86	18.42	2.36
Medical measurements	0.00	0.73	3.91	16.19	1.77	0.00	0.12	3.59	18.89	2.00
Self-completion questionnaires	0.00	0.80	3.88	16.92	1.76	0.00	0.13	3.47	18.31	1.75

(c) Calibration with alternative auxiliary information structure for the health interviewed

Prevalence estimates of chronic cardiac insufficiency, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**0.08**	**0.49**	**4.77**	**15.02**	**1.87**	**0.06**	**0.18**	**4.84**	**20.97**	**3.02**
Unweighted sample estimates										
Sampled	0.00	0.80	4.80	20.10	2.64	0.00	0.11	4.09	18.37	3.70
Prevalence estimates of chronic cardiac insufficiency, % - weighted by calibrated weights based on:										
Design weights										
Health interviewed:										
Demographic variables	0.00	0.74	4.99	18.62	2.14	0.00	0.17	4.31	18.28	2.65
Socio-economic variables	0.00	0.86	4.64	19.55	2.14	0.00	0.14	4.39	17.75	2.62
Income inequality variables	0.00	0.73	4.71	17.61	2.03	0.00	0.13	4.48	18.19	2.68
Design weights and sequential logit response probability										
Health interviewed:										
Demographic variables	0.00	0.70	4.54	16.44	2.09	0.00	0.16	3.68	18.18	2.32
Socio-economic variables	0.00	0.87	4.27	17.38	2.26	0.00	0.12	3.91	17.53	2.63
Income inequality variables	0.00	0.75	4.32	15.51	2.17	0.00	0.13	3.64	15.20	2.50
Design weights and multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	0.00	0.74	4.77	17.15	2.24	0.00	0.15	3.88	18.86	2.66
Socio-economic variables	0.00	0.89	4.38	18.03	2.33	0.00	0.12	3.96	18.16	2.89
Income inequality variables	0.00	0.79	4.46	15.61	2.26	0.00	0.13	3.75	15.56	2.76
Sequential logit response probability										
Health interviewed:										
Demographic variables	0.00	0.71	4.54	16.48	2.09	0.00	0.16	3.68	18.16	2.32
Socio-economic variables	0.00	0.85	4.26	16.78	2.22	0.00	0.13	3.92	17.28	2.62
Income inequality variables	0.00	0.73	4.32	15.02	2.13	0.00	0.13	3.69	15.08	2.51
Multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	0.00	0.74	4.77	17.17	2.24	0.00	0.16	3.88	18.83	2.65
Socio-economic variables	0.00	0.87	4.37	17.42	2.29	0.00	0.12	3.97	17.87	2.88
Income inequality variables	0.00	0.77	4.46	15.04	2.22	0.00	0.13	3.79	15.43	2.76

APPENDIX 6.9 Distribution of the ratio estimator for prevalence of connective tissue diseases, rheumatoid arthritis and comparable diseases[4] by weighting methods

(a) Inverse probability weighting

Prevalence estimates of connective tissue diseases, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**0.63**	**1.30**	**2.44**	**2.34**	**1.45**	**0.99**	**2.14**	**4.56**	**4.81**	**2.78**
Unweighted sample estimates										
Sampled	0.86	1.31	3.12	4.41	1.80	1.42	3.23	5.87	5.53	3.89
Health interviewed	0.87	1.46	3.07	4.86	1.92	1.59	3.18	6.08	5.43	3.91
Symptom interviewed	0.93	1.41	3.13	5.13	1.92	1.69	3.29	5.72	5.53	3.86
Medical measurements	0.96	1.45	2.97	6.45	1.93	1.74	3.37	5.11	5.71	3.68
Full response	1.05	1.52	2.84	8.25	1.99	1.83	3.28	4.97	5.96	3.59
Weighted by inverse of response probabilities from simple logit model										
Sampled	0.76	1.34	3.14	3.67	1.89	1.47	3.28	6.25	4.55	4.06
Health interviewed	0.83	1.54	3.14	4.01	2.08	1.67	3.17	6.54	4.26	4.01
Symptom interviewed	0.92	1.53	3.26	4.59	2.14	1.78	3.29	6.13	4.70	4.05
Medical measurements	0.96	1.59	3.04	6.54	2.20	1.84	3.38	5.13	5.01	3.80
Self-completion questionnaires	1.05	1.56	2.91	8.69	2.26	1.98	3.25	4.98	5.56	3.76
Weighted by inverse of response probabilities from sequential logit model										
Sampled	0.87	1.27	2.80	4.04	1.63	1.41	3.20	5.72	5.39	3.65
Health interviewed	0.85	1.41	2.81	4.72	1.74	1.59	3.18	5.80	5.94	3.71
Symptom interviewed	0.89	1.35	2.91	5.11	1.75	1.67	3.27	5.50	6.15	3.70
Medical measurements	0.91	1.39	2.78	5.97	1.75	1.71	3.34	5.11	6.13	3.60
Self-completion questionnaires	1.00	1.49	2.72	7.43	1.83	1.80	3.27	4.99	6.01	3.50
Weighted by inverse of response probabilities from multinomial logit model										
Sampled	0.80	1.34	3.09	3.71	1.88	1.50	3.33	6.26	4.62	4.11
Health interviewed	0.88	1.55	3.08	4.05	2.08	1.71	3.21	6.58	4.31	4.06
Symptom interviewed	0.97	1.53	3.19	4.60	2.12	1.82	3.33	6.15	4.83	4.12
Medical measurements	1.00	1.59	2.99	6.63	2.19	1.88	3.43	5.15	4.95	3.83
Self-completion questionnaires	1.10	1.55	2.85	8.89	2.24	2.02	3.30	5.03	5.56	3.80
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL)										
Sampled	0.82	1.28	2.97	3.87	1.66	1.42	3.24	5.85	5.95	3.81
Health interviewed	0.83	1.44	2.90	4.53	1.77	1.61	3.18	6.09	5.88	3.82
Symptom interviewed	0.88	1.37	3.02	4.64	1.77	1.70	3.29	5.68	5.89	3.75
Medical measurements	0.90	1.41	2.84	5.76	1.76	1.75	3.37	5.15	6.17	3.63
Self-completion questionnaires	0.99	1.49	2.76	7.17	1.83	1.84	3.31	5.01	6.10	3.53
Weighted by inverse of response probabilities from multilevel logit model (MCMC)										
Sampled	0.83	1.29	2.98	3.85	1.67	1.41	3.24	5.84	5.94	3.80
Health interviewed	0.83	1.44	2.91	4.51	1.77	1.60	3.18	6.07	5.86	3.81
Symptom interviewed	0.88	1.37	3.03	4.63	1.77	1.70	3.29	5.66	5.89	3.74
Medical measurements	0.91	1.41	2.84	5.74	1.76	1.75	3.37	5.14	6.17	3.63
Self-completion questionnaires	0.99	1.50	2.76	7.15	1.83	1.84	3.30	4.99	6.10	3.53

[4] Disease which entitled the patient to receive reimbursement of medicine costs under the Higher Special Refund Category; Disease code 202.

(b) With design weight adjustment

Prevalence estimates of connective tissue diseases, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**0.63**	**1.30**	**2.44**	**2.34**	**1.45**	**0.99**	**2.14**	**4.56**	**4.81**	**2.78**
Weighted by design-weights										
Sampled	0.88	1.30	3.25	4.43	1.83	1.42	3.20	5.71	5.54	3.83
Health interviewed	0.91	1.45	3.19	4.88	1.95	1.59	3.14	5.91	5.47	3.85
Symptom interviewed	0.97	1.39	3.26	5.14	1.95	1.69	3.24	5.57	5.61	3.81
Medical measurements	1.00	1.43	3.11	6.46	1.96	1.73	3.33	4.98	5.72	3.63
Self-completion questionnaires	1.10	1.50	2.99	8.24	2.02	1.83	3.23	4.82	6.01	3.54
Weighted by design-weights with non-response adjustment										
Sampled	0.87	1.33	3.22	4.58	1.99	1.44	3.18	5.67	5.57	4.04
Health interviewed	0.90	1.49	3.15	5.04	2.14	1.62	3.12	5.87	5.50	4.04
Symptom interviewed	0.97	1.43	3.23	5.23	2.13	1.72	3.22	5.52	5.69	3.99
Medical measurements	1.00	1.47	3.08	6.56	2.17	1.77	3.31	4.94	5.76	3.79
Self-completion questionnaires	1.10	1.54	2.96	8.36	2.27	1.87	3.22	4.78	5.98	3.69
Weighted by inverse of response probabilities from simple logit model and design weight										
Sampled	0.75	1.35	3.02	3.62	1.86	1.46	3.34	6.40	4.54	4.10
Health interviewed	0.80	1.54	3.03	3.96	2.04	1.65	3.23	6.71	4.20	4.05
Symptom interviewed	0.88	1.54	3.14	4.54	2.11	1.77	3.35	6.28	4.60	4.08
Medical measurements	0.92	1.60	2.92	6.43	2.16	1.82	3.44	5.28	5.02	3.86
Self-completion questionnaires	1.00	1.58	2.79	8.59	2.23	1.96	3.31	5.17	5.55	3.82
Weighted by inverse of response probabilities from sequential logit model and design weight										
Sampled	0.85	1.29	2.69	4.00	1.61	1.41	3.24	5.90	5.35	3.72
Health interviewed	0.80	1.43	2.70	4.69	1.71	1.59	3.22	5.99	5.86	3.78
Symptom interviewed	0.85	1.38	2.80	5.08	1.73	1.68	3.31	5.68	6.06	3.76
Medical measurements	0.87	1.41	2.65	5.95	1.73	1.72	3.39	5.26	6.09	3.66
Self-completion questionnaires	0.95	1.51	2.58	7.42	1.80	1.81	3.32	5.16	5.91	3.56
Weighted by inverse of response probabilities from multinomial logit model and design weight										
Sampled	0.78	1.35	2.98	3.67	1.86	1.49	3.38	6.42	4.60	4.15
Health interviewed	0.84	1.55	2.98	4.00	2.05	1.69	3.27	6.74	4.24	4.10
Symptom interviewed	0.93	1.54	3.07	4.57	2.09	1.81	3.39	6.31	4.72	4.15
Medical measurements	0.96	1.60	2.87	6.53	2.15	1.87	3.49	5.31	4.97	3.90
Self-completion questionnaires	1.05	1.57	2.73	8.81	2.21	2.00	3.36	5.21	5.55	3.87
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL) and design weight										
Sampled	0.80	1.30	2.85	3.84	1.64	1.42	3.28	6.03	5.94	3.87
Health interviewed	0.78	1.46	2.79	4.50	1.74	1.61	3.23	6.27	5.82	3.88
Symptom interviewed	0.84	1.39	2.92	4.62	1.74	1.70	3.33	5.86	5.80	3.81
Medical measurements	0.86	1.43	2.71	5.75	1.73	1.75	3.42	5.29	6.14	3.69
Self-completion questionnaires	0.94	1.52	2.63	7.18	1.80	1.84	3.36	5.17	6.01	3.59
Weighted by inverse of response probabilities from multilevel logit model (MCMC) and design weight										
Sampled	0.81	1.30	2.86	3.82	1.65	1.41	3.28	6.02	5.93	3.86
Health interviewed	0.79	1.46	2.81	4.49	1.75	1.60	3.23	6.26	5.80	3.88
Symptom interviewed	0.84	1.39	2.92	4.61	1.75	1.70	3.33	5.84	5.80	3.81
Medical measurements	0.86	1.43	2.71	5.73	1.73	1.75	3.42	5.28	6.15	3.69
Self-completion questionnaires	0.94	1.52	2.63	7.17	1.80	1.84	3.35	5.16	6.01	3.59

(c) Calibration with alternative auxiliary information structure for the health interviewed

