Can Survey Participation Alter Household Saving Behavior?*

Thomas F. Crossley† Jochem de Bresser‡ Liam Delaney§ Joachim Winter¶

August 11, 2014

Abstract

Much empirical research in economics is based on data from household surveys. Panel surveys are particularly valuable for understanding dynamics and heterogeneity. A possible concern with panel surveys is that survey participation itself may alter subsequent behavior. We provide novel evidence of survey effects on a central life-cycle choice: household saving. We exploit randomized assignment to survey modules within the LISS Panel, an internet panel survey which is representative of the Dutch population. We find that households that respond to detailed questions on expenditures and needs in retirement reduced their non-housing saving rate by 3.5 percentage points, on average. This mean effect is driven by high-education households which have the highest pension and housing wealth. Our saving measure is based on linked administrative wealth data. Thus we can rule out the possibility that the effect is on reporting, rather than on the underlying saving behavior. One interpretation is that the survey acted as a salience shock, possibly with respect to reduced housing costs in retirement.

Key words: Survey effects; savings; administrative data; retirement
JEL codes: C83; D14

*The authors thank the participants at the 2012 MESS workshop for helpful comments on an early version of this paper. We also thank Rob Alessie, James Banks, Marcel Das, Pierre-Carl Michaud, Martin Salm, Daniel Schunk, Arthur van Soest, Frederic Vermeulen; and seminar and conference participants at Essex University, Tilburg University, the Universities of Groningen, Mannheim, and Stirling, the Munich Institute for the Economics of Aging (MEA), the 2013 Canadian Research Data Centre Network Conference, the 2014 Netspar International Pension Workshop and the 2014 Royal Economic Society Annual Conference. Financial support for data access was provided by Netspar and CentERdata. Crossley acknowledges support from the ESRC through the ESRC-funded Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies (CPP, reference RES-544-28-5001).

†Essex University & Institute for Fiscal Studies, London; e-mail: tcross@essex.ac.uk
‡University of Groningen; e-mail: J.de.Bresser@rug.nl
§University of Stirling; e-mail: liam.delaney@stir.ac.uk
¶University of Munich; e-mail: winter@lmu.de

1 Introduction

Much empirical research in economics analyzes data from surveys of individuals and households. The development of panel surveys has allowed researchers to assess and account for heterogeneity and dynamics in economic behavior. However, repeated data collection from the same individuals or households brings a risk of "survey effects" – the possibility that questioning individuals about their actions or attitudes in a particular domain can alter their later behavior. Finding significant survey effects in important areas of economic research would require a rethinking of data collection strategies. More positively, finding such effects might also provide insight into the cognitive processes underlying broader economic behavior.

In this paper we test for survey effects in a central domain of economic research: household saving behavior. In particular, we ask whether being asked questions about retirement income needs leads to changes in household saving behavior. Recent work in behavioral finance suggests possible mechanisms for survey effects in saving behavior. Limited attention means that individuals tend to overlook some of the consequences of their decisions (DellaVigna, 2009). If those unnoticed consequences materialize in the future, as do the benefits of saving today, this results in biases that are similar to those induced by limited self-control (Karlan et al., 2012). However, in contrast to self-control problems, limited attention suggests that behavior might be corrected by focusing individuals' attention on the aspects they are missing. To the extent that surveys on retirement planning can direct participants' attention, they may have behavioural effects, though the direction of such effects, is not clear *a priori*.

Our research design has a number of critical features. First, we study survey effects in large representative survey of a developed country population (the Netherlands). Second, we study effects on a key life-cycle choice: the level of savings. Third, the randomized allocation of members of an internet panel to a survey module on retirement income needs provides for clean identification of the causal effect of participating in that survey module. Fourth, we measure household saving with linked administrative data – not with responses in subsequent panel surveys. This allows us to rule out the possibility that any observed effect is on reporting, rather than on the underlying saving behavior. No prior study combines all these attributes.

In an environment in which public and occupational pension schemes implied high income replacement rates in retirement (the Netherlands in 2008), we find robust evidence that exposure to the retirement needs module subsequently led to *lower* average household saving. This mean effect is driven by older and more educated households. These households have the highest expected pension wealth and higher housing wealth. They are also much more likely, in the survey, to report that their housing costs will fall in retirement (perhaps because they anticipate paying off their mortgage debt). Our interpretation is that the survey module directed attention to aspects of retirement preparedness and needs that were not otherwise salient, and for wealthy households this apparently implied that they should save less.

The survey methods literature has long been concerned with “panel conditioning”: the way in which experience in a panel survey affects subjects’ responses. Several studies have examined panel conditioning in domains such as subjective well-being (Van Landeghem, 2012), marital satisfaction (Glenn, 1998), and preferences (Binswanger et al., 2013). Much of this literature compares experienced panel members with refreshment samples or other novice respondents. As pointed out by Das et al. (2011), disentangling panel conditioning from unobserved factors influencing attrition is a complex task. Das et al. (2011) conclude that, after controlling for unobserved attrition factors, there are significant panel conditioning effects in knowledge questions but not in other types of questions. Our design avoids, based on random invitations to a survey module, concerns about attrition. In a recent review paper, Warren and Halpern-Manners (2012) note that the survey research literature on panel conditioning generally failed to employ randomized designs. They also point out that to date, there is little systematic evidence on panel conditioning in large-scale longitudinal social science surveys. Our paper addresses this concern as well.

The psychology and marketing literatures have documented a number of related behavioral phenomena. The “question-behavior effect” refers to the observation that asking respondents to predict future behavior results in more of the behavior under consideration (Dholakia, 2010). For example, Spangenberg (1997) shows that asking people to predict their workout behavior induces them to visit the gym more often. Where the behavior in question is normatively desirable this is sometimes called the “self-prophecy effect” (Sherman, 1980), and where it is normatively ambiguous, the "mere measurement effect" (Morwitz et al., 1993). Several papers have also shown that surveying people about risky behaviors can increase the propensity to engage in those behaviors (Fitzsimmons and Moore, 2008; Fitzsimmons and Shiv, 2001; Dholakia, 2010). The public health literature reports some evidence that supports the use of self-recording of behavior as a behavioral change intervention (Michie et al., 2011a,b; Burke et al., 2011), though recently Axinn et al. (2014) found little evidence of behaviorial effects of keeping survey diaries of sexual behavior for extended periods (up to 30 months).

