

COGNITIVE AGING AND SURVEY MEASUREMENT

by

Sonja Irene Ziniel

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Survey Methodology)
in The University of Michigan
2008**

Doctoral Committee:

**Professor Norbert W. Schwarz, Chair
Professor Robert M. Groves
Emeritus Professor Robert L. Kahn
Associate Professor Frederick G. Conrad
Research Scientist Steven G. Heeringa**

© SONJA IRENE ZINIEL

2008

ACKNOWLEDGEMENTS

This dissertation would have never been accomplished without my terrific dissertation committee. Every one of its members was indispensable in helping me finish it by playing a special role. First of all, I want to thank my chair, Norbert Schwarz, for his reliable guidance and unconditional support no matter where he was in the world at the time. He made sure that my focus stayed on the important things and that I kept the dissertation in perspective. If not for him, the dissertation process would have been much more difficult. Bob Groves has had the most significant influence on my development as a researcher and especially as a survey methodologist. Without him I would not be where I am today. He has challenged my thoughts for the better and supported me in developing my career in every way possible. Steven Heeringa is one of the most patient teachers I ever had and he has an extraordinary gift for guiding students through problems. Because of his calm manner and analytical but practical approach I always left his office more confident in my abilities to solve the given problem no matter how unsolvable it seemed at first. I am also extremely grateful to Fred Conrad and Norman Brown for believing in my ideas and guiding and supporting me through the process of my own data collection. Their ability to motivate me gave me the confidence to carry on as the challenges of the dissertation increased. Last but not least, I want to thank Bob Kahn for his wisdom and his enthusiasm in advising me on my dissertation. He was truly an inspiration. All of my committee members provided valuable comments on earlier drafts of the chapters that definitely improved them. I am also grateful for receiving the Regula Herzog Award and the National Science Foundation's dissertation research grant (0648709). Without these awards I would not have been able to field the study that is now part of my dissertation. Mary-Beth Ofstedal and David Weir were very

supportive of my research by providing some of the data and answering any questions I had about the Health and Retirement Study.

My sentence in the “dissertation dungeon” would have been unbearable without Kevin Walmsley, Elisabeth Schneider, Kristen Olson and Dani Fromm. They all helped me through the difficult times and their unconditional support kept me going. I am extremely grateful to Kevin for the 24-hour dissertation support program and his unshakable belief in my success from the day that I started the Ph.D. Program in Survey Methodology. He always supported me and my academic career and did not once complain about the work load that I faced during my Ph.D. program. I can’t acknowledge that enough. I am also grateful to Elisabeth who as one of my best friends was always willing to listen and made sure that once in a while I gave my mind a break and did some fun stuff. Kristen, as throughout our friendship, was always available to listen to me and discuss my research ideas. She provided valuable suggestions more than once. Dani was willing to be my dissertation buddy and essential in helping me set specific as well as realistic goals from day to day. Even though an ocean away, my family was always there when I needed them and showed their support throughout my academic career.

I would also like to take the opportunity to thank the entire faculty of the Michigan Program in Survey Methodology and other researchers at the Institute for Social Research, particularly Jim Lepkowski and Eleanor Singer, for all their guidance and support over the past five years. I appreciate now more than ever at the end of my passage through the program how much effort and time they contribute. I am truly grateful for the interest and patience they show foreign students, such as me, in classroom discussions and when reading papers. It must be as tiring for them to read our papers as for us to write them. Thanks for making it work anyway!

People who are usually not acknowledged but definitely should be are the support staff of the program: Jill Esau, Elisabeth Schneider, Nick Prieur, Rod Perkins, Patsy Gregory, Nancy Oeffner, and Sumi Raj. From the first day I was at

the ISR they were welcoming and indispensable in helping me understand how things work here in the United States. They patiently listened when I had difficulties expressing myself in English and made great efforts to answer whatever questions I had.

The Michigan Program in Survey Methodology with its entire faculty, staff and fellow students and the Institute for Social Research provided a very special environment for studying survey methodology, and it has been a great honor to work with all of them.

Danke für alles!

TABLE OF CONTENTS

Acknowledgements	ii
List of Tables	viii
List of Figures	ix
Abstract	x
Chapter 1 Introduction.....	1
Chapter 2 Do Older Respondents Use Response Strategies Differently to Answer Behavioral Frequency Questions and Does This Change the Accuracy of the Estimate?	6
2.1 Introduction	6
2.2 Answering Behavioral Frequency Questions	8
2.2.1 The Multiple Strategy Perspective.....	8
2.2.2 Factors Influencing the Choice of Response Strategy.....	12
2.2.3 Cognitive Aging, Response Strategy Choice, and Accuracy of Frequency Reports.....	14
2.2.4 Hypotheses	18
2.3 Analyses of Two Studies.....	20
2.3.1 Study 1: Analysis of Secondary Dataset.....	20
2.3.1.1 Data	20
2.3.1.2 Measures and Procedures.....	21
2.3.1.3 Response Strategy Selection.....	24
2.3.2 Study 2: Record Check Study	28
2.3.2.1 Measures and Procedures.....	28
2.3.2.2 Response Strategy Selection.....	33
2.3.2.3 Accuracy of Frequency Estimates.....	43
2.4 Discussion and Conclusion.....	49

Chapter 3 What Influences Panel Attrition Among Older Respondents?.....	51
3.1 Introduction	51
3.2 Predictors of Panel Attrition.....	53
3.2.1 Variable and Situational Circumstances.....	54
3.2.2 Respondent and Household Characteristics	60
3.2.3 Survey Protocol and Interview Situation.....	64
3.2.4 Summary	66
3.3 Disentangling the Influences of Physical Decline, Cognitive Decline and Changes in Social Networks on Panel Attrition in Studies of the Elderly	67
3.3.1 The Health and Retirement Study	67
3.3.2 Measures Used in the Analyses	68
3.3.2.1 Dependent Variable	68
3.3.2.2 Independent Variables	69
3.3.3 Data and Methods	71
3.3.3.1 Analytic Sample	71
3.3.3.2 Method of Analysis and Creation of the Analytic Dataset	72
3.3.4 Analyses.....	75
3.3.4.1 Attrition Rates	75
3.3.4.2 Characteristics of Panel Attritors.....	75
3.3.4.2 Multivariate Analysis of Panel Attrition	80
3.3.5 Limitations of the Analyses.....	87
3.4 Discussion and Conclusion.....	88
Chapter 4 Time-Of-Day of Interview, Changes in Circadian Rhythm And Data Quality: When Is The Best Time To Interview The Elderly?	89
4.1 Introduction	89
4.2 Circadian Rhythms and Aging.....	91
4.3 Data Quality	95
4.4 Methods	96
4.4.1 Data.....	96
4.4.2 Measures.....	97
4.5 Analysis.....	98

4.5.1 “Don’t Know” Responses.....	100
4.5.2 Cognitive Measures.....	101
4.6 Discussion and Conclusion	102
Chapter 5 Conclusion	104
References	106

LIST OF TABLES

Table 2.1: Frequency Questions Used by Conrad et al. (1998).....	21
Table 2.2: Multinomial Models Comparing Different Response Strategies With Categorical Age Variable.....	26
Table 2.3: Type of Frequency Questions Used in the Record Check Study.....	29
Table 2.4: Comparisons of Response Strategies Using Self-Reported Frequency	37
Table 2.5: Comparisons of Response Strategies Using True Frequency Counts	38
Table 2.6: Comparing of Response Strategies Using Self-Reported Frequency and Cognitive Functioning Measure	39
Table 2.7: Comparison of Response Strategies Using True Frequency Count and Cognitive Functioning Measure	40
Table 2.8: Comparison of Response Strategies Using Self-Reported Frequency, Cognitive Measures, and Age.....	41
Table 2.9: Comparison of Response Strategies Using True Frequency, Cognitive Measures, and Age	42
Table 2.10: Logistic Regression Models Predicting Inaccurate Frequency Reports	46
Table 2.11: Logistic Regression Models Predicting Underreports	49
Table 3.1: Structure of Analytic Dataset	73
Table 3.2: Wave 1 Characteristics of Panel Attritors by Attrition Type.....	77
Table 3.3: Goodness-of-Fit Statistics for Multinomial Logistic Regression Models with Different Sets of Variables.....	82
Table 3.4: Final Multinomial Logistic Regression Models	85

LIST OF FIGURES

Figure 2.1: Visual Summary of the Main Response Strategies Presented in the Previous Literature by Conrad et al. (1998), Brown (1995), and Blair and Burton (1987).....	11
Figure 2.2: Structure of the Interview.....	31
Figure 2.3: Scatter Plot of Record Counts and Self-Reported Frequencies	44
Figure 2.4: Scatter Plot of Record Counts and Self-Reported Frequency by Age Group	47
Figure 3.1: Final Cumulative Status of Wave 1 Respondents Across Waves 2 to 7	76

ABSTRACT

Cognitive Aging and Survey Measurement

by

Sonja Irene Ziniel

Chair: Norbert W. Schwarz

Increases in life expectancy and progress in medical treatments will dramatically change the age distribution of Western societies over the next few decades. The U.S. population of people age 65 and older is expected to double from 36 million in 2003 to 72 million in 2030, representing an increase from 12% of the population in 2003 to 20% in 2030. These demographic shifts pose a major challenge for survey methodologists. Normal aging is associated with a decline in

many cognitive abilities that play a prominent role in the processes underlying respondents' answers to survey questions. Hence, normal cognitive aging may be associated with increased difficulties in answering survey questions, resulting in poorer data quality. This dissertation addresses this possibility and explores when and how cognitive aging can introduce survey errors. It consists of three essays.

The first essay addresses whether age-related decline in cognitive functioning increases the likelihood that respondents rely on cognitively less taxing response strategies when answering behavioral frequency questions. The results show that older respondents are more likely to use strategies associated with overreporting, although reliance on these strategies did not produce overall differences in response accuracy. The second essay attempts to disentangle the influence of cognitive aging, decline in physical health and changes in social networks on panel attrition in studies of the elderly. It shows that cognitive aging as well as physical decline increase the likelihood of a proxy-interview compared to a self-interview in the next wave but exert no influence on the likelihood of a refusal. The use of proxy interviews seems to be an important tool to minimize panel attrition bias. The third essay explores how diurnal cycles influence the quality of older respondents' survey answers. In general, older adults show better cognitive performance early rather than late in the day, suggesting that time-of-day of the interview may affect data quality. This hypothesis received no support, nor could the usually obtained diurnal differences in cognitive functioning be observed in the survey context.

CHAPTER 1

INTRODUCTION

Sample surveys are an important source of information about the state of society and people's lives. Moreover, they provide an efficient way to identify and monitor trends over time, from economic developments and health care needs to public opinion. Not surprisingly, the validity of any conclusions drawn from survey statistics depends on the quality of the underlying data. While many general determinants of data quality are well understood (for reviews see Biemer and Lyberg (2003); De Leeuw (1992)), survey methodologists face a social trend that provides new challenges for the collection of reliable and valid survey data: the aging of society.

Increases in life expectancy and progress in medical treatments will dramatically change the age distribution of Western societies over the next few decades. The U.S. population of people age 65 and older is projected to double from 36 million in 2003 to 72 million in 2030, and to increase from 12% in 2003 to 20% of the population over the same time frame (Wan, Velkoff et al. 2005). The U.S. Census Bureau (2006) estimates that the numbers of very old people, aged 85 and older, will nearly triple from about 2.3 million today to about 7.3 million in 2020. When members of the baby boom cohorts reach old age by 2040 the numbers of very old people are projected to be about 15 million. Waite (2004) points out that "if the Census Bureau is correct, by 2050 one American in 20 will be 85 years old or older, compared to one in 100 today." This increase will also lead to a higher proportion of elderly in any given sample that is drawn for a survey representing the adult U.S. population.

Why should survey methodologists be concerned if their samples will include a higher proportion of elderly? Aging is defined as "the sequential or

progressive change in an organism that leads to an increased risk of debility, disease, and death” (“aging” in Encyclopaedia Britannica (Online 2006)). These physiological changes do not only occur in body parts such as organs, cells and bones but also in the brain, affecting cognition. Luckily, it seems that not every dimension of human cognition is affected by the aging process (Park 2000). For example, memory studies testing general knowledge have found that there is no significant age difference in remembering facts of general interest (Hoyer and Verhaeghen 2006). However, psychological research into age-related changes in cognitive functioning has documented systematic declines across the lifespan in processing speed and working memory function (for reviews see Craik (2000); Park (2000)). Park and Gutchess (2000) emphasize that “in situations that demand controlled processing and mental effort, the age-related declines in processing resources will be of great importance and older adults will evidence impairment in their behavior.” These age-related declines are likely to affect numerous survey outcomes, from unit and item nonresponse to the cognitive and communicative processes involved in the question answering process.

In fact, previous research has shown that these basic cognitive abilities play important roles in the question-answering process and negatively affect the quality of survey data (see the contributions in Schwarz, Park et al. (1998); Knäuper, Schwarz et al. (2004)). To date, survey methodological research into these issues has mostly focused on age differences in the emergence of response order and question order effects on attitude questions (Knäuper (1999); for a review see Schwarz, Knäuper et al. (in press)). If one fifth of a population sample drawn in 2030 will be over 65 years old and will likely show problems with participation in the survey and understanding the survey instrument, survey methodologists cannot neglect the possible increase in nonresponse and measurement error, in particular when this increase differentially affects survey estimates for different segments of the population.

People of age 65 and older are a growing proportion of the population and reliable data about this segment is urgently needed for many policy decisions. To ensure that high-quality data are collected from the elderly, survey

methodologists need to adapt their survey designs so that cognitively aged people are willing and able to participate in the survey and to provide data with a minimum amount of measurement error. To facilitate this adaptation, survey methodologists need to understand how age-related cognitive decline interacts with the survey process. Once a general understanding is reached, methods can be developed in order to minimize survey error for this subgroup.

This dissertation contributes to this goal. It consists of three related essays that aim at answering whether and when age-related cognitive decline interacts with the survey process. The three studies presented will help to identify what survey features need to be tailored to the elderly to minimize measurement and nonresponse error due to age-related cognitive decline. In addition, this dissertation uses findings of cognitive aging research to predict as well as explain survey relevant outcomes. Its results inform research in survey methodology through the use of cognitive aging theories and inform the field of cognitive psychology about how useful and predictive laboratory findings are for understanding the influence of cognitive aging in real-life situations. Therefore, this dissertation follows the line of earlier successful undertakings to enhance our understanding of the survey process through the use of other disciplinary knowledge bases (as one example, see Tourangeau, Rips et al. (2000)).

Chapter 2 assesses whether cognitive aging affects the response strategies that older compared to younger respondents use to answer behavioral frequency questions. Some response strategies, especially episodic enumeration, seem to be cognitively more demanding than others, such as general impression strategies. Elderly respondents experiencing cognitive decline should therefore be more likely to choose “easier” strategies. A secondary analysis of data collected by Conrad, Brown et al. (1998) and an analysis of data collected in a record check study are used to test this hypothesis. The availability of true values allows for addressing if older respondents are more likely to overreport behavioral instances since cognitively less demanding strategies have been shown to be associated with overreports. This result would be surprising and important for the field of survey methodology,

which currently assumes that elderly respondents are overall more likely to underreport due to the general decline in memory. This assertion, however, does not account for any differences in response strategy choice between young and old respondents and the type of errors associated with each strategy.

As indicated above, previous research has shown that cognitive aging influences the question answering process and can therefore cause measurement error (for examples see Knäuper (1998); Knäuper, Schwarz et al. (2004)). It can also be expected that cognitive aging influences the decision to participate in surveys especially if the request for participation comes repeatedly, as in longitudinal studies. Chapter 3 examines if cognitive aging influences the likelihood of different types of panel attrition: refusal and proxy-interviews compared to self-interviews in later waves of the Health and Retirement Study (HRS). While previous studies have mostly addressed either physical or cognitive decline the dataset used in this study allows differentiating between the influences of cognitive functioning and physical health decline as well as changes in social network structure on various types of panel attrition. Should panel attrition be selective with respect to key survey variables, the results can provide insights into processes that lead to panel attrition and allow researchers to develop better nonresponse adjustments or develop mechanisms to minimize the occurrence of panel attrition.

Laboratory studies in psychology as well as studies in survey research have shown that the decline in cognitive functioning affects the quality of survey data. Interestingly, studies have also shown that the cognitive functioning decline itself varies during the day. This finding has been confirmed for elderly but not for younger subjects. While elderly seem to be able to compensate for the cognitive decline to some extent in the morning they don't seem to be able to do so in the evening (Yoon, May et al. 1999). Chapter 4 addresses if it is possible to replicate these laboratory findings in a survey context. If data from elderly respondents show more measurement error when collected in the evening, the down time in their circadian rhythm, then survey researchers could improve data quality by conducting survey interviews with elderly in the morning.

To summarize, the purpose of this dissertation is to generate more knowledge about whether and when cognitive aging influences survey errors. Two of its studies illustrate possible effects on measurement error. The third one examines if cognitive aging leads to selective panel attrition and increases, therefore, the likelihood of nonresponse error in certain statistics of interest. All three studies are examples of when cognitive aging could be expected to influence data quality.

CHAPTER 2

DO OLDER RESPONDENTS USE RESPONSE STRATEGIES DIFFERENTLY TO ANSWER BEHAVIORAL FREQUENCY QUESTIONS AND DOES THIS CHANGE THE ACCURACY OF THE ESTIMATE?

2.1 INTRODUCTION

Behavioral frequency questions, such as “How often have you seen your doctor in the past month?”, are very common in surveys. They are used to assess the frequency with which respondents perform behaviors in a given time frame. The knowledge of how often people do something is of special importance to research in the health sciences such as intervention studies. It was recognized long ago that the accuracy of frequency reports is compromised through the fact that these self-reports rely on the accuracy of the memory of respondents (Neter and Waksberg 1964). The forgetting and telescoping of instances of the behavior have been the focus of research that assumed until recently that behavioral frequency reports are based on remembering single episodes of the behavior. It is, however, also possible that respondents use rate-based response strategies or strategies based on general impressions to answer behavioral frequency questions (Menon 1993). Some of these strategies, such as episodic enumeration, are more likely to lead to underreports of the true frequency while others like rate-based and general impression strategies are more likely to produce overreports. Even though characteristics of the behavior, of the question, and the interview setting have been shown to influence the choice of the type of response strategy it seems that an important factor has been

disregarded from research so far: the quality and effectiveness of cognitive functioning. Answering survey questions is a cognitive process and we would expect that the quality of cognitive functioning also has an impact on the quality of the cognitive process. It has been recognized that the quality of cognitive functioning diminishes with normal aging. Based on stereotypes in our society we would expect that older people are more likely to underreport behavioral frequencies since their memory worsens and makes it more likely that episodes are forgotten. This assumes that younger and older respondents are equally likely to choose episodic enumeration, the strategy generally associated with underreports. It seems more likely, however, that the decline in cognitive functioning also influences the choice of response strategy towards elderly using less cognitively demanding response strategies, controlling for all other known influence factors. Response strategies that are among the least demanding are general impression and rate-based strategies. If cognitive aging indeed influences the choice of response strategy we would expect that behavioral frequency reports are more likely to be overreports than underreports because the cognitively less demanding response strategies are also the ones that are associated with overreports. Confirmation of this expectation would also warrant the development of new procedures in surveys that improve the accuracy of behavioral frequency reports since current tools minimizing the inaccuracy of behavioral frequency reports, such as bounded recall (Sudman, Finn et al. 1984), assume the use of episodic enumeration and should therefore be less useful for elderly respondents, a group that will represent a higher percentage of sample respondents in the years to come.

This study will allow us to examine if elderly respondents are more likely to choose cognitively less demanding strategies and are therefore more likely to provide overreports than underreports compared to younger respondents. We will first review in Section 2.2 the multiple strategy perspective on answering behavioral frequency questions, what factors are known to influence the choice of response strategy and how cognitive aging is expected to affect the response strategy choice as well as the accuracy of frequency reports. We then test our

hypotheses by analyzing two studies (Section 2.3): a secondary dataset collected earlier by Conrad et al. (1998) (Section 2.3.1) and data from a record check study (Section 2.3.2) that allows the assessment of accuracy of responses from older compared to younger respondents in addition to determining if cognitive aging influences the choice of response strategy. Our hypothesis that elderly compared to younger respondents are less likely to choose episodic enumeration rather than other strategies is generally confirmed if the age of the respondent is used in the models. The effect can, however, not be replicated if chronological age is replaced with a measure of cognitive functioning. Against our expectations, the analyses with regard to the accuracy of the frequency reports revealed that elderly respondents are not more likely to provide less accurate frequency reports compared to younger respondents. Elderly respondents are also not more likely to overreport frequencies compared to younger respondents even after controlling for the type of response strategy used.

2.2 ANSWERING BEHAVIORAL FREQUENCY QUESTIONS

2.2.1 The Multiple Strategy Perspective

Survey researchers are interested in quantitative facts about respondents' behaviors and hope that respondents answer survey questions as accurately as they can. With regard to behavioral frequency questions survey researchers have believed for a long time that remembering episodes of the behavior and counting them for the reference period is far more accurate than any other way to arrive at frequency reports. In reality, however, it is not uncommon and sometimes even superior with regard to the accuracy of frequency reports that respondents have to use inferences to construct a numeric answer based on incomplete memory simply because they cannot remember distinct episodes of the behavior (Bradburn, Rips et al. 1987). When respondents attempt to answer frequency questions they can only rely on meta-memory and the information which they can

retrieve from their memory. This information from memory might be different from what actually happened because failures may occur during the encoding or retrieval process ((Burton and Blair 1991); (Lee, Brittingham et al. 1999)). Therefore the frequency reports respondents provide as well as the response strategy they use to answer questions is dependent on what information they remember but not necessarily on the information that was originally available when the behavior occurred.

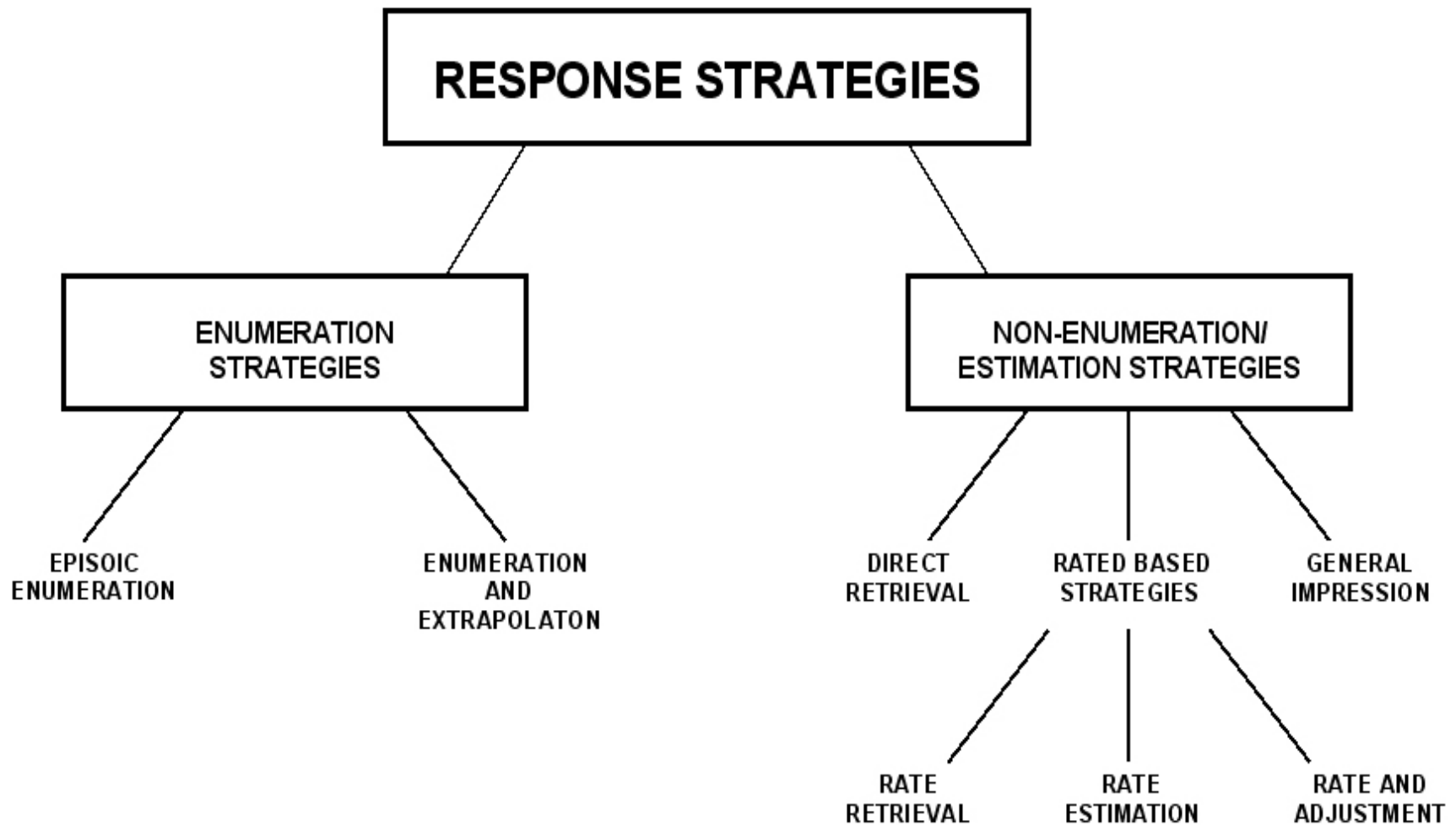
Brown (1995) and Conrad et al. (1998) assessed the various response strategies that respondents use to answer behavioral frequency questions in surveys and classified them into two main groups: strategies that are based on enumeration and those that are not (non-enumeration) (Figure 2.1). Within the enumeration category they distinguish between two strategies. Respondents can, as mentioned before, recall and count each instance for the whole reference period (episodic enumeration) or they can enumerate instances for a sub-period of the reference period and then extrapolate this count to the whole reference period (enumeration and extrapolation).

