Experimental analysis of survey response bias over the internet: Some results from the Retirement Perspectives Survey

Daniel McFadden
Department of Economics, UC Berkeley
mcfadden@econ.berkeley.edu

Joachim Winter
Sonderforschungsbereich 504, University of Mannheim
Department of Economics, UC Berkeley
winter@econ.berkeley.edu

October 30, 2001
(incomplete and preliminary)

Abstract: There is overwhelming empirical evidence that cognitive limitations and social interactions lead to biases in responses to survey questions. In addition, there is evidence that some of the underlying processes are moderated by age. The purpose of the Berkeley Internet Virtual Laboratory (IVLab) is to study these phenomena using experimental surveys conducted over the internet. The internet is a cost-effective and flexible way to conduct experiments on survey response bias. There is also a good chance that field surveys will be conducted over the internet in the near future. However, selection problems are a major concern, in particular when the target population consists primarily of older individuals. In this paper, we present a research strategy that allows to address selection problems and the analysis of survey response biases in a consistent framework. For illustration, we report some preliminary results based on data from a pilot study for the Retirement Perspectives Survey (RPS), an experimental survey conducted over the internet with a target population of older Americans.

Keywords: survey methods; measurement error; selection models; cognitive biases

JEL classification: C42; C81

We are grateful to the UC Retirement Center and the UC Alumni Association for their cooperation in recruiting participants for our online survey. We received helpful suggestions and comments from Rob Alessie, Marcel Das, Bas Donkers, Mike Hurd, Danny Kahneman, Arie Kapteyn, Chuck Manski, Stan Panis, Charlie Plott, Norbert Schwarz, Jim Smith and Arthur van Soest. Peter Adams and Rebecca Hellerstein provided valuable research assistance. Financial support for this project was provided by the National Institute of Aging (NIA) through a grant to the National Bureau of Economic Research (NBER), and by the European Union through the TMR project "Savings, pensions, and portfolio choice."

1. Introduction

There is overwhelming empirical evidence that cognitive limitations and social interactions lead to biases in responses to survey questions. A well-documented example is the "unfolding brackets" technique. While this question design can successfully reduce survey non-response, it introduces anchoring effects that might distort responses significantly (see, *inter alia*, Hurd, *et al*., 1998; Hurd 1999). Similarly, in questions on subjective expectations about future income, health conditions, or survival probabilities, the format and the wording of questions can influence responses heavily (for example, Hurd, McFadden, and Gan, 1998).[1] However, at least with respect to large socio-economic field surveys, there are only few systematic attempts to reduce and control for such biases.

The purpose of the Internet Virtual Laboratory (IVLab) is to acquire a better understanding of the cognitive processes that govern survey response behavior, in particular in economic surveys, and to develop and test question formats that help to alleviate response biases. While traditional approaches to survey experimentation and to pre-testing usually require expensive computer-aided personal interview techniques, the internet offers cost-effective and fast access to test subjects and allows for very flexible experimental survey design.

However, sample selection is a major concern when it comes to internet surveys. An important aim of the project is therefore to learn about the characteristics of individuals who are willing to participate in economic surveys over the internet, and to understand aspects of response behavior that are specific to internet-based surveys. This will allow us to design statistical methods for correcting biases introduced by selective participation in internet surveys. In addition to improving survey methodology, these results will also help to decide whether and to what extent economic field surveys themselves can eventually be administered over the internet.

The use of the internet for collecting survey data is attractive because of speed, cost effectiveness, and technical opportunities such as visual presentation of questions that are available only with the aid of a computer. However, internet surveys are also likely to be subject to potentially serious de-

[1] There is a large literature on survey design in cognitive and social psychology. We refer the interested reader to Krosnick (1999) and Schwarz (1999) for recent surveys, and to Moxey and Sanford (1993) and Sudman, Bradburn, and Schwarz (1996) for more details.

sign problems, sample selectivity and mode effects being the most obvious candidates. Much research is still needed in both areas. In section 2 of this paper, we discuss internet access in the US and implications for conducting representative surveys over the internet. With respect to mode effects and design opportunities in internet surveys, we refer to Couper, *et al.* (2001) for recent evidence and to Dillman (2000) for a comprehensive, hands-on discussion.

The aim of this paper is to develop a research strategy that allows to address selection problems and the analysis of survey response biases in internet surveys in a consistent framework. For illustration, we present some preliminary results based on a pilot study for the Retirement Perspectives Survey (RPS), an experimental survey conducted over the internet with a target population of U.S. citizens over 40.

The remainder of this paper is structured as follows. Section 2 contains some general remarks on internet access in the U.S. population and the differential impact of demographic characteristics. In section 3, we describe a pilot version of Retirement Perspectives Survey (RPS), and internet survey that we have conducted over the last few months with a target population of older Americans in mind. Section 4 contains a few preliminary results. Section 5 concludes.

2. Internet access in the U.S.

As pointed out by Couper (2000), coverage and sampling error are currently the biggest threat to inference from internet surveys. There are two related problems: First, currently only 59% of the U.S. population have access to the internet, and those who have access are significantly different from those how have no internet access. These differences occur along many important socio-demographic dimensions. Second, and at a more practical level, constructing a sample frame for internet surveys is much more difficult than in, say, a telephone survey where methods such as random digit dialing (RDD) are available. Currently, no mechanism similar to RDD is available for the internet. In this section, we concentrate on the first problem, describing the demographic characteristics of those who currently have internet access. We return to the more practical problem of sampling in Section 5.

In the following, we present some findings about internet access of the U.S. population based on supplements to the Current Population Survey (CPS) in October 1997, December 1998, and August 2000. In addition, we analyze questions on whether any household member owned a personal com-

puter that were part of the November 1994 CPS and the three later internet access supplements. Since the CPS is very close to being representative of the U.S. population, these numbers are arguably the most reliable information available on internet use to date. A main advantage of the CPS is its wealth of demographic variables. Bikson and Panis (1999) present a detailed and careful analysis of the October 1997 CPS data, together with earlier data on internet access and use obtained from other sources. Our analysis for 1997, 1998, and 2000 can be seen as an update. Data from the most recent and similar future CPS waves will also be extremely useful when it comes to computing population weights for data obtained from internet surveys.

For interpretation of the following figures it is important that they refer to individuals as reference persons, but the questions on computer and internet access and usage refer to the entire household. This implies that a person will be classified as having internet access even if he or she does not use the internet at all (but some other member of the household does). The wording of the computer ownership questions varies substantially across the four waves. Question on internet access in the three later waves do not match exactly, but are comparable. Details can be found in Adams and Winter (2001).

Figure 1 shows computer ownership in private homes in the U.S. population as a function of age. The increase in computer ownership rates across all age groups is apparent: In November 1994, ownership rates were below 40% for all age groups, in August 2000, they are above 50% for persons aged through 60, and well above 60% for persons younger than 22 and between 30 through 53. For the elderly, the fraction of persons with a computer in the household drops below 20% at about age 80.

Figures 2 and 3 focuses on overall internet access, combining access at home and elsewhere (more detailed figures are in Adams and Winter, 2001). Figure 2 shows internet access rates by education and as a function of age. There is a strong education effect in internet access; relative to the population average, numbers for those with a college or higher degree are much higher, and for those without a high-school degree much lower. This feature of the data has not changed between the October 1997 and August 2000 waves (right and left panels of figure 2, respectively). While internet access rates increased at all ages and educational levels, they are still only around 20% for persons with low education. More importantly, the age profile for persons with low education is rather flat. It is therefore unlikely that access rates increase over time due to a pure age effect (which is

certainly at work for the highly educated). These findings confirm a more detailed analysis of the dynamics of internet access through 1997 by Bikson and Panis (1999).