Prevalence estimates of connective tissue diseases, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**0.63**	**1.30**	**2.45**	**2.34**	**1.45**	**0.99**	**2.15**	**4.57**	**4.82**	**2.79**
Unweighted sample estimates										
Sampled	0.86	1.31	3.12	4.41	1.80	1.42	3.23	5.87	5.53	3.89
Prevalence estimates of connective tissue diseases, % - weighted by calibrated weights based on:										
Design weights										
Health interviewed:										
Demographic variables	0.62	1.53	3.56	3.65	1.87	1.52	3.35	6.26	5.11	3.92
Socio-economic variables	0.86	1.67	3.29	3.80	1.93	1.78	3.28	6.03	5.19	3.89
Income inequality variables	0.99	1.43	3.14	5.49	1.86	1.53	3.12	5.91	5.57	3.76
Design weights and sequential logit response probability										
Health interviewed:										
Demographic variables	0.56	1.50	3.24	4.10	1.83	1.54	3.35	6.07	5.70	3.87
Socio-economic variables	0.76	1.64	2.75	4.66	1.86	1.82	3.21	6.01	5.64	3.92
Income inequality variables	0.93	1.45	2.56	4.59	1.76	1.52	3.28	5.52	4.68	3.69
Design weights and multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	0.53	1.53	3.37	3.75	1.85	1.54	3.39	6.34	5.62	3.98
Socio-economic variables	0.74	1.65	2.84	3.93	1.85	1.81	3.20	6.34	5.64	4.00
Income inequality variables	0.91	1.49	2.68	4.23	1.79	1.53	3.31	5.69	4.81	3.79
Sequential logit response probability										
Health interviewed:										
Demographic variables	0.56	1.49	3.23	4.11	1.82	1.54	3.33	6.03	5.71	3.85
Socio-economic variables	0.81	1.62	2.88	4.75	1.90	1.82	3.18	5.81	5.70	3.85
Income inequality variables	0.99	1.43	2.67	4.68	1.80	1.53	3.23	5.37	4.75	3.64
Multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	0.52	1.52	3.36	3.74	1.84	1.54	3.37	6.31	5.63	3.96
Socio-economic variables	0.78	1.64	2.96	4.01	1.88	1.82	3.16	6.14	5.69	3.94
Income inequality variables	0.96	1.48	2.77	4.28	1.82	1.55	3.26	5.54	4.87	3.73

APPENDIX 6.10 Distribution of the ratio estimator for prevalence of chronic asthma and similar chronic obstructive pulmonary diseases[5] by weighting methods

(a) Inverse probability weighting

Prevalence estimates of chronic asthma, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**2.07**	**2.66**	**6.82**	**7.18**	**3.66**	**2.97**	**4.28**	**7.41**	**5.69**	**5.00**
Unweighted sample estimates										
Sampled	2.45	3.60	8.64	9.80	4.86	4.60	6.12	9.25	6.34	6.67
Health interviewed	2.60	3.59	8.54	10.81	4.98	4.50	5.92	9.93	6.59	6.76
Symptom interviewed	2.64	3.60	8.55	11.54	5.00	4.78	5.93	9.72	6.49	6.74
Medical measurements	2.72	3.70	8.62	12.90	5.07	4.78	5.88	9.99	6.79	6.81
Full response	2.11	3.68	8.70	13.40	4.89	4.74	5.86	9.82	5.50	6.60
Weighted by inverse of response probabilities from simple logit model										
Sampled	2.18	3.59	8.31	8.73	4.94	4.50	6.21	9.36	6.14	6.64
Health interviewed	2.36	3.66	8.17	9.55	5.19	4.47	5.93	9.99	6.61	6.82
Symptom interviewed	2.48	3.71	8.21	10.88	5.30	4.76	5.95	9.82	6.34	6.76
Medical measurements	2.57	3.86	8.41	13.81	5.48	4.67	5.93	10.22	6.93	6.95
Self-completion questionnaires	1.99	3.71	8.50	11.33	4.99	4.68	5.89	10.05	4.97	6.58
Weighted by inverse of response probabilities from sequential logit model										
Sampled	2.54	3.61	8.70	10.50	4.78	4.72	6.05	9.30	6.64	6.66
Health interviewed	2.77	3.60	8.62	12.26	4.92	4.59	5.89	9.92	7.25	6.76
Symptom interviewed	2.76	3.56	8.65	12.54	4.90	4.83	5.89	9.75	7.03	6.74
Medical measurements	2.82	3.64	8.68	13.83	4.97	4.86	5.84	9.95	7.25	6.76
Self-completion questionnaires	2.20	3.63	8.76	14.68	4.83	4.82	5.82	9.84	6.18	6.62
Weighted by inverse of response probabilities from multinomial logit model										
Sampled	2.17	3.53	8.40	9.00	4.95	4.44	6.19	9.36	6.22	6.65
Health interviewed	2.34	3.61	8.17	9.81	5.19	4.40	5.92	10.01	6.64	6.81
Symptom interviewed	2.43	3.65	8.21	11.65	5.31	4.71	5.92	9.85	6.23	6.73
Medical measurements	2.52	3.80	8.37	14.87	5.48	4.63	5.90	10.24	7.08	6.97
Self-completion questionnaires	1.93	3.70	8.46	11.85	4.97	4.62	5.87	10.02	5.11	6.57
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL)										
Sampled	2.53	3.60	8.51	10.04	4.72	4.70	6.07	9.10	6.81	6.64
Health interviewed	2.71	3.59	8.41	11.75	4.85	4.58	5.83	9.87	7.36	6.72
Symptom interviewed	2.72	3.55	8.46	11.98	4.82	4.83	5.83	9.70	7.29	6.71
Medical measurements	2.79	3.65	8.62	13.60	4.92	4.84	5.76	9.90	7.54	6.72
Self-completion questionnaires	2.16	3.64	8.71	14.09	4.77	4.79	5.74	9.79	6.25	6.56
Weighted by inverse of response probabilities from multilevel logit model (MCMC)										
Sampled	2.54	3.61	8.52	10.06	4.73	4.71	6.07	9.10	6.80	6.64
Health interviewed	2.72	3.59	8.41	11.78	4.85	4.58	5.83	9.88	7.34	6.73
Symptom interviewed	2.72	3.55	8.46	12.02	4.82	4.84	5.83	9.71	7.29	6.71
Medical measurements	2.79	3.65	8.62	13.69	4.92	4.84	5.76	9.91	7.53	6.72
Self-completion questionnaires	2.16	3.65	8.71	14.21	4.77	4.79	5.74	9.80	6.25	6.56

[5] Disease which entitled the patient to receive reimbursement of medicine costs under the Higher Special Refund Category; Disease code 203.

(b) With design weight adjustment

Prevalence estimates of chronic asthma, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**2.07**	**2.66**	**6.82**	**7.18**	**3.66**	**2.97**	**4.28**	**7.41**	**5.69**	**5.00**
Weighted by design-weights										
Sampled	2.36	3.65	8.57	9.72	4.83	4.66	6.00	9.25	6.25	6.61
Health interviewed	2.49	3.63	8.47	10.70	4.95	4.55	5.79	9.91	6.46	6.69
Symptom interviewed	2.53	3.64	8.49	11.53	4.98	4.82	5.81	9.69	6.41	6.68
Medical measurements	2.60	3.74	8.52	13.06	5.04	4.83	5.76	9.98	6.69	6.75
Self-completion questionnaires	2.02	3.74	8.65	13.55	4.90	4.84	5.73	9.78	5.40	6.54
Weighted by design-weights with non-response adjustment										
Sampled	2.39	3.67	8.59	9.70	5.11	4.73	5.98	9.23	6.31	6.58
Health interviewed	2.52	3.65	8.50	10.68	5.28	4.61	5.77	9.91	6.53	6.68
Symptom interviewed	2.56	3.66	8.53	11.49	5.33	4.89	5.78	9.68	6.48	6.66
Medical measurements	2.63	3.76	8.57	13.10	5.41	4.89	5.73	9.96	6.77	6.75
Self-completion questionnaires	2.01	3.77	8.71	13.54	5.24	4.89	5.71	9.72	5.46	6.46
Weighted by inverse of response probabilities from simple logit model and design weight										
Sampled	2.27	3.56	8.39	8.94	5.01	4.46	6.34	9.36	6.30	6.73
Health interviewed	2.46	3.63	8.25	9.78	5.26	4.44	6.07	9.99	6.81	6.92
Symptom interviewed	2.59	3.67	8.26	11.11	5.34	4.74	6.07	9.83	6.43	6.83
Medical measurements	2.68	3.82	8.51	13.81	5.52	4.63	6.06	10.21	7.06	7.02
Self-completion questionnaires	2.07	3.66	8.54	11.07	4.98	4.57	6.02	10.08	5.06	6.64
Weighted by inverse of response probabilities from sequential logit model and design weight										
Sampled	2.66	3.55	8.77	10.50	4.79	4.65	6.17	9.29	6.68	6.71
Health interviewed	2.91	3.55	8.69	12.31	4.95	4.54	6.01	9.93	7.34	6.81
Symptom interviewed	2.90	3.50	8.72	12.47	4.93	4.78	6.01	9.78	7.12	6.80
Medical measurements	2.96	3.59	8.79	13.66	4.99	4.80	5.95	9.96	7.32	6.81
Self-completion questionnaires	2.31	3.56	8.82	14.58	4.82	4.72	5.94	9.88	6.27	6.67
Weighted by inverse of response probabilities from multinomial logit model and design weight										
Sampled	2.25	3.51	8.48	9.22	5.02	4.40	6.32	9.36	6.41	6.75
Health interviewed	2.44	3.59	8.24	10.05	5.26	4.37	6.05	9.99	6.87	6.92
Symptom interviewed	2.54	3.62	8.25	11.94	5.37	4.68	6.04	9.84	6.32	6.79
Medical measurements	2.63	3.76	8.46	14.91	5.53	4.59	6.02	10.22	7.23	7.03
Self-completion questionnaires	2.01	3.64	8.50	11.58	4.95	4.52	6.00	10.03	5.21	6.63
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL) and design weight										
Sampled	2.64	3.55	8.59	10.17	4.75	4.64	6.17	9.09	6.87	6.68
Health interviewed	2.85	3.54	8.49	11.94	4.88	4.52	5.94	9.87	7.47	6.78
Symptom interviewed	2.86	3.49	8.55	12.01	4.85	4.79	5.94	9.72	7.39	6.77
Medical measurements	2.93	3.59	8.74	13.49	4.95	4.78	5.87	9.90	7.64	6.76
Self-completion questionnaires	2.26	3.57	8.78	14.04	4.77	4.68	5.85	9.82	6.33	6.60
Weighted by inverse of response probabilities from multilevel logit model (MCMC) and design weight										
Sampled	2.65	3.55	8.59	10.19	4.76	4.64	6.17	9.09	6.86	6.68
Health interviewed	2.85	3.55	8.49	11.97	4.89	4.53	5.94	9.88	7.46	6.78
Symptom interviewed	2.86	3.50	8.54	12.05	4.85	4.80	5.94	9.73	7.39	6.77
Medical measurements	2.93	3.60	8.73	13.59	4.95	4.79	5.87	9.91	7.63	6.77
Self-completion questionnaires	2.26	3.58	8.79	14.16	4.77	4.68	5.85	9.83	6.33	6.60

(c) Calibration with alternative auxiliary information structure for the health interviewed