In contrast, non-enumeration strategies, also called estimation strategies, include a variety of approaches to generate a frequency report. Respondents can simply retrieve the frequency of the behavior (direct retrieval). This usually occurs if a frequency has been generated before and is now remembered. They can also retrieve a rate for the whole reference time period (rate retrieval) or for a sub-period and extrapolate it (rate estimation). In addition, respondents can adjust frequency reports based on direct or rate retrieval because they can remember distinct episodes that do not fit into the retrieved rate schema (rate and adjustment). The last category of cognitive processes combines the strategies that are based on memory assessment (general impression). The frequency of a behavior can be generated based on an assessment of how easy it was to retrieve any behavioral instance (availability by ease) (Tversky and Kahneman 1974), how many instances were retrieved (availability by number) (Pandelaere and Hoorens 2006), how similar the retrieved instance/s is/are with what was asked for (availability by similarity) (Hintzman 1988), and how strongly

represented the behavior is in memory (availability by salience) (Lewandowsky and Smith 1983). Interestingly, in addition to respondents using these response strategies to report their own behavior there has been evidence that these different response strategies are even used when respondents serve as a proxy and are asked to report the behavior of others (Bickart, Phillips et al. 2006).

Cognitive psychologists and survey methodologists realized early on that different types of response strategies are linked to different types of response errors (Burton and Blair 1991): enumeration strategies, on the one hand, are expected to produce a frequency report that is lower than the actual frequency of the behavior due to forgetting of episodes; on the other hand, estimation strategies are linked to frequency reports that are higher than the actual frequency. With the knowledge that different response strategies produce, in general, different response errors and that respondents can use a variety of strategies to arrive at a frequency report survey methodologists embarked on a search for factors that influence response strategy choice. The more information about what type of response error might be prevailing in a given survey estimate provides opportunities to either account for the bias or to develop methods to diminish its existence in the first place. Knowledge about the factors that influence response strategy choice is therefore indispensable in attempts to improve the validity of behavioral frequency reports.

Figure 2.1: Visual Summary of the Main Response Strategies Presented in the Previous Literature by Conrad et al. (1998), Brown (1995), and Blair and Burton (1987).



2.2.2 Factors Influencing the Choice of Response Strategy

Which response strategy is used when answering a frequency question is not usually a conscious choice by respondents. It depends, rather, on a number of factors that influence how behavioral instances are encoded in and retrieved from their memory. Response strategies are dependent on the level of detail that can be retrieved from memory. Enumeration strategies rely on specific information about a behavioral instance, such as context information about the time, the location, the specific steps and the outcome. These strategies, therefore, can only be used if specific information, that distinguishes instances from each other, can be retrieved (Brown 1997). Non-enumeration strategies, however, do not rely on specific information about a behavioral instance or an event. It seems that the information needed to generate a response is restricted to more general information, such as the fact that a behavior has been performed, the steps on which the behavior is based, their sequence and some information about temporal distribution of the behavioral instances. Psychologists call these structured representations of actions scripts and they represent a generic knowledge of actions (Abelson 1981). Generic knowledge of actions is always present and is independent of the amount of instance-specific context information. It is, therefore, possible that non-enumeration strategies can be used even if specific context information is available.

Literature on the influences on the choice of response strategies can be grouped into four areas: characteristics of the behavior, the question, the interview setting and the respondent. The factors that most influence how behavioral instances are encoded and stored in memory are characteristics of the behavior itself. Blair and Burton (1987) showed that the frequency of the behavior has an influence on the choice of response strategy because episodic enumeration is too difficult for high frequency events and because specific context information for behavioral instances is more likely to merge into a general impression about the frequency of the behavior for higher-frequency behaviors

(Strube 1987). The salience of specific behavioral instances, however, has been shown to facilitate the use of episodic enumeration response strategies (Menon 1993). If behavioral instances show some time-related regularity in their occurrence or if behavioral instances are very similar to each other context-specific information is less likely to be encoded or useful in dissemination (Menon 1993).

In addition to characteristics of the behavior influencing what specific information is encoded and remains stored in memory, other factors –as explained above- have been shown to influence which of the response strategies the respondent uses: characteristics of the question, the interview setting and the respondent. It is not fully understood how these factors interact with the level of detail available in memory of behavioral instances. A more theory-based approach is clearly needed to understand when and how these factors influence the choice of response strategy.

Question characteristics that have been shown to influence the choice of response strategy include the length of the reference period (Neter and Waksberg 1964), the wording of the question “how many times” versus “how often” (Blair and Burton 1987), the description of the reference period using the adjectives “past” versus “typical” (Chang and Krosnick 2003), and the use of open-ended versus closed questions ((Schwarz, Hippler et al. 1985); (Menon, Raghubir et al. 1995); (Schwarz 1999)).

Regarding characteristics of the interview setting, face-to-face versus telephone interviews are hypothesized to influence the choice of response strategy for answering behavioral frequency questions by setting a different pace for answering questions ((Sudman and Bradburn 1973); (Burton and Blair 1991)). The training of the interviewers also seems to influence what response strategy the respondent chooses ((Cannell, Miller et al. 1981); (Billiet and Loosveldt 1988)).

The last group of factors includes characteristics of the respondents. Notably, the respondents’ motivation to answer a frequency question as accurately as possible affects what response strategy they employ (Neter and

Waksberg 1964). Cognitive skills, measured through the respondents' level of education, have been shown to influence whether people can distinguish between behavioral events that happen typically or have actually happened (Chang and Krosnick 2003).

2.2.3 Cognitive Aging, Response Strategy Choice, and Accuracy of Frequency Reports

All of the research regarding behavioral frequency questions has assumed that the cognitive processes and memory used to answer frequency questions are comparable across all respondents. While there has been research on how the behavior itself, the respondent, the question, and the interviewer can influence response strategy choice, influences from the quality and effectiveness of cognitive processes themselves have, surprisingly, been disregarded by survey researchers even though the cognitive account of the question answer process itself is solely dependent on such processes (Tourangeau, Rips et al. 2000). Research in cognitive psychology has shown that cognitive processes are not static, neither between people nor within a person. Cognition is influenced, for example, by sleep quality, chronic diseases such as high blood pressure, exercise, and motivation (Spirduso, Poon et al. 2008). Variation in cognition can also be observed within a day and across the life span (Hasher, Goldstein et al. 2005).

One factor widely accepted as having an influence on the quality of cognitive processes is the decline in cognitive capabilities due to normal aging. While there are large inter-individual differences in mental functioning in old age, cognitive psychologists have shown, through testing in laboratory settings, that cognitive decline due to aging begins around age 20 (Hoyer and Verhaeghen 2006). No matter what the national, linguistic or cultural background, age-related decline affects everyone (Crook III, Youngjohn et al. 1992). Both cross-sectional as well as longitudinal studies arrived, in general, at the same conclusions with regard to cognitive decline, even though the magnitude of the effects might vary.

In everyday life situations most of this decline is not noticeable until old age because people can usually compensate for it (Park 1999). As people age, they develop intuitive strategies, such as daily routines, to cope with this decline in cognitive functioning and try, in general, to avoid cognitively demanding situations.

A survey interview, however, is far from being an everyday life situation. Respondents are placed in an unusual situation with an unknown person who asks them a plethora of questions about their life experiences and their opinions. Except for participants in later waves of panel surveys, respondents have usually never experienced a survey interview and, thus, have not developed routines they could use to minimize the cognitive burden of the situation. Survey interviews can sometimes last more than an hour and even a young or middle-aged person can be tired at the end. Given that a survey interview and each individual survey question are highly cognitively demanding situations ((Schwarz 1998); (Tourangeau, Rips et al. 2000)) for which older people have not been able to develop coping mechanisms it seems plausible that cognitive decline due to aging will have a strong impact on the survey measurement process with regard to the reliability as well as the accuracy of the measurements.

Cognitive decline is known to occur in all memory systems except for verbal knowledge (Hoyer and Verhaeghen 2006). Working memory, short- and long-term episodic memory, as well as speed of processing are noticeably affected by aging after age 60. The most robust relationship between aging and impairments in memory systems is observed for episodic memory (Siedlecki 2007). Episodic memory involves conscious recollection of specific event instances together with their temporal, spatial, and other contexts. A decline in episodic memory leads to an impairment in remembering both the context of event instances and also event instances themselves. People either remember that they have done something but can't remember when and where, or they falsely remember event instances. The deficit in temporal, spatial and event context has been hypothesized to be due to less successful encoding of information (Park, Smith et al. 1989) as well as increased difficulty in retrieving

encoded information ((Craig, Govoni et al. 1996); (Craig and McDowd 1987); (Whiting and Smith 1997)). These changes in cognitive abilities due to aging have been shown to influence survey measurements. Older respondents were less subject to question order effects but more to response order effects compared to younger respondents (Knäuper 1998). Frequency scales were more likely to be used by older people compared to younger people when generating a frequency estimate for a non-salient behavior (Knäuper, Schwarz et al. 2004).

While there is no research on how age and its related cognitive decline influence the choice of response strategy, a few studies have examined if old age has an effect on the accuracy of frequency estimates. Auriat (1993), for example, asked subjects to recall and date their moving history and events such as births and deaths and compared it to the national population register in Belgium. Respondents over 50 years seemed to be less accurate in recalling and dating these events compared to younger respondents. In addition to the expectation that elderly people will be less accurate in recalling these events, we would also expect, given stereotypes about the forgetfulness of the elderly, that they are more likely to underestimate the true frequency of a given behavior when relying on episodic memory. Laboratory and real-life studies that have examined the influence of cognitive aging on accuracy, however, show mixed results. Most of the studies seem to support that frequency reports of elderly people are more likely under- than overreports. A study on the number of crimes respondents experienced showed that older respondents are less likely to forward telescope events into the reference period (Schneider 1981). This study hypothesized that memory decay in elderly respondents is responsible for the decline in forward telescoping. In a laboratory study, Mutter and Goedert (1997) presented respondents with target words at different frequencies within lists of other non-target words but only specified to the respondents that some would appear more often and that their memory for these words would be tested at the end. The results of this study indicated that the frequency reports of older respondents deviated more from true frequencies at which the word was presented as that frequency increased. The authors concluded that older respondents

underreported the true frequency more than younger respondents. In contrast, Morwitz (1997) concluded in her study about PC purchases that young people between 18 and 31 years and people over 60 years are more likely than middle-aged people to forward telescope events into the reference period indicating that these age groups also seem to be prone to misdating of remembered events in addition to not remembering events.

Most of the studies described above address the relationship between age of the respondent and the likelihood of telescoping events into the reference period. They make, therefore, an implicit assumption that their subjects will use episodic enumeration to answer the frequency question and hence any overestimate of the true frequency is due to forward telescoping. Overreports can, however, also be caused when subjects use general impression or rate-based response strategies (Sudman, Bradburn et al. 1996). In the same way, underreports do not necessarily have to be episodes that the respondent forgot to count but can also represent errors that were made using non-enumeration strategies (Tourangeau, Rips et al. 2000). It is only possible to characterize reasons for over- and underreports if we know what response strategy the respondent used when answering the frequency question. The information on what response strategy respondents used was, however, not collected in any of these studies. We can, therefore, only speculate what reasons might have contributed to the different results.

Nevertheless, it seems plausible that age and the related cognitive decline have an effect on the inaccuracy of frequency reports through response strategy choice because different types of response strategies are linked to different types of response errors. In general, we would expect the elderly to underreport the frequency of a behavior due to their decline in memory. Underreports are expected to arise from the use of enumeration strategies. Episodic enumeration is considered to be one of the most cognitively demanding response strategies (Blair and Burton 1987) because it relies highly on episodic as well as working memory both of which are greatly affected by cognitive decline (Hoyer and Verhaeghen 2006). Older respondents might be more likely to avoid taxing

response strategies and use less demanding strategies such as rate-based and general impression strategies given that these rely on more general knowledge about the behavior that should not be as dramatically affected by the decline in cognitive functioning. If older respondents are indeed more likely to use cognitively less taxing response strategies we would expect, contrary to first intuitions, that frequency reports are more likely to be biased through overreporting because the less taxing response strategies are associated with overreports ((Brown 1995); (Conrad, Brown et al. 1998)). This study can address if age and related cognitive decline influences what response strategy is chosen and how this influences the accuracy of frequency reports.

2.2.4 Hypotheses

Findings about changes to cognitive processes due to aging lead to a number of predictions about response strategy choice as well as the accuracy of frequency estimates. Strategy selection is likely influenced by the match between the cognitive demands of the response strategies and the cognitive resources of the respondents while controlling for other known influences such as the remembered frequency, regularity and similarity of the behavior. The more taxing a response strategy the less likely it is that respondents with diminished cognitive resources will choose it. Episodic enumeration is the strategy that is most demanding on cognitive functioning (Blair and Burton 1987), followed by rate-based strategies and availability heuristics. There is no theoretical reason to expect that rate-based strategies are cognitively more or less demanding than general impression strategies. The hypothesis regarding response strategy choice that can thus be deduced from our theoretical expectations is:

- H1.1 Compared to younger respondents, older respondents are less likely to use episodic enumeration than general impression and rate-based strategies, controlling for other known influences on the choice of response strategy.

Chronological age is used in these analyses as a proxy indicator for cognitive functioning that has been shown to decline with normal aging. In a newly fielded record check study we also measured the cognitive functioning of the respondent. Our hypothesis regarding the choice of response strategy when including a more objective measure of cognitive functioning is:

H1.2 Compared to highly cognitively functioning respondents, respondents having experienced cognitive decline are less likely to use episodic enumeration than general impression and rate-based strategies, controlling for other known influences on response strategy choice.

Besides using less taxing response strategies more often than young people we would also expect that older respondents have difficulty with correctly reporting the true frequency regardless of what response strategies they use. As reviewed above, older respondents show more problems with the correct encoding and retrieval of behavioral instances. They are also, therefore more likely to provide less accurate frequency reports than younger people. In addition, if our hypotheses that older respondents are more likely to use less taxing response strategies are correct we would expect that their frequency reports are rather overreports of the true frequency, given that those strategies produce, in general, more overreports of the frequency of a behavior. Thus, our second set of hypotheses is as follows:

H2.1 Compared to younger respondents, older respondents will provide less accurate frequency reports.

H2.2 Compared to younger respondents, older respondents are more likely to provide overreports than underreports.

The first hypotheses, H1.1 and H1.2, will be tested in two ways: through the analysis of a secondary data set and the newly fielded record check study. Because testing the two last hypotheses, H2.1 and H2.2, requires true values to which the frequency reports of older and younger respondents can be compared they can only be tested with the data collected from the record check study.

2.3 ANALYSES OF TWO STUDIES

Assessing how age influences the choice of response strategies requires, at minimum, knowledge of the strategy the respondent used to answer the frequency questions, the respondent's age, and self-reported frequency, and similarity, and regularity ratings of the behavior in order to control for the most influential factors on response strategy choice. An already existing dataset from a study by Conrad et al. (1998) fulfilled these requirements and was therefore used as a first step in testing the postulated hypotheses (Study 1). Ideally, the dataset should also contain the true frequency of each behavior for which the respondents provide a frequency report. Since the existing dataset did not contain such measurements we fielded our own study (Study 2). This study was designed as a record check study to allow us to address the impact of cognitive aging on the accuracy of frequency estimates. In the following sections we will describe the methods used to collect the data for these two studies, the measures included in each of them, and the results of the tested hypotheses for the choice of response strategy and the accuracy of frequency reports.

2.3.1 Study 1: Analysis of Secondary Dataset

2.3.1.1 DATA

The secondary data were collected in a study by Conrad et al. (1998). The paper published using these data examined the multiple strategy perspective of

answering behavioral frequency questions. Three interviewers called a national sample of 250 telephone numbers that was stratified by rural and metropolitan areas as well as Census region and population. Information about the classification of respondents within these categories has not been retained in the dataset after the study. One-hundred and six respondents agreed to participate in the short survey which yielded a response rate of 42%. Each interview was recorded with the consent of the respondent.

2.3.1.2 MEASURES AND PROCEDURES

The secondary dataset is based on a study that collected self-reported frequency estimates for a number of behaviors as well as similarity and regularity ratings by the respondent for each of the behaviors. The age of the respondent was also recorded but has not been used in any previous analyses.

Table 2.1: Frequency Questions Used by Conrad et al. (1998)

During the last month, how many times did you...
1) conduct a transaction with an Automated Teller Machine?
2) perform a transaction with a teller in a bank?
3) shop in a grocery store?
4) shop in a department store?
5) purchase gas for your car?
6) pay to have your car repaired?
7) receive subscription magazines by mail?
8) receive catalogues by mail?
9) eat ice cream?
10) eat spicy food?

The interviewer asked the respondents to report the frequency of ten behaviors for the past month (Table 2.1). After each behavioral frequency question, the respondent provided a retrospective think aloud protocol on how they arrived at the answer. The recording of the interviews made it possible to

code the retrospective reports for the response strategies used by respondents to arrive at answers to the behavioral frequency questions. If coders disagreed a consensus was reached on how to code a certain answer. Answers for one of these behaviors, the subscription of magazines, had to be excluded because coders could not agree on the type of strategies that were used. About 14% of all answers were uncodable because the respondent did not provide a sufficient retrospective report and therefore had to be excluded from all analyses (n=130). Overall, coders identified seven response strategies that can be classified into two groups: the strategies used to arrive at non-zero frequency responses and those for zero-frequency responses. Response strategies used for reported zero-frequency estimates were classified into attempted enumeration and rate retrieval. Rate retrieval was the response strategy chosen if it was clear that the respondent had never done the behavior and the rate of the behavior was therefore known. If the respondent had been exhibiting the behavior, then an attempted enumeration strategy was used to determine if the behavior took place in the reference period of one month. Zero-frequency answers to the questions were excluded from all further analyses (n=248). For non-zero frequency estimates coders identified five different strategies that have been described above: Episodic enumeration, rate retrieval, rate estimation, rate and adjustment and general impression. The range of self-reported frequencies extends from one to 100, with a mean of 7.9 and a median of four. The distribution of self-reported frequencies is right skewed (Skewness = 4.33). When answering the frequency questions respondents most often used episodic enumeration (34%, n=181), followed by general impression (22%, n=121), rate retrieval (19%, n=100), rate estimation (14%, n=76), and rate and adjustment (11%, n=60).

After frequency estimates and retrospective protocols were collected the interviewers asked the respondent to rate the regularity and similarity of each reported behavior, on a four-point scale. The respondent was able to classify the regularity of the reported behavior as very irregular (1), somewhat irregular (2), somewhat regular (3), or very regular (4), and the similarity of the instances of the reported behavior as very different (1), somewhat different (2), somewhat

similar (3), or very similar (4). Respondents judged that the majority of their behavior was very regular (47%, n=252), followed by somewhat regular (22%, n=120), somewhat irregular (18%, n=94), and very irregular (13%, n=71). The same pattern can be observed with regard to the similarity of the instances of a given behavior: forty-seven percent were judged to be very similar (n=254), followed by somewhat similar (31%, n=165), somewhat different (13%, n=70), and very different (9%, n=47).

At the end of the interview respondents also provided their gender and age. About 41% of the respondents were male and these respondents provided only 37% of the 538 answers. The mean age of respondents was 39.9 years with a somewhat lower median age of 35 years. The youngest respondent was 18 years old whereas the oldest respondent interviewed was 87 years old. Only seven respondents were 65 years of age or older. Due to the small number it was therefore not possible to compare respondents 65 years and older with younger respondents. We decided to categorize the age variable into four categories with each category representing a quartile of the respondents. The first quartile includes respondents 29 years and younger, the second quartile respondents from age 30 to 36 (included), the third quartile respondents from age 37 to 48 (included) and the fourth quartile represents all respondents 49 years of age and older. To assess the sensitivity of our analyses to this categorization of age we also performed a second set of analyses where age was included as a continuous variable.

After ten answers were dropped because of missing values for age, regularity, or similarity variables, the final answer-level dataset contained data for 105 respondents producing a total of 538 behavioral reports.

2.3.1.3 RESPONSE STRATEGY SELECTION

2.3.1.3.1 Analysis Methods

Multinomial logistic regression models were used to determine if age has an influence on the choice of response strategy controlling for other known influence factors. This type of logistic regression model allows use of categorical variables with three or more categories as the dependent variable. All categories can be compared with each other and evaluated with regard to the independent variables.

We included indicator variables for the different behaviors in our models to control for the fact that some behaviors are inherently more likely to be reported by a certain strategy over others. For example, car repairs, as a very infrequent behavior, are expected to be most often reported using episodic enumeration while grocery shopping, as a more frequent behavior, should likely be reported using other strategies. A X^2 -test shows a significant association between the type of behavior for which a frequency should be reported and the type of response strategy ($X^2(32)=234.45$; $p=0.000$) and confirms that indicator variables for the types of behaviors should be included in the model.

There was concern that the frequency of the behaviors asked is different across younger and older age groups and, therefore, influenced the response strategy choice. An interaction effect between age and the reported frequency was included in the model and proved to have a significant influence on strategy choice. In addition to the different indicators for the type of behavior and the interaction between age and the self-reported frequency, we also included in the model similarity and regularity of the behavior as well as gender of the respondent.

The frequency of car repairs was enumerated by all respondents. Thus, the indicator variable successfully predicted the outcome variable. We decided to

combine this indicator with the indicator for the question about the frequency of eating ice cream. Eating ice cream was very similar to car repairs with regard to the mean frequency (car repairs: 1.58; ice cream: 5.4), the mean rating of the similarity of the behavioral instances (car repairs: 2.48; ice cream: 2.23), and the mean rating of the behavior's regularity (car repairs: 2.48; ice cream: 3.08). We used eating spicy food as a comparison category for the indicator variables because it showed similar, low average ratings on regularity and similarity.

At first, the model included all five categories of the dependent variable. A Wald test addressing whether certain categories of the dependent variable could be combined revealed that the strategies "rate and adjustment" as well as "rate estimation" are not distinguishable with respect to the independent variables in the model and can therefore be combined ($\chi^2(15) = 14.185; p=0.512$). That is, all coefficients except the intercepts associated with these two strategies are zero in the model. The new dependent variable differentiates only between four response strategies: episodic enumeration, rate retrieval, rate estimation and adjustment (further referred to as rate estimation), and general impression. There are six unique pairs of response strategies that can be compared and used in testing our hypotheses and we will evaluate all of these comparisons next.

2.3.1.3.2 Results

Table 2.2 shows the different models using the categorical age variable. We also fit the models with a continuous age variable which did not change our conclusions. The construction of the final model with a categorical and a continuous age variable by adding sets of indicators one at a time. The directional statement of the hypothesis H1.1 allows the use of one-sided tests for the regression coefficients for the predictor variables representing the respondents' age. Significance levels reported throughout this section for the age effects are therefore based on one-sided significance tests.

Table 2.2: Multinomial Models Comparing Different Response Strategies With Categorical Age Variable

		Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
		EE vs GI		RR vs GI		RE vs GI		EE vs RR		EE vs RE		RR vs RE	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Constant		4.07***	1.032	-4.67***	0.923	-2.58*	1.054	8.74***	8.74	6.64***	1.199	-2.10	1.182
Item	ATM	-3.01***	0.712	0.98	0.835	0.77	0.744	-3.99***	0.796	-3.89***	0.766	0.21	0.868
	Teller	0.16	0.523	2.40**	0.739	1.43*	0.686	-2.25**	0.686	-1.28*	0.613	0.97	0.8
	Grocery	-1.57**	0.560	1.82**	0.682	1.52**	0.561	-3.39***	0.614	-3.09***	0.639	0.30	0.724
	Store	-1.69***	0.471	-0.84	0.852	0.03	0.600	-0.86	0.821	-1.72***	0.610	-0.86	0.943
	Gasoline	-1.51*	0.642	2.19**	0.691	1.60*	0.614	-3.70***	0.625	-3.12***	0.631	0.58	0.706
	Catalogue	-0.25	0.663	-0.41	1.024	0.66	0.702	0.16	0.977	-0.91	0.780	-1.07	1.11
	Spicy Food	-0.71	0.506	0.36	0.719	0.50	0.640	-1.07	0.692	-1.20	0.679	-0.14	0.781
Item Characteristics	Regularity	0.33*	0.138	0.90***	0.195	0.57**	0.193	-0.57**	0.189	-0.24	0.190	0.33	0.226
	Similarity	-0.05	0.159	0.09	0.180	0.01	0.171	-0.14	0.191	-0.05***	0.208	0.08	0.188
	Frequency	-0.9***	0.231	-0.03	0.021	-0.02	0.017	-0.87***	0.227	-0.88	0.226	-0.01	0.014
Respondent Characteristics and Age	Male	-0.76*	0.381	-0.12	0.405	-0.09	0.373	-0.63	0.397	-0.67	0.405	-0.03	0.294
	30 to 36 Years	-1.22	1.166	0.93	0.649	0.21	0.592	-2.15*	1.187	-1.43	1.103	0.72	0.538
	37 to 48 Years	0.61	1.099	2.19***	0.721	0.80	0.533	-1.58	1.079	-0.19	1.094	1.39**	0.57
	49 to 87 Years	-2.65***	0.963	0.22	0.561	-0.09	0.519	-2.87***	0.876	-2.57***	0.907	0.31	0.447
Interaction Effects	Frequency[^] 30 to 36 Years	0.43	0.335	-0.02	0.048	0.03	0.027	0.45	0.331	0.40	0.328	-0.05	0.042
	Frequency[^] 37 to 48 Years	0.08	0.329	-0.20*	0.092	0.01	0.023	0.28	0.311	0.07	0.329	-0.21*	0.089
	Frequency[^] 49 to 87 Years	0.58*	0.253	0.002	0.032	-0.01	0.021	0.58*	0.250	0.59*	0.248	0.01	0.029

*** p < 0.001

** p < 0.01

* p < 0.05

EE: Episodic Enumeration

RR: Rate Retrieval

RE: Rate Estimation

GI: General Impression

Our first hypothesis stated that older respondents compared to younger respondents are less likely to use episodic enumeration but more likely to use rate-based or general impression strategies, controlling for other known influences on response strategy choice. Comparisons 1, 4 and 5 in Table 2.2 show how age and other variables affect the likelihood of choosing episodic enumeration versus other strategies. Older people, age 49 to 87, are significantly less likely to use episodic enumeration than general impression strategies (Comparison 1), rate retrieval (Comparison 4), and rate estimation (Comparison 5) compared to respondents 29 years and younger when controlling for other characteristics of the behavior and the respondent. These comparisons between the different response strategies confirm our hypothesis H1.1 that older respondents are more likely to avoid the generally most cognitively taxing strategy, episodic enumeration. With regard to rate-based strategies, it seems that older respondents do not have a preference for choosing rate retrieval over rate estimation and vice versa.