With respect to the potential for internet surveys of the elderly, we should stress that the age profiles of internet access rates show a peak at very young ages (under 20), and then a hump between the ages of 30 and 60. They are sharply declining with increasing age after retirement. But even for over 90 year olds, there is still a fraction of 10% who have access to the internet (possibly via other household members).

These findings are generally well in line with those reported by Rodgers and Willis (2000). Their analysis is based on a question on internet access that was asked in the HRS 1999 and therefore restricted to individuals aged 50 and above. With respect to the age effects shown in our figures, it is important to stress that our cross-sectional data does not allow to disentangle pure age, and cohort, and time effects. It is very likely that these age profiles reflect mostly cohort effects, and that internet access of the elderly will increase over the next decades as these profiles shift to the right (see also Rodgers and Willis, 2000). However, it is hard to predict what overall internet access will be in ten or twenty years. It might well be that a number around 80% or so is an upper bound on internet access for all age groups, just like 95% has proved to be a stable upper bound for the proportion of households with own telephones (see also Couper, 2000). According to The Harris Poll, a commercial research institute that focuses on internet use, the growth of the internet population has slowed down for the first time since the first of their polls in 1959.[2]

Finally, Figure 3 shows a strong effect of household income on internet access. Again, this effect is present in all three waves. In the most recent August 2000 data, internet access rates are above 80% for high-income persons and around 20% for persons with low incomes. The spike at very low incomes, visible in all three waves but become more distinct over time is due to college students with own households and no or only low income.

Additional demographic effects are revealed by the logit regression of a zero-one variable on a number of demographic characteristics that are available for CPS respondents. Table 1 contains regression results for the most recent data. In addition to the age and education effects already mentioned, the effects of race are striking. The black population in the U.S. is much less likely to have

internet access than the white population (even controlling for other characteristics such as age and income). The population of metropolitan areas does not have more internet access than others, maybe surprisingly. Also, being unemployment does not have a significant effect on internet access.

In summary, results from two largely representative samples of the U.S. population – our analysis of CPS data, together with results by Bikson and Panis (1999), and that of Rodgers and Willis based on the HRS – confirm that the internet population is not representative of the population at large, and that selection effects along several important dimensions such as age, income, education, and race are severe. The strong age, income and other demographic effects in internet access rates present a serious obstacle to conducting surveys which are representative of the population over the internet. The data do not suggest that these problems disappear in the near future – internet access increases only very slowly for groups with low access rates. It is more likely that the internet can be used to collect information from specific groups of the population with high access rates, while traditional modes will continue to be needed for other groups. This perspective stresses the importance of a systematic analysis of mode effects of internet vs. traditional survey methods.

When internet surveys are used, the availability of representative data on internet access in the CPS allows to control for selection effects in internet surveys with appropriate statistical techniques (e.g., Lancaster and Imbens, 1990; McFadden, 1997, Horowitz and Manski, 1998). The analysis of the selection process into internet survey participation and the development of methods for re-weighting data are therefore important aspects of internet survey research. While these methods are unlikely to completely eliminate problems of a nonrandom sample, it is worth pointing out that experimental questions on response bias and survey design are, to some extend, unaffected by selection problems.

However, results on survey design obtained from experimental internet surveys will be conditional on the cognitive ability of participants, and it stands to reason that persons with internet access have higher cognitive ability than the population at large, in particular at older ages. Weir, *et al.* (2001) present an analysis of the effects of cognitive ability on survey response bias among the elderly, but more research is required in this area as well before data collection over the internet with elderly populations can be seriously considered.

---

[2] The Harris Poll reported that according to their phone survey conducted in August/September 2000, 59% of U.S. Adults used the internet from home, work or other locations. The numbers reported by the Harris Poll are in line with

3. The RPS pilot survey

This section summarizes our first experiences with conducting an experimental survey over the internet with a population of older participants. This survey was conducted from June 2000 through March 2001, and it is pilot version of the Retirement Perspectives Survey (RPS). The RPS is a major experimental survey aimed at understanding the selection process of participants into internet surveys and survey response bias, with special respect to older individuals.

3.1 Structure and software

The structure of the RPS is summarized in Table 2. The RPS consists of different modules, roughly ordered by increasing sensitivity of the information requested from participants. We ask for a few demographic characteristics up front so that we can investigate selection into the survey even for those who do not complete.

The RPS is administered over the internet server of the Econometrics Laboratory at Berkeley. The software for the current pilot version of the RPS is based on CGI and Perl scripts. Conditional branching and treatment assignment is realized with cookies. The answers to survey questions entered by participants are transmitted via HTTP forms, processed and written to a file. The upcoming field version of the RPS will use a different and more powerful software with full database support (realized with CGI on the internet server, a Perl parsing script developed for the special needs of the IVLab, and a mySQL relational database).

3.2 Samples

The RPS pilot survey was conducted in two waves. From June 12 through September 13, 2000, access was restricted to members of the UC Retirement Center who were solicited by an invitation in the UCRC newsletter; we refer to these participants as sample 1. During this period, 76 individuals participated in the survey. From September 14, 2000 through September 28, 2001, the survey was open to the public, and participants were solicited primarily by a link on Dan McFadden's internet

---

those that we computed based on the Current Population Survey.

homepage (sample 2); a total of 139 individuals participated during this period.[3] A few additional observations from individuals younger than 40 years were excluded.

The demographic characteristics of RPS participants are summarized in Table 3. The mean age is 58.1 years overall (65.9 and 54.0 years in the two sub-samples). Figure 3 shows the age distributions. While we have a few participants who are over 80, the absolute numbers are small given our sample size. Slightly more than half of the participants (57.2%) are male. Obviously, the sample is extremely selected with respect to education and income. Details can be found in Table 3, together with more descriptive statistics.

3.3 Response behavior

First, a few general remarks on our experiences with the pilot version of the RPS. Even though the software we used was by now means a professional solution, we have not experienced any major technical problems in administrating the survey over the internet.[4] Only very few potential participants with internet access were unable to answer the survey because of problems with their browsers; one had an old version of Netscape (2.02), the other experienced extremely slow response times, but we were not able to determine the reason for this particular problem. In addition, we received a few phone calls, mainly of people who couldn't access the survey because they typed the URL incorrectly.[5] Overall, running the survey over the internet did not seem to involve any particular technical problems.

More than three quarters of the participants have completed the entire survey, and none of the modules appears to have caused a significantly higher number of drop-outs (see Figure 4). However, the two questions in Module 2 – opinion questions on Social Security and Medicare – already resulted in 10 drop-outs, which is worse than we had hoped.

From the feedback we received, it is our impression is that those people who chose to participate were quite interested in the survey and willing to answer all questions (which we would of course

---

[3] The survey is still open to participants. Effective March 22, 2001, we implemented a revised version of the RPS with some modifications to survey design and software that allow collection of additional data on anchoring effects.

[4] During the first few days, we experienced some software problems related to the treatment structure which resulted in participants getting stuck at some point in the survey. When reported, such problem could be fixed it in a matter of a few minutes, and we believe that all significant bugs were fixed after a few days.

[5] One caller, aged 80, said she would be willing to answer such a survey, but she neither has nor otherwise uses a computer.

not want to generalize to broader samples). Our *ex ante* estimate of 30 minutes for the time required to complete the survey was right on target – the mean time reported by those whose completed the survey was slightly below 30 minutes. with one person as fast as 12 minutes, quite a few participants with 15 minutes, a median of 25 minutes, and a maximum of 90 minutes. Not surprisingly, there is a strong focal point effect in the answers to these questions – all but four stated times are multiples of five minutes.