Prevalence estimates of chronic asthma, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**2.10**	**2.67**	**6.83**	**7.19**	**3.67**	**3.00**	**4.30**	**7.43**	**5.70**	**5.02**
Unweighted sample estimates										
Sampled	2.45	3.60	8.64	9.80	4.86	4.60	6.12	9.25	6.34	6.67
Prevalence estimates of chronic asthma, % - weighted by calibrated weights based on:										
Design weights										
Health interviewed:										
Demographic variables	2.73	3.39	8.99	14.20	4.91	4.26	6.11	10.05	4.86	6.74
Socio-economic variables	2.53	4.07	8.27	9.84	4.89	4.55	6.18	9.76	7.52	6.95
Income inequality variables	2.57	3.67	8.52	11.28	4.81	4.51	5.80	10.06	6.64	6.79
Design weights and sequential logit response probability										
Health interviewed:										
Demographic variables	3.02	3.46	8.97	14.41	5.13	4.24	6.00	9.97	6.17	6.72
Socio-economic variables	2.97	3.99	8.67	11.55	5.28	4.46	6.47	9.82	7.92	7.12
Income inequality variables	3.07	3.79	8.61	10.61	5.16	4.48	6.39	9.04	6.29	6.76
Design weights and multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	2.98	3.43	8.72	13.81	5.05	4.22	5.94	9.91	6.26	6.66
Socio-economic variables	2.93	3.91	8.35	11.47	5.12	4.48	6.35	9.74	8.06	7.05
Income inequality variables	3.04	3.77	8.39	10.41	5.09	4.46	6.29	9.02	6.51	6.72
Sequential logit response probability										
Health interviewed:										
Demographic variables	3.02	3.46	8.94	14.36	5.12	4.23	6.03	10.00	6.16	6.73
Socio-economic variables	2.84	4.06	8.59	11.57	5.27	4.51	6.36	9.80	7.81	7.07
Income inequality variables	2.92	3.83	8.58	10.67	5.14	4.55	6.27	9.08	6.24	6.73
Multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	2.98	3.42	8.68	13.76	5.03	4.21	5.96	9.94	6.25	6.68
Socio-economic variables	2.80	3.98	8.25	11.32	5.10	4.53	6.24	9.73	7.93	7.00
Income inequality variables	2.89	3.81	8.33	10.35	5.05	4.52	6.18	9.06	6.44	6.69

APPENDIX 6.11 Distribution of the ratio estimator for prevalence of chronic coronary heart disease[6] by weighting methods

(a) Inverse probability weighting

Prevalence estimates of chronic coronnary heart disease, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**0.08**	**2.62**	**17.35**	**23.78**	**6.21**	**0.02**	**0.81**	**11.15**	**18.91**	**4.91**
Unweighted sample estimates										
Sampled	0.00	4.29	22.57	27.45	8.86	0.12	0.74	13.26	21.14	6.74
Health interviewed	0.00	4.39	22.70	26.49	9.11	0.13	0.62	13.78	22.09	6.76
Symptom interviewed	0.00	4.30	23.22	28.85	9.22	0.14	0.64	14.15	22.60	6.55
Medical measurements	0.00	4.42	22.88	33.06	9.13	0.00	0.66	14.30	23.93	6.03
Full response	0.00	4.56	21.74	30.93	8.62	0.00	0.70	14.16	24.77	5.66
Weighted by inverse of response probabilities from simple logit model										
Sampled	0.00	4.19	22.46	26.86	10.03	0.18	0.70	13.54	20.87	9.59
Health interviewed	0.00	4.35	22.72	24.93	10.34	0.20	0.61	14.22	21.63	9.71
Symptom interviewed	0.00	4.25	23.07	28.04	10.40	0.22	0.63	14.78	21.86	9.02
Medical measurements	0.00	4.42	22.54	32.85	10.12	0.00	0.65	15.13	25.73	8.20
Self-completion questionnaires	0.00	4.63	21.45	33.27	9.58	0.00	0.69	14.99	25.54	7.56
Weighted by inverse of response probabilities from sequential logit model										
Sampled	0.00	4.24	22.23	30.14	8.31	0.04	0.78	12.87	22.62	5.51
Health interviewed	0.00	4.41	22.47	28.74	8.59	0.04	0.65	13.32	24.35	5.46
Symptom interviewed	0.00	4.28	22.82	30.27	8.66	0.05	0.66	13.48	24.71	5.42
Medical measurements	0.00	4.38	22.54	32.50	8.59	0.00	0.68	13.56	24.40	5.12
Self-completion questionnaires	0.00	4.51	21.46	30.06	8.21	0.00	0.71	13.49	25.70	4.87
Weighted by inverse of response probabilities from multinomial logit model										
Sampled	0.00	4.26	22.44	26.85	9.99	0.22	0.70	13.55	21.50	9.92
Health interviewed	0.00	4.42	22.75	25.01	10.36	0.25	0.61	14.25	22.31	10.09
Symptom interviewed	0.00	4.30	23.03	27.63	10.28	0.26	0.63	14.78	22.78	9.36
Medical measurements	0.00	4.46	22.44	31.41	9.92	0.00	0.65	15.11	27.07	8.52
Self-completion questionnaires	0.00	4.63	21.39	31.68	9.36	0.00	0.69	14.96	26.64	7.77
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL)										
Sampled	0.00	4.28	22.46	28.32	8.37	0.08	0.76	12.72	21.74	5.77
Health interviewed	0.00	4.38	22.59	26.97	8.57	0.09	0.63	13.26	23.63	5.66
Symptom interviewed	0.00	4.22	22.87	29.92	8.60	0.09	0.65	13.49	24.20	5.49
Medical measurements	0.00	4.34	22.51	32.83	8.48	0.00	0.66	13.55	25.26	5.15
Self-completion questionnaires	0.00	4.48	21.46	30.55	8.13	0.00	0.70	13.48	26.54	4.88
Weighted by inverse of response probabilities from multilevel logit model (MCMC)										
Sampled	0.00	4.28	22.41	28.17	8.35	0.08	0.76	12.71	21.72	5.77
Health interviewed	0.00	4.39	22.54	26.77	8.55	0.09	0.63	13.25	23.61	5.65
Symptom interviewed	0.00	4.24	22.84	29.88	8.59	0.09	0.65	13.48	24.22	5.48
Medical measurements	0.00	4.36	22.48	32.79	8.47	0.00	0.66	13.54	25.23	5.14
Self-completion questionnaires	0.00	4.49	21.42	30.56	8.12	0.00	0.70	13.46	26.52	4.88

[6] Disease which entitled the patient to receive reimbursement of medicine costs under the Higher Special Refund Category; Disease code 206.

(b) With design weight adjustment

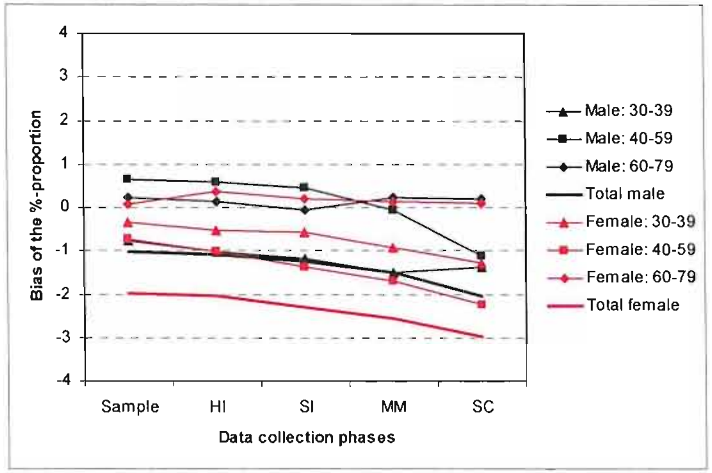
Prevalence estimates of chronic coronnary heart disease, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**0.08**	**2.62**	**17.35**	**23.78**	**6.21**	**0.02**	**0.81**	**11.15**	**18.91**	**4.91**
Weighted by design-weights										
Sampled	0.00	4.26	22.46	27.29	8.78	0.10	0.72	13.11	20.97	6.67
Health interviewed	0.00	4.35	22.59	26.40	9.04	0.12	0.61	13.58	21.90	6.67
Symptom interviewed	0.00	4.24	23.04	28.82	9.12	0.12	0.63	13.93	22.38	6.46
Medical measurements	0.00	4.35	22.68	32.76	9.03	0.00	0.65	14.10	23.66	5.96
Self-completion questionnaires	0.00	4.51	21.46	30.67	8.53	0.00	0.68	13.98	24.44	5.60
Weighted by design-weights with non-response adjustment										
Sampled	0.00	4.27	22.46	27.48	9.77	0.10	0.70	13.11	21.07	8.45
Health interviewed	0.00	4.37	22.55	26.61	10.01	0.11	0.59	13.59	22.03	8.48
Symptom interviewed	0.00	4.25	22.98	29.03	10.12	0.12	0.61	13.91	22.43	8.09
Medical measurements	0.00	4.36	22.61	33.01	10.03	0.00	0.62	14.09	23.67	7.31
Self-completion questionnaires	0.00	4.51	21.36	30.90	9.33	0.00	0.66	13.97	24.31	6.82
Weighted by inverse of response probabilities from simple logit model and design weight										
Sampled	0.00	4.22	22.45	26.76	10.08	0.20	0.72	13.68	20.82	9.60
Health interviewed	0.00	4.39	22.69	24.81	10.38	0.23	0.62	14.40	21.56	9.73
Symptom interviewed	0.00	4.31	23.17	27.77	10.45	0.25	0.64	14.99	21.79	9.06
Medical measurements	0.00	4.49	22.68	32.85	10.23	0.00	0.66	15.32	25.69	8.22
Self-completion questionnaires	0.00	4.68	21.74	33.45	9.69	0.00	0.70	15.16	25.50	7.57
Weighted by inverse of response probabilities from sequential logit model and design weight										
Sampled	0.00	4.26	22.42	30.40	8.41	0.04	0.79	13.00	22.86	5.57
Health interviewed	0.00	4.45	22.66	28.88	8.67	0.05	0.66	13.50	24.60	5.53
Symptom interviewed	0.00	4.35	23.08	30.46	8.77	0.05	0.68	13.68	25.02	5.50
Medical measurements	0.00	4.45	22.80	32.99	8.71	0.00	0.70	13.75	24.75	5.20
Self-completion questionnaires	0.00	4.56	21.78	30.43	8.32	0.00	0.73	13.66	26.17	4.93
Weighted by inverse of response probabilities from multinomial logit model and design weight										
Sampled	0.00	4.29	22.43	26.69	10.05	0.25	0.72	13.69	21.44	9.93
Health interviewed	0.00	4.46	22.73	24.84	10.39	0.28	0.62	14.44	22.23	10.11
Symptom interviewed	0.00	4.37	23.13	27.24	10.32	0.30	0.65	14.99	22.73	9.40
Medical measurements	0.00	4.53	22.59	31.30	10.02	0.00	0.67	15.30	27.00	8.53
Self-completion questionnaires	0.00	4.68	21.69	31.74	9.47	0.00	0.71	15.12	26.60	7.77
Weighted by inverse of response probabilities from multilevel logit model (2nd order PQL) and design weight										
Sampled	0.00	4.30	22.62	28.50	8.45	0.09	0.78	12.85	21.82	5.82
Health interviewed	0.00	4.41	22.75	27.02	8.64	0.10	0.64	13.45	23.70	5.72
Symptom interviewed	0.00	4.28	23.12	30.02	8.71	0.11	0.66	13.69	24.36	5.57
Medical measurements	0.00	4.41	22.76	33.35	8.60	0.00	0.68	13.74	25.58	5.21
Self-completion questionnaires	0.00	4.53	21.76	30.99	8.24	0.00	0.72	13.65	26.93	4.94
Weighted by inverse of response probabilities from multilevel logit model (MCMC) and design weight										
Sampled	0.00	4.31	22.57	28.35	8.43	0.09	0.78	12.84	21.81	5.82
Health interviewed	0.00	4.43	22.69	26.82	8.62	0.10	0.64	13.43	23.68	5.72
Symptom interviewed	0.00	4.30	23.08	29.99	8.70	0.11	0.66	13.68	24.39	5.56
Medical measurements	0.00	4.43	22.72	33.31	8.59	0.00	0.68	13.73	25.55	5.21
Self-completion questionnaires	0.00	4.55	21.72	31.02	8.23	0.00	0.72	13.63	26.92	4.93

(c) Calibration with alternative auxiliary information structure for the health interviewed