Other interesting effects that our hypotheses did not address were observed in these models. Respondents in the oldest age group are not significantly more or less likely to use rate retrieval than rate estimation (Comparison 6) or general impression (Comparison 2). Older age does also not seem to have any effect on the choice between rate estimation and general impression strategies (Comparison 3). Respondents between age 37 and 48 are significantly more likely than younger respondents to choose rate estimation compared to general impression strategies (Comparison 3). This age group is also significantly more likely to use rate retrieval when answering frequency questions than rate estimation (Comparison 6). This might indicate that this age group, compared to younger people, has a more scheduled time table and is, therefore, generally more likely to use rate-based strategies due to the regularity of their behavior.

Even though the main effects comparing episodic enumeration to other strategies are negative, as hypothesized, the interaction effect for this age group and the self-reported frequency of the behavior is positive, a finding which is

somewhat counterintuitive. As the frequency of a behavior is reported to increase, older people should even be decreasingly likely to use episodic enumeration than they already are. The positive and significant interaction coefficients, however, show the exact opposite pattern. It is not clear why this effect occurs. Unfortunately, it was also not possible to explore it further with this secondary data set.

2.3.2 Study 2: Record Check Study

In addition to the secondary data analysis, we fielded our own study in order to assess how the cognitive aging of the respondent influences the accuracy of frequency estimates. This study was designed as a record check study (Groves 1989) and approved by the University of Michigan Institutional Review Board for the Behavioral Sciences (HUM00003573).

2.3.2.1 MEASURES AND PROCEDURES

The respondents for this study were recruited as a convenience sample through newspaper ads as well as flyers. The study was advertised as an hour-long interview about consumer behavior and the promised reward could be up to \$60 depending on the number of records the respondent was able to provide. The newspaper ad as well as the flyer stated that only unmarried people or those not living with a partner were eligible for the study so that behaviors identified on the records could be uniquely attributed to the respondents. Respondents called a telephone number at the University of Michigan to indicate their willingness to participate in the study. The study staff screened them for their eligibility with regard to their age (20-30 or 65 and older) and their living arrangements (living alone or not with anyone likely to share financial accounts with the respondents, being widowed or divorced). If the subject qualified one of two interviewers then conducted an in-person interview in the homes of the respondents. Thirty-seven

people between the ages of 20 and 30 and 38 people aged 65 and older were interviewed between the months of December 2007 and June 2008. Before the interview began the interviewer conducted a short cognitive test with respondents 65 and older to check that their cognitive status enabled them to be interviewed (Brodaty, Kemp et al. 2004). All interviews were digitally recorded on two recording devices with the consent of the respondent.

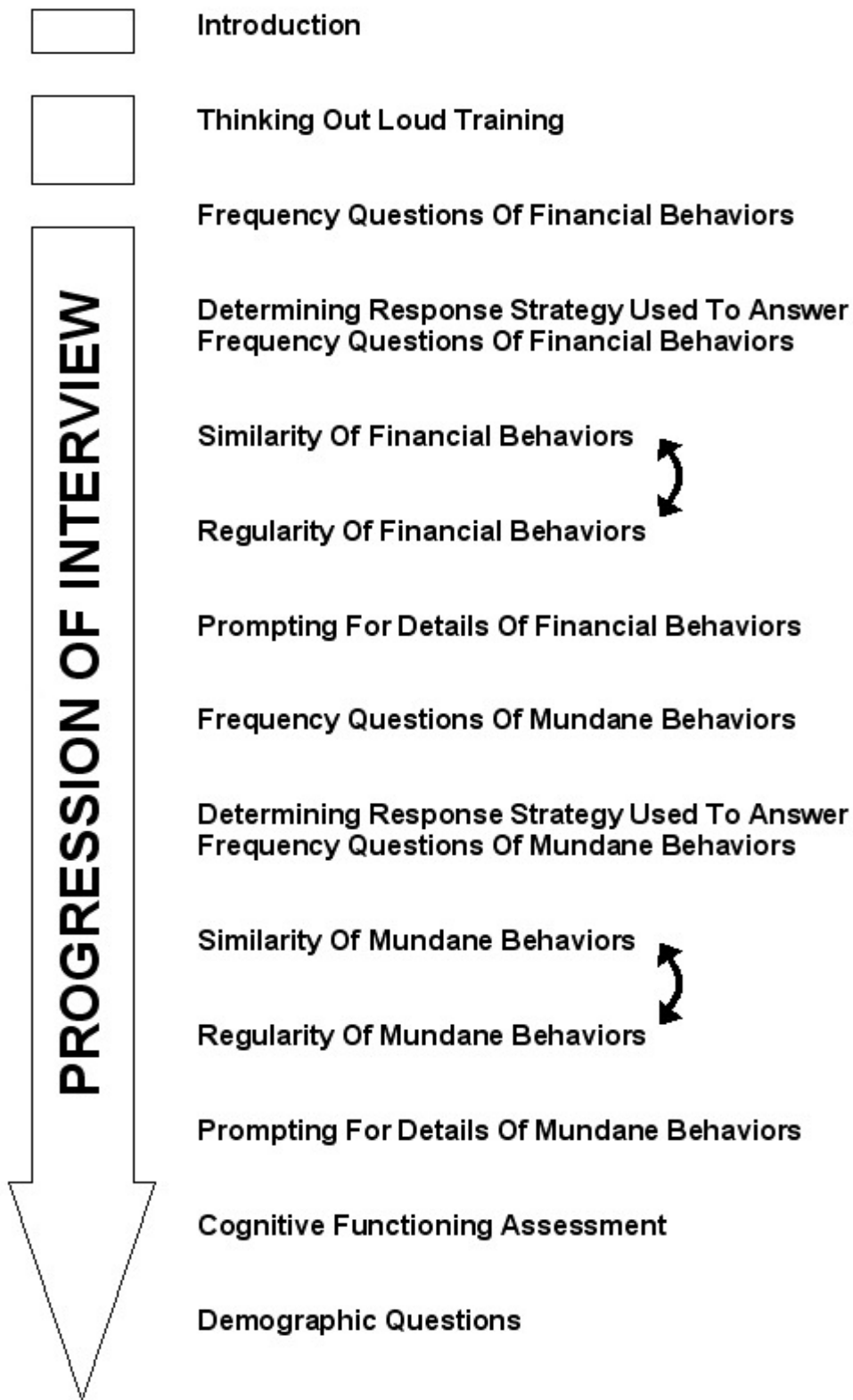
Table 2.3: Type of Frequency Questions Used in the Record Check Study

How many times did you...
1) withdraw money from an ATM in the last 30 days?
2) write paper checks in the last 30 days?
3) deposit a check in person at your bank in the last 30 days?
4) use your debit card(s) in the past 30 days to pay for goods or services?
5) use your credit card(s) in the past 30 days to pay for goods or services?
6) purchase gasoline in the past 30 days?
7) go out to eat in the past 30 days?
8) go grocery shopping in the past 30 days?

The interview included questions concerning the frequencies of financial as well as mundane behaviors with regard to the 30 days prior to the interview (Table 2.3). Respondents were asked and trained at the beginning of the interview to think out loud when answering behavioral frequency questions. The interviewer was instructed to probe if the respondent did not think out loud. After answering each of the frequency questions about financial behaviors respondents were handed a show card that displayed descriptions of the response strategies included in the multiple strategy perspective. The interviewer then replayed the recorded response to each frequency question to remind respondents of their answers and asked respondents to indicate which of the response strategies they used when answering the question. After indicating the response strategies for all the financial questions, respondents also judged the regularity and the similarity of the behavioral instances for each of them. The

regularity and similarity scales were the same as Conrad et al. (1998) used in their study. As a last step in this sequence the interviewer asked respondents to remember and describe for each behavior as many instances with as much detail as possible. This sequence: the interviewer asking the frequency questions, the respondent indicating the response strategy used, rating the regularity and similarity of the instances and providing details about the instances, was then repeated for mundane behaviors. After the sections on frequency questions had been concluded, the interviewer administered a cognitive test that was based on the test used in the Health and Retirement Study (Ofstedal, Fisher et al. 2005). The cognitive test measured the respondent's self-perception about memory through self rating, working memory functioning using a serial 7 subtraction task, the respondent's mental status through backwards counting, date, object and person naming, numeracy through performance of mathematical operation, as well as episodic memory using immediate and delayed word recall. The last questions of the interview were of a demographic nature. The overall structure of the interview is displayed in Figure 2.2. To avoid order effects we created two different sequences of items that were held constant across the questionnaire when asking about frequency, similarity, regularity and episodic details. We also changed the order of similarity and regularity questions: in half of the interviews similarity questions were asked before regularity questions while in the other version the opposite was true. This created four versions of the questionnaire that were randomly assigned to respondents.

Figure 2.2: Structure of the Interview



As in the study by Conrad et al. (1998) the audio recordings of the answers to the frequency questions were also coded. The coders used the more detailed classification used in Study 1 instead of the one that was presented to respondents on the showcards. For secondary data analysis we restructured the data into a person-question dataset with 573 records. More than one fourth of the reports could not be classified into one of the response strategies because there were some respondents that had difficulties with providing a concurrent verbal report (n=109) as in the study by Conrad et al. (1998). Elderly respondents were significantly more likely to provide an uncodable verbal report ($X^2(1) = 6.33$; $p=0.012$). Besides those records we also excluded all reports of zero-frequency behavior (24.61%, n=141). The initial analytic dataset therefore included 76 respondents producing a total of 323 behavioral reports.

The self-reported frequency of the behavior has a mean of 9.31, a median of 5, and a range between 1 and 120. In contrast, the distribution of the true frequency of the behaviors based on respondents' records shows a mean of 10.61, a median of 6, and a range between 0 and 55. It seems that even though the range of the true frequency is smaller than that of the reported frequency, the mean and the median are larger for the distribution of the true frequency. On average, frequency reports are more likely to be under- than overreports.

The response strategy used nearly half of the time is episodic enumeration (48%, n=155), followed by general impression strategies (17%, n=55), rate estimation (16%, n=50), rate retrieval (11%, n=37), and rate and adjustment (8%, n=26). This distribution is similar to that found in the study by Conrad et al. (1998) and might be due to the similarity of the behavioral frequency questions asked in both studies. Respondents rated over a third of the behaviors as very regular (36%, n=115), followed by regular (29%, n=90), irregular (19%, n=59), and very irregular (16%, n=52). Nearly two thirds of the behavioral instances (61%, n=189) were judged to be very similar. Nineteen percent were rated as similar (n=61), followed by dissimilar (12%, n=37), and very dissimilar (8%, n=25).

Less than one third of the respondents are males (29%). Only about 28% of the answers are provided by these male respondents. The younger group of respondents was between 20 and 30 years old with a mean of 22.4 and a median of 21 years of age. The older age group ranges from 65 to 87 years with a mean age of 73.4 and a median of 71. Respondents received a cognitive score between 0 and 1 indicating what percentage of the cognitive items the respondent answered correctly. The average cognitive score for the younger respondents is 0.72 while the one for the older respondents is 0.66 and significantly lower as would be expected ($t(294.22) = 5.913; p=0.000$).

A small number of items were missing for the regularity and similarity ratings. The final dataset for the analyses of the reported frequencies, constructed as a person-question dataset, includes 73 respondents and 307 items. The final analyses using the true frequency based on records include 39 respondents and 97 behavioral reports.

2.3.2.2 RESPONSE STRATEGY SELECTION

2.3.2.2.1 Analysis Methods

Multinomial logistic regression models were used for the data collected in this record check study as in the analyses of the secondary dataset. Our first model testing hypotheses H1.1 and H1.2 includes response strategies as a five category dependent variable and indicators for the type of behavior, its frequency, regularity and similarity ratings as well as gender and either age (young versus old) or the cognitive performance score as predictors. As frequency in this study was available as self-report by the respondents and as true count using the financial records we will test the hypotheses H1.1 and H1.2 using the frequency reports as well as counts.

Due to zero and low cell counts in a cross tabulation between the behavior indicators and the dependent variable a number of these indicators had to be

collapsed to allow the estimation of the model using the self-reported frequencies. Which categories were combined was decided based on similarity with regard to average frequency, regularity, and similarity ratings. In the end, the categories “writing checks”, “depositing checks and “ATM” were collapsed into one category and “debit card” and “credit card” into another. The “credit and debit card” category was chosen as the comparison group for the indicator variables of the behavior. The number of person-questions for the model using true counts was smaller than the one using the self-reported frequencies, 97 compared to 307. The cross tabulation between the behavior type and the dependent variable therefore showed an even larger number of zero or extremely small cell counts. No matter how we collapsed behavioral type indicators together for this model, the estimation of the model was not possible. Thus we left out any behavioral indicator for the models using the true frequency counts.

Once the models could be estimated we tested whether any of the categories of the dependent variable were indistinguishable from the others in the model. As in the model based on the secondary dataset a Wald test confirmed that two of the five categories, rate estimation and rate and adjustment, could be collapsed (self-reported frequencies: $X^2(9)=12.41$; $p=0.191$; true frequency counts: $X^2(6)=2.346$; $p=0.885$). The final dependent variable of this set of analyses was therefore the same as that used when analyzing the secondary data.

Age is only included as a dichotomous variable in the analyses using the record check study comparing respondents 65 years and older to younger respondents between the age of 20 and 30 years. In addition to age and gender this study also collected the education, the number of hours of sleep in the past night, and the employment status (half-time or more compared to less than half-time) of the respondent. Likelihood ratio tests showed that neither education, the number of hours of sleep in the past night nor the employment status significantly improved the fit of the models (self-reported frequencies: LR $X^2(8)=6.22$; $p=0.6222$; true frequency counts: LR $X^2(9)=7.70$; $p=0.5651$). We decided, therefore, to exclude these variables in order to allow the comparison of the

results based on the record check study to those from the secondary data analyses.

2.3.2.2.2 Results

All of the following analyses account for the clustering of observations within each person in the person-question dataset by using Taylor linearization to estimate robust variances. We can use, again, one-sided significant tests for the regression coefficient related to the respondents' age or cognition because the hypotheses H1.1 and H1.2 are directional. Table 2.4 shows the six different unique comparisons of the response strategies with the self-reported frequency of the behavior in the model.

Our first hypothesis predicts that older respondents are less likely to choose episodic enumeration than rate-based estimation or general impression strategies compared to younger respondents (H1.1). Comparisons 1, 4 and 5 in Table 2.4 compare episodic enumeration with other strategies using self-reported frequency. While all these comparisons indicate that older respondents are less likely to use episodic enumeration with regard to the other strategies, the only age indicator yielding statistical significance is the one comparing episodic enumeration with rate estimation. Our first hypothesis is therefore only partially confirmed using the data collected in the record check study.

As the next step, we predicted the response strategy choice using the true frequency counts from the records that respondents provided (Table 2.5). The comparisons of the response strategies show that older respondents are less likely to use episodic enumeration than other response strategies. The coefficients reach statistical significance for comparing episodic enumeration and general impression strategies as well as rate retrieval, but not for episodic enumeration compared to rate-estimation. Hypothesis H1.1 is only partly confirmed. As in the other analyses, this model also shows the counterintuitive effect that elderly respondents are more likely to choose cognitively more demanding strategies as the true frequency of the behavior increases compared

to younger respondents. The effects are significant for episodic enumeration versus general impression strategies and rate estimation.

In all previous analyses we used age as a proxy indicator for age-related cognitive decline. In the record check study, however, we also measured cognitive functioning and can therefore evaluate if it is cognitive decline due to normal aging that is responsible for the differences in choosing response strategies between young and older respondents. Table 2.6 shows the response strategy comparisons from the multinomial logistic regression models using self-reported frequency and the cognitive functioning measure. While the conclusions from the models using the chronological age of the respondent as a proxy for cognitive functioning were in general straightforward the picture seems to be more complicated if age is replaced by a measure of cognitive functioning itself.

Based on our hypothesis H1.2 we would expect that the higher the cognitive functioning the more likely it is that respondents use the cognitively more demanding strategies, such as episodic enumeration. While the direction of the effect is confirmed through the positive coefficient of cognitive functioning comparing the odds of episodic enumeration to general impression (Comparison 1), episodic enumeration is less likely to be chosen as a response strategy compared to rate retrieval (Comparison 4) and rate estimation (Comparison 5) as cognitive functioning increases. None of these coefficients are, however, significant. Our hypothesis H1.2 is therefore not confirmed if chronological age is replaced by a measure of cognitive functioning and self-reported frequency is used as a measure in the multinomial logistic regression model.

Table 2.7 shows the results of the model including the cognitive functioning measure and the true frequency count from the financial records. We can draw the same conclusions as from the model presented in Table 2.6 that includes the cognitive functioning measure and the self-reported frequency by the respondents. Both models using a direct measure of cognitive functioning therefore do not confirm our hypothesis H1.2.

Table 2.4: Comparisons of Response Strategies Using Self-Reported Frequency

		Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
		EE vs GI		RR vs GI		RE vs GI		EE vs RR		EE vs RE		RR vs RE	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Constant		5.17***	1.14	-4.97**	1.836	-1.31	1.003	10.14***	1.831	6.49***	1.165	-3.66	1.873
Item	ATM	-1.64	0.88	2.74	1.404	0.11	0.773	-4.38**	1.5	-1.75*	0.811	2.63	1.406
	Write Checks and Eat Out	0.5	0.583	1.91	1.201	0.21	0.458	-1.41	1.248	0.29	0.537	1.7	1.214
	Deposit Checks and Gasoline	-0.33	0.694	1.96	1.350	0.55	0.689	-2.29	1.362	-0.87	0.627	1.42	1.366
	Groceries	1.33	1.194	4.37***	1.587	2.65*	1.15	-3.04*	1.316	-1.32	0.704	1.72	1.282
Item Characteristics	Regularity	-0.28	0.226	1.18**	0.316	0.03	0.208	-1.45***	0.244	-0.3	0.205	1.15***	0.280
	Similarity	0.05	0.216	0.13	0.295	0.33	0.245	-0.08	0.264	-0.28	0.245	-0.2	0.307
	Frequency	-0.048***	0.089	-0.3*	0.142	0.01	0.017	-0.18	0.163	-0.5***	0.085	-0.31*	0.143
Respondent Characteristics	Old	-1.72	0.888	-0.68	1.028	0.02	0.623	-1.05	0.912	-1.75*	0.782	-0.7	0.907
	Male	-0.45	0.595	-0.17	1.007	0.25	0.546	-0.27	0.856	-0.7	0.526	-0.43	0.907
	Frequency[^] Old	0.23*	0.111	0.22	0.155	-0.03	0.029	0.09	0.18	0.27*	0.107	0.25	0.156

*** p < 0.001

** p < 0.01

* p < 0.05

EE: Episodic Enumeration

RR: Rate Retrieval

RE: Rate Estimation

GI: General Impression

Table 2.5: Comparisons of Response Strategies Using True Frequency Counts

		Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
		EE vs GI		RR vs GI		RE vs GI		EE vs RR		EE vs RE		RR vs RE	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Constant		4.04**	1.386	-2.04	2.290	-0.17	1.281	6.08**	2.170	4.21**	1.451	-1.87	2.341
Item Characteristics	Regularity	-0.51	0.301	0.69	0.539	0.11	0.321	-1.20*	0.511	-0.62	0.329	0.58	0.561
	Similarity	0.05	0.363	0.01	0.565	-0.06	0.391	0.04	0.523	0.12	0.393	0.08	0.584
	True Frequency	-0.25**	0.080	-0.18	0.113	0.01	0.030	-0.07	0.120	-0.26**	0.080	-0.19	0.113
Respondent Characteristics	Old	-2.62**	0.956	-0.12	1.299	-1.24	1.139	-2.51*	1.158	-1.38	1.073	1.12	1.406
	Male	0.15	0.764	1.02	0.957	0.14	0.841	-0.86	0.865	0.01	0.831	0.88	1.025
Interaction Effects	Frequency[^] Old	0.26*	0.096	0.04	0.158	0.06	0.057	0.22	0.163	0.20*	0.095	-0.02	0.160

*** p < 0.001

** p < 0.01

* p < 0.05

EE: Episodic Enumeration

RR: Rate Retrieval

RE: Rate Estimation

GI: General Impression

Table 2.6: Comparing of Response Strategies Using Self-Reported Frequency and Cognitive Functioning Measure

		Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
		EE vs GI		RR vs GI		RE vs GI		EE vs RR		EE vs RE		RR vs RE	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Constant		1.02	3.679	-11.87**	3.859	-7.42**	2.436	12.88**	4.156	8.43**	3.044	-4.45	3.704
Item	ATM	-1.28	0.796	2.62	1.357	0.45	0.811	-3.9*	1.465	-1.74*	0.758	2.16	1.422
	Write Checks and Eat Out	0.68	0.553	1.6	1.237	0.42	0.486	-0.91	1.254	0.27	0.502	1.18	1.248
	Deposit Checks and Gasoline	-0.02	0.711	1.83	1.348	0.98	0.699	-1.85	1.344	-1.0	0.615	0.85	1.372
	Groceries	1.59	1.18	4.3**	1.575	2.81*	1.174	-2.71*	1.29	-1.22	0.644	1.49	1.281
Item Characteristics	Regularity	-0.21	0.224	1.11**	0.329	0.09	0.223	-1.32***	0.25	-0.29	0.192	1.02**	0.289
	Similarity	0.15	0.183	0.42	0.276	0.36	0.221	-0.27	0.272	-0.21	0.215	0.06	0.289
	Frequency	-0.59	0.542	0.57	0.286	0.07	0.061	-1.16	0.592	-0.66	0.539	0.50	0.275
Respondent Characteristics	Old	3.67	5.187	8.79	4.846	8.1**	2.783	-5.12	5.164	-4.43	4.329	0.69	4.218
	Male	-0.35	0.469	-0.48	0.941	0.31	0.473	0.13	0.855	-0.66	0.431	-0.79	0.826
	Frequency^a Old	0.36	0.795	-1.08*	0.505	-0.09	0.091	1.44	0.895	0.44	0.789	-0.99*	0.487

*** p < 0.001

** p < 0.01

* p < 0.05

EE: Episodic Enumeration

RR: Rate Retrieval

RE: Rate Estimation

GI: General Impression

Table 2.7: Comparison of Response Strategies Using True Frequency Count and Cognitive Functioning Measure

		Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
		EE vs GI		RR vs GI		RE vs GI		EE vs RR		EE vs RE		RR vs RE	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Constant		-1.93	3.583	-7.99	5.079	-10.24*	4.831	6.06	4.559	8.3	4.747	2.24	5.91
Item Characteristics	Regularity	-0.38	0.295	0.70	0.515	0.23	0.392	-1.08*	0.478	-0.61	0.325	0.47	0.532
	Similarity	0.23	0.308	0.16	0.497	-0.11	0.328	0.07	0.473	0.34	0.322	0.27	0.512
	True Frequency	-0.32	0.299	0.29	0.494	0.35	0.243	-0.61	0.519	-0.68*	0.341	-0.07	0.521
Respondent Characteristics	Old	5.81	4.650	7.69	6.578	13.24*	6.216	-1.88	5.784	-7.43	5.95	-5.54	7.519
	Male	0.004	0.686	0.95	0.917	0.14	0.776	-0.95	0.807	-0.14	0.716	0.81	-0.949
Interaction Effects	Frequency^^a Old	0.28	0.395	-0.64	0.731	-0.45	-0.454	0.92	0.755	0.73	0.445	-0.18	0.759

*** p < 0.001

** p < 0.01

^a p < 0.05

EE: Episodic Enumeration

RR: Rate Retrieval

RE: Rate Estimation

GI: General Impression

Table 2.8: Comparison of Response Strategies Using Self-Reported Frequency, Cognitive Measures, and Age

		Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
		EE vs GI		RR vs GI		RE vs GI		EE vs RR		EE vs RE		RR vs RE	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Constant		4.33	4.228	-10.50*	4.213	-7.78**	2.607	14.84**	4.587	12.11**	3.391	-2.72	4.103
Item	ATM	-1.38	0.923	2.65	1.388	0.38	0.841	-4.03**	1.497	-1.76*	0.815	2.28	1.423
	Write Checks and Eat Out	0.77	0.570	1.59	1.239	0.41	0.495	-0.82	1.249	0.35	0.528	1.18	1.256
	Deposit Checks and Gasoline Purchases	-0.05	0.725	1.79	1.385	0.87	0.716	-1.85	1.337	-0.92	0.643	0.92	1.380
	Groceries	1.49	1.218	4.23*	1.586	2.74*	1.167	-2.74*	1.291	-1.24	0.693	1.50***	1.281
Item Characteristics	Regularity	-0.22	0.251	1.17**	0.336	0.08	0.224	-1.39***	0.270	-0.3	0.206	1.09	0.285
	Similarity	0.06	0.221	0.19	0.302	0.37	0.249	-0.13	0.288	-0.31	0.251	-0.19	0.331
	Self Reported Frequency	-1.08*	0.526	0.25	0.345	0.07	0.064	-1.34*	0.609	-1.15*	0.521	0.19	0.339
Respondent Characteristics	Male	-0.37	0.557	-0.23	0.956	0.34	0.495	-0.14	0.851	-0.71	0.534	-0.57	0.873
	Cognition	0.92	4.871	8.15	5.004	8.39**	3.011	-7.22	5.219	-7.46	3.805	-0.24	4.349
	Old	-1.62	1.000	-0.18	1.106	0.51	0.649	-1.44	1.004	-2.13**	0.852	-0.69	1.003
Interaction Effects	Cognition* Self Reported Frequency	0.79	0.658	-0.83	0.507	-0.07	0.100	1.63	0.814	0.87	0.653	-0.76	0.488
	Old* Self Reported Frequency	0.28*	0.121	0.17	0.159	-0.03	0.028	0.11	0.189	0.31**	0.115	0.20	0.159

*** p < 0.001 ** p < 0.01 * p < 0.05

EE: Episodic Enumeration

RR: Rate Retrieval

RE: Rate Estimation

GI: General Impression

Table 2.9: Comparison of Response Strategies Using True Frequency, Cognitive Measures, and Age

		Comparison 1		Comparison 2		Comparison 3		Comparison 4		Comparison 5		Comparison 6	
		EE vs GI		RR vs GI		RE vs GI		EE vs RR		EE vs RE		RR vs RE	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Constant		0.05	4.301	-14.26*	5.894	-9.86	5.539	14.31*	4.323	9.91*	4.505	-4.40	6.489
Item Characteristics	Regularity	-0.32	0.314	1.07*	0.426	0.26	0.394	-1.39*	0.375	-0.57	0.341	0.81	0.438
	Similarity	-0.04	0.304	-0.19	0.433	-0.19	0.407	0.15	0.370	0.15	0.418	-0.001	0.446
	True Frequency	-0.40	0.292	0.83	0.646	0.35	0.254	-1.23	0.669	-0.75*	0.325	0.48	0.728
Respondent Characteristics	Male	0.24	0.756	1.30	1.173	0.31	0.835	-1.07	1.078	-0.08	0.923	0.99	1.299
	Cognition	5.30	5.259	15.79*	7.102	13.07	7.154	-10.48	4.768	-7.76	5.900	2.72	7.715
	Old	-1.98*	0.955	1.55	1.223	-0.45	1.185	-3.53**	0.968	-1.54	1.040	1.99	1.134
Interaction Effects	Cognition* True Frequency	0.18	0.368	-1.41	1.047	-0.46	0.340	-10.48	4.768	0.64	0.433	-0.95	1.181
	Old* True Frequency	0.24*	0.090	-0.10	0.112	0.05	0.057	0.35	0.103	0.20*	0.094	-0.15	0.119

*** p < 0.001

** p < 0.01

* p < 0.05

EE: Episodic Enumeration

RR: Rate Retrieval

RE: Rate Estimation

GI: General Impression

In addition to using the respondents' age and the cognitive functioning measure separately in the models we also explored if chronological age has explanatory power when cognitive functioning is already included in the model. Again, this model was fit using self-reported frequency (Table 2.8) and the true frequency counts from the records (Table 2.9). Likelihood-ratio tests showed that age still significantly contributes to the explanation of the variance if the true frequency counts from the records are included (LR $X^2(6)=15.39$; $p=0.0174$) but not if the self-reported frequency is used in the model (LR $X^2(6)=12.33$; $p=0.0551$).