Table 4 contains a logit regression for the binary variable describing whether participants completed the survey. The likelihood of completing the survey is increasing with education, and there is a negative effect of being male. For those who completed the survey, Table 4 also reports Poisson regressions of the time required to complete. Answering time increases in age (see also the right panel of Figure 5), decreases with education and income, and is slightly longer for males. (Excluding a few obvious outliers does not change these results.)

In general, the frequency of "Don't know" (DK) and "Refuse to answer" (RF) responses to survey question is relatively low. Table 5 contains frequencies for some selected questions. Also, there are no apparent age effects in non-response rates.

The possibility to give feedback was quite appreciated. This might not be surprising since we deal with a sample of retired college professors and employees, many of whom might themselves have some experience with empirical research in the social sciences. Participants could give feedback in two ways. They could fill in a small form on the last page of the survey (restricted to 200 characters), or they could send e-mail to an administrative address. Three of the more frequent comments are worth mentioning. The majority of explicit comments explained why some question was not applicable to the participant's personal situation. A few participants commented on percentage questions, stressing that they are hard to answer. A fair number of participants (about 10%) said that they never gamble or wager and that they are therefore unable to answer such questions. The comments on gambling questions suggest that a significant share of the older population feels strongly about this issue.

4. Some preliminary results

Our discussion of specific results is preliminary and incomplete. We concentrate on a few questions that can be analyzed even with relatively small sample and highly selected sample. The analysis of many other questions has to be deferred until we have data from the RPS field version.

4.1 Qualitative questions on income, savings, and portfolio choice

The RPS contains a sequences of qualitative questions on income, savings and portfolio choice. Most of these questions are difficult to interpret given the small and highly selected sample of our pilot study. However, it was our impression bases on feedback we received that participants generally like these questions and think that they are important. A few descriptive statistics on these questions can be found in Table 5, but we refer from specific interpretation due to the non-representative nature of our sample.

4.2 Expectations about income, consumption, and health

The RPS contains a sequence of questions on expectations about income, consumption and planning for health expenses in old age. Again, most of these questions have been used in other surveys such as the Survey of Economic Expectations. We have included random variations in numerical clues in order to investigate anchoring effects. However, current sample sizes in the RPS pilot survey are too small for reliable analysis of the built-in experimental variation.

The RPS contains three questions on subjective survival probabilities at the 10, 15, and 20 year horizons. Similar questions are asked in the HRS, and prior research has shown that subjective survival probabilities in stated by older individuals aggregate closely to life table values with some tendency to overly optimistic responses, and to vary appropriately with known risk factors (e.g., Hurd and McGarry, 1997; Hurd, McFadden and Merrill, 1998).

Our analysis is of course quite restricted due to the small sample size. Figure 7 plots smoothed (running mean) survival curves for males and females at the three different horizons. For older ages, these smoothed curves depend on very few observations, and they are therefore not statistically reliable. However, one interesting result is the dip in the curves around age 50 (which cannot be attributed just to sample size). One explanation might be that in our sample of retired people those who are younger than 55 might actually have health problems that are reflected in lower subjective survival probabilities.

In Figure 8, we take a close look at subjective survival probabilities, again for males and females. The top and bottom panels contain probabilities at the 10-year and 20-year horizons, respectively, together with current (1997) their life-table counterparts. We smooth our data we a running-mean procedure so that the smoothed curve roughly represents mean probabilities by age. We find, again, that RPS participants between 50 and 60 are pessimistic relative to population values while those over 60 tend to be overly optimistic. As can be seen from the raw observations, our sample size is too small for reliable analysis a older ages. One interesting, and hard to interpret finding is the relatively large proportion of participants that state 100% survival probabilities even at the 20-year horizon.

4.3 Linking expectations, preferences, and behavior

A central aim of the RPS field version, and of many other surveys such as the HRS, is to provide data that allow to investigate the link between people's subjective expectations (about income changes and important health events), their preferences (such as the attitude towards risk), and behavior (such as retirement planning). The RPS pilot survey contained a number of questions that allow such analysis. The primary goal of the pilot survey was to explore how older people react to answering such questions over the internet, and the sample is relatively small and highly selected. As an illustration of the kinds of questions that can be explored with such data, we present a multivariate analysis of the link between subjective survival expectations, risk attitudes, and the decision to purchase long-term care insurance. The ordered logit model presented in Table 6 shows that in our sample, these factors, in addition to age, have the expected signs and are significant. While we do not intend to push this result to much given the nature of our data, we still find this result comforting in that the responses by our pilot survey participants are in line with economists' priors.

5. Outlook

At this point, we return to the important issue of selection effects in internet samples. The RPS pilot study is quite obviously not representative. With much larger sample size in the field version, we can check whether selection affects inference based on these stated survival curves by comparing raw and weighted data, were population weights are obtained from the CPS described in Section 2.

The next step in this research program is conducting a field version of the RPS. To this end, the questions in the RPS itself will be improved. For example, the demographic variables will be

brought in line with the CPS to facilitate the construction of weighting schemes. In addition to experiments on anchoring effects that document the resulting errors (if any), we will design experiments that allow to these hypotheses about the cognitive processes that govern response behavior in survey, and their relationship to age.

In conducting the field version of the RPS, we plan to co-operate with the AARP in one of its regular member surveys. The AARP conducts these surveys twice a year with a target sample of 2000 members aged 50 and older. We would increase the target sample size and provide a subsample with the option to answer the survey over the internet rather than return the paper version by mail. Such a design would allow to analyze the sample selection process that governs internet surveys. This combined mail-internet survey design addresses also the more practical problem of establishing sampling frames for internet surveys (see also Couper, 2000). It is our view that after careful observation of the initial selection process (along the lines of McFadden, 1997), internet panels can be extremely useful for scientific research in the behavioral sciences.

References

Adams, P., M. D. Hurd, D. McFadden, A. Merrill, and T. Ribeiro (2001): Healthy, wealthy, and wise? New evidence from AHEAD wave 3. Forthcoming, *Journal of Econometrics*.

Adams, P., and J. Winter (2001): The dynamics of internet access in the United States. Technical documentation, University of California, Berkeley.

Bikson, T. K. and C. W. A. Panis (1999): *Citizens, Computers, and Connectivity: A Review of Trends*. Santa Monica, CA: RAND.

Bosnjak, M., T. L. Tuten, and W. Bandilla (2001): Participation in web surveys: A typology. *ZUMA-Nachrichten*, 48, 7-17.

Couper, M. P. (2000): Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64, 464-494.

Couper, M. P., M. W. Traugott, and M. J. Lamias (2001): Web survey design and administration. *Public Opinion Quarterly*, 65, 230-253.

Dillman, D. A. (2000): *Mail and Internet Surveys: The Tailored Design Method*. New York:Wiley.

Dominitz, J. (1996): A comparison of subjective expectations elicitation methods in the Health and Retirement Study, the Panel of Income Dynamics, and the Survey of Economic Expectations. HRS/AHEAD Working Paper No. 96-043, Institute for Social Research, University of Michigan.

Dominitz, J. and C. F. Manski (1997): Using expectations data to study subjective income expectations. *Journal of the American Statistical Association*, 92, 855–867.

Dominitz, J. and C. F. Manski (2000): The Survey of Economic Expectations, waves 1-8, with data from the UW Survey Center's National Survey. 1st ed. Madison, WI: Data and Program Library Service.

Donkers, B., B. Melenberg, and A. van Soest (1999): Estimating risk attitudes using lotteries: A large sample approach. Discussion Paper No. 99-12, CentER for Economic Research, Tilburg University.