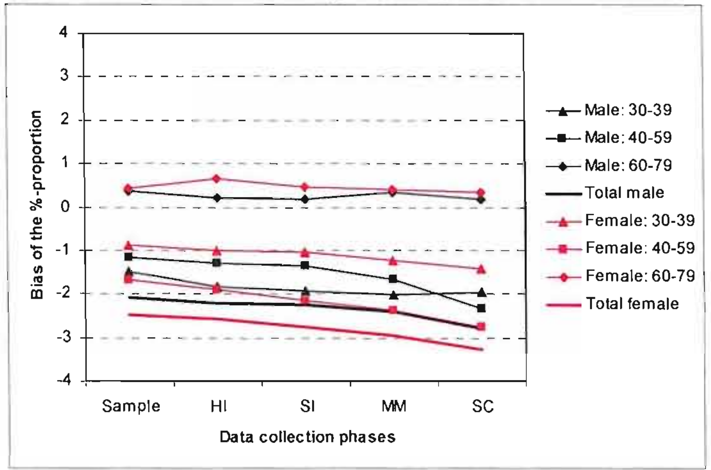
Prevalence estimates of chronic coronnary heart disease, %	Male 30-39	Male 40-59	Male 60-79	Male 80+	Male Total	Female 30-39	Female 40-59	Female 60-79	Female 80+	Female Total
True prevalence from registers										
Target population from register	**0.08**	**2.62**	**17.36**	**23.79**	**6.22**	**0.02**	**0.81**	**11.16**	**18.91**	**4.91**
Unweighted sample estimates										
Sampled	0.00	4.29	22.57	27.45	8.86	0.12	0.74	13.26	21.14	6.74
Prevalence estimates of chronic coronnary heart disease, %- weighted by calibrated weights based on:										
Design weights										
Health interviewed:										
Demographic variables	0.00	4.49	23.70	24.77	8.67	0.17	0.78	13.67	20.87	5.78
Socio-economic variables	0.00	4.55	22.89	28.21	8.61	0.18	0.64	14.76	21.31	6.06
Income inequality variables	0.00	4.31	22.69	26.82	8.40	0.12	0.65	13.93	21.56	5.84
Design weights and sequential logit response probability										
Health interviewed:										
Demographic variables	0.00	4.61	23.59	27.28	9.14	0.06	0.78	12.99	23.78	5.50
Socio-economic variables	0.00	4.88	22.70	29.45	9.40	0.07	0.68	14.59	24.25	6.46
Income inequality variables	0.00	4.74	22.22	24.25	9.00	0.05	0.74	12.51	20.59	5.85
Design weights and multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	0.00	4.58	23.60	25.50	9.07	0.12	0.76	12.96	22.66	5.66
Socio-economic variables	0.00	4.76	22.60	27.48	9.09	0.13	0.66	14.42	23.30	6.48
Income inequality variables	0.00	4.66	22.28	22.91	8.93	0.10	0.71	12.44	19.64	5.94
Sequential logit response probability										
Health interviewed:										
Demographic variables	0.00	4.58	23.55	27.33	9.11	0.06	0.78	12.99	23.73	5.49
Socio-economic variables	0.00	4.82	22.58	29.53	9.34	0.06	0.66	14.40	24.00	6.39
Income inequality variables	0.00	4.70	22.15	24.34	8.95	0.04	0.72	12.39	20.46	5.79
Multilevel sequential logit (MCMC) response probability										
Health interviewed:										
Demographic variables	0.00	4.55	23.56	25.56	9.04	0.12	0.76	12.96	22.62	5.66
Socio-economic variables	0.00	4.69	22.53	27.64	9.04	0.12	0.64	14.22	23.22	6.42
Income inequality variables	0.00	4.62	22.23	23.01	8.88	0.09	0.70	12.33	19.65	5.90

APPENDIX 6.12 Bias of estimated proportion of people receiving pension on disability for work

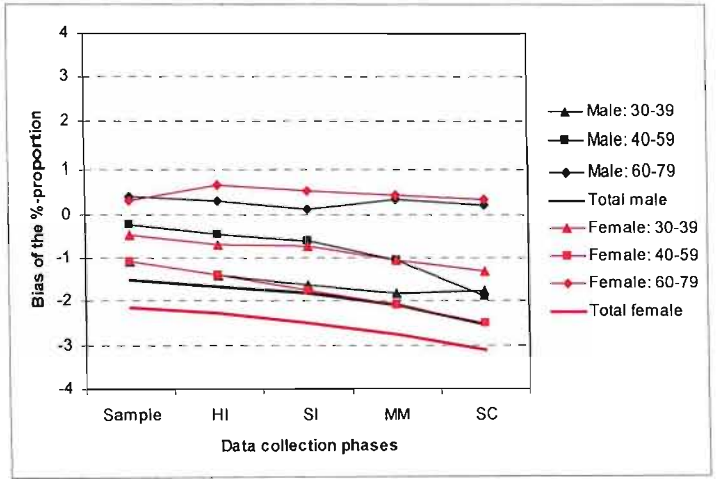
(a) Design weight



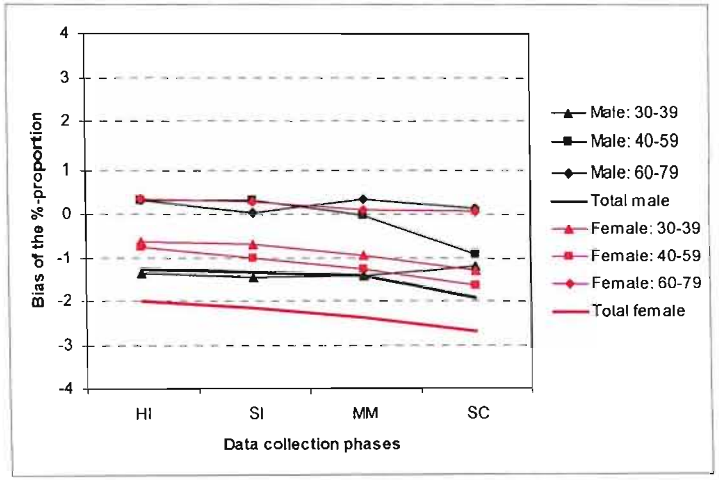
(b) Inverse probability weight (IPW)



(c) Inverse probability weight with interviewer effects (MCMC)



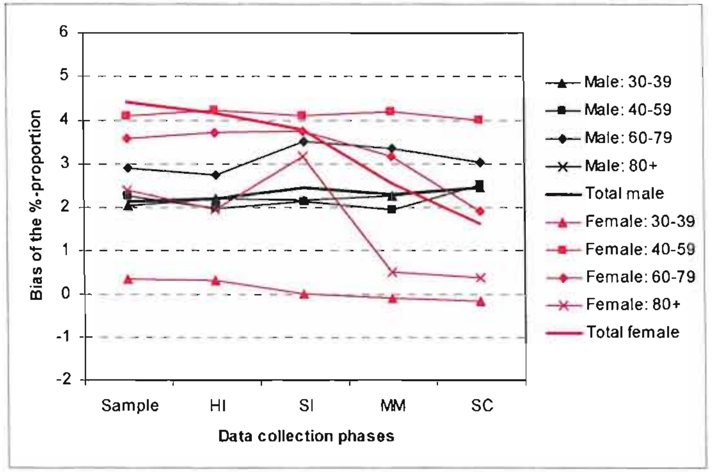
(d) Calibration of IPW with interviewer effects



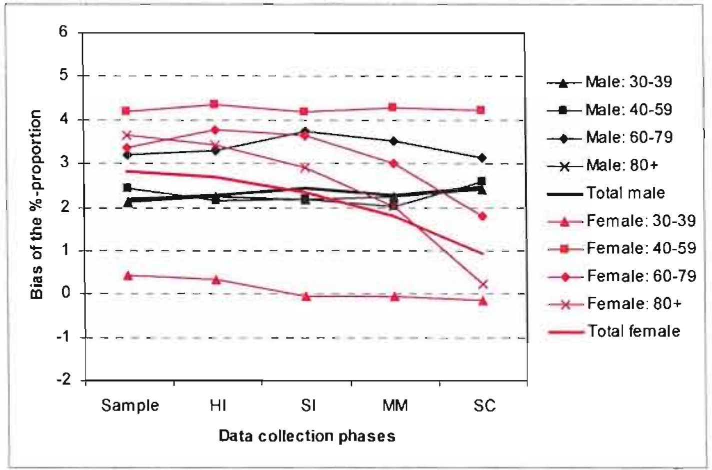
Source: Individual level data linkage from the register of The Social Insurance Institution and Taxation register for the sample, and target population totals derived from the taxation register file

APPENDIX 6.13 Bias of estimated proportion of people receiving medical reimbursement on chronic hypertension

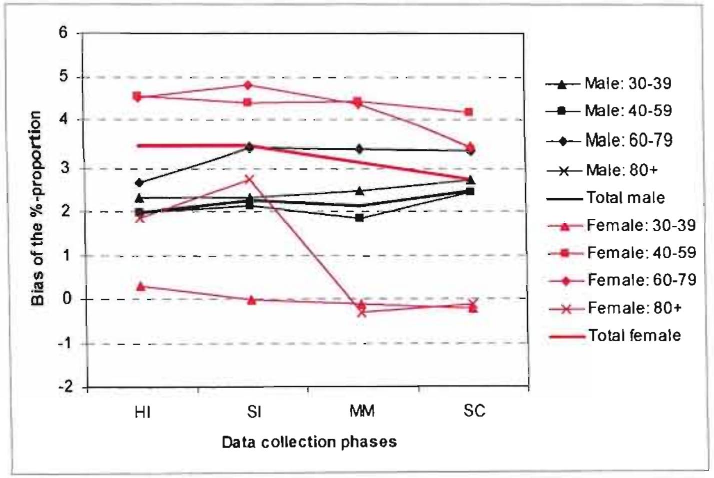
(a) Design weight



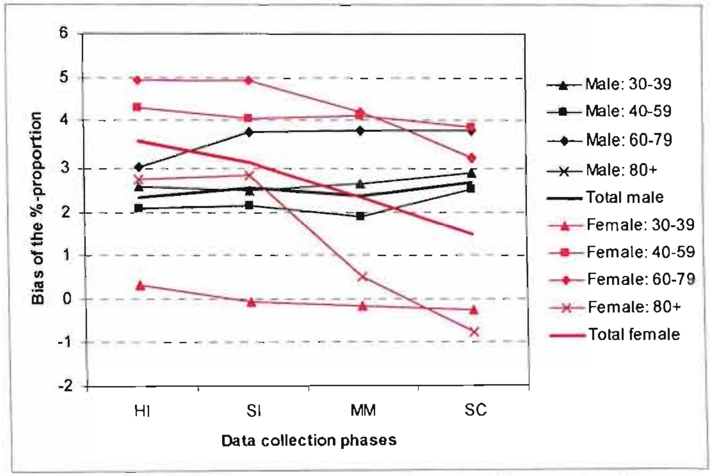
(b) Inverse probability weight (IPW)



(c) Inverse probability weight with interviewer effects (MCMC)



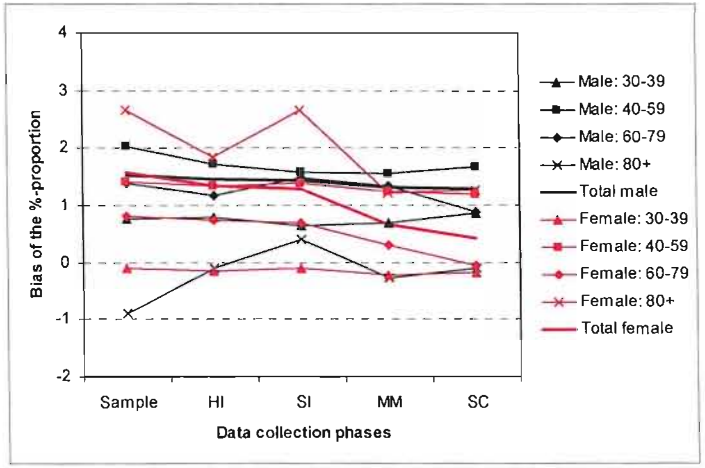
(d) Calibration of IPW with interviewer effects