The effects documented in these literatures connect questions about the intention to, or likelihood of, engaging in specific behaviors, or the recording of specific behaviors, to the propensity to engage in those behaviors. The survey effects that we study are less direct. We investigate the effect of being surveyed about topics related to the behavior of interest, rather than direct questions about that behavior. The question module that is our treatment has no specific questions about saving behavior.

Two further relevant studies are Zwane et al. (2011) and Stango and Zinman (2013). Like us they consider survey effects that result from being surveyed on a topic related to the behavior of interest, rather than from direct questions about that behavior, or self-recording of that behaviour. Furthermore, they use administrative data to measures outcomes, in order to isolate genuine changes in behavior from changes in reporting style.

Zwane et al. (2011) study the effect of randomly assigning subjects to extra survey monitoring in the context of five randomized controlled trials of interventions in developing countries. They find that survey monitoring significantly increases the probability of water treatment product usage and medical insurance usage. They do not find effects for micro-lending take-up or renewal.

Stango and Zinman (2013) study participants in a marketing research panel. Panel members are invited to complete periodic surveys, which vary in content. While participants self-select into individual surveys, they do so without knowledge of the content. Stango and Zinman (2013) show that participants that enter a survey with general questions about overdrafts and overdrafts fees are less likely to incur such fees in the same month, and in subsequent months. Their interpretation is that the surveys with overdraft questions act as "salience shocks" which help respondents avoid costs arising from limited attention.

Our paper differs from Zwane et al. (2011) and Stango and Zinman (2013) in that the analysis is based on a large population-representative survey used by social science researchers, and that the survey effects we study concern a central life-cycle choice – the level of saving. It may be much more difficult to shift choices of important real quantities like consumption, saving or labor supply, than it is to induce small changes in the timing of transactions or portfolio composition, or to encourage the avoidance of fees. Our analysis further differs from Zwane et al. (2011) in that the data are drawn from developed economy, and from Stango and Zinman (2013) in having a randomized design. Moreover, there is to date no evidence on survey effects with respect to core life-cycle choices like the level of savings or consumption.

Our analysis is based on the Dutch LISS Panel, a population-representative internet panel survey of households that has been conducted in the Netherlands since 2007. Members of the LISS Panel complete online surveys on a regular basis. These surveys collect data on a range of financial, health, and social topics. Crucially for this study, respondents are randomly assigned to eligibility for some additional survey modules. The treatment we study is a module of questions on expected needs in retirement and preferences for current versus retirement consumption (henceforth, the "retirement needs module"). This module was fielded in January 2008 and it was the first randomized content module in the LISS Panel. The module neither provides any information on the Dutch pension system in general nor on the pension entitlements of respondents. Nor does it contain direct questions about saving behavior.

We measure household saving through linkage to records in the Dutch national tax record system between 2007 and 2009. This system records detailed information on assets and debt across different asset classes at the beginning of the calendar year. We are thus able to construct a very accurate measure of saving for 2008 and 2007: the years immediately after and before treatment. As noted above, an independent measure of saving is necessary to be sure that observed effects represent genuine behavioral change, and not changes in survey reporting style.

Our experiment occurred just prior to the onset of the financial crisis. At this time, the combination of the first and second pillars of the Dutch pension system (the state pension and occupational pensions) offered a very high ratio of expected retirement income to peak working-life income to most Dutch workers. Bovenberg and Meijdam (2001) report an average after-tax replacement rate of 80 per cent.

We report both intention-to-treat (ITT) and instrumental variables (IV) estimates of the treatment effect. The latter use the survey invitation (which was random) as an instrument for survey completion. Survey take-up is very high, so there is little difference between ITT and IV estimates. The mean IV estimates indicate that survey participation *lowered* mean savings rate by 3.5 percentage points. Quantile IV estimates indicate that the effects of similar magnitude are present from across a wide range of the distribution. A falsification test on pre-survey savings behavior finds no effect.

Our findings present a significant challenge to survey designers and to the collection of longitudinal data in general, and data on the important economic topics of saving behavior and retirement preparation in particular. Measurement is not innocuous. Even a random sample of a population, which has not been subject to nonrandom attrition, can fail to be representative of the population under study if the act of measurement alters the behaviour of sample members. More positively, this study demonstrates the value of randomized survey content in allowing for the exploration of such effects.

Our results also significantly extend the evidence base for the importance of limited attention in household financial decision making (Karlan et al., 2012; Stango and Zinman, 2013). We show that survey questions can alter choices with respect to a key life-cycle variable, the overall level of saving. Our findings also demonstrate that salience shocks can shift behavior in different directions, depending on the context. While Karlan et al. (2012) find that deliberate and targeted reminders raise contributions to a goal-specific savings accounts in Bolivia, Peru and the Philippines, we find that exposure to a module of questions about retirement income needs lowered overall saving in the Netherlands. We also find quite different patterns of effect heterogeneity. While Stango and Zinman (2013) find larger effects of being surveyed among lower-educated subjects, we find the largest effects among the high-education group. We discuss this further below.

The rest of the paper is organized as follows. Section 2 outlines our research design and introduces the LISS survey data as well as the linked administrative data on financial assets. In section 3, we present the results as well as a number of robustness checks and falsification tests. Section 4 concludes.

2 Research design and data

2.1 The LISS Panel

The LISS Panel is a representative random sample from the Dutch population that was initiated during the autumn of 2007. The LISS Panel is administered by CentERdata, a survey research institute affiliated with Tilburg University, and follows close to 8,000 individuals from 5,000 households. Surveys are administrated over the internet. Though the Netherlands has a high rate of internet access (more than 80% of Dutch households are connected), CentERdata safeguards representativeness by providing sample households with an internet subscription and a simple computer when necessary. Scherpenzeel (2011) provides details on the design and sample.

All members of the LISS Panel respond to core surveys on subjects such as demographics, income, and assets on a regular basis. In addition, researchers have the possibility to design additional survey modules which are fielded throughout the year. Those customized modules are usually distributed only once and, to keep costs down, are often limited to a random subsample of the eligible sample. This distribution of modules to random subsets of the LISS Panel generates exogenous variation in survey participation which we exploit to estimate the effect of participation on household behavior.