2.3.2.3 ACCURACY OF FREQUENCY ESTIMATES

The record check study design allows addressing the accuracy of the frequency reports provided by old and young respondents by comparing the self-reported frequency of financial behaviors with the true counts based on checking account and credit card statements. Overall, true values are available for 59% ($n=101$) of the 171 financial behavioral reports. Forty-five percent of the behavioral reports ($n=45$) are from older respondents compared to 55% from younger respondents ($n=56$).

2.3.2.3.1 Analysis Methods

As reviewed above, older respondents are expected to have more problems with encoding and retrieval of behavioral instances. They should therefore also be more likely to provide less accurate frequency reports than young people. We will examine this using a logistic regression model predicting inaccurate reports while also controlling for the frequency of the behavior reported and the response strategy used to provide a frequency report.

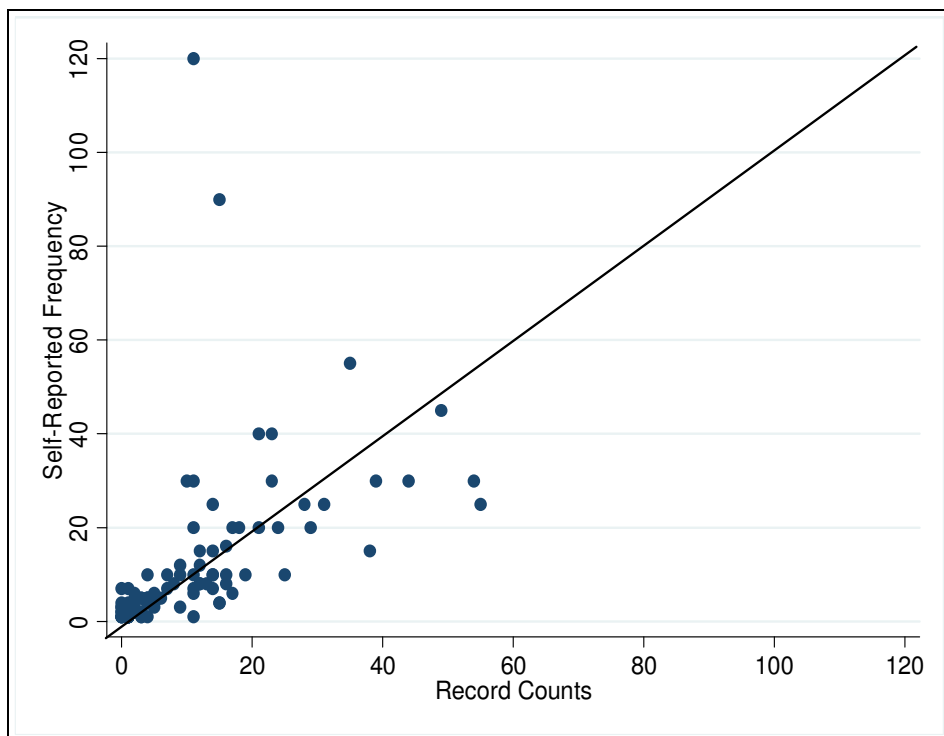
In addition, given that our hypothesis that older respondents are generally more likely to use less taxing response strategies has been confirmed we would

expect that their frequency reports are rather overreports of the true frequency given that those strategies produce more likely overreports of the frequency of a behavior. This will also be tested using a logistic regression model predicting under- compared to overreports. We will also control for the frequency of the behavior and the response strategy used to provide the inaccurate frequency report.

2.3.2.3.2 Results

Overall, twenty-two percent (n=22) of the 101 behavioral reports for which records were available were answered correctly. Figure 2.3 shows a scatter plot of the counts based on the records and the self-reported frequency. The line in the graph represents the location of the dots had there been only accurate behavioral reports. Dots below the line indicate underreports whereas those above the line represent overreports.

Figure 2.3: Scatter Plot of Record Counts and Self-Reported Frequencies



The percentage of inaccurate frequency reports is significantly different across the financial frequency questions ($X^2(4)=13.9859$; $p=0.007$): fifty-nine percent of self-reported frequencies for writing checks are inaccurate, followed by depositing checks with 63%, debit and credit card purchases with each 90% and ATM visits with 94% inaccurately reported frequencies. Accuracy of frequency reports is also significantly different across the various response strategies that respondents used to arrive at frequency reports ($X^2(3)=15.2048$; $p=0.002$). Episodic enumeration was the strategy that showed the smallest amount of inaccuracies (63%), followed by rate retrieval (70%) and rate estimation (90%). All frequency reports that were based on general impression strategies were incorrect.

Younger respondents were able to remember 30% ($n=17$) of the frequencies correctly while older respondents did this only for 11% of the questions ($n=5$). A X^2 -test confirmed that elderly respondents are significantly less likely to answer behavioral frequency questions accurately ($X^2(1)=5.4244$; $p=0.020$). Even though the bivariate relationship between age and correctly answering behavioral frequency questions was significant, logistic regression models additionally including the response strategy used and the frequency of the behavior show a different picture (Table 2.10). Each of the logistic regression models is only based on 76 evaluations of the accuracy of the behavioral reports because all reports based on general impression were excluded predicting inaccuracy perfectly. Both regression models, either including the self-reported frequency (LR $X^2(4)=13.91$; $p=0.0076$; Pseudo $R^2=0.1520$) or the true counts (LR $X^2(4)=13.85$; $p=0.0078$; Pseudo $R^2=0.1515$), show that older respondents are more likely to provide inaccurate frequency reports than younger people. The coefficients, however, do not reach significance. The models also show that higher self-reported frequency and true counts contribute positively, but not significantly, to inaccurate frequency reports as well. While rate estimation is identified in the model using self-reported frequency as the strategy contributing least to inaccuracy, episodic enumeration seems to be the one for the model using true counts. Again, none of the coefficients are significant. Our hypothesis

H2.1 that older respondents are more likely to provide inaccurate reports can only be confirmed in a bivariate context but not when controlling for the frequency of the behavior and the response strategy that was used.

Table 2.10: Logistic Regression Models Predicting Inaccurate Frequency Reports

Self Reported Frequencies			Record Counts		
	Coefficient	SE		Coefficient	SE
Self-Reported Frequency	0.14	0.086	Record Counts	0.10	0.057
Old	0.96	0.620	Old	0.81	0.633
Rate Retrieval	0.15	0.803	Rate Retrieval	0.09	0.814
Rate Estimation	-0.03	1.087	Rate Estimation	0.66	0.924
Constant	-0.24	0.428	Constant	-0.13	0.392

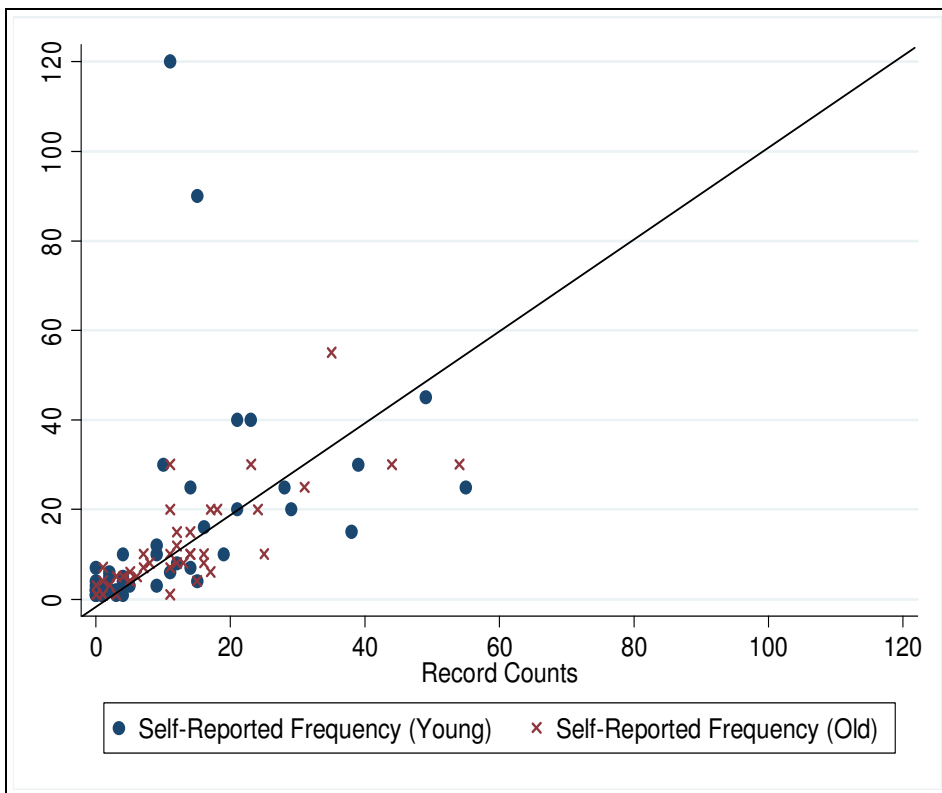
*** p < 0.001 ** p < 0.01 * p < 0.05

Even though our first hypothesis on the difference between older and younger respondents on reporting accurately could not be confirmed, it is still possible that older respondents are more likely to provide overreports than younger people. Overall, 37% of the frequency reports with available true counts were underreports (n=37) while 41% were overreports (n=42). The range of the signed errors (reported frequency – true frequency) goes from -30 to 109. We would expect based on previous findings in the literature that the type of response strategy is significantly associated with over- or underreports. The general pattern is as expected: episodic enumeration shows a smaller percentage of overreports than underreports (48%), followed by rate estimation with 57% overreports, general impression strategies producing 56% overreports and rate retrieval with 57% overreports. The association between the type of response strategy and the type of inaccuracy in frequency reports is, however, not significant ($X^2(3)=0.4449$; $p=0.931$). In contrast, the type of behavioral frequency reports is significantly associated with the type of inaccuracy in frequency reports

($X^2(4)=12.3514$; $p=0.015$): The frequency of writing checks and using credit cards is more likely to be under- than overreported (69% and 63%) while all other types of financial activities are more likely to be over- than underreported. Using ATMs and depositing checks show both 80% overreports followed by 53% of overreports when reporting the use of debit cards.

Figure 2.4 shows a scatter plot of the record counts and the self-reported frequency by the age group of respondents. As before, the line in the graph represents the location of the dots and crosses if all behavioral reports were accurate. Dots and crosses below the line indicate underreports whereas those above the line represent overreports.

Figure 2.4: Scatter Plot of Record Counts and Self-Reported Frequency by Age Group



Younger respondents had a higher percentage of overreports (56%) than underreports (44%) among the behavioral frequency reports that were inaccurate (n=39). Older respondents showed the same number of overreports (50%) as underreports (50%) among the 40 inaccurate frequency reports they provided. A X^2 -test confirmed that the respondents' age group is not significantly associated with over- or underreports ($X^2(1)=0.3259$; $p=0.568$).

The logistic regression models controlling for the frequency of the behavior and the response strategy used confirm that older respondents are neither more nor less likely to overreport than underreport when providing inaccurate frequency reports (Table 2.11). Each of the logistic regression models is based on the 79 inaccurate frequency reports because we excluded all correctly reported behavioral frequencies. Both regression models, either including the self-reported frequency (LR $X^2(5)=1.68$; $p=0.8917$; Pseudo $R^2=0.0154$) or the true counts (LR $X^2(5)=31.08$; $p=0.000$; Pseudo $R^2=0.2846$), show that older respondents compared to younger respondents are not more or less likely to provide underreports than overreports. In the model including the self-reported frequency none of the predictors shows any significant influence on providing under compared to overreports. This model does not fit the data. The model including the record counts shows, however, that a higher true frequency of the behavior is significantly more likely to produce under- compared to overreports. The response strategies rate estimation and general impression are significantly less likely to lead to under- than to overreports than episodic enumeration. Our hypothesis H2.2 that older respondents are more likely to provide overreports than underreports when providing inaccurate frequency reports can not be confirmed.

Table 2.11: Logistic Regression Models Predicting Underreports

Self Reported Frequencies			Record Counts		
	Coefficient	SE		Coefficient	SE
Self-Reported Frequency	-0.01	0.021	Record Counts	0.17***	0.044
Old	0.26	0.469	Old	0.24	0.580
Rate Retrieval	-0.44	0.861	Rate Retrieval	-0.66	1.0004
Rate Estimation	-0.004	0.680	Rate Estimation	-2.67**	1.012
General Impression	-0.17	0.595	General Impression	-1.93*	0.781
Constant	-0.02	0.438	Constant	-1.02	0.520

*** p < 0.001 ** p < 0.01 * p < 0.05

2.4 DISCUSSION AND CONCLUSION

This study explored if age-related decline in cognitive functioning influences the choice of response strategy and the accuracy of the reports when answering behavioral frequency questions. Our first hypothesis H1.1 stated that older respondents should be less likely to choose the cognitively demanding strategy “episodic enumeration” compared to other less demanding strategies such as rate-based estimation and general impression strategies, controlling for other factors known to influence the choice of response strategy. This hypothesis was confirmed in the analysis of data collected by Conrad et al. (1998). Analyses based on data from the record check study using the self-reported frequencies as well as true frequencies based on financial records showed that elderly are less likely to choose episodic enumeration compared to the other strategies. However, not every coefficient reached significance.

When chronological age as a proxy indicator was replaced by a cognitive functioning measure the effects were not as straightforward as before. The hypothesis H1.2 could not be confirmed. A possible reason for this effect could be that the cognitive test administered is not sensitive enough to measure exactly

those cognitive functions that are responsible for the change in response strategy preference.

The record check study allowed addressing whether older respondents are more likely than younger respondents to provide less accurate responses (Hypothesis H2.1). While this hypothesis could be confirmed in a bivariate context, we could not confirm this effect controlling for the frequency of the behavior and the response strategy used to answer the question.

The last hypothesis H2.2 addressed if older respondents are more likely to provide overreports than younger respondents given that they are more likely to use response strategies that are associated with overreports. Our analyses showed that this is also not the case.

To summarize, cognitive aging seems to influence the choice of response strategy when answering behavioral frequency questions. This, however, does not seem to lead to more inaccurate reports compared to young people. In addition, even though older respondents are more likely to choose strategies associated with overreports they are as likely as younger respondents to under- and overreport.

CHAPTER 3

WHAT INFLUENCES PANEL ATTRITION AMONG OLDER RESPONDENTS?

3.1 INTRODUCTION

Longitudinal studies have become indispensable for researchers focusing on intra-individual change or temporal relationships among variables. Besides logistical problems and high costs, panel studies can suffer from nonresponse over the duration of the study. While nonresponse in later waves can cause the loss of statistical power in subgroups due to decreasing sample size, researchers are more concerned about whether the attrition biases their key statistical estimates. If panel attrition is not random with respect to the key estimates, the internal and external validity of the analysis is in danger (Kalton, Kasprzyk et al. 1989). If the causes of panel attrition and statistics of interest are correlated then conclusions based on these might be biased. As in cross-sectional surveys, nonresponse can occur due to noncontact or refusal in any given wave. In addition, panel surveys can lose members for additional reasons because of their longitudinal survey design such as not locating respondents in subsequent waves. Respondents could also have left the population of inference by moving to an institution (Matthews, Chatfield et al. 2004), such as a nursing home or penitentiary, by immigrating to another country or by dying between waves (Jones, Koolman et al. 2006). Permanent loss of a panel member is known as “panel attrition” whereas temporary absence is defined as “wave nonresponse.” While death is a permanent loss, all other types can be of temporary nature. Whether these absences are permanent or temporary depends on the

respondents as well as on the panel design. It is necessary but not sufficient for minimizing attrition that the panel design previews efforts on part of the survey organization to locate, contact, and interview panel members in subsequent waves even though they were nonrespondents in a prior wave.

Techniques to decrease panel attrition have been central to research on panel surveys and in rigorous panel design have kept panel attrition to a minimum. One focus has been on improving the location of panel members that have moved between waves. The success of tracking study members depends on the general method of tracing, the time interval between panel waves, and the possibility of using information sources that could contain information about the whereabouts of the study members (Burgess 1989). In addition, tracking success is also highly dependent on the quality of contacts between the waves (Lepkowski and Couper 2002) as well as collecting contact information from close family or friends (Couper and Ofstedal 2006).

Another area of research has further remediated the loss of panel members due to refusal or inability to be interviewed by allowing proxy respondents to replace respondents for parts of the interview (Tennstedt, Dettling et al. 1992). While some questions, such as attitudes, cannot be answered by proxy respondents, factual questions can be answered with sufficient validity (Nelson, Longstreth et al. 1994) so that the use of proxy respondents allows for at least some data on panel members for more waves and therefore minimizes panel attrition.

Even though survey methodologists have been developing methods to minimize panel attrition, it will always remain a part of longitudinal studies. One subgroup of panel studies seems to be especially sensitive to selective attrition: panel studies of the elderly. Studies about health, retirement and care arrangements of the elderly are more and more common due to the aging of Western societies. Studying intra-individual change for this population subgroup has become especially interesting to researchers and policy makers who examine how the aging of Western society influences economic situations, health care needs and other issues in late life.

In addition to general predictors of attrition in population surveys, longitudinal studies of the elderly have to face influences that are related to the aging process itself: physical decline, increased morbidity, cognitive aging, and changes in the social network of respondents. Statistics based on panel studies of the elderly are more likely than those of the general population to be biased due to panel attrition because the aim of these studies is to measure physical decline, cognitive aging, health-related issues etc. – so factors of attrition and survey measures are highly correlated.

It is therefore extremely important to understand the mechanisms behind panel attrition so that knowledge about potential bias in the statistics of interest can be accumulated and used for nonresponse adjustments or further improvements of strategies to minimize panel attrition. Few longitudinal studies of the elderly, however, have distinguished between different types of panel attrition. Studies also rarely assess the marginal influence on attrition of physical health, and cognitive aging, and changes in the social networks of respondents. This chapter examines the differential influence of these factors on the likelihood of different types of nonresponse: non-contact, refusal, proxy-interview.

3.2 PREDICTORS OF PANEL ATTRITION

Most studies on panel attrition have searched for strong correlates to provide a better understanding of the mechanisms behind nonresponse in longitudinal surveys. Groves and Couper (1998) developed a conceptual framework that differentiates not only between various types of nonresponse but also provides a theoretical view on survey participation. They argued that a thorough understanding of the mechanisms leading to nonresponse is required to develop useful nonresponse adjustments to compensate for nonresponse and strategies to minimize panel attrition. While this framework also applies to the first wave of panel studies, Lepkowski and Couper (2002) have addressed nonresponse in the second wave of panel surveys on a more theoretical basis. They differentiate factors influencing the likelihood of location of respondents, of contact given

successful location and of cooperation given successful contact in the second wave of longitudinal surveys. Generally, factors influencing panel attrition can be categorized into a) situational or variable factor, b) stable characteristics of the respondent or the household and c) characteristics of the survey protocol, the interview situation and experiences in previous interviews. We will review the literature on these characteristics in panel studies of the population and the elderly and utilize it to gain insights into the different types of nonresponse in panel surveys of the elderly: non-location, non-contact, and non-cooperation.

3.2.1 Variable and Situational Circumstances

Variable and situational circumstances that influence panel attrition include lifestyle characteristics, employment status, physical and psychological health status, residential mobility, community involvement, and extent of social interactions (Lepkowski and Couper 2002). Having moved between previous waves, especially long distance moves and moves out of the respondents' state of origin, is associated with a higher likelihood to attrite and being located in later waves ((Zabel 1998); (Behr, Bellgardt et al. 2005); (Lepkowski and Couper 2002)). Moving during the duration of the panel also seems to be problematic with regard to panel studies of the elderly (Matthews, Chatfield et al. 2004). While residential mobility is usually linked to changes in employment and family status, such as the birth of a child or a divorce, older respondents move for different reasons. Declines in health seem to be the main reason for the increased likelihood to move closer to family members, retirement communities or nursing homes. Respondents who own the apartment or house they live in compared to renting it are less likely to move and therefore to drop out ((Zabel 1998); (Harris-Kojetin and Tucker 1998)).

Being employed has a negative relationship with attrition in later waves ((Zabel 1998); (Behr, Bellgardt et al. 2005)). Respondents that are unemployed and looking for a job are more likely to move than those respondents that have a stable employment situation. For studies of the elderly the relationship between

employment status and panel attrition can be expected to be somewhat different given the high percentage of retired respondents in the sample. Kapteyn, Michaud et al. (2006) showed that retired people are more likely to drop out than those who are employed.

Strong social integration and community attachment decrease the likelihood of panel attrition (Lepkowski and Couper 2002). Activities, such as volunteering and a high number of friends and relatives in the community, seem to have a positive effect on remaining in the panel. Being altruistic and more responsive to civic duty has also been shown to increase panel participation. For example, Loosveldt and Carton (2002) showed that the more individualistic people are the more likely they are to drop out.

In addition to the just reviewed situational and variable factors, panel attrition in studies of elderly respondents seems to be influenced by three additional factors unique to this segment of the population: physical health problems, changes in cognitive functioning, and changes in social networks.

Physical health problems: Physical health problems and conditions that have previously been examined as predictors for panel attrition span from acute health problems to more chronic and long-term conditions. Generally, studies conclude that respondents who report to be currently in poor health –however it might be defined- are more likely to drop out of the in future waves than those who exhibit good health. Most panel studies include only self-reported health measures rather than objective measures due to cost constraints and increased burden for the respondent. Poor health in old age is usually not an isolated and short-term phenomenon but rather a sign of general physical deterioration. An indication of a health status rated less than average is therefore predictive of future health problems that might force respondents to move to get better care for health problems or because independent living is no longer feasible (Mihelic and Crimmins 1997). A common measure for health status is to ask respondents to rate their current health on a scale that ranges most often from “poor” to “excellent.” Respondents who drop out rate their health significantly worse

((Sharma, Tobin et al. 1986); (Van Beijsterveldt, Van Boxtel et al. 2002); (Jones, Koolman et al. 2006); (Young, Powers et al. 2006); (Gray, Campanelli et al. 1996); (Goldberg, Chastang et al. 2006); (Anstey and Luszcz 2002); (Matthews, Chatfield et al. 2006); for an exception see (Kautter, Khatutsky et al. 2006)) or report health problems in general (Jones, Koolman et al. 2006). The accumulation of health problems and related use of medication can make respondents less willing to endure the burden of survey interviews or to talk about their declining health (Gray, Campanelli et al. 1996). A higher number of chronic conditions, such as heart attack, stroke, arthritis, cancer etc., decrease the likelihood of staying in the panel ((Albert, Wilson et al. 1975); (Chatfield, Brayne et al. 2005); (Goldberg, Chastang et al. 2006); (Martin, Haren et al. 2007); for an exception see (Matthews, Chatfield et al. 2004)). Van Beijsterveldt et al. (2002) also showed that the higher the number of prescription medications that respondents take the greater the likelihood of dropping out of the panel.

Panel studies of the elderly have also been concerned about the number of disabilities, frailty and difficulties with functional abilities and their impact on participation in future waves. Being disabled and the extent of the disabilities can be a sign of a constant need of medical care and a declining ability to live independently in the future and is related to panel attrition through non-location, the inability and non-cooperation to be interviewed ((Jones, Koolman et al. 2006); (Kempen and van Sonderen 2002); (Tennstedt, Dettling et al. 1992); for an exception see (Krishnan, Murtagh et al. 2004)). Another way of assessing functional abilities in panel studies has been the use of the “Activities of Daily Living (ADL)” and the “Instrumental Activities of Daily Living (IADL)” scales. While the ADL scale focuses more on basic physical functioning abilities such as walking, dressing, eating, getting in and out of bed, using the toilet, and bathing (Fonda and Herzog 2004), the IADL scale assesses functions that enable independent living: preparing meals, grocery shopping, managing money, making phone calls etc. The numbers of these functions that respondents have difficulties with or are not able to do seems to be related to panel attrition

((Mihelic and Crimmins 1997); (Jones, Koolman et al. 2006); for an exception see (Matthews, Chatfield et al. 2004)).

Only a few studies have included more objective criteria about the respondents' health status in assessing the predictors of panel attrition. Goldberg et al. (2006) reported that the number of sick leaves male respondents have taken the less likely they are to respond in the next wave. Physician based examinations of the functional impairment and disability status of respondents revealed that respondents with higher levels of impairment related to the need of assistance or even hospital or institutional care are more likely to refuse (Levin, Katzen et al. 2000) and more likely to dropout in general (Strömgren, Sjogren et al. 2005).