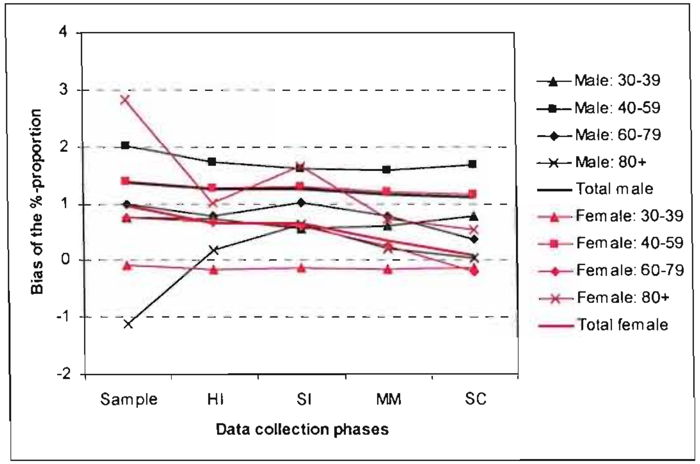
Source: Individual level data linkage from the register of The Social Insurance Institution for the sample, and target population totals derived from the taxation register file

APPENDIX 6.14 Bias of estimated proportion of people receiving medical reimbursement on diabetes mellitus

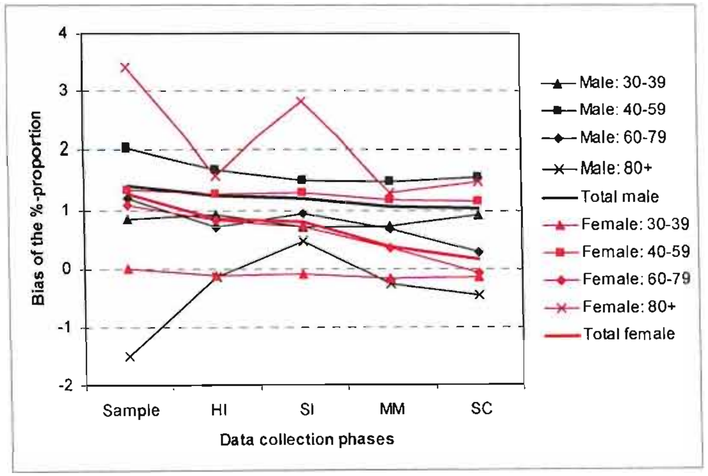
(a) Design weight



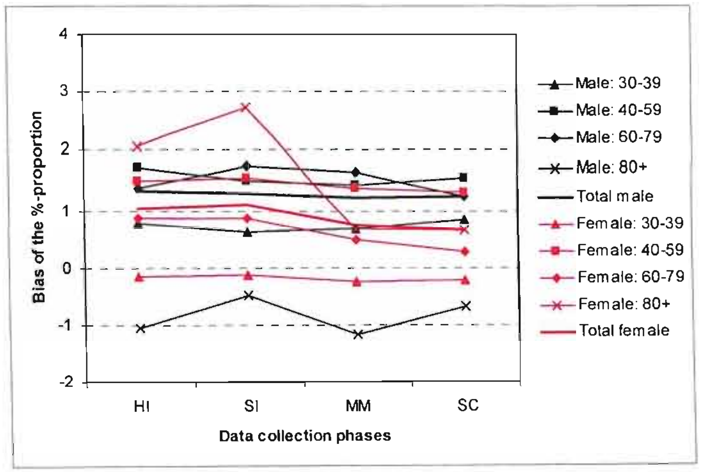
(b) Inverse probability weight (IPW)



(c) Inverse probability weight with interviewer effects (MCMC)

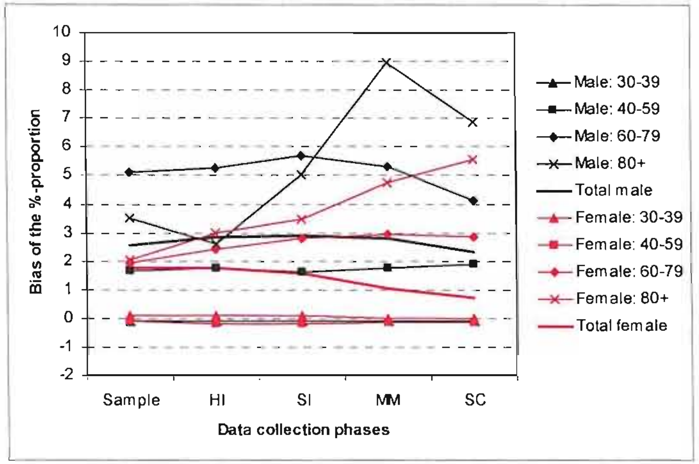


(d) Calibration of IPW with interviewer effects

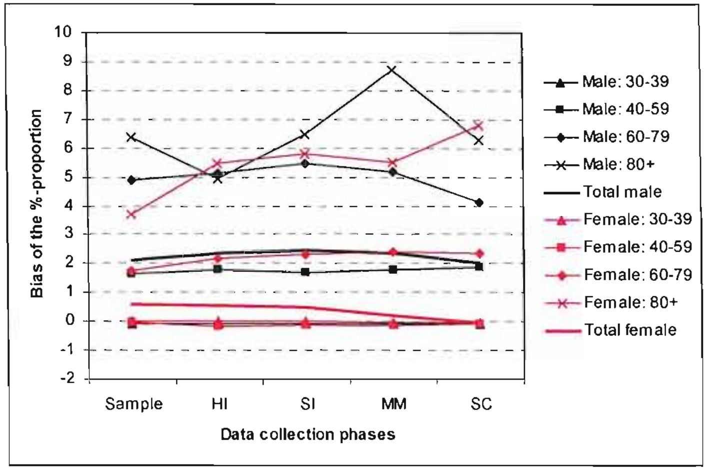


APPENDIX 6.15 Bias of estimated proportion of people receiving medical reimbursement on chronic coronary heart disease

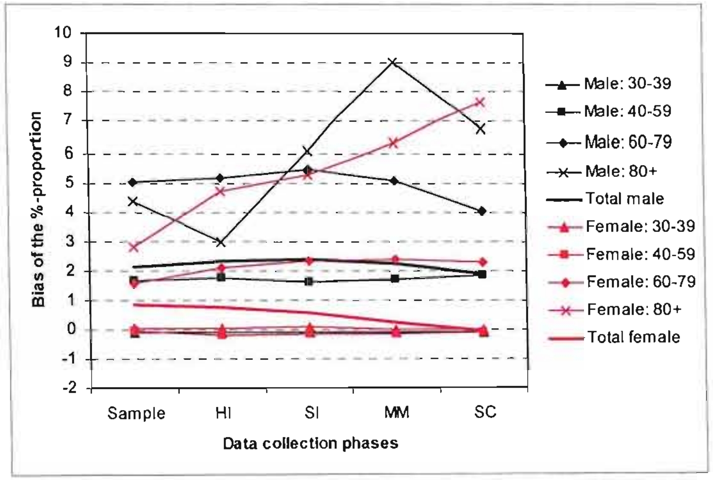
(a) Design weight



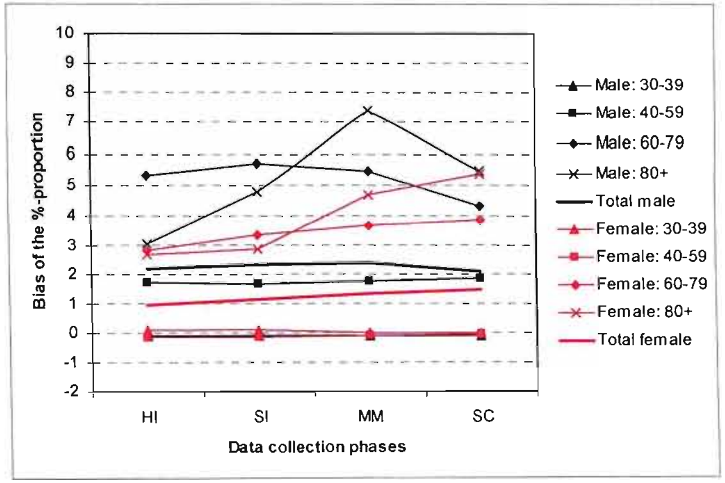
(b) Inverse probability weight (IPW)



(c) Inverse probability weight with interviewer effects (MCMC)

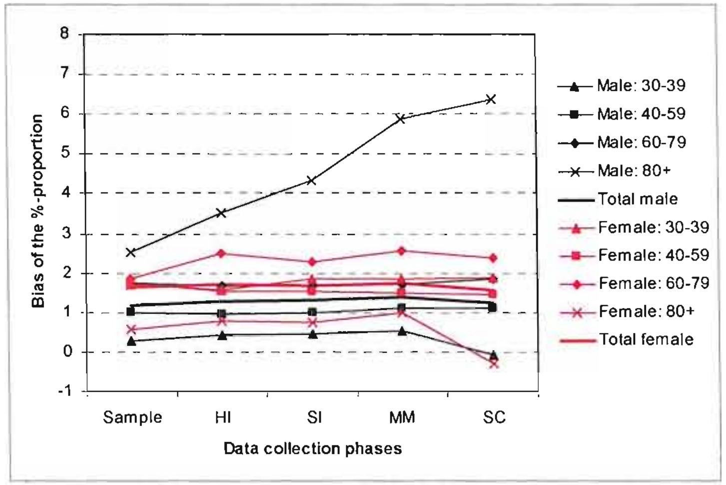


(d) Calibration of IPW with interviewer effects

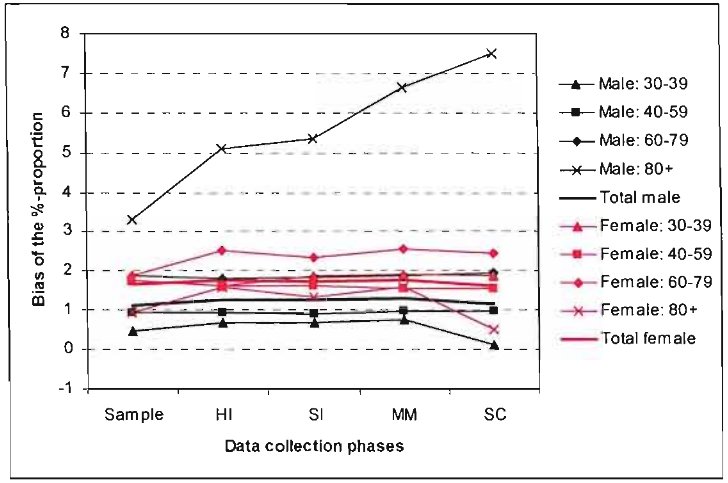


APPENDIX 6.16 Bias of estimated proportion of people receiving medical reimbursement on chronic asthma

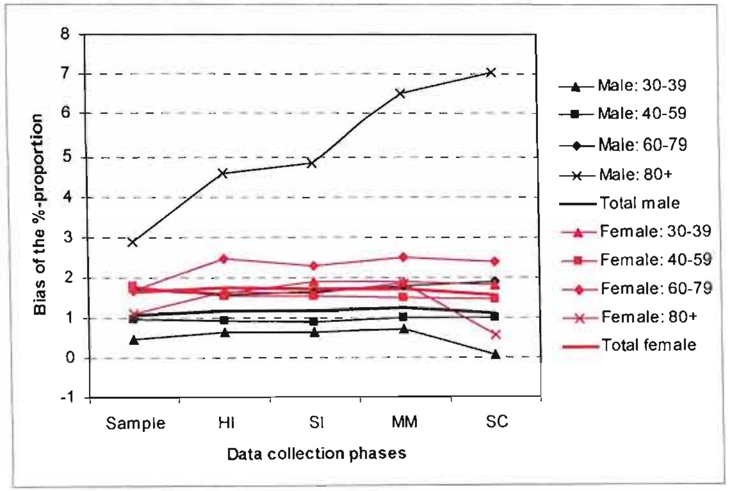
(a) Design weight



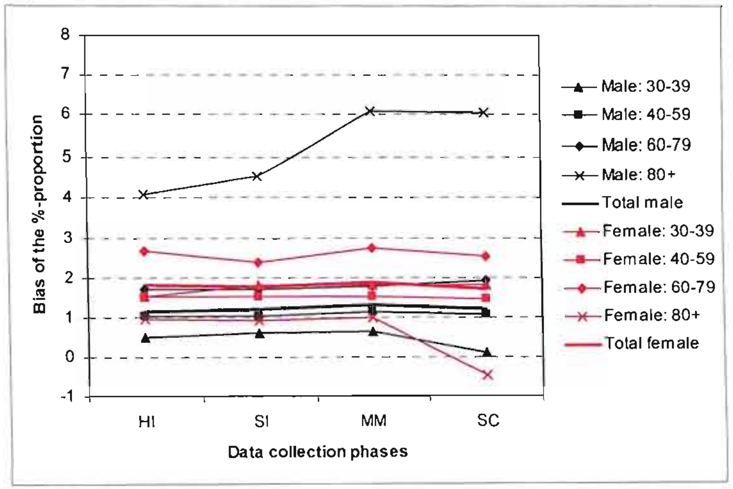
(b) Inverse probability weight (IPW)



(c) Inverse probability weight with interviewer effects (MCMC)



(d) Calibration of IPW with interviewer effects



List of references

AAPOR (2000). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* The American Association for Public Opinion Research. Ann Arbor, Michigan.