Donkers, B. and A. van Soest (1999): Subjective measures of household preferences and financial decisions. *Journal of Economic Pyschology*, 20, 613–642.

Donkers, B. (2000): Explaining time preference anomalies: A quantification. Unpublished manuscript, Erasmus University, Rotterdam.

Fischhoff, B., N. Welch, and S. Frederick (1999): Construal processes in preference assessment, *Journal of Risk and Uncertainty*, 19, 139–164.

Green, D., K. E. Jacowitz, D. Kahneman, and D. L. McFadden (1998): Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics*, 20, 85–116.

Horowitz, J. L. and C. F. Manski (1998): Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations. *Journal of Econometrics*, 84, 37–58.

Hurd, M. D. (1999): Anchoring and acquiescence bias in measuring assets in household surveys. *Journal of Risk and Uncertainty*, 19, 111–136.

Hurd, M. D., D. L. McFadden, H. Chand, L. Gan, A. Merrill, and M. Roberts (1998): Consumption and savings balances of the elderly: Experimental evidence on survey response bias. In D. Wise (ed.), *Frontiers in the Economics of Aging*, 353–387. Chicago, IL: University of Chicago Press.

Hurd, M. D., D. L. McFadden, and L. Gan (1998): Subjective survival curves and life-cycle behavior. In D. A. Wise (ed.), *Inquiries in the Economics of Aging*, 259–305. Chicago, IL: University of Chicago Press.

Hurd, M. D., D. L. McFadden, and A. Merrill (2001): Predictors of mortality among the elderly. Forthcoming in D. A. Wise (ed.), *Themes in the Economics of Aging*. Chicago, IL: University of Chicago Press.

Jacowitz, K. E. and D. Kahneman (1995): Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21, 1161–1166.

Kaptyen, A. and F. Teppa (2000): Hypothetical intertemporal consumption choices. Unpublished manuscript, Tilburg University.

Krosnick, J. A. (1999): Survey research. *Annual Review of Psychology*, 50, 537–567.

Lancaster, T. and G. Imbens (1990): Choice-based sampling of dynamic populations. In J. Hartog, G. Ridder, and J. Theeuwes (eds.): *Panel Data and Labor Market Studies*. Amsterdam: Elsevier. 21–43.

Lewbel, A., O. Linton and D. McFadden (2001): Estimating features of a distribution from binomial data. Manuscript, Brandeis University, LSE and UC Berkeley.

McFadden, D. L. (1994): Contingent valuation and social choice. *American Journal of Agricultural Economics*, 76, 689–708.

McFadden, D. L. (1997): On the analysis of endogenously recruited panels. Unpublished manuscript, University of California, Berkeley.

McFadden, D. L. (1999): Rationality for Economists? *Journal of Risk and Uncertainty*, 19, 73–105.

Moxey, L. M. and A. J. Sanford (1993): *Communicating Quantities.* Mahwah, NJ: Erlbaum.

Mussweiler, T. and F. Strack (2000): Numeric judgements under uncertainty: The role of knowledge in anchoring. *Journal of Experimental Social Psychology*, 36, 495–518.

Poterba, J. and L. Summers (1986): Reporting errors and labor market dynamics. *Econometrica*, 54, 1319–1338.

Rodepeter, R. and J. K. Winter (1999): Rules of thumb in life-cycle savings models. Discussion Paper No. 99-81, Sonderforschungsbereich 504, University of Mannheim.

Rodgers, W. and R. J. Willis (2000): A note on access to the internet by older Americans: The effects of age, education, and cognition. Manuscript, University of Michigan.

Schwarz, N. (1999): Self-reports. How the questions shape the answers. *American Psychologist*, 54:2, 93–105.

Schwarz, N. and S. Sudman (eds.) (1995): *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco, CA: Jossey-Bass.

Sudman, S., N. M. Bradburn, and N. Schwarz (1996): *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco, CA: Jossey-Bass.

Smith, V. K. and D. H. Taylor, Jr., and F. A. Sloan (2001): Longevity expectations and death: Can people predict their own demise? *American Economic Review*, 91(4), 1126-1134.

Tversky, A. and D. Kahneman (1974): Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124–1130.

Wallsten, T. S. and D. V. Budescu (1983): Encoding subjective probabilities: A psychological and psychometric review. *Management Science*, 29, 151–173.

Wallsten, T. S., D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth (1986): Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, 115, 348–365.

Wallsten, T. S., D. V. Budescu, R. Zwick, and S. M. Kemp (1983): Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*, 31, 135–138.

Weir, D., R. Willis, and W. Rodgers (2001): Cognitive abilities and quality of response to economic surveys: The HRS experience. Unpublished manuscript, Survey Research Center (SRC), University of Michigan, Ann Arbor.

Table 1: Internet access and demographic characteristics of the U.S. population

	Internet access (yes/no) Logit regression	
Demographic characteristics	Coefficient	Std. error ***
Age (years, truncated at 90)	-.0330	0.001 ***
Total household income, last 12 months ($1000)	0.033	0.000 ***
Total household income below $7,500 (dummy)	0.231	0.057 ***
High school graduation diploma or equivalent (dummy)	0.516	0.025 ***
Some college, but no degree (dummy)	0.122	0.028 ***
Associate degree, occupational/vocational (dummy)	1.106	0.046 ***
Associate degree, academic program (dummy)	1.432	0.050 ***
Bachelor's degree (dummy)	1.793	0.033 ***
Master's degree (dummy)	2.085	0.051 ***
Professional school or doctoral degree (dummy)	2.081	0.075 ***
Unemployed (dummy)	-0.012	0.050
Not in labor force, retired (dummy)	-0.353	0.034 ***
Not in labor force, disabled (dummy)	-0.416	0.049 ***
Not in labor force, other (dummy)	-0.099	0.026 ***
Metropolitan area (dummy)	-0.008	0.020
Black (dummy)	-0.700	0.029 ***
American Indian, Aleut, Eskimo (dummy)	-0.482	0.071 ***
Asian or Pacific Islander (dummy)	-0.081	0.045 *
Intercept	-0.356	0.037 ***
N	80259	
Log likelihood	-41141.155	
Pseudo R^2	0.252	

Source: Current Population Survey (CPS), Internet and Computer Use Supplement, August 2000; own calculations.

Notes: Reference groups are as follows: for education, no degree; for labor force status, employed; for race, white. The CPS sample covers the U.S. population aged 15 and older and is reported to be representative.