2.2 Treatment and eligibility

We define treatment as participation in the survey module entitled "What is an adequate old age income?" – the retirement needs module. This module was fielded in January 2008, and it was the first randomized module in the LISS Panel. The module was designed by Binswanger and Schunk to study preferences and attitudes relating to living standards in retirement. The module was first fielded in an older internet panel (the CentERpanel) and the resulting data are analysed in Binswanger and Schunk (2012). Binswanger et al. (2013) the compare responses to the module across the two panels to study how differences in responses between experienced (CentERpanel) and novice (LISS) respondents. They find greater nonresponse among novice respondents, but little difference between experienced and novice panel members conditional on response. Note that the survey effects they study are very different from the effects studied in this paper. They compare novice and experienced panel members, all of whom receive the retirement needs module, to estimate the effect of part experience with an internet panel of survey responses (particularly reposes to this module.) We compare novice panel members who received the retirement needs module to novice panel members who did not, in order to estimate the effect or receiving this module on actual savings behaviour (measured independently of the survey).

It is important to emphasise that all the subjects we study were novice panel members, and none had been previously been exposed to randomised survey content (as noted above, the retirement needs module was the first randomised module in the LISS panel). Panel members who were not offered the retirement needs module were *not* offered another module, and random assignment to the offer the retirement needs module was independent of assignment to all subsequent modules. Thus the effect we study is receiving the module versus not receiving it, and not a comparison between this module and some other module, or sequence of modules.

The retirement needs module consists of around 60 items that concern desired expenditure levels in retirement, the tradeoff between current and future consumption, and risk attitudes with respect to income after retirement.

- The module starts by asking how much respondents have thought about retirement and whether they would be willing to cut down on housing expenditures when they stop working.

- The next two questions elicit expectations with regard to the evolution of housing costs during the first decade following retirement: the general direction, decrease/roughly equal/increase, followed by the expected change in euro per year.

- After having been primed in this way to consider housing as an important, and potentially changing, category of expenditures, respondents are asked what the *minimum* expenditure level is that they would never want to fall below during retirement. Respondents then compare this minimal expenditure level with their current expenditures and indicate the reasoning behind their answer (e.g. summing projected expenditures in different categories or taking a certain fraction of current income or expenditures).
- After reporting their *minimal* expenditures during retirement, a series of multiple choice questions elicits *desired* expenditures by means of choices between different expenditure paths during working life and retirement.
- The questions on desired expenditure levels are followed by a series of choices between lotteries that involve income streams during retirement, to measure (domain-specific) risk preferences, and a vignette question in which respondents indicate whether they agree that one hypothetical lifetime expenditure path is preferable to another.
- Finally, respondents are asked a question about their willingness to pay for prevention of climate change and they evaluate how difficult they found the questionnaire.

Importantly, the module neither provided respondents with any information about the Dutch system of retirement income provision in general, nor about respondents' personal entitlements in particular. Thus, the randomized survey module did not constitute an information shock. Since the survey did not include any questions on predicted or intended savings either, the randomized survey module could not induce a question-behavior effect; respondents were not asked to predict their own behavior.

Eligibility for the survey was limited to all LISS members that were 25 years or older, who had a net household income of at least 800 euro per month, and who were either the head of the household or his/her partner (children or other household members were excluded from participation). This lead to a total eligible sample of 5,435 individuals, 2,755 of which were selected at random and were offered the retirement needs module. The take-up rate among those who received the offer was 74% at the individual level and 79% at the household level (at which our analysis is performed).

2.3 Outcome measures

We investigate the effect of survey participation on household savings. Though the LISS data include elaborate biannual surveys on assets and debt, we prefer to use administrative wealth records for two reasons. First, we want to rule out the confounding influence of differences in reporting behavior between those who were and were not offered the retirement needs module. If we found an effect of survey participation on self-reported savings, one could argue that the survey changed reporting styles rather than behavior. Deriving our outcome measures from administrative data eliminates that concern. Second, there is a general concern about the quality of self-reported survey data on assets (Bound et al., 2001).

An important concern when using combined survey and administrative data is incomplete linkage. Respondent refusal to consent is one reason for incomplete linkage (Sakshaug et al., 2012). In our study, this problem is exacerbated by panel attrition. Informed consent for the match of LISS data with administrative records was elicited only in September of 2011, almost four years after the retirement needs module was fielded. Out of the 3,125 households that contain at least one member that was eligible for the retirement needs module, we could match only 1,602. De Bresser and Knoef (2013) show that this loss of data is mostly due to panel attrition rather than refusals: only 10% of the respondents to the retirement needs module that were still in the LISS Panel in 2011 did not provide consent to the administrative data linkage.

The administrative assets data are taken from the Complete Asset Data of the Netherlands (Integraal Vermogensbestand, CAD), which was constructed by Statistics Netherlands. The CAD is based on tax records, which are supplemented with information from banks. The CAD contains a detailed decomposition of household-level wealth for the entire Dutch population. It measures assets on the first of January for the years 2007, 2008 and 2009 (data for more recent years are not yet released at the time of writing). Available records thus allow us to compute yearly savings during 2007 and 2008 as the differences between wealth stocks in consecutive years. We compute these wealth stocks net of the value of the primary residence, because we want to focus on pure savings and housing has an important consumption component. The use of administrative data on changes in wealth to measure savings and consumption is becoming more common; see Browning et al. (2013) for a recent example and Browning et al. (2014) for a brief survey of the advantages and disadvantages of such data.

Many studies of household savings behavior are limited by having only data on specific assets or accounts.[1] Thus, changes in contributions may represent portfolio reshuffling rather than changes in net saving. In contrast, a strength of our data is that we observe an almost complete measure of wealth. The categories of assets that we observe are checking and saving accounts, bonds, stocks, property, other real estate, business capital and other tangibles. For debt, the CAD distinguishes between mortgage and other debt. We miss savings held in small accounts because banks are not obliged to report accounts with a balance of less than 500 euro or less than 15 euros in interest payments. We also do not observe debt for households without capital income. Finally, we miss savings held in tax-exempt private retirement ("third pillar") pensions.[2] Such accounts are taxed only during the payout phase and are therefore invisible in tax records up to retirement; in the period we study, such accounts were held by very few households and contributed only 7% of household retirement income. While we do not observe holdings in such accounts in the administrative data, we do have information from the LISS assets survey (one of the core surveys that are answered by all panel households). Thus we do have a check on changes in saving in these accounts.