Decline in cognitive functioning: Selective subject attrition due to cognitive decline has long been of concern to longitudinal studies that aim at measuring cognitive functioning change of older adults. Decline in cognitive functioning can have a number of causes. First, it can be normal cognitive aging that everyone experiences to a certain extent in old age. Second, it can be a first preclinical sign of dementia. Finally, late-onset depression has shown to significantly decrease levels of cognitive functioning (Steffens and Potter 2008). It can be difficult to determine prospectively what causes cognitive decline. The studies that assess how cognitive decline influences panel attrition usually do not include measures that could best distinguish what predicts cognitive decline as well as panel attrition. Comparisons of remaining panel members and those that left the panel on various measures of cognitive functioning revealed that cognitively impaired panel members are more likely to drop out even when controlling for chronological age. In contrast to measures of physical health cognition is largely assessed through the administration of cognitive functioning tests that have been shown to measure certain cognitive dimensions and therefore provides a more objective assessment of the cognitive status of respondents. A common cognitive test that assesses general cognitive functioning is the Mini-Mental State Exam (MMSE). The large number of studies that included the MMSE in their interview agreed that respondents with higher scores indicating better cognitive functioning

are less likely to leave the panel in future waves than those with lower scores ((Anstey and Luszcz 2002); (Gerstorf, Herlitz et al. 2006); (Matthews, Chatfield et al. 2006); (Chatfield, Brayne et al. 2005); (Matthews, Chatfield et al. 2004); for an exception see (Van Beijsterveldt, Van Boxtel et al. 2002)). The most common reason that respondents with lower MMSE scores do not stay in the panel seems to be the development of dementia which makes people unable to be interviewed and more likely to refuse (Chatfield, Brayne et al. 2005).

Other measures used to assess cognitive functioning are different tests of general intelligence and intellectual ability. As for the MMSE score selective subject attrition is confirmed when comparing the intelligence score of those staying in the panel with those who have left: the intellectually superior respondents remain ((Gray, Campanelli et al. 1996); (Sullivan and Corkin 1984); (Siegler and Botwinick 1979); (Rabbitt, Lunn et al. 2006); (Ritchie and Tuokko 2007)). Cognitive tests that also seem to predict panel attrition are immediate and delayed recall tests assessing the respondents' memory. Kennison and Zelinski (2005) as well as Sliwinski et al. (2003) conclude that bad performance on these memory tests identifies the respondents that are less likely to participate in the next wave of the panel. A clear relationship between memory performance and panel attrition due to refusal has also been established ((Van Beijsterveldt, Van Boxtel et al. 2002); (Levin, Katzen et al. 2000)). While the measures described are those commonly found in studies addressing the influence of cognitive aging on panel attrition there are a large number of less used measures that reveal the same relationship between the decline in cognition in older adults and the likelihood of panel attrition: measures of executive functioning (Levin, Katzen et al. 2000), processing speed (Anstey and Luszcz 2002); (Van Beijsterveldt, Van Boxtel et al. 2002)), memory of symbols and pictures (Anstey and Luszcz 2002), general cognitive impairment (Chatfield, Brayne et al. 2005), etc.

Research has shown that cognitive aging is a natural process that no one can escape (Raz 2000). The degree of cognitive aging, however, can be very different between individuals. Besides preclinical symptoms of dementia late-

onset depression has been shown to negatively influence cognitive functioning (Steffens and Potter 2008). In addition, several studies that have looked at the effect of depression on panel attrition hypothesizing that depression can interfere with respondents' motivation as well as willingness to be interviewed in future waves ((Matthews, Chatfield et al. 2004); (Mirowsky and Reynolds 2000); (Levin, Katzen et al. 2000)). The conclusions that these studies draw are mixed. Mirowsky and Reynolds (2000) stated that survey participation is negatively influenced by high levels of depression. This result was confirmed by Levin et al. (2000) who controlled for both cognitive aging and age itself. Matthews et al. (2004) did not find that depression increased the likelihood of refusal but stated that it could be related to the likelihood of moving and non-contact.

Changes in social networks: Toh and Yu (1996) found in their study that the number of persons in the household had the most influence on whether or not a respondent remained in the panel. They hypothesized that married people are less likely to drop out due to their more stable family dynamics. Even if the respondent is in the need of care the likelihood of moving and non-contact is minimized by the presence of household members who also can be easily recruited as proxy respondents if necessary. Living alone has also shown to increase the likelihood of refusal (Matthews, Chatfield et al. 2004). Respondents with social ties should therefore be in general more likely to stay in the panel. In a study of multiple generations weaker emotional ties of elderly with their children were predictive of drop out of the elderly respondents in later waves (Feng, Silverstein et al. 2006). It seems that a strong social network increases the likelihood that elderly stay active and interested. Old age is, however, characterized by weakened ties to society. Husbands and wives, siblings and friends of the respondents die, minimizing their social ties. Having children and grandchildren could alleviate the extent to which older respondents feel isolated from society.

There is significant evidence that age-related change in physical health and cognitive functioning as well as social networks increases the likelihood of

panel attrition. While most of these studies look only at one of these three mechanisms at a time the current study aims to determine the differential influence each of these factors has on the different types of panel attrition. This allows researchers collecting panel data as well as those that use panel data in analyses to get a better understanding where bias in key statistics can be expected and to provide more information for improved nonresponse adjustments. Before we assess the differential influence of these three factors in the Health and Retirement Study we will briefly review response, household, and survey characteristics that have been shown to be important correlates of panel attrition. We will include them as control variables in our analyses.

3.2.2 Respondent and Household Characteristics

Socio-demographic characteristics are commonly used as predictors of panel attrition. In general, household and respondent characteristics are included to approximate more specific factors that are associated with panel attrition. Socio-demographic characteristics are repeatedly found to be significant predictors of panel attrition in general as well as for specific types of nonresponse (Lepkowski and Couper 2002). More stable respondent and household characteristics that are reviewed in the following sections are the respondent's age, gender, race, ethnicity, marital status, education, economic status, geographical location and degree of urbanicity.

Findings with regard to the respondents' age are fairly consistent in the literature. Older people seem to be generally more likely to participate in later waves of the panel compared to middle-aged people ((Behr, Bellgardt et al. 2005); (Hart, Rennison et al. 2005)). Respondents who are 65 years old or older are most likely retired and might have more discretionary time at their hands to do surveys. They seem also to be more likely to be at home at different times during the day which increases the likelihood of contact. While Harris-Kojetin and Tucker (1998) did not directly predict attrition but rather nonresponse of any kind in later waves they also found that nonresponse is less likely to occur the older

the respondents are. Elderly are hypothesized to have more sense of civic duty and feel that it is important to support research in a continuous manner. In addition, older respondents that feel alone can look forward to survey interviews as a welcomed change to their daily life routines. In contrast, longitudinal studies of the elderly have found that chronological age is one of the most (Ghisletta and Spini 2004) powerful predictors for panel attrition ((Ghisletta and Spini 2004); (Sliwinski, Hofer et al. 2003); (Kempen and van Sonderen 2002); (Sharma, Tobin et al. 1982); (Ritchie and Tuokko 2007); (Toh and Hu 1996); (Van Beijsterveldt, Van Boxtel et al. 2002); (Mihelic and Crimmins 1997); (Chatfield, Brayne et al. 2005); (Matthews, Chatfield et al. 2004)). The reason for these opposite conclusions with regard to the relationship between age and panel attrition seems to be that age is used in these studies as a proxy variable for different characteristics of the respondent. In studies of the elderly, described above, age is strongly associated with declines in physical or mental health as well as death. This association can be hypothesized to be weaker or even missing in studies of the general population where the majority of the sample is less than 65 years old.

Nearly all studies agree that males are more likely to drop out of panel studies ((Behr, Bellgardt et al. 2005); (Hart, Rennison et al. 2005); (Harris-Kojetin and Tucker 1998); (Watson 2003); (Gray, Campanelli et al. 1996); for an exception see (Pickery, Loosveldt et al. 2001)). Most panel studies of the elderly also find, like panel studies of the population, that females are less likely to drop out ((Ghisletta and Spini 2004); (Krishnan, Murtagh et al. 2004); (Van den Berg, Van der Velden et al. 2007); (Mirowsky and Reynolds 2000); for no effect see (Mihelic and Crimmins 1997)).

African Americans or other races compared to Caucasians are, in all studies reviewed for this research, significantly more likely to drop out ((Zabel 1998); (Hart, Rennison et al. 2005); (Harris-Kojetin and Tucker 1998); (Gray, Campanelli et al. 1996); (Jones, Koolman et al. 2006)). The same phenomenon can be observed for Hispanics ((Hart, Rennison et al. 2005); (Harris-Kojetin and Tucker 1998)). This might be linked to the mobility of these subgroups. The effects of race and ethnicity found in general population studies are generally

confirmed in studies of the elderly. Krishnan et al. (2004) showed that non-Caucasians are more likely to drop out. Kapteyn et al. (2006) found that female Hispanics are more likely to be wave-nonrespondents and attrite compared to non-Hispanic females and that African American males also have higher chances of wave-nonresponse and attrition. A different study, however, showed that neither race nor ethnicity had an effect on attrition (Mihelic and Crimmins 1997).

With regard to marital status of the respondents, being married is significantly associated with a higher likelihood of staying in the panel ((Behr, Bellgardt et al. 2005); (Hart, Rennison et al. 2005); (Harris-Kojetin and Tucker 1998); (Watson 2003)). Married people are more likely to be socially integrated and more stable with regard to their living situation. Single or never married people are generally more likely to leave the panel in future waves ((Behr, Bellgardt et al. 2005); (Hart, Rennison et al. 2005); (Harris-Kojetin and Tucker 1998); (Gray, Campanelli et al. 1996); (Jones, Koolman et al. 2006)) and some studies find that the same is true for widowed and divorced respondents ((Behr, Bellgardt et al. 2005); (Hart, Rennison et al. 2005); (Gray, Campanelli et al. 1996); (Watson 2003)). These latter subgroups are more mobile and might also have fewer ties to the community they currently live in. Results with regard to the association between marital status and panel attrition are more mixed in studies of the elderly than in those of the general population. Most of the studies found that single respondents are more likely to attrite than married people ((Van den Berg, Van der Velden et al. 2007); (Kapteyn, Michaud et al. 2006)). Others found that in addition to single people being less likely to participate in later waves of the panel separated, widowed and divorced respondents are also more likely than married respondents to leave the panel ((Kautter, Khatutsky et al. 2006); (Young, Powers et al. 2006)). Goldberg et al (2006) and Martin et al. (2007), however, found no effect of marital status on panel attrition.

The presence of children in the household seems to decrease the likelihood of attrition (Zabel 1998). Children attending school can provide more social integration as well as a more stable living situation. Instead of the presence of children in the household, studies of the elderly use household size

as an indicator for establishing if the respondents have social ties within the households. The general finding is that respondents living alone are more likely to drop out than respondents that live with one or more adults in the household ((Mihelic and Crimmins 1997); (Kautter, Khatutsky et al. 2006); (Kapteyn, Michaud et al. 2006)).

All studies reviewed here have shown that education has a significant effect on panel attrition. Higher education, usually a college degree or higher, increases the likelihood of participation in later waves of the panel for some studies ((Hart, Rennison et al. 2005); (Harris-Kojetin and Tucker 1998); (Fitzgerald, Gottschalk et al. 1998); (Watson 2003); (Jones, Koolman et al. 2006)). As in longitudinal studies of the general population, higher education has found to be related to staying in the panel in later waves ((Kapteyn, Michaud et al. 2006); (Krishnan, Murtagh et al. 2004); (Mihelic and Crimmins 1997); (Kautter, Khatutsky et al. 2006); (Young, Powers et al. 2006); (Goldberg, Chastang et al. 2006); (Van den Berg, Van der Velden et al. 2007); (Levin, Katzen et al. 2000); (Kempen and van Sonderen 2002); (Sharma, Tobin et al. 1982); (Sullivan and Corkin 1984); (Ritchie and Tuokko 2007); for an exception see (Martin, Haren et al. 2007)).

Results on the association between the respondents' economic status and panel attrition are mixed. Fitzgerald et al. (1998), Gray et al. (1996) and Watson (2003) show that those who dropped out later in the panel were usually at the lower end of the socioeconomic distribution. Behr et al. (2005) found either a significant relationship between economic status and panel attrition and/or a positive relationship between unemployment and panel attrition for different European countries. Gray et al. (1996) confirmed this finding showing that retired respondents are more likely to drop out of the panel. Watson (2003), however, found the exact opposite. While some studies do not find an effect of employment status or income on panel attrition in studies of the elderly (Martin, Haren et al. 2007) most studies confirm the relationship observed in longitudinal studies of the general population. Respondents of low socio-economic status or with low income are more likely to drop out than people with high income ((Ghisletta and

Spini 2004); (Young, Powers et al. 2006)). Kautter et al. (2006) found this effect not only for low-income respondents but also for those with high income compared to respondents with an average income. Van den Berg et al. (Van den Berg, Van der Velden et al. 2007) found that respondents who have a paid job are more likely to stay in later waves than those with no job. Retired respondents seem to be more likely to leave the panel ((Kapteyn, Michaud et al. 2006); (Goldberg, Chastang et al. 2006)) than people still in the workforce. Unskilled workers also have a higher likelihood of attrition than skilled workers or managers (Goldberg, Chastang et al. 2006).

All analyses of this study will include indicators of respondent and household characteristics as control variables.

3.2.3 Survey Protocol and Interview Situation

It has been shown in the literature on survey nonresponse in cross-sectional surveys that the survey protocol, the interview situation as well as the interviewer can influence participation (Groves and Couper 1998). A number of survey protocol characteristics have also been assessed in panel surveys.

With regard to panel surveys of the elderly Levin et al. (2000) point out that keeping older respondents in a panel might need more flexibility, empathy and persistence than with the normal population. In addition elderly respondents that have one or more chronic illnesses might require special types of incentives and services to reduce nonresponse due to nonlocation, noncontact or non-cooperation and keep them as respondents in the next wave.

Interviewing frail or disabled elderly might also add another challenge to panel surveys because contact is usually made with a healthier spouse or a caretaker. While these gatekeepers seem to reason that the selected respondents are too ill or cognitively impaired they seem to be quite willing to be a proxy-interviewer if they feel knowledgeable enough about the respondent (Tennstedt, Dettling et al. 1992). Keeping close contact to the proxy-respondents

in those cases between waves seemed to help decrease the likelihood of refusal in future waves.

In addition, due to more frequent temporary illnesses and possibly required stays in hospitals Tennstedt et al. (1992) recommends that the field period is extended beyond the normal time line for general population surveys.

Possible problems with vision and hearing and the need for more and repeated explanation of the survey process and the questions has triggered several studies on how the mode of data collection can influence panel attrition. As in cross-sectional surveys respondents are generally more likely to participate in future waves if the data collection mode is face-to-face (Hart, Rennison et al. 2005). The impact of the mode of data collection with regard to panel attrition has been mixed. Earlier studies have shown that using the telephone for studies with the elderly results in higher refusal rates compared to face-to-face ((Herzog and Rodgers 1988); (Herzog, Rodgers et al. 1983); for no effect see (Zabel 1998)). The reason for higher refusal rates in telephone studies is most likely no disinterest but the difficulty of answering the survey request over the telephone due to hearing difficulties. Tennstedt et al. (1992), however, showed that the telephone mode can be used successfully when converting refusals in studies of the elderly.

Assigning the same interviewer to respondents across waves of the panel decreases significantly the likelihood of attrition in panel surveys of the general population and the elderly ((Zabel 1998); (Behr, Bellgardt et al. 2005); (Hill and Willis 2001)). Interview length has been shown to be negatively related to attrition in later waves ((Zabel 1998); (Hart, Rennison et al. 2005)). Even though intuitively a longer interview seems to be more burdensome it can also be a sign that respondents and interviewers enjoyed the interview and the interactions. In general, the respondents' experiences have a big influence on participation in the next wave. Martin et al. (2007) showed that respondents who complained about the burden of the interview and told the interviewer that they didn't want to participate anymore were significantly more likely to attrite in the next wave. This seems to be similar in panel studies of the elderly (Steinhauser, Clipp et al.

2006). Interestingly, older respondents seem to suffer less from participation fatigue and are more likely to stay in the panel even if the number of data collection points is higher (Toh and Hu 1996). The influence of the first or previous interview experience has also been confirmed by Pickery et al. (2001): The interviewer of the first wave had a higher influence than the interviewer in the second wave.

Unfortunately, the design of the Health and Retirement Study does not allow for easily examining the influences of these factors in panel studies of the elderly. Some of these items have been measured in some waves but not all, and other information is not readily available for data analysis. The only factor that influence can be controlled for is the mode of data collection. Because the assignment of the mode of data collection to respondents is not random but dependent on the wave number, the respondents' choice and their cognitive status, caution should be used when interpreting the coefficients in our models.

3.2.4 Summary

This study will close several gaps that exist in the literature on attrition in panel studies of the elderly. First, it examines the differential influence of decline in physical health, cognitive aging, and changes in social networks by using multivariate analyses that control for the associations between the factors contributing to panel attrition. As reviewed above, earlier studies usually looked at each of these factors in isolation and it is not clear if they all contribute equally to the likelihood to drop out or if one set of factors will be more important than others in predicting panel attrition. Second, by using multinomial logistic regression models it is possible to examine if there are differential influences of these factors on the different types of panel attrition (death, non-contact, refusal, proxy-interview). Most previous studies that have examined attrition in studies of the elderly generally did not differentiate between these different types. From a survey methodologist perspective, however, this information is extremely important to adapt the survey design and process to minimize attrition. It also

fosters understanding at what point in the survey process possible nonresponse bias is introduced in key statistics and can help to develop better nonresponse adjustments. Third, this study includes data from seven waves of the Health and Retirement Study and provides a unique opportunity to gain insight into the mechanisms of panel attrition in later waves of the panel.

3.3 DISENTANGLING THE INFLUENCES OF PHYSICAL DECLINE, COGNITIVE DECLINE AND CHANGES IN SOCIAL NETWORKS ON PANEL ATTRITION IN STUDIES OF THE ELDERLY

3.3.1 The Health and Retirement Study

The Health and Retirement Study (HRS) was launched in 1992 as a longitudinal study on the health, work and retirement, income and wealth, family characteristics and intergenerational transfers of older adults in the United States. Respondents are interviewed every two years. Respondents within a household are not selected so that if a household contains a married couple both persons are interviewed. Today the HRS follows over 20,000 men and women 50 years and older. The participants are part of one of five birth cohorts included in the study up to now: the cohort that was originally included in the Aging and Health Dynamics study (AHEAD) which was later merged into the Health and Retirement Study and represents people born before 1923; the original cohort from HRS including men and women born from 1931 until 1941; the Children of the Depression Age from 1923 to 1930 (CODA), the War Babies from 1942 to 1947 (WB), and the Early Baby Boomers born between 1948 and 1953 (EBB). Due to the availability of key predictors this study will only include two of the five cohorts, the AHEAD and CODA cohorts. A number of indicators were continuously measured only on respondents 65 years and older to minimize the survey burden as much as possible and because changes in these measurements were not observed for younger participants. Respondents of the

AHEAD and CODA cohort were both 65 years and older when their cohort was interviewed the first time and have been administered the key measurements in each wave. The population of inference for this study contains therefore all men and women that were born in and before 1930 that were eligible to be interviewed according to AHEAD/HRS criteria at the time of the first wave their cohort was interviewed.

3.3.2 Measures Used in the Analyses

3.3.2.1 DEPENDENT VARIABLE

The final response status of a respondent in any given wave was classified into five categories: self-interview, proxy-interview, refusal (by the respondent or a proxy), non-contact, and death. In the analyses that follow a respondent is assumed to have left the panel when any status other than self-interview is assigned or in other words when the pattern of self-response is broken or the observation is censored. We realize that this simplifies reality because the Health and Retirement Study will attempt to interview the respondents again even though they might not have been contacted or they refused or they were replaced by a proxy-respondent in the previous wave.

Even though it is theoretically possible to explore the influence of the explanatory variables on all of the categories of the dependent variable our analyses will focus on those for which the respondent is at least somewhat involved in making a decision with regard to survey participation. The first comparison is proxy-interview versus interview, where respondents prefer or require that other persons answer for them. This usually happens due to health reasons. The second comparison looks at refusal versus interview. The last comparison is proxy-interview versus refusal where respondents decline to be interviewed but allow that a proxy-interview is conducted and the proxy-respondent is also willing to be interviewed instead of the respondent. Only a

small percentage of refusals originate from designated proxies refusing to be interviewed.

3.3.2.2 INDEPENDENT VARIABLES

The analyses include four sets of independent variables: general background variables, and variables measuring cognitive functioning, physical health and extent of social network. The general background variables consist of the respondent's gender, race (African American versus Caucasian and other races), ethnicity (Hispanic versus non-Hispanic), age and education. Education is measured as the number of years the respondent attended school. We also included the number of respondents in the household (one or two), and an indicator for the data collection mode (face-to-face versus telephone) and study cohort membership (AHEAD versus CODA) to control for differences in essential survey conditions between data collection modes as well as respondent cohorts.

Cognitive functioning indicators: Cognitive functioning of the respondent was measured through subjective judgments as well as objective measures. The respondents were asked to rate their current memory on a 5-point scale (Poor, Fair, Good, Very good, Excellent) and how their current memory compares to one year ago/the last interview on a 3-point scale (Worse, About the same, Better). The objective measures that were included in HRS covered cognitive functions such as memory, working memory, and mental status. Memory was assessed through the number of correct words respondents remembered immediately (immediate recall) and five minutes (delayed recall) after the interviewer read ten words. Scores could range from 0, with no word correctly remembered, to 10, with every word correctly remembered. The Serial 7's test, a subsequent subtraction task, was used to measure the quality of the respondent's working memory. The values could range from 0 to 5, where higher values indicate a better functioning of working memory. The mental status of the respondent was estimated through a composite score, TICS, that included measures such as backwards count from 20, date naming, object naming and

president and vice president naming. Respondents could receive between 0 and 10 points with higher points indicating a higher mental status.

Physical health indicators: Physical health was assessed only based on self-report measures. As for cognitive functioning, respondents rated their current health on a 5-point scale (Poor, Fair, Good, Very good, Excellent) and used a 3-point scale to compare how their health changed since one year ago/the last interview (Worse, About the same, Better). Respondents also indicated whether they ever had been diagnosed with one or more of the following chronic diseases: high blood pressure, diabetes, heart problems, lung problems, stroke, cancer, and psychological problems. All of these measures were summarized into one score indicating the number of chronic diseases a respondent ever experienced with a value range of 0 to 7. A number of questions assessed if the respondent experienced any of the following acute symptoms of chronic diseases at the time of the interview: swollen feet, shortness of breath, dizziness, back pain, persistent headache, fatigue as well as coughing up phlegm or wheezing. These indicators were also summarized into one score that can range from 0 to 7 with higher values indicating more acute symptoms. HRS also included measurements of difficulties with activities of daily living (ADL) and instrumental activities of daily living (IADL). Activities of daily living include walking across the room, getting dressed, bathing, eating, getting out of bed, using the bathroom. A summary score was created with a range from 0 to 6 where higher values indicated more difficulties with activities of daily living. IADLs included items about preparing meals, getting groceries, making telephone calls, taking medications, managing money, and reading a map. The IADL difficulties score ranged from 0 to 6 with higher values representing more difficulties with instrumental activities of daily living. HRS also included items that addressed the mobility of the respondent: walking one block, walking several blocks, sitting, getting up, climbing one flight of stairs, climbing several flights of stairs, lifting heavy things, kneeling, picking up a dime, extending arms, and pulling heavy objects. We combined these items again into one score indicating the difficulty of performing these actions. The mobility difficulties score ranges from 0 to 11

where higher numbers indicate more difficulties with mobility. Respondents were also asked if they engaged in the past 12 months in vigorous physical activity averaging three times a week (yes versus no). Lastly, respondents also rated their vision using glasses if necessary and hearing using hearing aids if necessary on a 5-point scale (Poor, Fair, Good, Very good, Excellent). In addition, the Body-Mass-Index was also calculated based on self-reported height and weight.

Social network indicators: The variables measuring the size of the respondent's social network include the number of living children, the number of siblings, and the number of people in the household. HRS also asked for the existence of relatives as well as friends nearby (yes versus no). Finally, respondents estimated the frequency with which they get together with friends or neighbors just to chat or for social visits.

3.3.3 Data and Methods

3.3.3.1 ANALYTIC SAMPLE

The analytic sample for this study includes, as previously described, only members of the AHEAD and CODA cohort who were eligible according to the AHEAD/HRS criteria. Spouses that entered the panel later due to marriage or partnership with an AHEAD or CODA member were excluded because their decision process to participate in the survey could be different as they join a household of an established panel member. Cases that had never been interviewed were dropped from the analytic sample as well. In addition, due to the lack of information on the panel members before their first wave it is not possible to predict nonresponse in the respondent's first wave. Only cases that were interviewed during their first wave were therefore included in the analytic sample. The final number of respondents in Wave 1 who are included in the analyses is n=8,637. Seventy-six percent (n=6,588) originate from the AHEAD

cohort while 24 percent (n=2,049) belong to the CODA cohort. Members of the AHEAD cohort can contribute up to seven waves of measurements (beginning in 1993) whereas the members of the CODA cohort were only measured for five waves (beginning in 1998). Nearly two thirds of the sample are female (n=5,345; 61.88%) and more than one third male (n=3,292; 38.12%). The mean number of years that respondents attended school amounts to 11.2. The range of school years represented in the sample extends from 0 to 17 with a median of 12. African-Americans account for 12.32% (n=1,064) of the sample. The other respondents (87.68%; n=7,573) are of Caucasian and other races. Only 5.18% of the respondents classify themselves as Hispanics (n=447) whereas 94.82% (n=8,190) do not.

3.3.3.2 METHOD OF ANALYSIS AND CREATION OF THE ANALYTIC DATASET

The method of analysis is described by Allison (1982) and allows us to model longitudinal data measured in discrete time intervals with time-varying covariates as well as with a dependent variable with more than two categories. This discrete-time hazard model is estimated through a multinomial logistic regression. The number of separate observations for a person depends on how many waves a person experienced before either censoring occurred or the person transferred into an absorbing state such as death, non-contact, refusal, and proxy-interview. We disregard in our analysis the year in which the interviews occurred but focus instead on the wave number that interviews in a given year represent for the AHEAD and the CODA cohorts. In other words, wave 1 in our dataset represents the interviews completed in 1993 for AHEAD members but for CODA members those completed in 1998 (Figure 3.1).

Table 3.1: Structure of Analytic Dataset

	WAVE 1	WAVE 2	WAVE 3	WAVE 4	WAVE 5	WAVE 6	WAVE 7
AHEAD	1993	1995	1998	2000	2002	2004	2006
CODA	1998	2000	2002	2004	2006		

As previously mentioned, AHEAD members were able to participate in a maximum of seven waves until they were censored whereas censoring for CODA members occurred already after five waves. The person-wave dataset for this study amounts to 25,961 person-waves.