Table 2: Modules of the Retirement Perspectives Survey

Module	Description
	Registration page
1	Socio-demographic variables I (age, gender, family status, education, employment status)
2	Opinion questions on Social Security and Medicare
3	Qualitative questions on income, savings, and portfolio choice
4	Preferences and attitudes towards risk
5	Expectations about income, consumption, and health Socio-demographic variables II
6	Unfolding brackets questions on income, savings, and wealth
	Feedback questions

Source: Retirement Perspectives Survey (June 2000–September 2001)

Table 3: Demographic characteristics of RPS participants

		Sample 1			Sample 2			Samples 1 & 2		
Item	Answer	N	%	Mean (st.d.)	N	%	Mean (st.d.)	N	%	Mean (st.d.)
Age	[Fill-in]	76		65.86 (9.61)	139		54.01 (8.47)	215		58.14(10.48)
Gender	Male	39	51.32		84	60.43		123	57.21	
	Female	37	48.68		55	39.57		92	42.79	
Years of education	[Fill-in]	73		15.43(5.14)	138		17.18(7.86)	211		16.57(7.07)
	RF	3			1			4		
Employment status	Working full-time	18	23.68		85	61.15		103	47.91	
	Working part-time	9	11.84		18	12.95		27	12.56	
	Not working	0	0.00		4	2.88		4	1.86	
	Registered unemployed	0	0.00		0	0.00		0	0.00	
	Fully retired	34	44.74		25	17.99		59	27.44	
	Retired with part-time job	15	19.74		7	5.04		22	10.23	
State	California	66	86.84		62	44.60		128	59.53	
	Other U.S. states	10	13.16		67	48.20		77	35.81	
	Other countries	0			10	7.19		10	4.65	
U.S. Citizen	Yes	56	91.80		105	92.92		161	92.53	
	No	5	8.20		8	7.08		13	7.47	
	RF	0	0.00		0	0.00		0	0.00	
	Unknown	15			26			41		
Family status	Married	37	60.66		79	69.91		116	66.67	
	Living with a partner	3	4.92		3	2.65		6	3.45	
	Married, living separated	1	1.64		2	1.77		3	1.72	
	Divorced	7	11.48		15	13.27		22	12.64	
	Widowed	4	6.56		4	3.54		8	4.60	
	Single, never married	7	11.48		9	7.96		16	9.20	
	RF	2	3.28		1	0.88		3	1.72	
	Unknown	15			26			41		
Household status	Living alone	16	26.23		25	22.12		41	23.56	
	Living with spouse/partner	34	55.74		54	47.79		88	50.57	
	With other family	5	8.20		12	10.62		17	9.77	
	With spouse/partner and other family	5	8.20		22	19.47		27	15.52	
	RF	1	1.64		0	0		1	0.57	
	Unknown	15			26			41		

Table 3, cont'd: Demographic characteristics of RPS participants

		Sample 1			Sample 2			Samples 1 & 2		
Item	Answer	N	%	Mean (st.d.)	N	%	Mean (st.d.)	N	%	Mean (st.d.)
Household size	[Fill-in]	59		1.90 (0.71)	110		2.18 (0.94)	169		2.08 (0.88)
	RF	2			3			5		
	Unknown	15			26			41		
Number of children	0	19	31.15		33	29.20		52	29.89	
	1	8	13.11		23	20.35		31	17.82	
	2	17	27.87		34	30.09		51	29.31	
	3	13	21.31		11	9.73		24	13.79	
	4	0	0.00		10	8.85		10	5.75	
	5+	4	6.56		2	1.77		6	3.45	
	RF	0	0.00		0	0.00		0	0.00	
	Unknown	15			26			41		
Education level	High school	5	8.20		6	5.31		11	6.32	
	Bachelor degree	19	31.15		26	23.01		45	25.86	
	Master degree	20	32.79		43	38.05		63	36.21	
	Ph.D.	15	24.59		34	30.09		49	28.16	
	Other degree	2	3.28		3	2.65		5	2.87	
	None	0	0.00		0	0.00		0	0.00	
	RF	0	0.00		1	0.88		1	0.57	
	Unknown	15			26			41		
Household income (gross)	< $16.000	0	0.00		0	0.00		0	0.00	
	$16.000 ... $39.000	8	13.11		9	7.96		17	9.77	
	$39.000 ... $75.000	13	21.31		36	31.86		49	28.16	
	> $75.000	39	63.93		67	59.29		106	60.92	
	DK	0	0.00		0	0.00		0	0.00	
	RF	1	1.64		1	0.88		2	1.15	
	Unknown	15			26			41		

Source: Retirement Perspectives Survey (June 2000–March 2001); own calculations.

Notes: Some of these questions appeared towards the end of the survey. "Unknown" refers to participants who have not completed the survey and thus have not answered these questions. Sample 1: Members of the UC Retirement Center; Sample 2: publicly open survey.

Table 4: Answering behavior of RPS participants

	Survey completed (yes/no) Logit regression		Time required (minutes) Poisson regression			
			Specification 1		Specification 2	
Demographic characteristics	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error
Age	0.0115	0.0166	0.0124	0.0014	0.0123	0.0014
Gender = male	-0.8882	0.3698	0.1883	0.0311	0.2142	0.0344
Years of education	0.0678	0.0341	-0.0045	0.0021	-0.0041	0.0021
Gross HH income $39,000 – $75,000					-0.1470	0.0563
Gross HH income > $75,000					-0.1885	0.0530
Intercept	0.0866	1.1999	2.5407	0.0932	2.6906	0.1079
N	211		159		159	
Log likelihood	-105.81		-720.14		-711.58	
Pseudo R^2	0.0437		0.0793		0.0862	

Source: Retirement Perspectives Survey (June 2000–March 2001); own calculations.

Table 5: Descriptive statistics for selected RPS questions

	Item	Answer	N	%	Mean (st.d.)
POL1	Which of these actions would you want the government to take?	Expand Medicare coverage in place of private insurance.	94	44.98	
		Reduce Medicare coverage in favor of private insurance.	17	8.13	
		No change.	48	22.97	
		DK	41	19.62	
		RF	9	4.31	
		Unknown	6		
POL2_1	What should the U.S. Government do with its budget surplus? (1) Pay off national debt. (2) Invest the surplus. (3) Return to taxpayers. What is the best option?	Best option: 1	134	14.65	
		Best option: 2	43	20.10	
		Best option: 3	21	10.24	
		DK	4	1.95	
		RF	3	1.46	
		Unknown	10		
POL2_2	What should the U.S. Government do with its budget surplus? (1) Pay off national debt. (2) Invest the surplus. (3) Return to taxpayers. What is the second best option?	Second best option: 1	49	23.90	
		Second best option: 2	92	44.88	
		Second best option: 3	44	21.46	
		DK	13	6.34	
		RF	7	3.41	
		Unknown	10		
INCQ3_1	Including any income that you expect from Social Security or pensions, will you have enough savings to maintain your current living standard after retirement?	Yes	58	47.93	
		No	38	31.40	
		DK	24	19.83	
		RF	1	0.83	
		Unknown / NA	94		

Table 5, cont'd: Descriptive statistics for selected RPS questions

	Item	Answer	N	%	Mean (st.d.)
INCQ3_7	How do you expect you standard of living in retirement to be, compared to your present standard of living?	Substantially higher	4	3.31	
		About the same	85	70.25	
		Substantially lower	22	18.18	
		DK	6	4.96	
		RF	3	2.48	
		Unknown / NA	95		
SAVQ_1	Thinking of your saving behavior over the past 20 or 30 years, do you think now that what you saved was about right, too little, or too much?	Too little	84	43.30	
		About right	99	51.03	
		Too much	4	2.06	
		DK	7	3.61	
		RF	0	0.00	
		Unknown	19		
SAVQ_2	If you feel that you saved too little, how much more should you have saved?	[Fill-in, percent]	79	94.05	60.29 (111.36)
		DK / RF	5	5.95	
		Unknown / NA	131		
HEP1_1	Have you considered purchasing insurance that covers the cost of long-term care in a nursing home or in your home?	I have purchased such an insurance.	36	20.45	
		I have considered purchasing insurance but decided not to.	57	32.39	
		I have not considered purchasing such an insurance.	77	43.75	
		DK	3	1.70	
		RF	3	1.70	
		Unknown	39		

Table 5, cont'd: Descriptive statistics for selected RPS questions

	Item	Answer	N	%	Mean (st.d.)
HEP2_1	How familiar are you with Medicaid and the protection it may provide to individuals whose nursing home expenses exhaust their assets?	Not familiar.	57	32.39	
		Somewhat familiar.	87	49.43	
		Very familiar.	30	17.05	
		DK	2	1.14	
		RF	0	0.00	
		Unknown	39		
HEP3_1	How familiar are you with the spend-down provisions in Medicaid nursing home coverage in your state?	Not familiar.	91	52.00	
		Somewhat familiar.	56	32.00	
		Very familiar.	27	15.43	
		DK	1	0.57	
		RF	0	0.00	
		Unknown	40		
HEP4_1	In making financial plans for your retirement years, have you made adjustments to take into account the possibility that you may incur nursing home expenses down the road?	Yes	80	45.98	
		No	92	52.87	
		DK	0	0.00	
		RF	2	1.15	
		Unknown	41		
HEP4_2	If you have made such financial adjustments, have you adjusted the form or legal ownership of your assets to make it easier to qualify for Medicaid should you eventually need assistance with nursing home costs?	Yes	16	9.20	
		No	136	78.16	
		DK	13	7.47	
		RF	9	5.17	
		Unknown	41		

Source: Retirement Perspectives Survey (June 2000–September 2001); own calculations.