Aasland, A., and Fløtten, T. (2001). Ethnicity and social exclusion in Estonia and Latvia. *Europe-Asia Studies.* Vol 53. No 7. 1023–1049.

Agresti, A. (1990). *Categorical Data Analysis.* New York: John Wiley & Sons.

Agresti, A. (2002). Links Between Binary and Multi-Category Logit Item Response Models and Quasi Symmetric Loglinear Models. *Annales de la Faculté des Sciences Toulouse.* Vol XI. No 4. 443–454.

Allison, P. D. (1982). Discrte-Time Methods for the Analysis of Event-Histories. In S. Leinhardt (ed). Sociological Methodology. San Francisco: Jossey-Bass. 61–98.

Allison, P. D. (1984). *Event History Analysis. Regression for Longitudinal Event Data.* Sage University Paper 46. SAGE Publications Ltd.

Andersson, C., and Nordberg, L. (1998). *A User's Guide to CLAN.* Örebro: Statistics Sweden.

Aromaa, A., and Koskinen, S. (eds.) (2002). *Terveys ja toimintakyky Suomessa. Terveys 2000 –tutkimuksen perustulokset.* Publication of the National Public Health Institute B3/2002. Helsinki. (In Finnish only; *Health and functional capacity in Finland. Baseline results of the Health 2000 health examination survey*).

Aromaa, A., Koskinen, S. and the project groups (2004). Population and methods. In Aromaa, A., and Koskinen, S. (eds.). *Health and functional capacity in Finland. Baseline results of the Health 2000 Health Examination Survey.* Publication of the National Public Health Institute B12/2002. Helsinki. (Translation of the revised version of the previous B3/2002 publication).

Atkinson, R., and Davoudi, S. (2000). The Concept of Social Exclusion in the European Union: Context, Development and Possibilities. *Journal of Common Market Studies.* Vol 38. No 3. 427– 448.

Atkinson, T. (1998). Social Exclusion, Poverty and Unemployment. In Atkinson, T., and Hills, J. (eds). *Exclusion, Employment and Opportunity.* CASEpaper. CASE/4. Centre for Analysis of Social Exclusion. London School of Economics.

Berridge, D. M. (1992). A Review of methods for modelling ordered categorical survey data. In Westlake, A., Banks, R., Payne, C., & Orchard, T. (eds) *Survey and Statistical Computing.* Elsevier Science Publishers B. V.

Bethlehem, J. (2002). Weighting nonresponse adjustments based on auxiliary infromation. In Groves, R. M., Eltinge, J. L., and Little, R. J. A. *Survey Nonresponse.* New York: John Wiley & Sons. 275–288.

Bethlehem, J. G., and Kersten, H. M. P. (1985). On the Treatment of Nonresponse in Sample Surveys. *Journal of Official Statistics.* Vol 1. No 3. 287–300.

Biemer, P. P., and Lyberg, L. E. (2003). *Introduction to Survey Quality.* Hoboken, New Jersey: John Wiley & Sons.

Bradburn, N. M. (1992). A response to the nonresponse problem. *Public Opinion Quarterly.* Vol. 56. 391–397.

Brehm, J. (1993). *The Phantom Respondent: Opinion Surveys and Political Representation.* Ann Arbor: University of Michigan Press

Breslow, N. E., and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association.* Vol. 88. No. 421. 9–25.

Breslow, N. E., and Lin, X. (1995). Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika.* Vol. 82. No. 1. 81–91.

Brick, J. M., and Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research.* 5. 215–238.

Brown, C. C. (1975). On the Use of Indicator Variables for Studying the Time Dependence of Parameters in a response-Time Model. *Biometrics.* 31. 863–872.

Brown, P. R., and Bishop, G. F. (1982). *Who refuses and resists in telephone surveys? Some new evidence.* Paper presented at the MAPOR Annual Conference, November.

Browne, W.J. (1998). *Applying MCMC Methods to Multi-level Models.* PhD dissertation, Department of Mathematical Sciences, University of Bath, UK.

Campanelli, P., and O'Muircheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey non-response. *Quality & Quantity.* 33. 59–76.

Campanelli, P., and Sturgis, P. (1997). Separating interviewer and area effects. In Campanelli, P., Sturgis, P., and Purdon, S. (eds) *Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates*, London: SCPR.

Carter, W. B., Elward, K., Malmgren, J., Martin, M. L., and Larson, E. (1991). Participation of older adults in health programs and research: A critical review of the literature. *The Gerontologist.* Vol 31. No 5. 584–592.

Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician.* Vol. 46. No. 3. 167–174.

Clayton, D., Spiegelhalter, D., Dunn, G., and Pickles, A. (1998). Analysis of binary longitudinal data from multiphase samples. *Journal of the Royal Statistical Society.* Series B. 60. 71–87.

Cobb, S., King, S., and Chen, E. (1957). Differences Between Respondents and Nonrespondents in a Morbidity Survey Involving Clinical Examination. *Journal of Chronic Diseases.* Vol 6. No 2. 95–108.

Cohen, G., and Duffy, J. C. (2002). Are nonrespondents to health surveys less healthy than respondents? *Journal of Official Statistics.* 18. 13–23.

Collett, D. (2003). *Modelling Binary Data.* 2nd ed. CRC, Florida: Chapman & Hall.

Couper, M. P. and Groves, R. M. (1992). The Role of the Interviewer in Survey Participation. *Survey Methodology.* 18. 263–278.

Couper, M. P. (1997). Survey introductions and data quality. *Public Opinion Quarterly.* 61. 317–338.

Couper, M. P., and Groves, R. M. (1992). The Role of the Interviewer in Survey Participation. *Survey Methodology.* 18. 263–277.

Couper, M. P., Singer E., and Kulka, R. A. (1997). Participation in the Decennial Census: Politics, Privacy, Pressures. *American Politics Quarterly.* 26. 59–80.

Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society.* Series B. 34. 187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman and Hall.

de Haan, A. (1999) *Social exclusion: towards a holistic understanding of deprivation*, Villa Borsig Workshop Series. Department for International Development (DFID). UK.

de Heer, W. F., and Israëls, A. Z. (1992). Response Trends in Europe. *Proceedings of the Survey Research Methods Section.* American Statistical Association (ASA).

de Leeuw, E. D. (1992). Data quality in mail, telephone, and face-to-face surveys. Amsterdam: TT-Publicaties.

de Leeuw, E. D., Hox, J. J., Snijkers, G., and de Heer, W. (1998). Interviewer opinions, attitudes and strategies regarding survey participation and their effect on response. in Koch, A., and Porst, R. (eds.) *Nonresponse in Survey Research.* ZUMA Nachrichten Spezial, 4. Mannheim: ZUMA. 239–248.

de Leeuw, E., and de Heer, W. F. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. in Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R., J., A. (eds). *Survey Nonresponse.* New York: John Wiley & Sons. 41–54.

de Leeuw, E., Hox, J., and Huisman, M., (2003). Prevention and Treatment of Nonresponse. *Journal of Official Statistics.* 19. 153–176.

de Marco, Verlato G., Zanolin E., Bugiani M., and Drane J.W. (1994). Nonresponse bias in EC Respiratory Survey in Italy. *European Respiratory Journal.* 7. 2139–2145.

Deepchand, K., and Laiho, J. (1999). Methodology and response. In Scottish Health Survey 1998. Volume 2. Chapter 15. Eds. Shaw, A., McMunn, A., and Field, J. Joint Health Surveys Unit, at The National Centre for Social Research and Department of Epidemiology & Public Health, UCL.

DeMaio, T. J. (1980). Refusals: Who, Where and Why. *Public Opinion Quarterly.* 44. 223–233.

Deming, W. E. (1953). On a Probability Mechanism to Attain an Economic Balance Between the Resultant Error of Response and the Bias of Nonresponse. *Journal of the American Statistical Association.* 48. 264. 743–772.

Deville, J. C. (1988). Estimation linéaire et redessement sur informations auxiliaires d'enquêtes par sondage. in Monfort, A. and Laffond, J. J. (eds). *Essais en l'honneur d'Edmond Malinvaud.* Paris: Economica. 915–927.

Deville, J. C., and Särndal, C. E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association.* 87. 376–382.

Deville, J. C., Särndal, C. E., and Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association.* 88. 1013–1020.

Diggle, P. J., Heagerty, P., Liang, K.-Y., Zeger, S. L. (2002). *Analysis of Longitudinal Data.* 2nd edition. Oxford: Oxford University Press.

Djerf, K. (2001). *Properties of Some Estimators Under Unit Nonresponse.* Research Reports 231. Statistics Finland.

Djerf, K. (2004). Nonresponse in time: A Time series analysis of the Finnish Labour Force Survey. *Journal of Official Statistics.* Vol. 20. No. 1. 39–54.

Djerf, K., Laiho, J., and Härkänen, T. (2004). Asetelmapohjainen estimointi. (Design-based estimation.) in Laiho, J. and Nieminen, T. (eds) Terveys 2000-tutkimus. Aikuisväestön haastatteluaineiston tilastollinen laatu. (*Health 2000-survey. Statistical quality of the health interview data on adult population.*) (in Finnish only). 48–60.

Duhart, D., Bates, N., Williams, B. T., Diffendal, G., and Chiu, P. L. (2001). *Are Late/Difficult Cases in Demographic Survey Interviews Worth the Effort? A Review of Several Federal Surveys.* Report. US Bureau of the Census.

Durbin, J., and Stuart, A. (1951). Differences in response rates of experienced and inexperienced interviewers. *Journal of the Royal Statistical Society.* Series A. 114. 163–206.

Ekholm, A., and Laaksonen, S. (1991). Weighting via Response Modeling in the Finnish Household Budget Survey. *Journal of Official Statistics.* Vol 7. No 3. 325–337.

Elliot, D. (1991). *Weighting for non-response.* A survey researcher's guide. OPCD. Social Survey Division.

Eurostat. (2001). *Summary Report from the Leadership Group (LEG) on Quality.* 31 July 2001.

Fienberg, S. E. (1980). The Analysis of Cross-classified Categorical Data. 2nd edition. M.I.T. Press, Cambridge, MA.

Fitzgerald, J., Gottschalk, P., and Moffitt, R., (1998). An Analysis of Sample Attrition in Panel data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources.* Vol 33. No 2. 251–299.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association.* Vol. 85. No. 410. 398–409.

Glenn, N.G. (1969). Aging, Disengagement and Opinionation. *Public Opinion Quarterly.* Vol 33. No 1. 17–33

Goldstein, H. (1995). *Multilevel Statistical Models*. 2nd ed. London: Edward Arnold.

Goldstein, H. (1998). *Consistent estimators for multilevel generalised linear models using an iterated bootstrap.* Multilevel Models Project. Working Paper. December, 98. Institute of Education. London, UK

Goldstein, H., Browne, W., and Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics*. 1. 223–232.