In our analysis of savings, we look both at levels, in euro per year, and rates, which are levels divided by yearly disposable income. The data on the yearly disposable income of households are also taken from tax records. We use the Complete Household Income Data of the Netherlands (Integraal Huishoudens Inkomstenbestand, CHID), assembled by Statistics Netherlands. The measure for primary income in the CHID is quite complete: in addition to labor income it includes income from entrepreneurship and from assets (interest payments and imputed rent for

[1] For example, in Karlan et al. (2012), Duflo et al. (2006), and in many studies of individual retirement accounts, the outcome variables are contributions to one account; see Crossley et al. (2012) for a discussion.

[2] We discuss the four pillars of the Dutch pension system below.

homeowners). Disposable income is defined as primary income plus government transfers that the household received minus the transfers and taxes paid by the household. The administrative income measure that we use is likely to be more accurate than survey measures of income since information about the various income streams is provided electronically by employers and financial institutions to the tax authority.

2.4 Institutional context

The Dutch system of income provision during retirement is easily understood in terms of four categories or "pillars". The first pillar is that of the public pension, which provides everybody who lived in the Netherlands between the ages of 15 and 65 with a subsistence income. Coverage of the public pension is close to universal, since uninterrupted residence in the country is the only criterion (benefits are cut by two percent for each year spent abroad).[3] The level of the public pension is set in reference to the minimum wage. Since public pensions only provide a minimum income, almost all employees accumulate additional entitlements in occupational pensions (the second pillar). Such arrangements cover 90% of all employees and are usually organized at the level of the sector or of the company (Bovenberg and Meijdam, 2001). Participation in the first two pillars is mandatory and together they replace 70 percent of gross last earned income on average, which translates into replacement rates net of taxes above 80 percent (Kapteyn and De Vos, 2008; Bovenberg and Meijdam, 2001). The third pillar contains all private savings vehicles that are aimed specifically at retirement, such as life annuities. Such voluntary arrangements are especially important for individuals that cannot rely on occupational pensions, such as the self employed. Finally, all other forms of wealth that can be drawn down to generate additional income after retirement, such as savings accounts, investments, and real estate, make up the fourth pillar.

It is important to stress that the Dutch pension system in 2008 was characterized by arrangements that were almost universal and provided extremely generous income replacement. Given this institutional environment, it is not surprising that the first two pillars, public old age pensions

[3]Technically, one is covered by the public pension if one's income is subject to Dutch income taxes. Residence abroad does not affect the accumulation of entitlements as long as income is taxed within the Dutch system.

and occupational pensions, together provided 95% of income in retirement (Kapteyn and De Vos, 2008). The final 5% was accounted for by private pension products and other savings.

De Bresser and Knoef (2013) use responses to the retirement needs module in the LISS panel (our treatment) combined with wealth data from the CAD and other data to analyze whether the Dutch pension system succeeds in providing an adequate retirement income to its contributors. They compare projected annuities from pensions and non-pension savings with the self-reported minimal and desired expenditure levels from the retirement needs module. Respondents report rather high minimum expenditure levels, on average 50% higher than the highest official poverty line of Statistics Netherlands. Nevertheless, a majority of close to 70% can still expect to exceed their own *minimum* expenditure level using their funds in the first two pillars of the system. The extent of over-saving is substantial: the median difference relative to what would be required to maintain the self-reported minimum expenditure floor is 25% taking only the first two pillars of the retirement saving system into account; and this rises to 36% if they take non-pension, non-housing savings into account. Moreover, the median difference between the annuity from pensions and non-housing wealth and the self-reported desired *desired* expenditure level is 18%. Thus it seems that at the time the retirement module was fielded in LISS (in 2008), a large fraction of the population could significantly reduce their savings and still meet their post-retirement expenditure goals.

Another important aspect of the institutional context is that beginning in 2008, individuals could find detailed information on their personal pension entitlements in their Uniform Pension Overview (UPO). These UPOs provide all members of pension funds, in both the second and third pillars, with yearly updates on their current entitlements and projected future entitlements at age 65. UPOs have been mandatory for all financial institutions in the Netherlands since January 1st, 2008.

2.5 Threats to validity

Our analysis faces two threats to internal validity. The first problem is the issue of incomplete compliance with the treatment: Not every LISS Panel member who was offered the retirement needs module completed this specific survey. We apply two remedies. First, we perform an intention-to-treat (ITT) analysis that compares those who did receive the offer with those who did not (instead of comparing those who were treated with those who were not). Second, we perform instrumental variables (IV) analyses in which we use the random offer of treatment as an instrument for being treated. Both methods allow us to obtain estimates of treatment effects that are not affected by endogenous sample selection as a result of non-compliance (that is, non-participation in the retirement needs module), since they rely on exogenous variation in the module offer. In addition to IV regressions for the conditional mean of the savings distribution, we also estimate unconditional decile treatment effects in order to establish the robustness of our results. We use the estimator proposed in Frölich and Melly (2013).[4] Importantly, the research design we use is characterized by one-sided non-compliance: Those respondents who were not randomly selected for the offer of the module could not complete the module. Thus, the monotonicity requirement for the identification of local average treatment effects (Angrist et al., 1996) is satisfied.

The second threat to internal validity is that of selection into outcome measurement due to the substantial loss of observations when we match LISS observations with administrative records. As mentioned above, we could only link 1,602 out of 3,125 eligible households in the LISS Panel to administrative records, mostly because of attrition in the period between the retirement needs module (January 2008) and the match (September 2011). Therefore, it is important to verify that sample selection is not related to the offer of the retirement needs survey. We check this by testing for mean independence of sample selection with respect to the instrument.