Notes: "Unknown" refers to the number of participants that have not answered a question; either because they were not asked a particular question that was not applicable, or because they did not complete the survey.

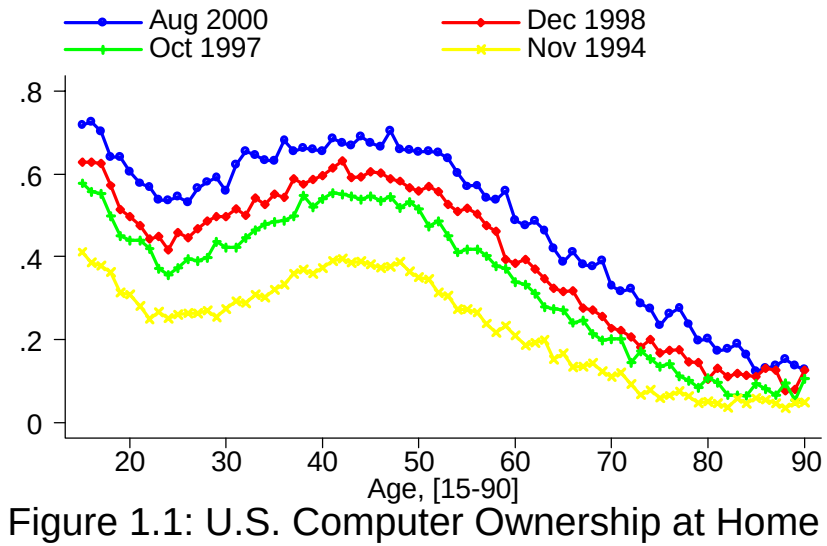
Table 6: Ordered logit estimates for decision to purchase long-term care insurance

	Coefficient	Std. error	p value
10-year subjective survival probability	-.0161950	.0097972	0.098
Risk aversion (dummy)	-.7096436	.3474078	0.041
Age	-.0958541	.0214766	0.000
Male (dummy)	-.9741466	.3687893	0.008
Years of education	.0087347	.0190975	0.647
Income between $39k and $75k (dummy)	-.5915305	.6329511	0.350
Income above $75k (dummy)	-1.397576	.5935055	0.019
Cut 1	-11.19817	2.210755	
Cut 2	-9.282325	2.137496	
N	139		
Log likelihood	-127.79		
Pseudo R^2	0.1232		

Source: Retirement Perspectives Survey (June 2000–September 2001); own calculations.

Notes: Dependent variable based on question HEP1_1: "Have you considered purchasing insurance that covers the cost of long-term care in a nursing home or in your home?" Values: (1) "I have purchased long-term care insurance." (2) "I have considered it carefully and decided not to buy long-term care insurance." (3) "I have not considered buying long-term care insurance carefully."

Figure 1: Computer ownership in private homes in the United States, 1994–2000



Source: Current Population Survey (CPS), Internet and Computer Use Supplements, November 1994, October 1997, December 1998, August 2000; own calculations.

Figure 2: Internet access in the U.S. population by age, 2000 vs. 1997

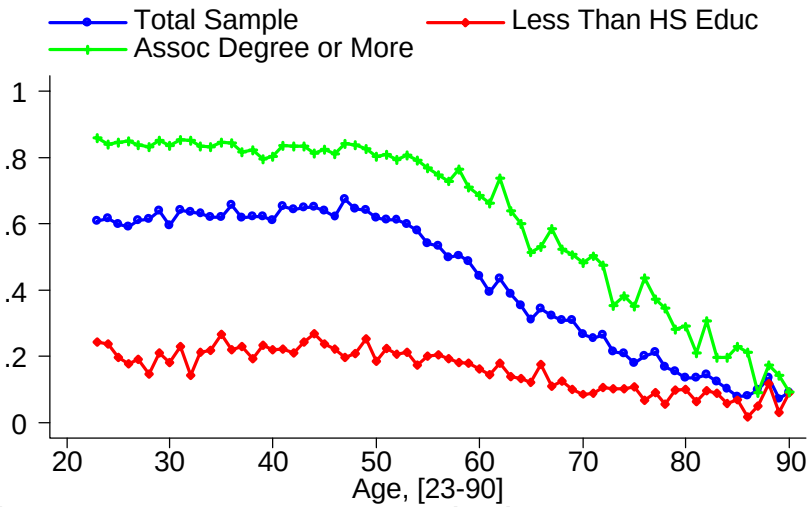


Figure 2.1: Internet Access in the U.S., Aug 2000

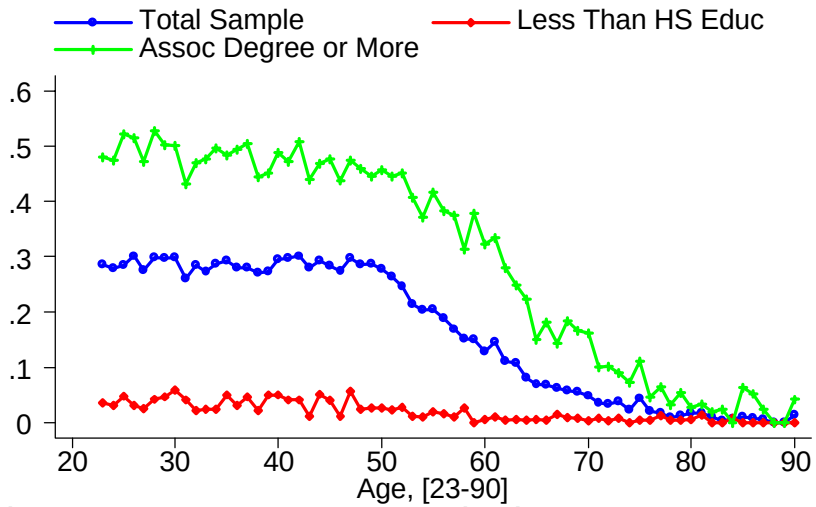
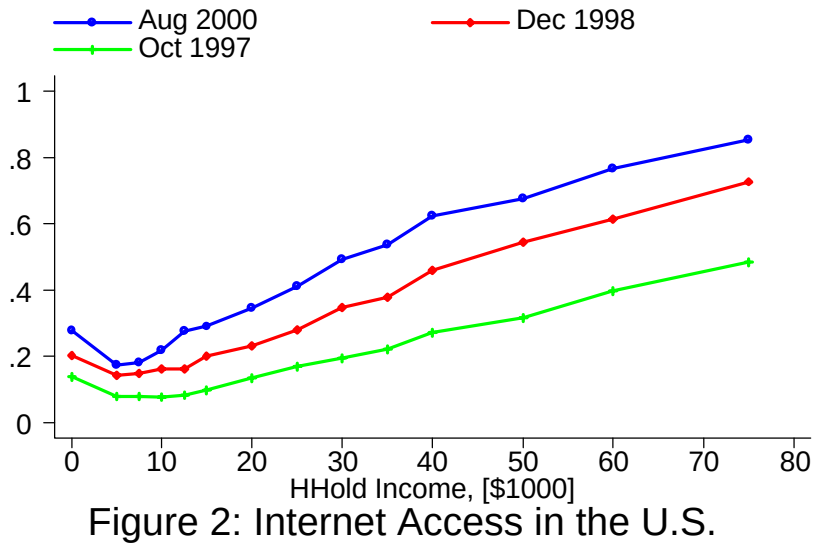