The dataset is structured so that the independent variables measured in wave 1 predict the response status of the person in wave 2, the independent variables measured in wave 2 then predict the response status of the person in wave 3 and so forth. Depending on the variables included in the model, we test different assumptions about each person's hazard rate to experience one of the competing outcomes. If only the independent variables are included in the model, we assume that a person's hazard rate to experience one of the events of the dependent variable does not change across waves by itself but only because of the explanatory variables that are taken into account. Including indicator variables for the waves allows, however, that a person's hazard rate changes across the waves of a panel without the influence of the independent variables. Moreover, interactions between the wave indicator variables and any independent variables account for differential effects of the independent variable across the waves. We will include wave indicator variables as well as interaction effects between them and the other independent variable to allow for estimation of differential effects of predictors across waves.

This analysis method is not unproblematic. The dataset is a person-wave structure where multiple observations represent one respondent and is subject to correlations due to multiple observations from the same respondent. This can

produce inefficient estimates and underestimation of standard errors with traditional estimators because the model's error terms are not independent (Allison 1982). In addition, the Health and Retirement Study is based on a complex sample design with strata and clusters. All of our analyses, therefore, account for the complex sample design as well as the artificial correlation introduced by creating multiple observations of the same record. In reality, however, we had to compromise by not accounting for the stratification of the sample because the number of predictor variables in the context of a multinomial model exceeded the available degrees of freedom when analyzing the data under a complex sample analysis setting. We were nevertheless able to adjust for the clustering of the data within respondent, household, and primary sampling unit (PSU) by using a Huber-White variance estimator that produces robust variances even if the data are subject to intra-cluster correlation ((Williams 2000); (Froot 1989)). The level of clustering that was chosen for reflection in variance computation for our model was the PSU-level. Even though we were not able to account fully for the complex sample design we can be confident in our estimates because they are more likely to be conservative given that the gains that would be reached through stratification are not taken into account.

As in every dataset missing values occurred in nearly all independent variables. In order to be able to use all cases and all waves in the analysis we multiply imputed the missing values as proposed by Little and Rubin (1987). There are two different types of missing values that can be differentiated in our study. The first type of missing values that were imputed includes refusals and "don't know" answers. The percentage of missing values in the person-wave dataset for this type was always under 1%. The second type of missing values was created when a measure was not included in the questionnaire in one of the waves, which was the case for the first wave of the AHEAD cohort. In this case no respondent had the opportunity to answer that question. Imputation for this kind of missing values can be justified on the one hand because the dataset includes data on this measure from all the other waves for a given respondent. On the other hand due to the structure of the dataset we also have

measurements on these variables in the “same” wave from the CODA cohort. The percentage of missing values due to this type is, at 50%, considerably higher because most of the variables were systematically not measured in the first wave of the AHEAD cohort. The data were imputed using IVEWare (Raghunathan, Lepkowski et al. 2001). Ten multiples were created and used in all analyses.

3.3.4 Analyses

3.3.4.1 ATTRITION RATES

Figure 3.2 shows the final status of respondents across waves 2 to 7. Each column represents all 8,637 respondents who are examined in this study. If respondents reach a state considered “attrition” in a wave they occupy, they will keep this category in all following waves. We can see a dramatic cumulative increase of the percentages of respondents from the original sample who died or were replaced by a proxy-respondent. The graph also shows the censoring that occurs for the CODA cohort in Wave 6 and 7 because it was introduced two waves later than the AHEAD cohort. The average number of waves for a respondent until attrition or censoring is 3.8.

3.3.4.2 CHARACTERISTICS OF PANEL ATTRITORS

Table 3.1 shows the percentages or means of all the variables in wave 1 that are included in our multivariate discrete-time hazard model, given the conceptual framework above. Descriptive statistics are reported for each attrition type as well as for the group of respondents still being respondents themselves after wave 7 where censoring of all cases occurred due to the availability of data. If attrition was truly a random process with respect to these variables, the means or percentages of characteristics of sample members would not be different across

the groups. At a first glance we see that attrition is selective for most of the characteristics measured in wave 1 of the HRS study.

Remaining respondents are well educated, are cognitively highly functioning and don't show poorer health. They have the highest number of school years in wave 1, supporting several earlier studies showing that respondents are better educated than panel attritors. This group also shows the lowest percentage of AHEAD members, African Americans and face-to-face interviews. Face-to-face is usually the data collection mode of choice in the first interview with respondents to build rapport and when the respondents' health circumstances make telephone interviews difficult. The remaining respondents' objective cognitive functioning measures, immediate and delayed recall, TICS and Series 7, have the highest scores as previous literature indicated. Remaining respondents also rate their health better. They have the lowest number of acute symptoms and difficulties with IADLs. While physical health indicators used in previous studies might differ from those in this study the general conclusion that non-attritors are healthier is also confirmed. In addition, they have the highest average number of living children and respondents in the household but the lowest average number of household members.

Figure 3.1: Final Cumulative Status of Wave 1 Respondents Across Waves 2 to 7

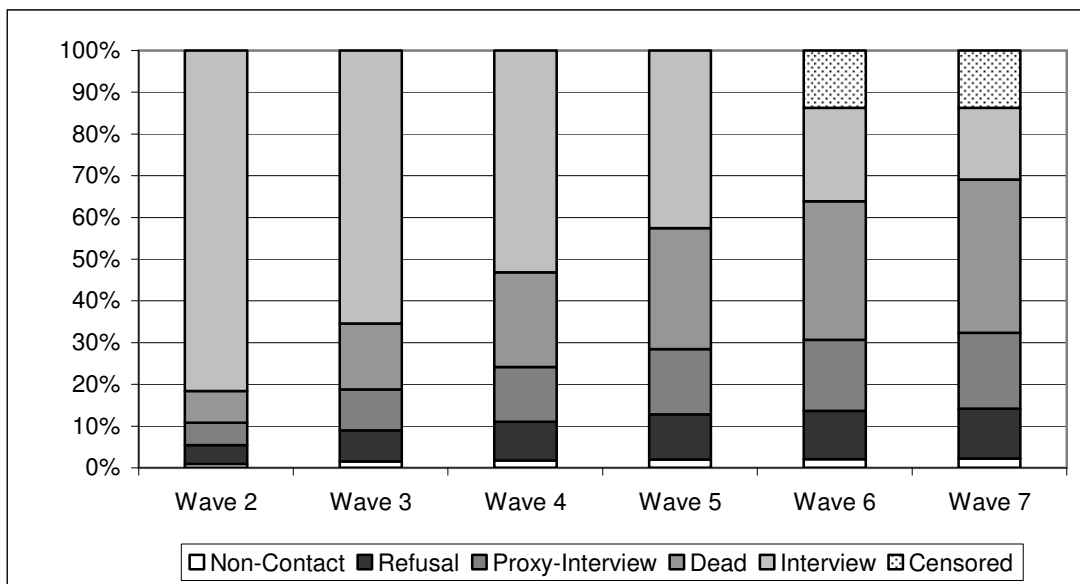


Table 3.2: Wave 1 Characteristics of Panel Attritors by Attrition Type

		Interview		Dead		Non-Contact		Refusal		Proxy Interview	
		n=4,601		n=1,993		n=157		n=772		n=1,114	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Respondent Characteristics	Male (%)	36.02	0.79	46.11	0.01	38.47	0.05	33.83	1.93	35.26	1.60
	Age	77.1	0.10	78.96	0.17	76.31	0.54	75.12	0.20	78.93	0.22
	African-American (%)	7.19	0.40	8.66	0.01	19.62	0.03	7.33	0.92	11.73	0.96
	Hispanics (%)	3.58	0.26	2.93	0.00	13.51	0.03	3.32	0.64	5.13	0.64
	AHEAD (%)	67.12	0.83	83.51	0.01	74.66	0.04	69.42	1.90	83.78	1.26
	# School Years	11.89	0.05	11.06	0.09	10.21	0.32	11.23	0.12	10.04	0.12
Data Collection Mode	Face-to-Face (%)	44.47	1.10	56.29	1.22	48.78	4.53	46.04	1.95	59.72	1.67
Cognitive Functioning Measures	Self Rated Current Memory	3.13	0.02	3.05	0.03	3.18	0.09	3.2	0.04	2.85	0.03
	Self Rated Past Memory	-0.15	0.01	-0.16	0.01	-0.11	0.04	-0.1	0.02	-0.19	0.02
	Immediate Word Recall	4.98	0.03	4.25	0.05	4.11	0.18	4.83	0.08	3.64	0.07
	Delayed Word Recall	3.73	0.04	2.8	0.05	2.64	0.18	3.49	0.10	2.12	0.08
	TICS	9.34	0.02	8.8	0.04	8.2	0.18	9.16	0.05	8.15	0.07
	Series 7	3.98	0.02	3.84	0.04	3.77	0.14	3.88	0.06	3.66	0.06

(Table continued on next page)

		Interview		Dead		Non-Contact		Refusal		Proxy Interview	
		n=4,601		n=1,993		n=157		n=772		n=1,114	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Physical Health Measures	Self Rated Current Health	3.15	0.02	2.55	0.03	2.97	0.11	3.1	0.04	2.75	0.04
	Self Rated Past Health	-0.14	0.01	-0.21	0.02	-0.07	0.06	-0.1	0.03	-0.15	0.02
	BMI	25.95	0.09	25.1	0.13	25.98	0.52	25.53	0.19	25.15	0.16
	# Chronic Conditions	1.46	0.02	1.84	0.04	1.55	0.11	1.28	0.04	1.54	0.04
	# Acute Symptoms	1.26	0.09	1.71	0.20	1.54	0.24	1.28	0.21	1.54	0.24
	Rating of Vision	3.15	0.02	2.9	0.03	3.02	0.11	3.18	0.04	2.83	0.04
	Rating of Hearing	3.24	0.02	3.11	0.03	3.35	0.10	3.29	0.05	3.04	0.04
	Vigorous Exercise (%)	32.70	0.88	23.05	1.31	30.51	4.40	36.77	2.40	27.58	2.21
	# Difficulties With Mobility	2.71	0.08	3.64	0.19	2.89	0.32	2.39	0.17	3.29	0.18
	# Difficulties With ADL	0.3	0.01	0.64	0.03	0.46	0.09	0.25	0.03	0.56	0.04
# Difficulties With IADL	1.12	0.33	2.2	0.04	2.54	0.17	2.27	0.08	2.46	0.06	
Social Network Indicators	# Living Children	2.63	0.34	2.36	0.06	2.58	0.25	2.42	0.09	2.46	0.07
	# Siblings	2.08	0.33	1.82	0.05	1.97	0.17	2.15	0.08	1.95	0.07
	Relatives Nearby (%)	32.14	0.84	31.99	1.85	29.53	5.62	31.53	2.48	31.03	2.03
	Friends Nearby (%)	75.83	0.89	73.25	1.55	68.40	5.14	73.07	2.16	76.63	1.78
	Frequency of Chatting With Neighbors	1.37	0.05	1.34	0.10	1.23	0.12	1.3	0.08	1.56	0.35
Household Characteristics	# Household Members	1.63	0.02	2.13	0.03	2.64	0.14	2.12	0.04	2.35	0.04
	# Respondents in HH	1.39	0.01	1.32	0.13	1.14	0.03	1.32	0.02	1.35	0.02

In contrast, those respondents who died during the course of the panel are on average the oldest and have the most physical health problems in wave 1. This group has the highest percentage of males because, in general, males die before women. Those who died have the lowest percentage of Hispanics. These respondents rate their current and past health lowest confirming past literature as well. They also have the lowest average BMI and the highest number of chronic conditions and acute symptoms indicating frailty and a general higher likelihood for more physical decline. Among all groups this one shows the lowest percentage of respondents doing vigorous exercise and the highest average number of difficulties with mobility and ADL most likely due to the indicated frailty. They also have the lowest number of living children and siblings.

A big percentage of those respondents who attrite as non-contacts lies among African Americans and Hispanics. This group tends to rate their past health and hearing acuity highest. They also show the highest BMI, which is protective and positive for the age groups in the HRS compared to a high BMI for the young- and middle-aged population. However, they have the most difficulties with IADLs. Their social network seems to be the weakest of all groups: Non-contacts have the lowest average percentage of relatives and friends nearby. They chat with neighbors rarely. Although this group has on average the highest number of household members, it has the lowest number of respondents in the household.

Refusals are the youngest group and show the smallest percentage of males. These respondents rate their current and past memory the highest. They have the lowest number of chronic conditions and difficulties with mobility and ADLs. They also report the best vision and the most exercise. All of this indicates that this group is highly functioning and not as initially expected a group of respondents who perceives itself to be too sick to participate in a survey. They also have the highest average number of siblings.

The respondents that have been replaced by proxy respondents have the highest percentage of AHEAD members and face-to-face interviews. They completed the lowest average amount of school years and show the worst

performance on all cognitive functioning measures. They rate their vision and hearing worst which might be one of the main reasons that proxy respondents are used for future waves. Interestingly, they report on average the highest number of nearby friends and chats with neighbors.

To summarize, bad cognitive functioning in wave 1 seems to influence the transition from a self to a proxy interview the most. Physical health does not seem to be a factor except for low sensory functioning that can be expected to make the interview difficult for the respondent. Refusals, on the other hand, are the most active and a very healthy group. Against our expectation, bad cognitive functioning does not seem to play a factor in the likelihood to refuse. Non-contacts have the highest group of African Americans as well as Hispanics suggesting that they might be more mobile or have different at-home patterns than non-minority respondents. Their social network seems to be fairly weak as well. Respondents who die during the course of the study have the worst health and are the oldest. In contrast, respondents that do not attrite have the highest level of education, function on a high cognitive level, have fewer complaints about their physical health and are more likely to have another respondent in the household.

3.3.4.2 MULTIVARIATE ANALYSIS OF PANEL ATTRITION

A multivariate analysis that includes cognitive functioning, physical health and social network variables allows us to differentiate between the influences of a certain set of factors on the likelihood of refusal and proxy interview compared to self-interview and on the likelihood of proxy-interview versus refusal while controlling for the other two sets as well as the demographic control variables. Most of the independent variables are introduced into the models as time-varying covariates except for the study membership, gender, education, race, and ethnicity. After fitting an initial full model we determined using likelihood-ratio tests that some of the independent variables did not significantly improve the overall fit of the model and were therefore dropped to minimize the number of

independent variables for more degrees of freedom. All models that are presented are weighted by the weight that each respondent received in his first wave. The variances take the clustering by PSU as well as the multiple imputations into account. In addition, we fit the models without a constant so that the coefficients can be interpreted as discrete-time hazard rates.

We also addressed the sensitivity of results to the clustering and compared estimates based on the PSU and the person as a cluster. The few differences that were found are small and would not have altered our conclusions described below. The same is true for the sensitivity of the estimates to the definition of the risk set, meaning the different final states respondents can reach. We fitted separate logistic regression models for two comparisons of the three categories of our dependent variable: refusal versus self-interview and proxy-versus self-interview. By censoring observations that dropped out of the panel due to non-contact or death before the failure wave all respondents were kept in the analyses. The results from the logistic regression models show that the estimates of the multinomial regression model are independent of the other categories of the dependent variable, death and non-contact, and can be expected to be relatively robust to different definitions of the risk set.

3.3.4.2.1 Comparisons of the Influences of Different Variable Groups

It is of interest for researchers to know which of the different variable groups are most important in explaining variability in panel attrition outcomes. We used a number of likelihood-ratio tests, scalar tests and information criteria to assess how each variable group and combinations of variable groups perform (Table 3.2). The first measure shows likelihood ratio χ^2 statistics comparing the various models to the base model that only includes the demographic variables. All different combinations of variable groups improve the model fit significantly compared to the base model.

Table 3.3: Goodness-of-Fit Statistics for Multinomial Logistic Regression Models with Different Sets of Variables

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
	Base Model	Model With Cognitive Functioning Variables	Model With Physical Health Variables	Model With Social Network Variables	Model With Cognitive Functioning and Physical Health Variables	Model With Cognitive Functioning and Social Network Variables	Model With Physical Health and Social Network Variables	Full Model
	(df) Statistic	(df) Statistic	(df) Statistic	(df) Statistic	(df) Statistic	(df) Statistic	(df) Statistic	(df) Statistic
LR χ^2 Test: Current Model Compared with Base Model		(18) 1387.18***	(33) 1530.37***	(3) 65.87***	(51) 2388.77***	(21) 1414.91***	(36) 1552.57***	(54) 2383.58***
LR χ^2 Test : Full Model Compared With Current Model	(54) 2383.58***	(36) 996.40***	(21) 853.21***	(51) 2317.71***	(3) 14.81**	(18) 831.01***	(33) 968.67***	
Mc Fadden's Adjusted R2	0.508	0.528	0.531	0.51	0.543	0.527	0.53	0.543
ML (Cox-Snell) R2	0.759	0.769	0.771	0.757	0.778	0.769	0.771	0.778
Cragg-Uhler (Nagelkerke) R2	0.807	0.82	0.822	0.807	0.83	0.82	0.823	0.83
AIC*n	35398.217	34607.1	33866.286	35334.945	33084.435	34043.309	33849.268	33074.872
BIC	35716.627	34532.46	34454.119	35677.847	33819.227	34533.17	34461.594	33834.156
Deviance	(25922) 35320.217	(25904) 33953.1	(25889) 33722.286	(25919) 35250.945	(25871) 32904.435	(25901) 33923.309	(25886) 33699.268	(25868) 32888.872

Although scalar measures as McFadden's Adjusted R^2 , Cox-Snell's R^2 , and Nagelkerke's R^2 are reviewed with some skepticism (Long 1997) they nevertheless can provide some general information about different models. All else being equal larger values are preferred even though there are no standards on how large these different R^2 s have to be to indicate a good model fit. Across all different types of R^2 s we can see that the base model and the model with the only remaining social network variable receive the lowest values followed by the model with the cognitive variables, the one with physical variables, the combination of both variable groups and the full model. All measures indicate a preference of the model only with the physical health variables over the one only with the cognitive functioning measures. The model combining both is, however, preferred over each of the variable groups alone. The addition of the social network variable does not increase any of the R^2 measure in a noticeable way.

The Akaike (AIC) and the Bayesian (BIC) information criteria both indicate a better fitting model through a smaller value (Long 1997). As expected, the base model shows the highest value for both criteria followed by the model including only the social network variable. As the R^2 measures indicated the physical variable group increases the fit on their own more than the group of the cognitive functioning measures alone. Both together provide a better fit than each of them alone. While including the social network variable decreases the AIC criterion slightly, the BIC criterion shows an increase suggesting that including the variable does not improve the model fit.

The various goodness-of-fit measures show that both the physical health and the cognitive functioning measures are important in predicting panel attrition and improving the fit of the model. The strong association between the physical health status and the death of respondents during the course of the panel that was revealed in the descriptive statistics seems to drive the importance of the physical health variables for a good model fit. Although the only social network indicator that was left in the final model still contributes significantly to improving the model fit according to the likelihood-ratio test it is not strong enough to increase the model fit as measured by the scalar measures.

After comparing the overall usefulness of the different groups of variables we will now assess the relationship between cognitive functioning, physical health, social network extent and different types of panel attrition. Socio-demographic variables are included in the model as covariates. Table 3.3 shows the results of the final multinomial logistic regression model.

3.3.4.2.2 Relationship between Cognitive Functioning and Attrition

Cognitive functioning is assessed in the final model through current self-rated memory, immediate and delayed word recall and the mental status test, TICS. It was surprising that the Series 7 test for working memory did not significantly improve the model fit and could be dropped from the model. Although a survey interview is very demanding and would be more difficult the worse the working memory is poor working memory does not influence the participation decisions in future waves of the panel. Self-rated memory and objective memory scores and general mental status seem to influence the survey participation decision. This could be due to a higher influence of memory measures and mental status on personal perception of the ability to participate.

Against our initial expectations based on the literature but confirming the descriptive statistics, none of these variables seems to have a significant effect on the likelihood of refusal compared to self-interview. With regard to comparisons of proxy-interview with self-interview or refusal, cognitive functioning seems to be a significant influence factor. A low score on memory measures and the mental health assessment, and a negative self-assessment of the quality of memory increases the likelihood of a proxy-interview in the next wave as predicted. While the memory measures' influence is stable across waves the influence of the mental status assessment increases significantly in later waves. The increase, however, is very small.

Table 3.4: Final Multinomial Logistic Regression Models

		Model 1		Model 2		Model 3		Model 4	
		Attrition Versus Interview		Refusal Versus Interview		Proxy-Interview Versus Interview		Proxy-Interview Versus Refusal	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Respondent Characteristics	Wave 3	-5.52***	0.79	-2.65	1.84	-6.74***	1.31	-4.09	2.25
	Wave 4, 5, 6, 7	-3.55***	0.33	-2.00*	0.85	-4.84***	0.53	-2.84**	1.01
	Face-To-Face Ahead	-0.09*	0.03	-0.16	0.09	-0.02	0.06	0.15	0.10
	Male	0.4***	0.06	0.11	0.13	0.48***	0.13	0.36	0.19
	# HH Members	0.29***	0.03	-0.08	0.06	-0.06	0.06	0.01	0.09
	# Respondents In HH	0.02	0.02	-0.00	0.05	0.07	0.04	0.07	0.06
	Age	-0.15***	0.04	-0.10	0.09	0.12	0.07	0.22*	0.11
	Age* Wave 3	0.05***	0.00	0.01	0.01	0.05***	0.00	0.04***	0.01
	Age* Wave 4, 5, 6, 7	0.03***	0.01	0.01	0.02	0.05***	0.01	0.04**	0.02
	# Years of School	0.01***	0.00	0.02***	0.00	0.02***	0.00	0.00	0.00
	African American	-0.02**	0.01	-0.05***	0.01	-0.03**	0.01	0.02	0.01
	Hispanic	-0.15*	0.06	-0.17	0.12	-0.23*	0.10	-0.06	0.14
			-0.29**	0.09	-0.43*	0.22	-0.27	0.15	0.16
Cognitive Functioning Measures	Self-Rated Current Memory	0.03	0.02	0.03	0.04	-0.09**	0.03	-0.12**	0.05
	Immediate Word Recall	-0.06***	0.02	0.00	0.04	-0.13***	0.03	-0.13**	0.04
	Delayed Word Recall	-0.08***	0.01	-0.05	0.03	-0.17***	0.02	-0.12***	0.03
	TICS	-0.21***	0.02	-0.07	0.04	-0.31***	0.03	-0.24***	0.04
	TICS* Wave 3	-0.04	0.04	0.01	0.09	0.03	0.05	0.02	0.10
	TICS* Wave 4, 5, 6, 7	0.11***	0.03	-0.08	0.08	0.17***	0.04	0.25**	0.08
	Self-Rated Current Health	-0.19**	0.02	-0.07	0.04	-0.10**	0.03	-0.03	0.05
Physical Health Measures	BMI	-0.04***	0.01	-0.02	0.01	-0.04***	0.01	-0.02	0.01
	BMI* Wave 3	0.04	0.01	0.02	0.02	0.04	0.02	0.02	0.03
	BMI* Wave 4, 5, 6, 7	0.01	0.01	0.00	0.02	0.02	0.02	0.02	0.03
	# Chronic Conditions	0.11***	0.02	-0.05	0.04	0.05	0.03	0.10*	0.05
	Rating of Vision	-0.04*	0.02	-0.03	0.03	-0.07*	0.03	-0.04	0.04
	Vigorous Exercise	-0.22***	0.05	0.01	0.09	-0.18	0.12	-0.20	0.15
	# Difficulties With ADL	0.14***	0.02	0.00	0.06	0.12***	0.03	0.11	0.06
	# Difficulties With IADL	-0.18***	0.03	-0.10	0.05	-0.19***	0.04	-0.08	0.05
	# Difficulties With IADL* Wave 3	0.15**	0.05	-0.04*	0.13	0.11	0.10	0.15	0.14
	# Difficulties With IADL* Wave 4, 5, 6, 7	0.2***	0.04	-0.13	0.16	0.26***	0.07	0.39*	0.17
Social Network	# Living Children	-0.03*	0.01	-0.05*	0.02	-0.02	0.02	0.03	0.02

*** p < 0.001 ** p < 0.01 * p < 0.5

3.3.4.2.3 Relationship between Physical Health and Attrition

The influences of measures of physical health differ across the various types of panel attrition. Physical health factors seem to be less influential in predicting the likelihood of refusal compared to interview. Only the number of IADL difficulties has a negative significant influence on respondents' refusal in the next wave compared to remaining a self-respondent. For the likelihood of proxy-interview compared to interview nearly all physical health measures seem to have an influence. When comparing the likelihood between proxy-interview and refusal only the number of chronic diseases ever had shows a significant influence. This pattern is rather surprising given that, based on the descriptive statistics, the differences on measures of physical health in wave 1 proxy respondents are more different from refusals than from interviews. Compared to respondents remaining in the panel, the respondents who are replaced by proxies showed a lower BMI, worse vision, less exercise, and more difficulties with ADL but less with IADL. They also rated their health lower. In addition to poor cognitive functioning, proxy-interviews are more likely than interviews when physical health is also negatively affected from the beginning of the panel.

3.3.4.2.4 Relationship between Social Network and Attrition

The only measure for the respondents' social network that remained in the final model is the number of living children. In contrast to cognitive functioning and physical health indicators the number of living children does not have a significant effect on the likelihood of a proxy-interview. It seems, however, to be negatively related to the likelihood to refuse compared to be interviewed. One possible explanation is that respondents with more living children and most likely grandchildren are more incorporated into society and more willing to report about their health and retirement status.

3.3.4.2.5 Summary

The different groups of variables show differential influences on the likelihood of the different types of attrition compared to staying as respondents in the panel. The likelihood of refusal in later waves does not seem to be greatly influenced by cognitive or physical decline against our initial expectations. Demographic and social network variables describe some sample subgroups that have a significantly lower likelihood to refuse. Hispanics, higher educated respondents, and respondents with a higher number of living children are less likely to refuse than be interviewed. Respondents that have participated more than three times are also significantly less likely to refuse.

The likelihood of being replaced by proxy respondents is, however, significantly influenced by all cognitive functioning measures included in the final model. The more cognition declines the more likely it is that respondents are replaced by proxies. This is true for the comparison to staying an active respondent as well as to refusing participation. Declining physical health seems also to increase in general the likelihood of proxy respondents. This is more the case for the comparison of staying a respondent versus refusing. The only social indicator that is left in the model does not seem to have any influence.