2.6 Matching LISS and administrative data

Our basic unit of analysis is the household since we measure both wealth and income at the household level. We classify a household as being offered the survey if at least one household member that was eligible for the retirement needs module received the offer to participate. Likewise, we classify all households as treated in which at least one member that was offered the module actually filled it out. The construction of our estimation sample starts with 3,125 households that

[4]The Stata code we used for implementing this estimator is discussed in Frölich and Melly (2010).

contain at least one member that was eligible for the retirement needs module according to the criteria mentioned in section 2.2. After matching the LISS respondents to administrative data, we obtain wealth records for 1,429, 1,437 and 1,449 households in the years 2007, 2008, and 2009, respectively. We drop those households for which all eligible members were retired in 2008, reducing the sample to 1,275 households. Even with accurately measured administrative data, there can be large outliers in ratio variables. We trim all households for which 2008 savings rates relative to after-tax household income were larger than 50% in absolute value, leaving us with an estimation sample of 999 households. We also tried trimming the sample at savings rates larger than 75% and 100% of net household income and found that our results are quantitatively similar.[5]

Table 1 presents descriptive statistics of the those variables that we use as controls in the regressions, for both the full sample and for the estimation sample, and separately for couples and singles. The full sample consists of 2,816 rather than 3,125 households, because we exclude households for which all eligible members were retired in 2008. For couples, defined as households in which two partners live together irrespective of their marital status, individual-specific attributes are reported for the head of the household. Table 1 allows one to compare the characteristics of all eligible LISS households with those of the final estimation sample. Both for couples and singles, these samples are very similar.

One difference between the full sample and the estimation sample that is not reported in Table 1 is that of compliance to the survey offer. As mentioned in Section 2.2, 74% of the individuals in the full sample who were offered the survey participated. Household-level compliance is around 5 percentage points higher: among those households for which at least one eligible member was offered the survey, at least one member filled it out in 79% of the cases. In the estimation sample, the corresponding compliance rates are 82% for individuals and 87% for households, which is 8 percentage points higher than in the full sample. It is not surprising that compliance is related to being observed in our final dataset, since non-compliers to the survey offer are also more likely to leave the LISS Panel altogether. As a result, non-compliers were less likely to give their permission for the match to administrative records and are lost from our estimation sample. However, this

[5]Estimates are available on request.

Table 1: Descriptive statistics of the covariates

	Couples				Singles			
	Full sample (LISS)		Estimation sample		Full sample (LISS)		Estimation sample	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Female	0.12	(0.32)	0.11	(0.32)	0.58	(0.49)	0.62	(0.49)
Age	47.4	(11.8)	46.9	(11.5)	45.4	(12.1)	44.6	(11.6)
Children	1.16	(1.15)	1.15	(1.16)	0.41	(0.81)	0.36	(0.78)
Homeowner	0.83	(0.38)	0.81	(0.40)	0.49	0.50)	0.42	(0.50)
Marital status								
Married	0.81	(0.39)	0.83	(0.38)	0.06	(0.23)	0.04	(0.19)
Separated/divorced	0.05	(0.23)	0.04	(0.20)	0.33	(0.47)	0.38	(0.49)
Widowed	0.002	(0.04)	0.001	(0.04)	0.10	(0.30)	0.07	(0.25)
Never married	0.13	(0.34)	0.11	(0.32)	0.51	(0.50)	0.52	(0.50)
Education								
Primary	0.08	(0.28)	0.08	(0.28)	0.09	(0.28)	0.09	(0.29)
Int. Secondary	0.23	(0.42)	0.25	(0.43)	0.24	(0.43)	0.26	(0.44)
Higher secondary	0.07	(0.26)	0.08	(0.28)	0.08	(0.27)	0.07	(0.25)
Int. vocational	0.25	(0.43)	0.25	(0.43)	0.23	(0.42)	0.24	(0.43)
Higher vocational	0.25	(0.43)	0.25	(0.43)	0.25	(0.43)	0.28	(0.45)
University	0.11	(0.32)	0.09	(0.29)	0.10	(0.30)	0.06	(0.25)
Most imporant activity								
Employed	0.72	(0.45)	0.75	(0.43)	0.69	(0.46)	0.70	(0.46)
Self employed	0.11	(0.31)	0.07	(0.26)	0.09	(0.29)	0.08	(0.28)
HH work	0.01	(0.11)	0.01	(0.11)	0.05	(0.21)	0.06	(0.23
Retired	0.11	(0.31)	0.11	(0.32)	0.00	(0.00)	0.00	(0.00)
Disabled	0.03	(0.16)	0.03	(0.17)	0.07	(0.26)	0.07	(0.25)
Other	0.03	(0.16)	0.02	(0.15)	0.10	(0.30)	0.09	(0.29)
N	2167 (77.0%)		768 (76.9%)		649 (23.0%)		231 (23.1%)	

For couples all individual-specific variables refer to the head of the household.

does not compromise our research design, so long as the instrument is orthogonal to this selection process. In Section 3.1, we show that the instrument is not correlated with selection into the estimation sample. Hence, the comparison of households based on the random survey offer is as valid there as it would be in the full sample.

2.7 Descriptive statistics for the dependent variables

Table 2 describes our administrative assets records for the years 2007, 2008, and 2009 (all in 2008 euros). The single most important category of assets is that of the primary residence, with an average value of around 200,000 euro. Savings accounts follow at great distance as the second most important type both in terms of mean (27,000 euro) and median (13,000 euro) value. Real estate other than the primary residence is also important, but only for a small minority: the mean value is around 7,000 euro though only 8% of the sample has any non-residential real estate. The mean value of risky assets, stocks and bonds, drops from 7,210 euro in 2007 to 4,857 euro in 2009 (median holdings are zero in all years). Business wealth and other wealth are the least important categories of assets with a mean value below 1,500 euro in all years.

Table 2: Descriptive statistics of the assets and debt variables

	2007			2008			2009		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Assets									
Saving accounts	25,551	12,799	40,870	26,728	12,728	42,568	28,008	13,196	44,963
Risky assets	7,210	0	23,974	6,627	0	22,560	4,857	0	17,468
Property	196,713	205,571	162,020	201,325	212,589	161,463	199,616	212,181	155,063
Real estate	9,808	0	53,750	6,906	0	41,722	7,689	0	44,442
Business	1,202	0	12,588	1,212	0	13,842	1,459	0	15,485
Other	861	0	9,914	959	0	10,871	1,014	0	10,779
Debt									
Mortgage	105,119	86,091	105,553	104,079	84,827	103,787	108,243	91,412	106,892
Non-mortgage debt	2,375	0	13,796	1,992	0	12,621	2,432	0	18,823
Net worth	133,852	78,760	172,072	137,687	81,551	170,305	131,968	77,021	167,938
Net housing wealth	91,594	40,984	131,904	97,246	47,128	128,481	91,374	43,949	124,627
Net worth excl. housing	42,258	16,292	82,970	40,440	15,643	76,797	40,594	15,232	77,325
N		983			999			999	

All assets are reported in 2008 euros.