Figure 2.3: Internet Access in the U.S., Oct 1997

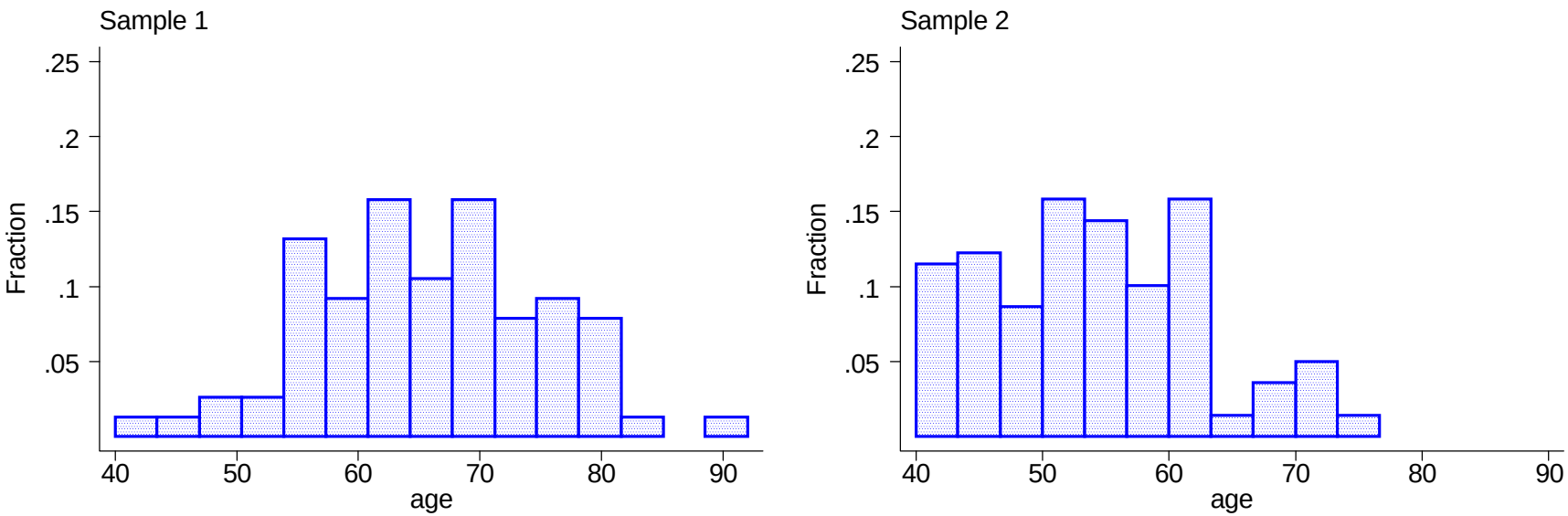
Source: Current Population Survey (CPS), Internet and Computer Use Supplement, October 1997, August 2000; own calculations.

Figure 3: Internet access in the U.S. population by income, 1997–2000



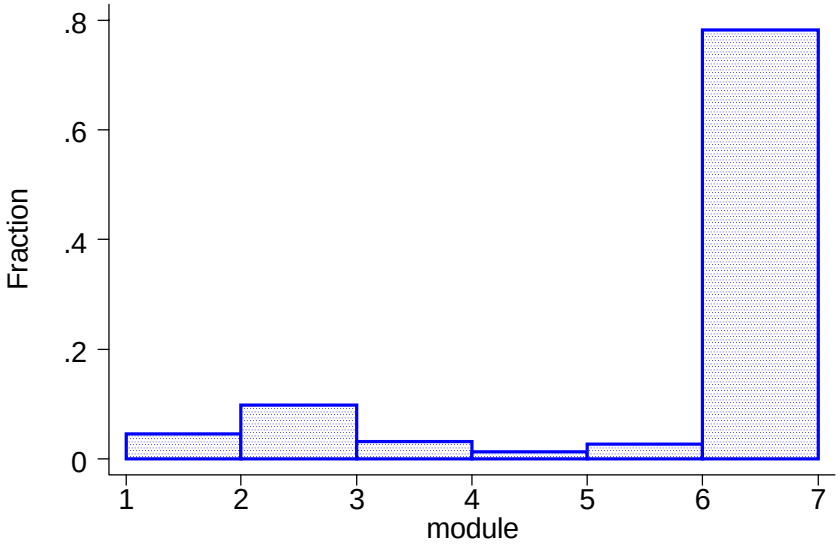
Source: Current Population Survey (CPS), Internet and Computer Use Supplement, October 1997, December 1998, August 2000; own calculations.

Figure 4: Age distribution of RPS participants



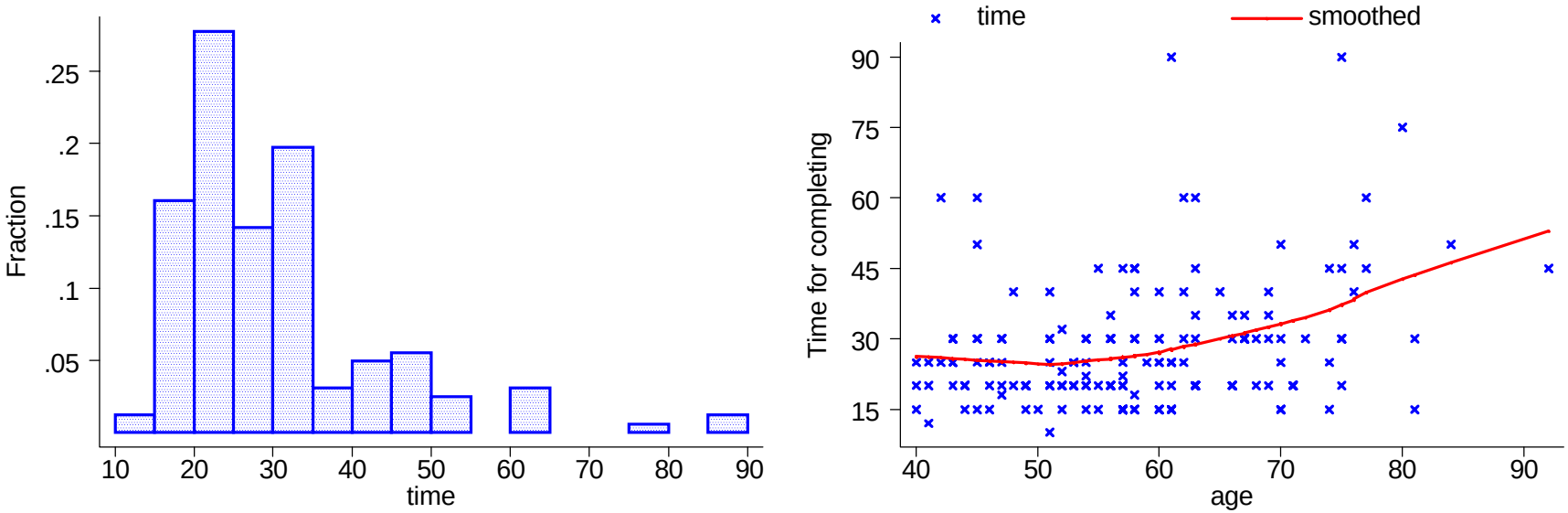
Source: Retirement Perspectives Survey (June 2000–September 2001); own calculations.