Goyder, J. (1987). *The Silent Minority: Nonrespondents on Sample Surveys*. Boulder, CO: Westview.

Goyder, J., Lock, J., and McNair, T. (1992). Urbanization effects on survey nonresponse: a test within and across cities. *Quality and Quantity*. Vol. 26. 39–48.

Gray, R., Campanelli, P., Deepchand, K., and Prescott-Clarke, P. (1996). Exploring Survey Non-response: The Effect of Attrition on a Follow-Up of the 1984-85 Health and Life Style Survey. *Statistician*. Vol. 45. Issue 2. 163–183.

Greenwood, M. (1926). The natural duration of cancer. *Reports on Public Health and Medicine Subjects*. 33. 1–26.

Groves, R. M. (1978). On the Mode of Administering a Questionnaire and Responses to Open-ended Items. *Journal of Marketing Research*. 15. 622–631.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons.

Groves, R. M., and Couper, M. P. (1992). *Respondent-interviewer interactions in survey introductions*. Paper presented at the 3rd International Workshop on Household Survey Nonresponse, CBS, Voorburg, Holland.

Groves, R. M., and Couper, M. P. (1995). Theoretical Motivation for Post-Survey Nonresponse Adjustment in Household Surveys. *Journal of Official Statistics*. Vol. 11 No. 1. 93–106.

Groves, R. M., and Couper, M. P. (1996). Contact-Level Influences on Cooperation in Face-to-Face Surveys. *Journal of Official Statistics*. Vol. 12 No. 1. 63–83.

Groves, R. M., and Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: John Wiley & Sons.

Groves, R. M., and Couper, M. P. (1995). Theoretical Motivation for Post-Survey Nonresponse Adjustment in Household Surveys. *Journal of Official Statistics*. Vol 11. No 1. 93–106.

Groves, R. M., Cialdini, R. B., and Couper, M. P. (1992). Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly*. Vol. 56. No 4. 475–495.

Groves, R. M., Fowler, F.J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2004). *Survey Methodology*. New York: John Wiley & Sons.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). *Sample Survey Methods and Theory*. New York: John Wiley & Sons.

Helasoja V., Prättälä R., Dregval L., et al. (2002). Late response and item nonresponse in the Finbalt Health Monitor Survey. *European Journal of Public Health.* Vol 12. 117–23.

Herzog, A. R., and Rodgers, W. L., (1988). Age and Response Rates to Interview Sample Surveys. *Journal of Gerontology: Social Sciences.* Vol 43. No 6. 200–205.

Hidiroglou, M. A., Drew, D., and Gray, G. B. (1993). A Framework for Measuring and reducing Nonresponse in Surveys. *Survey Methodology.* Vol. 19. No 1. 81–94.

Horvitz, D. G., and Thompson, D. J. (1952). A Generalization of Sampling without Replacement from a Finite Universe. *Journal of the American Statistical Association.* 47. 663–685.

Hosmer, D. W., and Lemeshow, S. (1989). *Applied Logistic Regression.* New York: John Wiley & Sons.

Hosmer, D. W., and Lemeshow, S. (2000). *Applied Logistic Regression.* (2nd Ed). New York: John Wiley & Sons.

House, J. S., and Wolf, S. (1978). Effects of Urban Residence on Interpersonal Trust and Helping Behavior. *Journal of Personality and Social Psychology.* Vol. 36. No 9. 1029-1043.

Hox, J. J. (2002). *Multilevel Analysis. Techniques and Applications.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hox, J. J., and de Leeuw, E. (1994). A Comparison of Nonresponse in Mail, Telephone, and Face-to-face Surveys. *Quality and Quantity.* Vol. 28. 329–344.

Hox, J. J., and de Leeuw, E. (2002). The Influence of Interviewers' Attitude and Behavior on Household Survey Nonresponse: An international Comparison. In Groves, R. M., Dillman, D. A., Eltinge, J. L., Little, R. J. A. (eds) *Survey Nonresponse.* New York: John Wiley & Sons. 103–120.

Hox, J. J., de Leeuw, E., and Kreft, G. G. (1991). The effect of interviewer and respondent characteristics on the quality of survey data: a multilevel model. in Biemer, P. et al. (eds). *Measurement errors in Surveys.* New York: John Wiley & Sons.

Hox, J., de Leeuw, E. (2002). The Influence of Interviewers' Attitude and Behavior on Household Survey Nonresponse: An international Comparison. In Groves, R. M., Dillman, D. A., Eltinge, J. L., Little, R. J. A. (eds) *Survey Nonresponse.* New York: John Wiley & Sons. 103–120.

Hyman, H. H. (1954). *Interviewing in Social Research.* Chicago: Chicago University Press.

Iannacchione, V.G. (2003). Sequential Weight Adjustments for Location and Cooperation Propensity for the 1995 National Survey of Family Growth. *Journal of Official Statistics.* Vol 19. No 1. 31–43.

Iisakka, L. (2004a). *The main social surveys of Statistics Finland since 1990 and their link to the social capital.* Unpublished review available from the internet. In Finnish only. (Tilastokeskuksen keskeisimmät survey-tutkimukset vuodesta 1990 alkaen ja niiden yhteys sosiaaliseen pääomaan). [www.stat.fi/org/tut/dthemes/papers/sospaaoma_surveyt.html]

Iisakka, L. (2004b). *Other central surveys with the link to social capital in Finland.* Unpublished review available from the internet. In Finnish only. (Muut keskeiset Suomessa tehdyt sosiaalista pääomaa sivuavat aineistot). [www.stat.fi/org/tut/dthemes/papers/sospaaoma_surveyt2.html]

Japec, L., and Lundqvist, P. (1999). *Interviewer strategies and attitudes.* Paper presented at the International Conference on Survey Nonresponse, Portland, Oregon, USA.

Johnson, T. P., O'Rourke, D., Burris, J., and Owens, L. (2002). Culture and Survey Nonresponse. In Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. A. (eds) *Survey Nonresponse.* New York: John Wiley & Sons. 55–69.

Jones, A. M., Koolman, X., and Rice, N. (2006). Health-related non-response in the British Household Panel Survey and European Community Household Panel: using inverse-probability-weighted estimators in non-linear models. *Journal of Royal Statistical Society.* Series A. Vol 169. Part 3. 543–569.

Kahn, R. L., and Cannell, C. F. (1957). *The Dynamics of Interviewing: Theory Techniques and Cases.* New York: John Wiley & Sons.

Kalsbeek, W. D. (1980). *A Conceptual Review of Survey Error due to Nonresponse.* Proceedings of the Survey Research Methods Section. American Statistical Association.

Kalton, G., and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics.* Vol 19. No 2. 81–97.

Kalton, G., and Kasprzyk, D. (1986). The Treatment of Missing Survey Data. *Survey Methodology.* Vol. 12. No 1. 1–16.

Kalton, G., and Maligalig, D.S. (1991). A Comparison of Methods of Weighting Adjustment for Nonresponse. Proceedings of the U.S. Bureau of the Census Annual Research Conference. 409–428.

Kane, E. W., and McCaulay, L. J. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly.* Vol. 57. 1–28.

Kelleher, C., Gabhainn, S. N., Friel, S., Corrigan, H., Nolan, G., Sixsmith, J., Walsh, O., and Cooke, M. (2003). *The National Health & Lifestyle Surveys.* Surveys of Lifestyle, Attitudes and Nutrition (SLAN) & The Irish Health Behaviour in School-Aged Children Survey (HBSC). Health Promotion Unit, Deaprtment of Health and Children, Dublin & Centre for Health Promotion Studies, National University of Ireland, Galway & Department of Public Health Medicine and Epidemiology, University College Dublin.

Kemsley, W. F. F. (1976) National Food Survey – a study of differential non-response based on a comparison of the 1971 sample with the Census. *Statistical News* No. 35. 18–22.

Kerola, O., Järvinen, P., Pelkonen, L., and Nyfors, H. (2003). Report by the Committee on reform of the payment policy and the payment system in social welfare and health care. Committee Report. *Publications of the Ministry of Social Affairs 2005:10.* Ministry of Social Affairs. (In Finnish only).

Kessler, R. C., Little, R. J. A., and Groves, R. M. (1995). Advances in strategies for minimizing and adjusting for survey nonresponse. *Epidemiologic Reviews*. Vol 17. No 1. 192–204.

Kleinbaum, D. G. (1996). *Survival Analysis*. New York: Springer Verlag.

Kojetin, B. A., and Tucker, C. (1999). Exploring the Relation of Economical and Political Conditions with Refusal rates to a Government Survey. *Journal of Official Statistics*. 2. 167–184.

Koponen P., and Aromaa A. (2001). *Health examination surveys (HES). Review of literature and inventory of surveys in the EU/EFTA Member States*. National Public Health Institute, Helsinki.

Koponen, P., and Aromaa, A. (2003). *Survey design and methodology in national health interview and health examination surveys*. Review of literature, European survey experiences and recommendations. Health surveys in the EU: HIS and HIS/HES evaluations and models. Phase 2/Subproject 3. European Commission.

Korn, E. L., and Graubard, B. I. (1995). Analysis of Large Health Surveys: Accounting for the Sampling Design. Journal of the Royal Society. Series A. Vol. 158. No. 2. 263–295.

Laiho, J. (1998). Varallisuustutkimus 1994. Laatuselvitys. Katsauksia 1998/2. Tilastokeskus. (*Quality report of the Wealth Survey 1994*. Reports 1998/2. Statistics Finland. In Finnish only).

Laiho, J. and Nieminen, T. (2004). *Terveys-2000 tutkimus. Aikuisväestön haastatteluaineiston tilastollinen laatu*. Research reports 239. Statistics Finland. (In Finnish only; *Health 2000 Survey. The statistical quality of the interview data of the adult population*)

Laiho, J., and Lynn, P. (1999). *Separating out the Phases of the Survey Participation Process*. Paper presented at the ICSN conference. Portland, USA.

Larsen, K., and Merlo, J. (2005). Appropriate Assessment of Neighbourhood Effects on Individual Health: Integrating Random and Fixed Effects in Multilevel regression. *American Journal of Epidemiology*. Vol. 161. No. 1. 81–88.

Larsen, K., Petersen, J. H., Budtz-Jorgensen, E., and Endahl, L. (2000). Intepreting Parameters in the Logistic Regression Model with Random Effects. *Biometrics*. Vol. 56. 909–914.

Lehtonen, R. (1996). Interviewer Attitudes and Unit Nonresponse in Two Different Interviewing Schemes. In Laaksonen, S. (ed) *International Perspectives on Nonresponse*. Research Reports 219. Helsinki: Statistics Finland. 130–140.

Lehtonen, R. and Pahkinen, E. (1996). *Practical Methods for Design and Analysis of Complex Surveys*. Revised Edition. Chichester: John Wiley & Sons.

Lehtonen, R., Djerf, K., Härkänen, T., and Laiho, J. (2003). Modelling complex health survey data: a case study. In Höglund, R., Jäntti, M., Rosenqvist, G. (eds.) *Statistics, econometrics and society: Essays in honour of Leif Nordberg*. Research reports 238. Statistics Finland.

Lemel, Y. (1976). Une Généralisation de la Méthode du Quotient pour le Redressement des Enquêtes par Sondage. Annales de l'INSEE. 22–23. 272–282.

Lepkowski, J. M., and Couper, M. P. (2002). Nonresponse in the Second Wave of Longitudinal Household Surveys. in Groves, R. M., Dillman, D. A., Eltinge, J.L., and Little, R. J. A. (eds) *Survey Nonresponse.* New York: John Wiley & Sons. 259–274.

Lessler, J., and Kalsbeek, W. (eds.). (1992). *Nonsampling Error in Surveys.* New York: John Wiley & Sons.

Lievesley, D. (1988). *Unit Nonresponse in Interview Surveys.* Unpublished Working Paper. London: Social and Community Planning Research.