On average, households have about 105,000–110,000 euro in mortgage debt and around 2,000–2,500 euro of non-mortgage debt. Non-mortgage debt is concentrated in a small minority of 6% of the sample, among which the mean non-mortgage debt is around 20,000 euro.

Taking assets and debt together, the mean net worth of the households in the sample is around 135,000 euro. Unsurprisingly, net worth is concentrated in the primary residence, which has a mean value net of mortgage of around 95,000 euro. Because of the consumption value of housing, we compute savings based on the remaining 40,000 euro of non-housing savings.

Table 3 presents summary statistics for net income and for the outcome variables (savings and saving rates) used in the analysis. Mean household income is 38,165 euro in 2008 and the median is 35,699 euro, both of which are slightly higher than the average of 33,100 euro for the Dutch population at large (Centraal Bureau voor de Statistiek (CBS), 2012). 2008 non-housing savings, computed as the difference between the non-housing wealth stocks of January 1st 2008 and 2009, has a sample mean of 154 euro and a median of 2 euro, showing that the distribution of savings is centered around zero. There is, however, considerable variation in savings: the standard deviation is 9,411 euro. We compute savings rates as the level of savings divided by after-tax income. The distribution of savings rates is centered around zero, but there is considerable variation, too: the standard deviation of the savings rate is 19 percentage points.

Table 3: Descriptive statistics of outcomes

			Percentiles				
	Mean	Std. dev.	0.05	0.25	0.5	0.75	0.95
HH income[a]	38,165	17,649	16,289	27,056	35,699	46,107	67,474
Non-housing savings[b]							
Levels (2008 euros)	154	9,411	-13,632	-3,221	2	3,084	14,860
Savings rates	-0.01	0.19	-0.40	-0.09	0.00	0.09	0.33
N				999			

[a] HH income net of taxes.
[b] Savings corrected for inflation and net of property value and mortgages.

3 Results

3.1 Validity and relevance of the instrument

As explained in section 2.6, we lose about half of our sample when we match LISS Panel records with administrative data. This loss of data would compromise the internal validity of our empirical strategy if the matching of the LISS Panel observations with the administrative data were related to the instrument, the offer of the retirement needs module. Table 4 shows estimates of a linear regression that uses our instrument, called "offer", to explain an indicator of inclusion in the estimation sample. We find that sample selection is not correlated with the offer of the retirement needs module, so the loss of data that results from matching survey participants to administrative data does not invalidate our instrument.

Table 4: Exogeneity of the instrument w.r.t. data linkage

	Dependent variable: indicator for estimation sample
Offered	-0.0209
	(0.0201)
Multiple eligibles	0.00700
	(0.0215)
Constant	0.364***
	(0.0207)
Number of selected HHs	999
N	2,816

Robust standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

In the regression in Table 4 and in all other models reported below, we control for the presence of multiple household members. This is necessary because the randomization of the offer was performed at the level of the individual while the outcome variables we analyze are measured at the household level. We classify a household as being offered the survey if at least one eligible member received the offer, so by construction households with multiple eligible members are more likely to receive the offer. Conditional on the number of eligible household members, however, randomization across individuals ensures that the offer is random at the household level. We checked whether the randomization was successful by regressing our instrument, an indicator for being offered the retirement needs module, on all socio-demographic variables listed in Table 1 (controlling for the presence of multiple household members). The covariates are jointly insignificant ($p = 0.901$). Hence, the covariates are balanced, as one would expect given the randomization.

Turning to instrument relevance, the first-stage regression shows that the instrument is highly significant: the F-statistic for the coefficient of the instrument in a model that controls for the presence of multiple eligibles is 4,818.37. Complete estimation results for the first stage are reported in Table 5.

Table 5: First stage

	Dependent variable: HH treated
HH offered	0.879***
	(0.0127)
Multiple eligibles	-0.0376**
	(0.0159)
Constant	0.0231**
	(0.00988)
R squared	0.688
F(1, n-(k+1))	4,818.37***
N	999

Robust standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

3.2 Effects of survey participation on household saving

Table 6 presents ITT estimates of the effect of survey participation on household saving, both for the mean and at various quantiles. The top panel uses 2008 non-housing savings as the outcome variable, while the bottom panel explains the 2008 savings rate (non-housing savings divided by household income). In the mean regression, we find a 3 percentage points reduction in the savings rate, or about 1500 euros of 2008 non-housing savings. The estimated saving rates effects at various deciles are of the same order of magnitude as the mean effect, between 2 and 5 percentage points.

Table 7 presents our main results, obtained from IV regressions of participation in the retirement needs module on the two savings measures (levels and rates). The leftmost column shows the estimated coefficients and accompanying standard errors for the treatment dummy in 2SLS

Table 6: The effect of survey participation on savings
(ITT effects)

	Mean	Deciles[a] 0.20	0.30	0.40	0.50	0.60	0.70	0.80
	Dependent variable: 2008 non-housing savings (thousands of euros)							
Offered	-1.478**	-1.502	-0.956**	-0.581	-0.499	-0.697*	-0.968**	-0.791
	(0.672)	(1.011)	(0.475)	(0.373)	(0.357)	(0.382)	(0.459)	(0.685)
Sample statistics	0.154	-4.935	-2.061	-0.583	0.002	1.060	2.245	4.393
N	999							
	Dependent variable: 2008 non-housing savings rate (1 = 100%)							
Offered	-0.0308**	-0.0495*	-0.0256*	-0.0197*	-0.0113	-0.0233**	-0.0243*	-0.0312
	(0.0127)	(0.0258)	(0.0148)	(0.0112)	(0.0107)	(0.0113)	(0.0131)	(0.0196)
Sample statistics	-0.01	-0.14	-0.07	-0.02	0.00	0.03	0.07	0.13
N	999							

[a] For decile models we report unconditional treatment effects.
We control for the presence of multiple eligibles.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

models where we instrument survey participation with the random offer of the survey. Participation in the survey caused households to save 1,683 euro less on average during 2008. This is a large effect, especially considering the sample average of 154 euro and the standard deviation of 9,411. When we express savings relative to household income, we also find a significant and negative effect. Survey participation caused households to save 3.5 percentage points less on average, compared with a sample average of -1% and standard deviation of 19 percentage points.