While proxy respondents seem to be truly chosen when physical and cognitive decline does not allow respondents to continue to be interviewed refusals themselves seem not to be a function of cognitive and physical decline even in later waves of the HRS study.

3.3.5 Limitations of the Analyses

The analyses presented in this study have some limitations that should to be taken into account when interpreting the models. The general problem of panel attrition has been simplified in two ways: first, we only included respondents that have been interviewed in the first wave of the panel and neglected, therefore, the effects of initial non-cooperation on the sample composition and changes of

statistics of interest (Kempen and van Sonderen 2002). Including non-respondents of the first wave in our analyses was not possible since no information is available about these individuals with regard to physical health, cognitive functioning, and social network. Second, we defined panel attrition as a break from continuous self-response either through non-contact, refusal or the use of a proxy respondent. In reality, however, respondents from the HRS are followed up over several waves even if no interviews were conducted in previous waves. In this study, about 80% of all respondents that did not participate in a wave were final panel attritors and never came back as panel respondents.

3.4 DISCUSSION AND CONCLUSION

The aim of this study was to better understand the mechanisms behind panel attrition by distinguishing between different types of panel attrition (death, non-contact, refusal and proxy interview) and assessing the marginal influence of physical health, cognitive aging, and changes in the social networks on attrition in panel studies of the elderly. Measures of physical health and cognitive functioning explain a large part of the variance when predicting panel attrition and should definitely be included in analyses of panel attrition.

While only a few respondent characteristics influence the likelihood of refusal compared to interview, cognitive functioning and physical health measures are significant in predicting a proxy- compared to a self-interview. The only indicator for social support that remains in the final model does not seem to be of major importance as a predictor for panel attrition.

Future research should address the magnitude of nonresponse bias that could be introduced through the selective attrition of respondents in panels of the elderly. Insights from this study can also help survey methodologists in developing tools to minimize panel attrition. Proxy-interviews as one example have proven their usefulness in this study by keeping respondents with significant cognitive and physical decline in the panel.

CHAPTER 4

TIME-OF-DAY OF INTERVIEW, CHANGES IN CIRCADIAN RHYTHM AND DATA QUALITY: WHEN IS THE BEST TIME TO INTERVIEW THE ELDERLY?

4.1 INTRODUCTION

The quality of the data measured by a survey is of utmost importance when using survey data as basis for decision making in policy settings. Although sampling, coverage, nonresponse and processing errors have been shown to contribute to the amount of bias observed in a statistic of interest ((Groves 1989); (Biemer and Lyberg 2003)) researchers still consider measurement error to be the most threatening factor to data quality. Accurate reporting of experiences and opinions is the pillar for survey research. There is, however, an extensive literature describing factors influencing the accuracy of survey reports (for a summary see (Tourangeau, Rips et al. 2000)). Besides the questionnaire and the interviewer, respondents themselves can provide inaccurate information, deliberately or without any intention to do so. Data quality, for example, seems to decrease with the age of the respondent due to the age-related decline in cognitive resources ((Knäuper, Belli et al. 1997); (Knäuper, Schwarz et al. 2004); (Knäuper 1998)). A great number of laboratory studies in the psychological literature support this age-related decline in cognitive functioning. Cognitive psychologists have shown in both cross-sectional and longitudinal laboratory studies that cognitive functions such as working memory, explicit recall and processing speed show declines from age 20 on (Hoyer and Verhaeghen 2006). Interestingly, a number of studies

have also suggested that this decline is moderated by the time of day when cognitive abilities are measured (e.g. (Hasher, Goldstein et al. 2005)). Elderly people seem to be able to compensate for the limitations of their cognitive capacities in the morning, but apparently not in the evening. Circadian rhythms have been shown to be responsible for this variation in the quality and effectiveness of cognitive functioning of the elderly: the peak arousal time for cognitive functioning for the elderly is in the morning (May, Hasher et al. 1993). Winocur and Hasher (2002), for example, provide evidence for a number of measures, such as word span, number of words forgotten and word recognition, that elderly people are performing significantly worse in the evening whereas younger subjects either improve their test score from morning to evening or perform at about the same level as in the morning. Younger people seem to still have the cognitive capacities that can be activated when tested at a non-optimal time for their age whereas the elderly have minimal or no cognitive resources that can be used during non-optimal times. If these findings based on laboratory research hold in a survey context, survey researchers may be able to improve data quality by paying attention to circadian rhythms and conducting the survey interview in the morning.

This study provides a test of this possibility. We will first review the literature on circadian rhythms, how they are linked to cognitive functioning and how they are affected by aging (Section 4.2). We then turn to a description of data quality and review the existing studies of the effect of aging on data quality (Section 4.3). As a next step we present our analyses that are based on data from the Health and Retirement Study (Section 4.4). Even though laboratory experiments confirm the sensitivity of cognitive measures to the time-of-day for elderly subjects, we were not able to confirm these results using real-world survey data.

4.2 CIRCADIAN RHYTHMS AND AGING

It has long been known to humans that activities of mammals, lower animals, and even plants follow a 24-hour cycle, adapting their behavior or movement to periods of light and darkness. Those following a diurnal cycle have a waking period during the day and a period of rest during the night. Nearly all of the processes in the human body are temporally organized through these circadian rhythms into sleep and waking phases and maximize the survival and reproduction of the species (Moore 1997). Processes that are generally known to oscillate between high and low performance are core body temperature, hormone production, heart rate, and blood pressure. The time of the day when performance reaches its height is usually called the optimal time versus the non-optimal time indicating when performance of these processes is very low.

Psychologists suspected early on that cognitive functions are also subject to circadian rhythms (for a summary of the recent literature see Schmidt et al. (2007)). Over 120 years ago, Ebbinghaus (1885) showed in an experiment that it is easier to learn nonsense syllables in the morning than in the evening. Laboratory experiments over the past decades demonstrated that indeed a number of different cognitive functions are subject to circadian rhythms such as alertness (West, Murphy et al. 2002), attention ((Blake 1967); (Yoon, May et al. 1999)), working memory (West, Murphy et al. 2002), recall speed ((Petros, Beckwith et al. 1990); (Anderson, Petros et al. 1991)), short-term memory (Baddeley, Hatter et al. 1970), recognition (May, Hasher et al. 1993), executive functioning (Yoon, May et al. 1999), and inhibitory processes ((May and Hasher 1998); (Hasher, Chung et al. 2002)). Yoon (1997) also showed that people are more likely to use schema-based processing strategies during their non-optimal time than strategies that are based on details. May and Hasher (1998) noted that cognitive functioning is at its best during the peak time of circadian arousal compared to the off-peak time where cognitive functioning seems to be at its worst and named this pattern the “synchrony effect.” This relationship between circadian rhythms and cognitive functioning has also been confirmed in animals

(Winocur and Hasher 2004) and seems to be a general phenomenon for mammals as well as lower animals.

Against the expectation that all of these physical and cognitive processes follow the same general circadian rhythm, there are a number of factors that can cause interindividual differences in circadian rhythms among people. First of all, people seem to have an inert feeling when their “best time” is during the course of a day and classify themselves as an early bird or a night owl. Horne and Ostberg ((1976); (1977)) developed a questionnaire, the Morningness-Eveningness Questionnaire, that classifies people based on 19 questions into five chronotypes ranging from “definitely morning types” to “definitely evening types.” It is also possible to be classified as being neither morning nor evening type. Scores from this test correlate with variations in physical measures such as body temperature, sleep-wake cycle and perceived alertness ((Tankova, Adan et al. 1994); (Hasher, Goldstein et al. 2005)). Second, circadian rhythms are also different for various cognitive tasks based on the cognitive domain they belong to, their duration and difficulty, the administration method, and the measured variable. May et al (2005), for example, found that explicit and implicit retrieval have different circadian rhythms. Bonnefond et al. (2003) also confirmed this assertion by finding an age effect for more complex tasks demanding additional attentional resources and memory capacity but not for a simple visual discrimination task. Finally, and of most importance to this study, circadian rhythms also change across the life span.

Taking the vast amount of literature based on the Morningness-Eveningness Questionnaire into account Hasher et al. (2005) conclude: “In general, most children prefer morning times, most younger adults prefer afternoon or evening times, and most older adults once again prefer morning times for both intellectual and physical activities.” This phenomenon seems to be independent from the cultural context of the subjects (Schmidt, Collette et al. 2007). It is, however, surprising that even though all the literature on differences in circadian rhythms across age separates young from old people and morning from evening, research has not yet addressed explicitly at what age this shift in

circadian arousal towards morningness occurs. Schmidt et al. (2007) suggest that the changes in circadian arousal patterns appear at the age of 50 years. “Young age” is usually operationalized in the laboratory experiments described above by using subjects between the ages of 17 and 32: 17 to 28 years (Intons-Peterson, Rocchi et al. 1998), 18 to 25 years (Intons-Peterson, Rocchi et al. 1999), 18 to 22 years (May, Hasher et al. 1993), 18 to 23 years (May, Hasher et al. 2005), 20 to 30 years (Bonfond, Rohmer et al. 2003), and 18 to 22 years (Yoon 1997). “Old age,” in contrast, is defined as subjects aged 55 years and older: over 55 years (Intons-Peterson, Rocchi et al. 1998), 60 to 90 years (Intons-Peterson, Rocchi et al. 1999), 66 to 78 years (May, Hasher et al. 1993), 60 to 75 years (May, Hasher et al. 2005), 50 to 60 years (Bonfond, Rohmer et al. 2003), 61 to 88 years (Bugg, DeLosh et al. 2006), and 65 to 79 years (Yoon 1997). As with regard to age research has also been sparse with regard to the time-of-day and transitions between optimal and non-optimal times and vice-versa. Laboratory studies defined morning sessions as occurring between 8am and 10:30am: around 9am (West, Murphy et al. 2002), at 8am or 9am (Yoon 1997), at 8am (Bugg, DeLosh et al. 2006), between 8am and 9am (May, Hasher et al. 2005), and before 10:30am (Intons-Peterson, Rocchi et al. 1999). Evening sessions, on the other hand, were scheduled after 3pm: at 5pm (West, Murphy et al. 2002), at 4pm or 5pm (Yoon 1997), at 5pm (Bugg, DeLosh et al. 2006), between 5pm and 6pm (May, Hasher et al. 2005), and from 3pm on (Intons-Peterson, Rocchi et al. 1999). To summarize, we do not know when optimal and non-optimal times start and end.

The impact of the shift in the peak time of circadian arousal across age groups on cognitive performance has been shown in a number of studies. Yoon (1997) concluded in her study that older adults improve their recognition accuracy dramatically in the morning compared to the evening. Elderly subjects seem also able to use more detailed-oriented processing strategies when interviewed in the morning whereas they relied on schema-based strategies in the evening. Hasher, Goldstein and May (2005) summarized previous studies and showed a consistent effect of declining performance for elderly subjects and

improving performance for younger subjects in the evening for measures such as correct recognition, forgetting of story material, correct stem cued recall of words as well as performance on stop signal trials. West et al. (2002) fielded a study to examine the effect of distractions during cognitive tasks on working memory performance and found that the number of intrusion errors was modulated by the time-of-day of testing for elderly subjects. False memories were also more likely in elderly subjects tested at non-optimal times (Intons-Peterson, Rocchi et al. 1999).

Differences in chronotypes, characteristics of the cognitive task, the subject's age as well as the time-of-day of testing seem to play a moderating role with regard to the relationship between the respondent's age and cognitive functioning as measured through tests we administer. In fact, Winocur and Hasher (2004) hypothesize that, based on their experience, since the cognitive testing in laboratory settings is usually performed in the afternoon the cognitive differences between young and old people might be at least slightly overestimated. Surveys focusing on the elderly, such as the Health and Retirement Study, routinely include measures of cognitive functioning to be able to assess the influence of cognitive decline on physical health and the ability to perform everyday life situations. These instruments are able to reliably measure the difference in cognitive functioning between younger and elderly respondents and it can therefore be hypothesized that the time-of-day when the interview is conducted shows the same influence on cognitive performance in a survey interview as in the laboratory studies. Moreover, survey methodologists should be concerned that the variation of cognitive performance across a day and the non-ability to compensate for cognitive decline by older respondents leads to a decrease in data quality when older people are interviewed at their non-optimal time in the evening.

4.3 DATA QUALITY

The quality of data collected through a survey is generally assessed through various types of survey errors that have a negative impact on the accuracy and the precision of the estimate (Groves 1989). While it is known that sampling procedures, coverage issues and survey nonresponse can have detrimental effects on survey statistics we will focus here – as outlined - on data quality in terms of indicators for measurement error. Measurement error can arise because of the respondent, the interviewer, the survey instrument, and the data collection method. In this study, however, we will concentrate on measurement error that arises due to the respondents' age and changes in cognitive functioning across the day.

Previous research has shown that the respondents' age is associated with a decline in data quality. Elderly people are more likely to refuse to answer specific questions (Colsher and Wallace 1989), they also exhibit more social desirability tendencies, they are more likely to show acquiescence ((Kogan 1961); (Knäuper and Wittchen 1994)) and more response but less question order effects than younger respondents (Knäuper 1999). Frequency reports of the elderly are also more influenced by answer scales presented than those of younger respondents (Knäuper, Schwarz et al. 2004). Knäuper et al. (1997) used direct measurements of a number of cognitive abilities instead of approximating the level of cognitive functioning through the respondents' age. They concluded that lower cognitive functioning is linked to worse data quality especially for difficult questions and confirmed the assumption of the negative influence of cognitive aging that previous researchers have made.

This study assesses data quality by predicting the likelihood of “don't know” responses for a number of questions. “Don't know” responses have been hypothesize to originate in an inadequate understanding of the question, low motivation, and the decision to withhold the retrieved information (Beatty and Herrmann 2002). Older respondents with age-related cognitive decline seem to be more at risk to not understanding a complicated question or to abort the

question answering process because the necessary cognitive processes are too burdensome given the cognitive functioning level at any given time. If survey interviews with older respondents are conducted at their non-optimal time of day it is plausible that they will be more likely to provide a “don’t know” answer for questions requiring more intense cognitive processing.

This study therefore addresses if the effects of changes in circadian rhythms combined with age-related cognitive decline on cognitive functioning can also be observed in the quality of survey data. The hypothesis we are going to test is as follows:

H4.1 The older the respondents are and the later the interview takes place the more likely they will provide “don’t know” responses.

4.4 METHODS

4.4.1 Data

The analyses are based on the Health and Retirement Study (HRS) which has been surveying more than 22,000 Americans over the age of 50 every two years. The study includes questions about the respondent’s health status, job history, family structure, disabilities, retirement plans, housing, and income. A test to measure cognitive functioning is included as well. Data are collected from five birth cohorts: the AHEAD (Aging and Health Dynamics) cohort born before 1923, the CODA (Children of the Depression Age) cohort born between and including 1923 and 1930, the original HRS cohort born between and including 1931 and 1941, the WB (War Babies) cohort born between and including 1942 and 1947, and the EBB (Early Baby Boomers) cohort born between and including 1948 and 1953. Age-eligible spouses of selected respondents were also included in the sample. Interviews were done either face-to-face or by telephone. The face-to-face mode is usually used in the first wave, for very old respondents and if the

respondent prefers to do the interviews of later waves in the same mode. Otherwise, respondents are interviewed by phone.

For each wave, interviewers contacted the respondents by phone to either make an appointment for the interview or to conduct the interview at that time. Interviewers used information on contact and interview time from previous waves to guide their decision on when to attempt to contact the respondent's household. All of the following analyses could therefore be subject by self-selection bias if the interviewing appointment is based on the respondent's preferences. People usually have a sense when their "good time" during the day is and might prefer to schedule the interview during that time (Hasher, Goldstein et al. 2005). Unfortunately, the information about who chose the final time of the interview is not recorded. In general, we can assume that most of the face-to-face interviews are conducted on an appointment basis while telephone interviews can be conducted with or without an appointment. The data for this study include at least an indicator if and how many appointments were made until the interview was completed that allows us to somewhat control for the possible self selection bias. The study design of the HRS allows the use of proxy respondents when the respondent is either unable to be interviewed or refuses to participate. Data from proxy respondents are, however, excluded from all analyses. The final dataset includes two waves of data collection: 2002 and 2004. The EBB cohort was introduced in 2004 and contributes therefore only to this wave of the data. Due to technical difficulties time-of-day of interview was not recorded for a small percentage of interviews in each wave (around 1%). We excluded these cases from all analyses.

4.4.2 Measures

We selected a number of measures based on three criteria: first, the measures should be cognitively demanding because we can expect that an interaction effect between time-of-day of interview and the respondents' age can be easily detected. Second, we chose only measures for which a "don't know" answer is

truly a sign of cognitive burden and not for any other reason such as privacy. Even though “don’t know” responses are very common for income questions we did not include measures like these because “don’t know” answers could either indicate that the respondents do not want to reveal their income to the interviewer or that it was indeed cognitively too demanding for the respondent to calculate their income. Third, a certain amount of don’t know responses for each measure was needed so that model estimation was possible. The types of measures that we selected for our assessment of the likelihood of don’t know responses included therefore numeracy questions, cognitive functioning measures, a behavioral frequency question, and likelihood assessments for future events.

In addition, we decided to replicate directly the findings from the laboratory settings described above. If the effects that are shown in the studies described above translate at all to the survey setting it is most likely that they are replicated with the same measures that were used to establish them before. The Health and Retirement Study includes the common cognitive measures for various cognitive dimensions, such as short-term and working memory. We selected subjective memory measures such as self-rated current and past memory as well as objective measures, such as immediate and delayed recall, the Series 7 test, and the counting backwards task.

The respondents’ age and the time-of-day of interview were included as continuous measures in the multivariate analyses.

4.5 ANALYSIS

Our analyses are focused on the interaction between the time-of-day of the interview and the respondents’ age. It is rare that the time-of-day of the interview is truly randomly assigned to respondents. Interviewers have information about the weekday and the time when respondents were contacted and interviewed in the previous waves and are encouraged to use this information when trying to contact respondents in the current wave. Nearly all first contact is made by

telephone. If respondents are supposed to be interviewed over the phone interviewers will try and conduct the interview when contact is established. Respondents also have the choice to make an appointment for the interview. If the mode of data collection is face-to-face interviewers have to schedule appointments when they can visit respondents at home to conduct the interview. To ease the burden on respondents, interviewers usually accommodate the respondents' wishes with regard to the time and the weekday of the interview. Respondents can prefer a certain time to be interviewed out of various reasons such as work schedules, activities or family responsibilities. Elderly respondents, however, could also choose the time for an interview based on their subjective judgment of when their best time is to be interviewed. Assuming that the time they choose is their optimal time of day with regard to cognitive functioning our analyses could be biased due to self selection. An indicator whether or not respondents were interviewed with an appointment is available and can be used as an indicator for possible selection bias.

χ^2 -tests show that the respondents' age and the time-of-day of interview are significantly associated with each other (2002: $\chi^2(4)=178.3166$, $p=0.000$; 2004: $\chi^2(4)=32.2537$, $p=0.000$; 2006: $\chi^2(4)=34.7215$, $p=0.000$). In all three years, older respondents are more likely to be interviewed later in the day although the differences compared to the other age groups are very small. The time-of-day of interview is also significantly associated with the indicator if an interview was conducted by appointment or immediately when respondents were contacted (2002: $\chi^2(2)=175.7066$, $p=0.000$; 2004: $\chi^2(2)=46.0022$, $p=0.000$; 2006: $\chi^2(2)=112.8191$, $p=0.000$). Interviews based on appointments are in all three years more likely to be conducted before noon. The age of the respondents and whether an interview was conducted with or without an appointment were significantly associated in two of the three years (2002: $\chi^2(2)=7.2839$, $p=0.026$; 2004: $\chi^2(2)=8.2520$, $p=0.016$; 2006: $\chi^2(2)=1.1271$, $p=0.569$). Respondents in the youngest age group were overall slightly more likely to be interviewed with an appointment than older respondents. Based on our assumptions what times older

respondents would prefer if they chose their preferred time of day it does not seem that self selection bias is present.

Nevertheless, we will use the appointment indicator described above to control to some extent for possible but unlikely selection bias. Even though this might not be the best indicator in theory, better and more detailed data on who chose the time of interview, respondent or interviewer, and for what reason is not readily available for analysis. We also include the number of school years completed by the respondent and the respondents' gender as control variables since studies have indicated that these can influence data quality.

4.5.1 “Don’t Know” Responses

All models predicting the occurrence of a “don’t know” answer are logistic regressions. The questions for which we will predict the likelihood of “don’t know” responses can be differentiated into numeracy questions, cognitive functioning measures, a behavioral frequency question, and likelihood assessments for future events. Numeracy questions in the 2002 wave of the Health and Retirement Study include the calculation of the likelihood of getting a disease given certain parameters and dividing a lottery win. Neither age, time, nor their interaction significantly predicts “don’t know” responses even after excluding those interviews that were conducted based on an appointment.

Cognitive measures include items that are used to assess the mental status of the respondents, such as naming the president and vice president, objects and today’s date. As for the numeracy questions no significant effects could be found for the 2002 and 2004 waves, even if only interviews without an appointment were included. The measure for working memory, the Serial 7’s task, is also not sensitive to the time-of-day of interview, the respondents’ age and their interaction, regardless of wave and if we exclude interviews scheduled by appointment.

The behavioral frequency question of how many hours in total respondents helped their friends or neighbors without pay was also not more

likely to be answered with “don’t know” by older respondents later in the day even though this question could involve extensive cognitive effort.

Older respondents interviewed later during the day were also not more likely to provide “don’t know” answers to questions asking for likelihood assessments of future events. Possible future events included the respondents being victims of biological terrorism, the United States being victims of biological terrorism, inflation, possible moves to a new residence, the respondents’ own death, getting Alzheimer’s disease or cancer. As before, the respondents’ age, the time-of-day of interview, and their interaction have no significant influence on the likelihood of don’t know answers to any of these questions in both waves and when excluding interviews with appointments.

To conclude, the interaction between time-of-day and the respondents’ age does not seem to have any influence on the quality of survey data with regard to the likelihood of providing “don’t know” responses. Our hypothesis H4.1 could therefore not be confirmed.

4.5.2 Cognitive Measures

We also decided to assess the influence time-of-interview and the respondents’ age have on cognitive measures themselves that have been used to measure cognitive functioning in a laboratory setting through linear regressions. If the effects that are shown in the studies described above translate at all to the survey setting it is most likely that they will be replicated with the same measures that were used to establish them before. Subjective measures of cognitive functioning, self-rated current and past memory, are not significantly influenced by the time-of-day of interview or the interaction with the respondents’ age. With regard to memory measures, such as immediate and delayed recall, we can see that the respondents’ age has a significant negative influence as we would have expected. Neither the time-of-day of the interview nor the interaction with age seem to influence these memory scores. The measure for working memory, the Series 7 test, seems not to be influenced by age, time-of-day or their interaction.

The older the respondents the less likely they are to correctly count back from 20, another test for working memory. Again, neither the time-of-day of interview nor the interaction with age seems to influence this score. The conclusions hold for measures in 2002 and 2004 as well as for analyses that include all cases or only those that were interviewed without an appointment. In short, we were not able to reproduce the time-of-day of interview and age interaction effects from laboratory settings in the same measures when administered in a survey and can therefore not confirm hypothesis H4.1.

4.6 DISCUSSION AND CONCLUSION

Studies using survey data have shown that the age-related decline in cognitive functioning affects the quality of survey data, confirming psychological experiments. Cognitive functioning itself varies across the day because it is subject, as are so many other body functions to a circadian rhythm. While this variation is barely detectable for younger respondents, the elderly show cognitive functioning deficits later in the day and score significantly lower on cognitive tests. We hypothesized, based on these findings, that the elderly should provide higher quality data when interviewed in the morning than in the evening. The analyses of this study have shown that the time-of-day when elderly people are interviewed does not have any effect on the likelihood of “don’t know” answers even if the questions are cognitively quite demanding. We were also not able to replicate the finding that elderly perform significantly worse on cognitive functioning tests when they are administered in the evening, though the Health and Retirement Study uses mostly the same measures that have also been employed in laboratory studies.

From the perspective of a survey methodologist these results are good news. The time-of-day does not seem to introduce any additional measurement error in survey data. It is possible that survey interviews are not as sensitive as laboratory studies to circadian effects in cognitive functioning. The presence of the interviewer or the more conversational aspects of a survey interview could

contribute to alleviate the effects found in the laboratory studies. It seems, however, to be too early to draw these conclusions without further research. We were not able to control very well in our study for possible self-selection effects by using an indicator if the interview was based on an appointment, or not. We also do not have any measures of whether the elderly have taken a nap in the afternoon and were therefore able to enhance their cognitive performance later during the day. In order to rule out any of these influences that could possibly bias our analyses a randomized experiment is needed that assigns younger and older respondents to being interviewed in the morning compared to the evening. Because these scheduling constraints are likely to increase nonresponse designing measures that can be included in analyses as control variables might be a more practical alternative.

CHAPTER 5

CONCLUSION

Aging of humans is not only characterized by changes in physical health but also in cognitive functioning. Psychological research into these changes has documented systematic decline across the lifespan in nearly all cognitive dimensions ((Craik 2000); (Park 2000)). Evidence of this cognitive decline can be expected especially in situations that are highly cognitively demanding and require the extensive use of cognitive resources, such as a survey interview. The question-answering process is known to be highly dependent on basic cognitive abilities (Tourangeau, Rips et al. 2000). It is therefore not surprising that the few studies conducted in this area of survey research have shown that old age and related cognitive decline interacts with the survey process and negatively affects the quality of survey data (see the contributions in (Schwarz, Park et al. 1998); (Knäuper, Schwarz et al. 2004)). With the continuous growth of older segments of the population, policy makers are more and more in need of high quality data. Thus, survey methodologists need to gain a deeper understanding on how age-related cognitive decline affects the survey process so that methods and adaptations can be developed to minimize survey errors that are unique to this population segment.

This dissertation addressed three possible areas in which interactions between the age-related cognitive decline and the survey process can be expected. The first study in Chapter 2 examined if cognitive decline influenced respondents' choice with regard to response strategies when answering behavioral frequency questions. We could confirm that older respondents compared to younger respondents are indeed more likely to choose rate-based or general impression strategies than episodic enumeration. Further analyses

revealed that, against initial expectations, older respondents are not more likely to provide inaccurate reports or more over- than underreports.