Figure 5: Modules completed by RPS participants



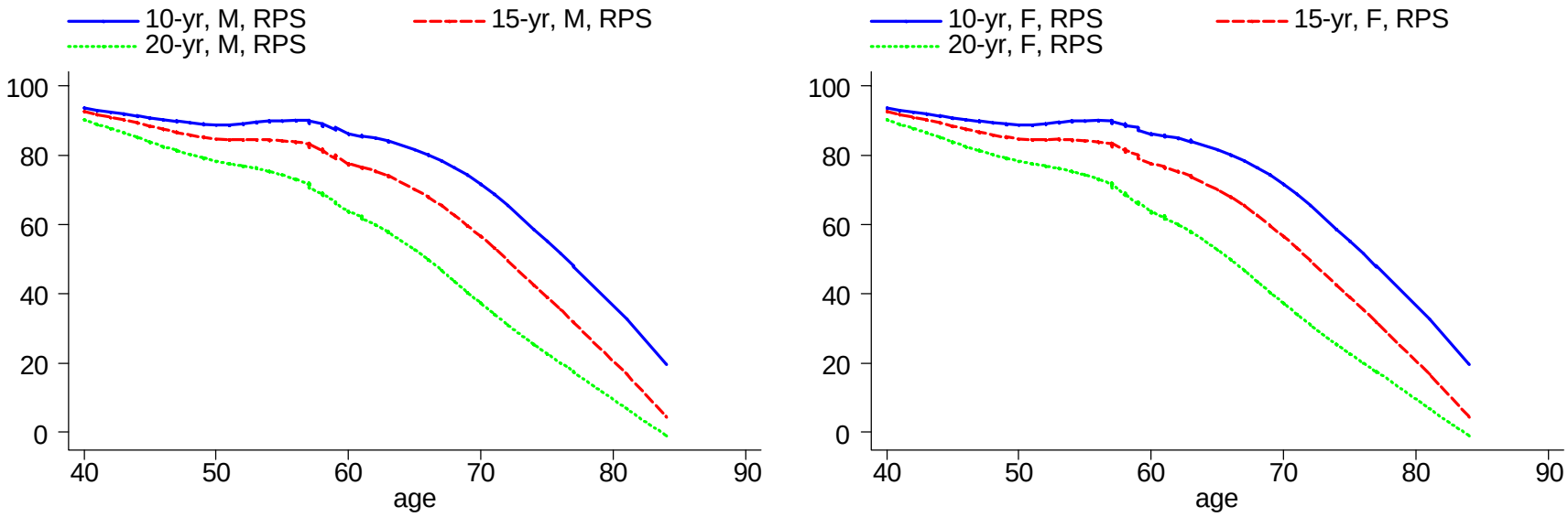
Source: Retirement Perspectives Survey (June 2000–September 2001); own calculations.

Figure 6: Time required for completing the survey



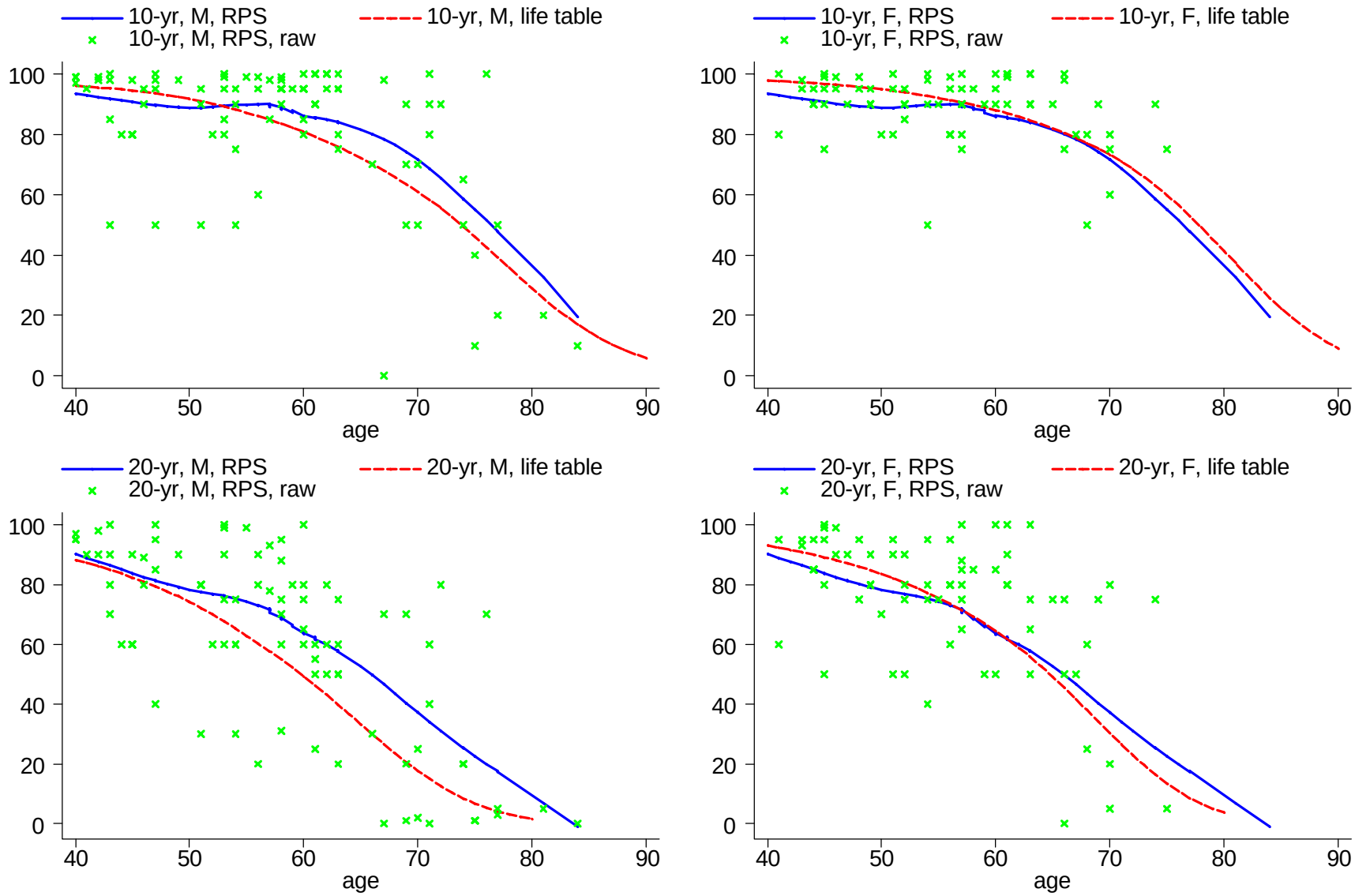
Source: Retirement Perspectives Survey (June 2000–September 2001); own calculations.

Figure 7: Subjective survival probabilities of RPS participants



Source: Retirement Perspectives Survey (June 2000–September 2001); own calculations.

Figure 8: Subjective survival probabilities of RPS participants vs. population values



Source: Retirement Perspectives Survey (June 2000–September 2001); 1997 life tables for the U.S.; own calculations.