The remaining columns of Table 7 report quantile IV effects; we report estimates for the second through eighth deciles. For the level of savings, we find significant and large effects for the third, sixth and seventh deciles. The estimated coefficients for the other deciles are also all negative. For the non-housing savings rate, we find strongly significant effects at the third and sixth deciles as well as marginally significant effects at the second and fourth deciles. These estimates show that large parts of the savings distribution were shifted by the survey, with similar effect sizes below and above the median. Alternative trimming rules yield quantitatively similar results, as

Table 7: The effect of survey participation on savings
(IV treatment effects)

		Deciles[a]						
	Mean	0.20	0.30	0.40	0.50	0.60	0.70	0.80
	Dependent variable: 2008 non-housing savings (thousands of euros)							
Treated	-1.683**	-1.792	-1.193**	-0.644	-0.474	-0.955**	-1.085**	-0.784
	(0.764)	(1.119)	(0.552)	(0.458)	(0.438)	(0.461)	(0.530)	(0.709)
Sample statistics	0.154	-4.935	-2.061	-0.583	0.002	1.060	2.245	4.393
Proportion compliers	0.875							
N	999							
	Dependent variable: 2008 non-housing savings rate (1 = 100%)							
Treated	-0.0351**	-0.0519*	-0.0337**	-0.0224*	-0.00922	-0.0352**	-0.0247	-0.0317
	(0.0144)	(0.0286)	(0.0166)	(0.0135)	(0.0130)	(0.0141)	(0.0153)	(0.0208)
Sample statistics	-0.01	-0.14	-0.07	-0.02	0.00	0.03	0.07	0.13
Proportion compliers	0.875							
N	999							

[a] For decile models we report unconditional treatment effects.
We control for the presence of multiple eligibles.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

did narrower definitions of wealth that include only risky assets and bank accounts.[6]

Our identification is based on the randomized offer of the retirement needs module to a subset of the eligible panel members. Randomization allows us to cleanly measure the causal effect of interest. Nevertheless, we added to the models reported in Table 7 all the covariates listed in Table 1 and found that all effects are robust to including these additional controls.[7]

3.3 Falsification test

As a falsification test, we estimate the same models with 2007 savings as the dependent variable, that is on behavior realized text itbefore exposure to the retirement needs module. Results are shown in Table 8. We find no evidence of any systematic differences in savings behavior, neither in terms of the average level of savings nor any of the deciles. Note that, in contrast to Table 7, the coefficients of the various deciles are not even of the same sign.

[6]Estimates available on request. Assets held in bank accounts and risky investments are provided directly by banks to the tax authority and as a result they are probably measured most accurately. Therefore, we also tried yearly savings in those categories as alternative outcome variables.

[7]Estimates available on request.

Table 8: Falsification tests
(IV treatment effects)

	Mean	Deciles[a] 0.20	0.30	0.40	0.50	0.60	0.70	0.80
	Dependent variable: 2007 non-housing savings (thousands of euros)							
Treated	-0.406	-0.582	0.130	0.158	-0.090	-0.393	-0.684	-1.792
	(0.749)	(0.889)	(0.460)	(0.414)	(0.417)	(0.478)	(0.832)	(1.251)
Proportion compliers	0.866							
N	1,014							
	Dependent variable: 2007 non-housing savings rate (1 = 100%)							
Treated	-0.0136	-0.00961	-0.00279	0.00197	-0.00198	-0.00296	-0.0211	-0.0257
	(0.0147)	(0.0249)	(0.0133)	(0.0121)	(0.0122)	(0.0136)	(0.0218)	(0.0298)
Proportion compliers	0.866							
N	1,014							

[a] For decile models we report unconditional treatment effects.
We control for the presence of multiple eligibles.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

3.4 Effect heterogeneity

We next investigate effect heterogeneity. One approach would be to run IV analyses on subsamples, but many variables that could be used for interesting splits of the sample are correlated. Examples are income and education, or income and age. Therefore, we prefer a regression approach, where we regress savings on an indicator equal to one if a household was offered the survey (our instrument in the preceding analysis); all covariates from Table 1; and interaction terms of covariates with the offer. Hence, we can interpret the results in this section as (heterogeneous) intention-to-treat effects. We investigate heterogeneity along the lines of income, education, and age. Note that, for reasons of sample size, the specification does not contain dummies for all cells defined by those variables, but only interactions of the separate variables with the offer indicator.

The upper panel of Table 9 displays coefficient estimates for the main effect and interaction terms of the model with the savings rate as the dependent variable. According to these estimates, the offer of the survey did not affect the average savings rate of young, income-poor households that are poorly educated. We find strong evidence for differential effects along the age and education dimensions: households with more highly educated or older heads reduced their savings more after having been offered the retirement needs module.

The lower panel of Table 9 shows the differences in savings between offered and non-offered households for subsamples defined along the age, education, and income categories; these estimates are functions of the estimated parameter values reported in the top panel of the table. Offered households with poorly educated heads who are younger than 40 and have a disposable income above the sample median actually saved close to 7 percentage points more than non-offered households with the same education and age. Households in the highest education category saved less regardless of the age of the head and their household income, but we find the strongest effects for older households: the intention-to-treat-effect is -3/-7 percentage points for households below age 40; -5/-10 percentage points for age 40-54; and -10/-15 for households aged 55 or older. We find similar heterogeneity in the survey effect on the level of savings.[8]

Table 9 shows that the size of the effect of survey participation on saving is much larger for the highly educated (college and university graduates). Indeed the coefficients imply that there is no statistically significant effect for low and middle education levels, while the point estimates for the high education level indicated falls of 8 percentage points or more, and are very statistically significant.

[8]Estimates available on request.