Lin, I.- F., and Scaeffer, N, C. (1995). Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly.* 59. 236-258.

Lindqvist, M., Pajunen, A., and Laiho, J. (2001). Kulutustutkimukset 1994-1996. Laatuselvitys. Katsauksia 2000/1. Tilastokeskus. (*Household Budget Surveys in 1994-1996. A quality report.* Reports 2000/1. Statistics Finland. In Finnish only).

Lindström, H. (1983). Non-Response Errors in Sample Surveys. *Urval.* No 16.

Little, R. J. A. (1986). Survey Nonresponse Adjustments for Estimates of Means. *International Statistical Review.* 54. 139–157.

Loosveldt, G., Pickery, J. and Billiet, J. (2002). Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. *Journal of Official Statistics.* Vol 18. No 4. 545–557.

Lundström, S., and Särndal, C.-E. (2001). Estimation in the Presence of Nonresponse and Frame Imperfections. *A Current Best Method (CBM) Manual.* Statistics Sweden.

Lynn, P., Beerten, R., Laiho, J., and Martin, J. (2002a). Towards standardisation of survey outcome categories and response rate calculations. *Research in Official Statistics.* Vol. 5. No. 1. 61–84.

Lynn P., Clarke, P., Martin, J., and Sturgis, P. (2002b). The effects of extended interviewer efforts on non-response bias. in Groves, R. M., Dillman, D. A., Eltinge, J.L., and Little, R. J. A. (eds) *Survey Nonresponse.* New York: John Wiley & Sons. 135–147.

Mare, R.D. (1980). Social Background and School Continuation Decisions. *Journal of the American Statistical Association.* 75. 295–303.

Martin, J., and Beerten, R. (1999). *The effect of Interviewer Characteristics on Survey Response Rates.* Paper presented at ICSN conference. Portland.

Mason, R., Lessler., V., and Traugott, M. W. (2002). Effect of Item Nonresponse on Nonresponse Error and Inference. In Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (eds) *Survey Nonresponse.* New York: John Wiley & Sons. 149–162.

Mathiowetz, N. A., Demaio, T. J., and Martin, E. (1991). *Political Alienation, Voter Registration and the 1990 Census.* Paper presented at the annual conference of the American Association for Public Opinion Research. Phoenix. AZ.

Mayes D. G., Berghman, J. and Salais, R. (2001). *Social exclusion and European Policy.* Cheltenham: Edward Elgar.

McCullagh, P. and Nelder, J. A. (1989). *Generalised Linear Models.* 2nd ed. Chapman and Hall.

Mercer, J. R., and Butler, E. W. (1967). Disengagement of the Aged Population and Response Differentials in Survey Research. *Social Forces.* Vol 46. No 1. 89–96.

Miethe, T., and Lee, G. R. (1984). Fear of crime among older people. A reassessment of the predictive power of criminal-related factors. *Sociological Quarterly.* Vol 25. No 3. 397–415.

Morton-Williams, J. (1993). *Interviewer Approaches.* Aldershot: Dartmouth Publishing Company.

Mulry, M.H., Spencer, B.D. (1991). Total Error in PES Estimates of Population. *Journal of the American Statistical Association.* 86. 839–855.

Myrskylä, P. (1991). Census by Questionniare – Census by Registers and Administrative Records: The experience of Finland. *Journal of Official Statistics.* Vol. 7. No 4. 457–474.

National Agency for Medicines and Social Insurance Institution (2005). Finnish Statistics on Medicines 2004.

Novo, M.., Hammarström, A., and Janlert, U. (1999). Does low willingness to respond introduce a bias? Results from a socio-epidemiological survey among young men and women. *International Journal of Social Welfare.* 8. 155–163.

O'Muircheartaigh, C., and Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society.* Series A. 161. 63–77.

Oh, J. L., and Scheuren, F. (1983). Weighting Adjustment for Unit Nonresponse. In: Madow, W. G., Olkin, I., and Rubin, D. B. (toim.). *Incomplete Data in Sample Surveys.* Vol. 2. New York: Academic Press. 143–184.

Osberg, L. (2000). Poverty in Canada and the United States: Measurement, Trends and Implications. *The Canadian Journal of Economics.* Vol. 33. No. 4. 847–877.

Owens, L., Johnson, T. P., and O'Rourke, D. (2001): *Culture and Item Nonresponse in Health Surveys.* In conference proceedings: Cynamon, M. L., and Kulka, R. A. (Eds) Seventh Conference on Health research Methods. Department of Health and Human Services. Centers for Disease Control and Prevention. National Center for Health Statistics. 69–74.

Pennel, S. G. (1990). *Evaluation of Advance Letter Experiment in 1988 New York Reproductive Health Survey.* Report to the Centers for Disease Control. Ann Arbor, MI: Survey Research Center.

Pickery, J., and Loosveldt, G. (2000). Modeling Interviewer Effects in Panel Survey: An application. *Survey Methodology.* Vol. 26. No. 2. 189–198.

Pickery, J., Loosveldt, G., and Carlton, A. (2001). The effects of the interviewer and respondent characteristics on response behaviour in panel surveys: A multilevel approach. *Sociological Methods and Research*. Vol. 29. No. 4. 509–523.

Pidgin, C. F. (1888). *Practical Statistics: A Handbook for the Use of the Statistician at Work*. Boston: The William E. Smythe Company.

Powers, D. A. and Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis*. San Diego: Academic Press.

Rasbash, J. and Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*. 19. 337–350.

Räty, S. ,Aromaa, A., and Koponen, P. (2003). *Measurement of physical functioning in comprehensive national health surveys - ICF as a framework*. Health Surveys in the EU: HIS and HIS/HES Evaluations and Models. Phase 2, Subproject 4. European Commission and National Public Health Institute, Department of health and functional capacity.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross sectional and longitudinal research. *Journal of Educational Statistics*. 18. 321–349.

Research Triangle Institute (2004). SUDAAN Language Manual, Release 9.0 Research. Triangle Park, NC: Research Triangle Institute.

Riedel-Heller, S. G., Busse, A., and Angermeyer, M. C. (2000). Are cognitively impaired individuals adequately represented in community surveys? Recruitment challenges and strategies to facilitate participation in community surveys of older adults. A review. *European Journal of Epidemiology*. Vol 16. 827–835.

Rodríguez, G., and Goldman, N. (1995). An Assessment of Estimation procedures for Multilevel Models with Binary Responses. *Journal of the Royal Statistical Society*. Series A. 158. 73–90.

Rucker, R. E. (1990). Urban crime: Fear of victimization and perceptions of risk. *Free Inquiry in Creative Sociology*. Vol 18. No 2. 151–160.

Ruotsalainen, K. (2002). Use of administrative data in population censuses – Definition of main type of activity as an example. *UNECE Works session on Statistical Editing*. 27–29 May 2002, Helsinki, Finland.

Särndal, C.-E., and Lundström, S. (2005). *Estimation in surveys with nonresponse*. Chichester: John Wiley & Sons.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Sautory, O. (1993). *La macro CALMAR. Redressement d'un échantillon par calage sur marges*. I.N.S.E.E., Série des documents de travail *n*° F 9310. Paris: I.N.S.E.E.

Scheuch, E. K. (1967). Das Interview in der Sozialforschung. (Interview in the Social Research, in German only). In Konig, R. (ed.) *Handbuch der Emprischen Sozialforschung.* Vol. 1. Stuttgart: F. Enke.

Sharp, L. M., and Frankel, J. (1983). Respondent Burden: A Test of Some Common Assumptions. *Public Opinion Quarterly.* Vol. 47. Issue 1. 36–53.

Singer, E. Frankel, M. R., and Glassman, M. B. (1983). The effect of Interviewer Characteristics and Expectations on Response. *Public Opinion Quarterly.* Vol 47. 68–83.

Singer, E., and Kohnke-Aguirre, L. (1979). Interviewer expectation effects: a replication and extension. *Public Opinion Quarterly.* Vol 43. 245–260.

Singer, E., Mathiowetz, N. A., and Couper, M. P. (1993). The Impact of Privacy and Confidentiality Concerns on Survey Participation. The Case of the 1990 U.S. Census. *Public Opinion Quarterly.* Vol. 57. 465–482.

Singer, E., von Thurn, D.R., and Miller, E.R. (1995). Confidentiality assurances and response. A Quantitative review of the experimental literature. *Public Opinion Quarterly.* Vol. 59. 66–77.

Singer, J. D., and Willett, J. B. (1993). It's about time: Using discrete-Time Survival Analysis to Study Duration and Timing of Events. *Journal of Educational Statistics.* Vol 18. No 2. 155–195.

Skinner, C. J., Holt, T. and Smith, T. M. F. (eds) (1989). *Analysis of Complex Surveys.* Chichester: John Wiley & Sons.

Smith, T. W. (1979). *Sex and the GSS.* General Social Survey Technical Report. No 17. Chigago: National Opinion Research Center.

Smith, T. W. (1983). The Hidden 25 Percent: An Analysis of Nonresponse on the 1980 General Social Survey. *Public Opinion Quarterly.* Vol.47. 386–404.

Smith, T. W. (2002). Developing Nonresponse Standards. In Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (eds) *Survey Nonresponse.* New York: John Wiley & Sons. 27–40.

Snijders, G., Hox, J., and de Leeuw, E. (1999). Interviewers' Tactics for Fighting Survey Nonresponse. *Journal of Official Statistics.* Vol. 15. No. 2. 185–198.

Snijders, T., and Bosker, R., (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling.* London: SAGE Publishing.

Southwell, P.L. (1985). Alienation and Nonvoting in the United States: A Refined Operationalization. *Western Political Quarterly.* Vol. 38. 663–674.

Stang, A., and Jöckel, K.-H., (2003) Studies with Low Response Proportions May Be Less Biased than Studies with High Response Proportions. *American Journal of Epidemiology.* Vol. 159. No. 2. 204–210.

Statistics Act (280/2004). National Statistical Act of Finland.

Statistics Finland. (2002). *Quality Guidelines for Official Statistics.* Handbooks 43b. Laiho, J. and Hietaniemi, L (eds). Helsinki: Statistics Finland.

Steeh, C. G. (1981). Trends in Nonresponse Rates, 1952–1979. *Public Opinion Quarterly*. Vol. 45. 40–57.

Sudman, S., and Bradburn, N. M. (1974). *Response Effects in Surveys*. New York. Academic Press.

Sudman, S., Norman, M., Blair, E., and Stocking, C. (1977). Modest expectations: the effects of the interviewers' prior expectations on response. *Sociological Research*. Vol. 6. 177–182.

Sugiyama, M. (1992). Responses and nonresponses. In Lebart, L. (ed) *Quality of Information in Sample Surveys*. Paris: Dunond.

Tremblay, M. (2005). *Canadian Health Measures Survey (CHMS): An introduction and overview*. Seminar Presentation at the National Public Health Institute. Helsinki, Finland.

Weatherford, M. S. (1991). Mapping the Ties That Bind: Legitimacy, Representation, and Alienation. *Western Political Quarterly*. Vol. 44. 251–276.

Wolter, K.M. (1986). Some Coverage Error Models for Census Data. *Journal of the American Statistical Association*. 81. 338–346.

Woods, R., Dobbs, L., Gordon, C., Moore, C., and Simpson, G. (2004). *Report of a thematic study using transnational comparisons to analyse and identify cultural polices and programmes that contribute to preventing and reducing poverty and social exclusion*. A report undertaken for the European Commission by: The Centre for Public Policy, Northumbria University, Newcastle upon Tyne, UK.

Xie, Y. (1994). The Log-Multiplicative Models for Discrete-Time, Discrete-Covariate Event History Data. *Sociological Methodology*. 24. 301–340.

Yamaguchi, K. (1991). *Event History Analysis*. Applied Social Research Methods Series. Volume 28. Newsbury Park: Sage.

Ylitalo, M. (2002). Väestötietojärjestelmän luotettavuustutkimus 2001. (*Reliability Study of the Population Information System* 2001, In Finnish only). Report 15.2.2002. Statistics Finland. Social Statistics. Interview and Survey Services.