Chapter 3 addressed how cognitive aging influences different types of panel attrition while controlling for decline in physical health and changes in social network. This study showed that cognitive aging, decline in physical health and changes in social networks have differential influence on refusal and proxy-interview compared to self-interview. Cognitive functioning and physical health measures were important predictors for proxy-interviewers but not for refusals. Refusals in later waves of a panel might be more likely linked to distinct experiences in previous waves that could not be included in this study.

The last chapter examined if the time-of-day of interview could have a significant effect on data quality of survey responses by older respondents. We were not able to find any effect of time-of-day of interview on the likelihood of “don’t know” responses. Furthermore, we were also not able to replicate the significant differences in cognitive functioning found by laboratory studies when testing elderly at different times throughout the day.

This dissertation has shown, confirming previous studies, that cognitive aging can influence survey errors. Because of the increase of older people in our society and therefore also in our surveys, more research is needed to better understand when cognitive aging causes survey errors, and when not.

REFERENCES

- Abelson, R. P. (1981). "Psychological status of the script concept." American Psychologist **36**(7): 715-729.
- Albert, J., E. Wilson, et al. (1975). "Attrition in longitudinal study of aged populations - characteristics of drop-outs." Gerontologist **15**(5): 99.
- Allison, P. D. (1982). "Discrete-Time Methods for the Analysis of Event Histories." Sociological Methodology **13**: 61-98.
- Anderson, M. J., T. V. Petros, et al. (1991). "Individual differences in the effect of time of day on long-term memory access." American Journal of Psychology **104**: 241-255.
- Anstey, K. J. and M. A. Luszcz (2002). "Selective non-response to clinical assessment in the longitudinal study of aging: implications for estimating population levels of cognitive function and dementia." International Journal of Geriatric Psychiatry **17**: 704-709.
- Auriat, N. (1993). "'My Wife Knows Best': A Comparison of Event Dating Accuracy Between the Wife, the Husband, the Couple, and the Belgium Population Register." Public Opinion Quarterly **57**(2): 165-190.
- Baddeley, A. D., J. E. Hatter, et al. (1970). "Memory and Time of Day." Quarterly Journal of Experimental Psychology **22**: 605-609.
- Beatty, P. and D. Herrmann (2002). To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse. Survey Nonresponse. R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little. New York, Wiley: 71-86.
- Behr, A., E. Bellgardt, et al. (2005). "Extent and Determinants of Panel Attrition in the European Community Household Panel." European Sociological Review **21**(5): 489-512.
- Bickart, B. A., J. M. Phillips, et al. (2006). "The effects of discussion and question wording on self and proxy reports of behavioral frequencies." Market Letters **17**: 167-180.

Biemer, P. P. and L. E. Lyberg (2003). Introduction to Survey Quality. Hoboken, N.J., John Wiley & Sons, Inc.

Billiet, J. and G. Loosveldt (1988). "Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training." Public Opinion Quarterly **52**(2): 190-211.

Blair, E. and S. Burton (1987). "Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions." The Journal of Consumer Research **14**(2): 280-288.

Blake, M. J. F. (1967). "Time of day effects on performance in a range of tasks." Psychonomic Science **9**: 349-350.

Bonnefond, A., O. Rohmer, et al. (2003). "Interaction of age with time of day and mental load in different cognitive tasks." Perceptual and Motor Skills **96**: 1223-1236.

Bradburn, N. M., L. J. Rips, et al. (1987). "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys." Science **236**(4798): 157-161.

Brodsky, H., N. M. Kemp, et al. (2004). "Characteristics of the GPCOG, a screening tool for cognitive impairment." International Journal of Geriatric Psychiatry **19**(9): 870-874.

Brown, N. R. (1995). "Estimation Strategies and the Judgment of Event Frequency." Journal of Experimental Psychology: Learning, Memory, and Cognition **21**(6): 1539-1553.

Brown, N. R. (1997). "Context Memory and the Selection of Frequency Estimation Strategies." Journal of Experimental Psychology: Learning, Memory, and Cognition **23**(4): 898-914.

Bugg, J. M., E. L. DeLosh, et al. (2006). "Physical Activity Moderates Time-of-Day Differences in Older Adults' Working Memory Performance." Experimental Aging Research **32**: 431-446.

Burgess, R. D. (1989). Major Issues and Implications of Tracing Survey Respondents. Panel Surveys. D. Kasprzyk, G. J. Duncan, G. Kalton and M. P. Singh. New York, John Wiley & Sons: 52-74.

Burton, S. and E. Blair (1991). "Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys." Public Opinion Quarterly **55**(1): 50-79.

- Cannell, C. F., P. V. Miller, et al. (1981). "Research on Interviewing Techniques." Sociological Methodology **12**: 389-437.
- Chang, L. and J. A. Krosnick (2003). "Measuring the Frequency of Regular Behaviors: Comparing the "Typical Week" to the "Past Week"." Sociological Methodology, vol. 33, pp. 55-80 **33**: 55-80.
- Chatfield, M., C. Brayne, et al. (2005). "A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies." Journal of Clinical Epidemiology **58**: 13-19.
- Colsher, P. L. and R. B. Wallace (1989). "Data quality and age: health and psychobehavioral correlates of item nonresponse and inconsistent responses." Journal of Gerontology **44**(2): P45-52.
- Conrad, F. G., N. R. Brown, et al. (1998). "Strategies for Estimating Behavioural Frequency in Survey Interviews." Memory **6**(4): 339-366.
- Couper, M. P. and M. B. Ofstedal (2006). Keeping in Contact with Mobile Sample Members. MOLS 2006: Methodology of Longitudinal Surveys. University of Essex, Colchester, UK.
- Craik, F. I. M. (2000). Age-related changes in human memory. Cognitive aging: a primer. D. Park and N. Schwarz. Philadelphia, PA, Psychology Press: 75-92.
- Craik, F. I. M., R. Govoni, et al. (1996). "The Effects of Divided Attention on Encoding and Retrieval Processes in Human Memory." Journal of Experimental Psychology: General **125**(2): 159-180.
- Craik, F. I. M. and J. M. McDowd (1987). "Age differences in recall and recognition." Journal of Experimental Psychology: Learning, Memory, and Cognition **13**(3): 474-479.
- Crook III, T. H., J. R. Youngjohn, et al. (1992). "Aging and Everyday Memory: A Cross-Cultural Study." Neuropsychology **6**(2): 123-136.
- De Leeuw, E. D. (1992). Data Quality in Mail, Telephone and Face-to-face Surveys. Amsterdam, TT-Publikaties.
- Ebbinghaus, H. (1885). Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Darmstadt, Wissenschaftliche Buchgesellschaft.
- Feng, D., M. Silverstein, et al. (2006). "Attrition of Older Adults in Longitudinal Surveys: Detection and Correction of Sample Selection Bias Using Multigenerational Data." Journal of Gerontology: Social Sciences **61B**(6): S323-S328.

Fitzgerald, J., P. Gottschalk, et al. (1998). "An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics." Journal of Human Resources **33**(2): 300-344.

Fonda, S. and A. R. Herzog (2004). Documentation of Physical Functioning Measured in the Health and Retirement Study and the Asset and Health Dynamics among the Oldest Old Study. Ann Arbor, University of Michigan.

Froot, K. A. (1989). "Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data." Journal of Financial and Quantitative Analysis **24**: 333-355.

Gerstorff, D., A. Herlitz, et al. (2006). "Stability of Sex Differences in Cognition in Advanced Old Age: The Role of Education and Attrition." Journal of Gerontology: Psychological Sciences **61B**(4): P245-P249.

Ghisletta, P. and D. Spini (2004). "An Introduction to Generalized Estimating Equations and an Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals." Journal of Educational and Behavioral Statistics **29**(4): 421-437.

Goldberg, M., J. F. Chastang, et al. (2006). "Health problems were the strongest predictors of attrition during follow-up of the GAZEL cohort." Journal of Clinical Epidemiology **59**: 1213-1221.

Gray, R., P. Campanelli, et al. (1996). "Exploring Survey Non-Response: The Effect of Attrition on a Follow-Up of the 1984-85 Health and Life Style Survey." The Statistician **45**(2): 163-183.

Groves, R. M. (1989). Survey Errors and Survey Costs. New York, John Wiley & Sons.

Groves, R. M. and M. P. Couper (1998). Nonresponse in Household Interview Surveys. New York, John Wiley & Sons.

Harris-Kojetin, B. A. and C. Tucker (1998). "Longitudinal Nonresponse in the Current Population Survey (CPS)." ZUMA Nachrichten Spezial: 264-272.

Hart, T. C., C. M. Rennison, et al. (2005). "Revisiting Respondent "Fatigue Bias" in the National Crime Victimization Survey." Journal of Quantitative Criminology **21**(3): 345-363.

Hasher, L., C. Chung, et al. (2002). "Age, Time of Testing, and Proactive Interference." Canadian Journal of Experimental Psychology **56**(3): 200-207.

Hasher, L., D. Goldstein, et al. (2005). It's About Time: Circadian Rhythms, Memory, and Aging. Human Learning and Memory: Advances in Theory and Application. C. Izawa and N. Ohta. Mahwah, NJ, Lawrence Erlbaum Associates: 199-218.

Herzog, A. R. and W. L. Rodgers (1988). "Age and response rates to interview sample surveys." Journal of Gerontology **43**(6): S200-205.

Herzog, A. R., W. L. Rodgers, et al. (1983). "Interviewing older adults: A comparison of telephone and face to face modalities." Public Opinion Quarterly **47**: 405-418.

Hill, D. H. and R. J. Willis (2001). "Reducing Panel Attrition: A Search for Effective Policy Instruments." The Journal of Human Resources **36**(3): 416-438.

Hintzman, D. L. (1988). "Judgments of Frequency and Recognition Memory in a Multiple-Trace Memory Model." Psychological review **95**(4): 528-551.

Horne, J. A. and O. Ostberg (1976). "A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms." International Journal of Chronobiology **4**(2): 97-110.

Horne, J. A. and O. Ostberg (1977). "Individual differences in human circadian rhythms." Biological Psychology **5**(3): 179-190.

Hoyer, W. J. and P. Verhaeghen (2006). Memory Aging. Handbook of the Psychology of Aging. J. E. Birren and K. W. Schaie. Burlington, MA, Elsevier Academic Press.

Intons-Peterson, M. J., P. Rocchi, et al. (1998). "Aging, Optimal Testing Times, and Negative Priming." Journal of Experimental Psychology: Learning, Memory, and Cognition **24**(2): 362-376.

Intons-Peterson, M. J., P. Rocchi, et al. (1999). "Age, Testing at Preferred or Nonpreferred Times (Testing Optimality), and False Memory." Journal of Experimental Psychology: Learning, Memory, and Cognition **25**(1): 23-40.

Jones, A. M., X. Koolman, et al. (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 the Royal Statistical Society A **169**(3): 543-569.

Kalton, G., D. Kasprzyk, et al. (1989). Nonsampling Errors in Panel Surveys. Panel Surveys. D. Kasprzyk, G. J. Duncan, G. Kalton and M. P. Singh. New York, Wiley: 249-270.

- Kapteyn, A., P.-C. Michaud, et al. (2006). Effects of Attrition and Non-Response in the Health and Retirement Study. Santa Monica, RAND.
- Kautter, J., G. Khatutsky, et al. (2006). "Impact of Nonresponse on Medicare Current Beneficiary Survey Estimates." Health Care Financing Review **27**(4): 71-93.
- Kempen, G. I. J. M. and E. van Sonderen (2002). "Psychological attributes and changes in disability among low-functioning older persons: Does attrition affect the outcomes?" Journal of Clinical Epidemiology **55**: 224-229.
- Kennison, R. F. and E. M. Zelinski (2005). "Estimating Age Change in List Recall in Asset and Health Dynamics of the Oldest-Old: The Effects of Attrition Bias and Missing Data Treatment." Psychology and Aging **20**(3): 460-475.
- Knäuper, B. (1998). Age Differences in Question and Response Order Effects. Cognition, Aging, and Self-Reports. N. Schwarz, D. Park, B. Knäuper and S. Sudman. Philadelphia, PA, Psychology Press: 341-364.
- Knäuper, B. (1999). "The impact of age and education on response order effects in attitude measurement." Public Opinion Quarterly **63**: 347-370.
- Knäuper, B., R. F. Belli, et al. (1997). "Question difficulty and respondents' cognitive ability. The impact on data quality." Journal of Official Statistics **13**(2): 181-199.
- Knäuper, B., N. Schwarz, et al. (2004). "Frequency reports across age groups." Journal of Official Statistics **20**(1): 91-96.
- Knäuper, B. and H.-U. Wittchen (1994). "Diagnosing major depression in the elderly: Evidence for response bias in standardized diagnostic interviews?" Journal of Psychiatric Research **28**(2): 147-164.
- Kogan, N. (1961). "Attitudes toward old people in an older sample." The Journal of Abnormal and Social Psychology **62**(3): 616-622.
- Krishnan, E., K. Murtagh, et al. (2004). "Attrition Bias in Rheumatoid Arthritis Databanks: A Case Study of 6346 Patients in 11 Databanks and 65,649 Administrations of the Health Assessment Questionnaire." The Journal of Rheumatology **31**(7): 1320-1326.
- Lee, L., A. Brittingham, et al. (1999). "Are Reporting Errors Due to Encoding Limitations or Retrieval Failure? Surveys of Child Vaccination as a Case Study." Applied Cognitive Psychology **13**: 43-63.

Lepkowski, J. M. and M. P. Couper (2002). Nonresponse in the Second Wave of Longitudinal Household Surveys. Survey Nonresponse. R. M. Groves, D. Dillman, J. L. Eltinge and R. J. A. Little. New York, Wiley & Sons: 259-272.

Levin, B. E., H. L. Katzen, et al. (2000). "Cognitive Decline Affects Subject Attrition in Longitudinal Research." Journal of Clinical and Experimental Neuropsychology **22**(5): 580-586

Lewandowsky, S. and P. W. Smith (1983). "The effect of increasing the memorability of category instances on estimates of category size." Memory & Cognition **11**: 347-350.

Little, R. J. A. and D. B. Rubin (1987). Statistical Analysis with Missing Data. New York, Wiley.

Long, J. S. (1997). Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, Sage.

Loosveldt, G. and A. Carton (2002). "Utilitarian Individualism and Panel Nonresponse." International Journal of Public Opinion Research **14**(4): 428-438.

Martin, S. A., M. R. Haren, et al. (2007). "The Florey Adelaide Male Ageing Study (FAMAS): Design, procedures & participants." BMC Public Health **7**: 126-141.

Matthews, F. E., M. Chatfield, et al. (2006). "An investigation of whether factors associated with short-term attrition change or persist over ten years: data from the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS)." BMC Public Health **6**: 185-195.

Matthews, F. E., M. Chatfield, et al. (2004). "Attrition and bias in the MRC cognitive function and ageing study: an epidemiological investigation." BMC Public Health **4**(12).

May, C. P. and L. Hasher (1998). "Synchrony effects in inhibitory control over thought and action." Journal of Experimental Psychology: Human Perception and Performance **24**: 363-379.

May, C. P., L. Hasher, et al. (2005). "Implicit Memory, Age, and Time of Day." Psychological Science **16**(2): 96-100.

May, C. P., L. Hasher, et al. (1993). "Optimal time of day and the magnitude of age differences in memory." Psychological Science **4**: 326-330.

Menon, G. (1993). "The Effects of Accessibility of Information in Memory on Judgments of Behavioral Frequencies." The Journal of Consumer Research **20**(3): 431-440.

- Menon, G., P. Raghuram, et al. (1995). "Behavioral Frequency Judgments: An Accessibility-Diagnosticity Framework." The Journal of Consumer Research **22**(2): 212-228.
- Mihelic, A. H. and E. M. Crimmins (1997). "Loss to Follow-up in a Sample of Americans 70 Years of Age and Older: The LSOA 1984-1990." Journal of Gerontology: Social Sciences **52B**(1): S37-S48.
- Mirowsky, J. and J. R. Reynolds (2000). "Age, Depression, and Attrition in the National Survey of Families and Households." Sociological Methods & Research **28**(4): 476-504.
- Moore, R. Y. (1997). "Circadian Rhythms: Basic Neurobiology and Clinical Applications." Annual Reviews in Medicine **48**: 253-266.
- Morwitz, V. G. (1997). "It Seems like Only Yesterday: The Nature and Consequences of Telescoping Errors in Marketing Research." Journal of Consumer Psychology **6**(1): 1-29.
- Mutter, S. A. and K. M. Goedert (1997). "Frequency Discrimination vs Frequency Estimation: Adult Age Differences and the Effect of Divided Attention." The Journals of Gerontology **52B**(6): P319-P328.
- Nelson, L. M., W. T. J. Longstreth, et al. (1994). "Completeness and Accuracy of Interview Data from Proxy Respondents: Demographic, Medical, and Life-Style Factors." Epidemiology **5**(2): 204-217.
- Neter, J. and J. Waksberg (1964). "A Study of Response Errors in Expenditures Data from Household Interviews." Journal of the American Statistical Association **59**(305): 18-55.
- Ofstedal, M. B., G. G. Fisher, et al. (2005). Documentation of Cognitive Functioning Measures in the Health and Retirement Study. Ann Arbor, University of Michigan.
- Online, E. B. (2006). "Aging." Retrieved November 27, 2006, from <http://www.britannica.com/eb/article-9110653>.
- Pandelaere, M. and V. Hoorens (2006). "The Effect of Category Focus at Encoding on Category Frequency Estimation Strategies." Memory & Cognition **34**(1): 28-40.
- Park, D. (1999). Cognitive Aging, Processing Resources, and Self-Report. Cognition, Aging, and Self-Reports. N. Schwarz, D. Park, B. Knäuper and S. Sudman. Philadelphia, PA, Psychology Press.

Park, D. (2000). The basic mechanisms accounting for age-related decline in cognitive function. Cognitive aging: a primer. D. Park and N. Schwarz. Philadelphia, PA, Psychology Press: 3-22.

Park, D. and A. H. Gutchess (2000). Cognitive aging and every day life. Cognitive aging: a primer. D. Park and N. Schwarz. Philadelphia, PA, Psychology Press: 217-232.

Park, D. C., A. D. Smith, et al. (1989). "Effects of Age and a Divided Attention Task Presented During Encoding and Retrieval on Memory." Journal of Experimental Psychology: Learning, Memory, and Cognition **15**: 1185-1191.

Petros, T. V., B. E. Beckwith, et al. (1990). "Individual differences in the effects of time of day and passage difficulty on prose memory in adults." British Journal of Psychology **81**: 63-72.

Pickery, J., G. Loosveldt, et al. (2001). "The Effects of Interviewer and Respondent Characteristics on Response Behavior in Panel Surveys: A Multilevel Approach." Sociological Methods & Research **29**(4): 509-523.

Rabbitt, P., M. Lunn, et al. (2006). "Understanding Terminal Decline in Cognition and Risk of Death: Methodological and Theoretical Implications of Practice and Dropout Effects." European Psychologist **11**(3): 164-171.

Raghunathan, T., J. Lepkowski, et al. (2001). "A multivariate technique for multiply imputing missing values using a sequence of regression models." Survey Methodology **27**(1): 85-95.

Raz, N. (2000). Aging of the brain and its impact on cognitive functioning: integration of structural and functional findings. The Handbook of Aging and Cognition. F. I. M. Craik and T. A. Salthouse. Mahwah, NJ, Lawrence Erlbaum Associates: 1-90.

Ritchie, L. J. and H. Tuokko (2007). "Neuropsychological prediction of attrition due to death." Journal of Clinical and Experimental Neuropsychology **29**(4): 385-394.

Schmidt, C., F. Collette, et al. (2007). "A time to think: Circadian rhythms in human cognition." Cognitive Neuropsychology **24**(7): 755-789.

Schneider, A. L. (1981). "Methodological Problems in Victim Surveys and Their Implications for Research in Victimology." The Journal of Criminal Law and Criminology **72**(2): 818-838.

Schwarz, N. (1998). Self-Reports of Behaviors and Opinions: Cognitive and Communicative Processes. Cognition, Aging, and Self-Reports. N. Schwarz, D. Park, B. Knäuper and S. Sudman. Philadelphia, PA, Psychology Press: 17-44.

Schwarz, N. (1999). "Self Reports: How the Questions Shape the Answers." American Psychologist **54**(2): 93-105.

Schwarz, N., H.-J. Hippler, et al. (1985). "Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments." Public Opinion Quarterly **49**(3): 388-395.

Schwarz, N., B. Knäuper, et al. (in press). Cognition, aging and self-reports. Cognitive aging: a primer. D. Park and N. Schwarz. Philadelphia, PA, Psychology Press.

Schwarz, N., D. Park, et al., Eds. (1998). Cognition, Aging, and Self-Reports. Philadelphia, PA, Psychology Press.

Sharma, S., J. D. Tobin, et al. (1982). "Factors Influencing the Attrition of Subjects in the Baltimore Longitudinal Study of Aging." Gerontologist **22**: 174.

Sharma, S. K., J. D. Tobin, et al. (1986). "Factors Affecting Attrition in the Baltimore Longitudinal Study of Aging." Experimental Gerontology **21**: 329-340.

Siedlecki, K. L. (2007). "Investigating the Structure and Age Invariance of Episodic Memory Across the Adult Lifespan." Psychology and Aging **22**(2): 251-268.

Siegler, I. C. and J. Botwinick (1979). "A Long-Term Longitudinal Study of Intellectual Ability of Older Adults: The Matter of Selective Subject Attrition." Journal of Gerontology **34**(2): 242-245.

Sliwinski, M. J., S. M. Hofer, et al. (2003). "Modeling Memory Decline in Older Adults: The Importance of Preclinical Dementia, Attrition, and Chronological Age." Psychology and Aging **18**(4): 658-671.

Steffens, D. C. and G. G. Potter (2008). "Geriatric depression and cognitive impairment." Psychological Medicine **38**(2): 163-175.

Steinhauser, K. E., E. C. Clipp, et al. (2006). "Identifying, recruiting, and retaining seriously-ill patients and their caregivers in longitudinal research." Palliative Medicine **20**: 745-754.

Strömberg, A. S., P. Sjogren, et al. (2005). "A Longitudinal Study of Palliative Care: Patient-Evaluated Outcome and Impact of Attrition." Cancer **103**(8): 1747-1755.

Strube, G. (1987). Answering survey questions: the role of memory. Social Information Processing and Survey Methodology. H.-J. Hippler, N. Schwarz and S. Sudman. New York, Springer.

Sudman, S. and N. M. Bradburn (1973). "Effects of Time and Memory Factors on Response in Surveys." Journal of the American Statistical Association **68**(344): 805-815.

Sudman, S., N. M. Bradburn, et al. (1996). Thinking about answer: The application of cognitive processes to survey methodology. San Francisco, Jossey-Bass.

Sudman, S., A. Finn, et al. (1984). "The Use of Bounded Recall Procedures in Single Interviews." Public Opinion Quarterly **48**(2): 520-524.

Sullivan, E. V. and S. Corkin (1984). "Selective Subject Attrition in a Longitudinal Study of Head-Injured Veterans." Journal of Gerontology **39**(6): 718-720.

Tankova, I., A. Adan, et al. (1994). "Circadian typology and individual differences. A review." Personality and Individual Differences **16**: 671-684.

Tennstedt, S. L., U. Dettling, et al. (1992). "Refusal Rates in a Longitudinal Study of Older People: Implications for Field Methods." Journal of Gerontology **47**(6): S313-S318.

Toh, R. S. and M. Y. Hu (1996). "Natural Mortality and Participation Fatigue as Potential Biases in Diary Panels: Impact of Some Demographic Factors and Behavioral Characteristics on Systematic Attrition." Journal of Business Research **35**: 129-138.

Tourangeau, R., L. J. Rips, et al. (2000). The Psychology of Survey Response. Cambridge, Cambridge University Press.

Tversky, A. and D. Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases." Science **185**(4157): 1124-1131.

U.S. Census Bureau. Information and Research Services, P. D., Population Projections Branch (2006). Interim Projections by Age, Sex, Race and Hispanic Origin. .

Van Beijsterveldt, C. E. M., M. P. J. Van Boxtel, et al. (2002). "Predictors of attrition in a longitudinal cognitive ageing study: The Maastricht Aging Study (MAAS)." Journal of Clinical Epidemiology **55**: 216-223.

Van den Berg, B., P. Van der Velden, et al. (2007). "Selective attrition and bias in longitudinal health survey among survivors of a disaster." BMC Medical Research Methodology **7**(8).

Waite, L. (2004). The Demographic Faces of Eldery. Aging, Health, and Public Policy: Demographic and Economic Perspectives. Supplement to Population and Development Review. L. Waite. New York, Population Council. **30**.

Wan, M. S., V. A. Velkoff, et al. (2005). 65+ in the United States. Current Population Reports. Washington, DC, U.S. Census Bureau.

Watson, D. (2003). "Sample Attrition between Waves 1 and 5 in the European Community Household Panel." European Sociological Review **19**(4): 361-378.

West, R., K. J. Murphy, et al. (2002). "Effects of Time of Day on Age Differences in Working Memory." Journal of Gerontology: Psychological Sciences **57B**(1): P3-P10.

Whiting, W. L. and A. D. Smith (1997). "Differential age-related processing resource limitations in recall and recognition tasks." Psychology and Aging **12**: 216-224.

Williams, R. L. (2000). "A note on robust variance estimation for cluster-correlated data." Biometrics **56**: 645-646.

Winocur, G. and L. Hasher (2002). Circadian Rhythms and Memory in Aged Humans and Animals. Neuropsychology of memory. L. R. Squire and D. L. Schacter. New York, The Guilford Press: 273-285.

Winocur, G. and L. Hasher (2004). "Age and time-of-day effects on learning and memory in a non-matching-to-sample test." Neurobiology of Aging **25**: 1107-1115.

Yoon, C. (1997). "Age Differences in Consumers' Processing Strategies: An Investigation of Moderating Influences." Journal of Consumer Psychology **24**: 329-342.

Yoon, C., C. P. May, et al. (1999). Aging, Circadian Arousal Patterns, and Cognition. Cognition, Aging and Self-Reports. N. Schwarz, D. Park, B. Knäuper and S. Sudman. Philadelphia, PA, Psychology Press: 117-145.

Young, A. F., J. R. Powers, et al. (2006). "Attrition in longitudinal studies: who do you lose?" Australian and New Zealand Journal of Public Health **30**(4): 353-361.

Zabel, J. E. (1998). "An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an

Application to a Model of Labor Market Behavior." The Journal of Human Resources **33**(2): 479-506.