Table 9: Heterogeneous intention-to-treat effects – savings rate

	Dependent variable: 2008 savings rate (1 = 100%)
Offered	0.0213
	(0.0300)
Offered × HH inc. high	0.0482*
	(0.0268)
Offered × educ. middle	-0.0150
	(0.0319)
Offered × educ. high	-0.100***
	(0.0319)
Offered × age 40-54	-0.0200
	(0.0294)
Offered × age 55+	-0.0706**
	(0.0330)
Controls	Yes
R-squared	0.0574
N	999

	Heterogeneous effects					
	Income below median			Income above median		
	Education			Education		
	Low	Middle	High	Low	Middle	High
Age <40	0.0213	0.00626	-0.0789***	0.0695*	0.0545*	-0.0307
	(0.0300)	(0.0260)	(0.0290)	(0.0377)	(0.0304)	(0.0298)
Age 40-54	0.00129	-0.0138	-0.0989***	0.0495	0.0344	-0.0507*
	(0.0300)	(0.0308)	(0.0330)	(0.0319)	(0.0280)	(0.0270)
Age 55+	-0.0493*	-0.0643*	-0.150***	-0.00111	-0.0162	-0.101***
	(0.0280)	(0.0348)	(0.0342)	(0.0349)	(0.0371)	(0.0336)

Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

In addition, we also checked whether the intention-to-treat effect of the survey offer differs depending on whether the individual(-s) who received the offer is a husband; a wife; both husband and wife; a single male; or a single female. However, with our sample size, we are unable to discern differential intention-to-treat effects depending on whether men or women were offered the survey, or whether one or two household members were offered the survey. These estimates are available on request.

Next we further explored potential sources of the significant effect heterogeneity across education groups. The various estimates underlying the results discussed below are not presented in tables, but again are available on request.

Households in the highest education group have higher rates of home ownership. They also have higher pension wealth in terms of a standardized annuity from the first two (mandatory) pillars of the pension system: the median predicted annuity, net of taxes, is 1,442 euro/month for poorly educated households, compared with 1,725 and 2,039 euro/month respectively for the higher education groups. However, the replacement rate of the projected annuity relative to current income is similar for all education groups, ranging from 79% to 81%. Conditional on education, neither home ownership nor either measure of pension wealth further significantly interacts with the treatment effect. This suggests that differences in financial circumstances may not drive the heterogeneity in treatment effects across education groups.

An alternative hypothesis is that the education groups differ in the way the survey affects attention to retirement saving and needs. However, there is no significant difference in their reported rates of thinking about retirement prior to the survey. The high education group answer the survey more quickly than the low education group (median difference 2 minutes, $p = 0.058$), but the difference between the high education and middle income groups is not statistically significant. Moreover, we find no significant interaction between treatment effect and survey response time.

A question in the retirement needs module elicited subjects' expectations of housing costs in retirement. Interestingly, relative to low education households, highly educated households are much more likely to expect a decrease in housing costs after retirement. These differences are economically and statistically significant ($p < 0.001$). Unfortunately, as this question was asked in the retirement needs module, i.e. only of treated sample, we cannot test for treatment effect heterogeneity in this specific dimension.

3.5 Evidence on portfolio effects

Finally, we consider whether the observed effect of survey participation on household saving might reflect portfolio reshuffling rather changes in the level of active saving. There are two concerns. First, as explained in section 2.3, the administrative wealth data do not contain information on a relatively small category tax-favored savings accounts data. Though these are not a major asset category in the Netherlands, it remains possible that our findings of a negative effect of survey participation on non-housing savings results from a re-allocation of assets to those accounts. Second, our wealth-based measure of saving contains both active saving and capital gains. A possibility is that treated households shifted saving to risky assets and then experienced significant losses with the onset of the financial crisis in 2008. Changes in active saving and changes in portfolio allocation would both be a survey effect, but it is important and interesting to distinguish the nature of the effect.

With respect to the first point, we used survey data from the LISS assets module to look at investments in these tax-favored saving accounts. We tested for survey effects on accounts on both extensive and intensive margins. We find no effect on participation in such accounts, change in participation, balances conditional on participation or unconditional balances[9]. While the survey data may quite noisy at the intensive margin, we believe the participation is relatively well-measured. Turning to the second point, we use the administrative wealth data to test for an effect of survey participation on the portfolio share of risky assets, participation in risky assets, or changes in either shares or participation.[10] We do not find any evidence to suggest that capital gains for risky assets drive the survey effects that we document. Moreover, we also find survey effects when we limit our attention to changes in the balance of saving accounts, an asset class that does not exhibit large price-driven variation in value.

[9]In all these regressions, we find $p > 0.3$ for the effect of the treatment dummy; estimates are available on request.
[10]Estimates are available on request.

Our conclusion then is that participation in the survey had an effect on the level of active saving.

4 Conclusion

In this paper, we show that participating in a non-informative survey module on retirement needs led Dutch households to save significantly less. Our analysis uses administrative wealth data to calculate clean measures of savings that are not contaminated by the reporting styles of survey respondents, which themselves might have been affected by the intervention. Participation in the survey is instrumented by randomized assignment of invitations to participate in the module, so our estimates are unaffected by endogenous compliance. Estimated effects are large: the saving rate (saving as a fraction of disposable income) is 3.5 percentage points lower among treated households. Quantile IV models show effects across a wide range of the savings distribution. Falsification checks reveal no effect on savings prior to the survey. We find evidence for heterogenous treatment effects. The mean effect is driven by older and highly educated households. These households have the highest expected pension wealth and higher housing wealth. They are also much more likely, in the survey, to report that their housing costs will fall in retirement (perhaps because they anticipate paying off their mortgage debt). Our wealth data only allow us to compute savings during 2007 and 2008. Naturally we will investigate the durability of the effects when more data becomes available.

Our results are consistent with limited attention (DellaVigna, 2009). The survey may have made aspects of retirement needs and retirement planning more salient to participants. After reflecting on their expenditure needs in retirement, older and highly educated households concluded that they can afford to save less while the young and poorly educated marginally increased their savings. Asymmetric costs of adjustment may also be relevant. If the survey led some participants to conclude that they were saving too much, and some to conclude they were saving too little, the former may have found it easier to adjust.

These results are likely specific to the context of the experiment. At the time the retirement needs module was fielded in 2008, mandatory pensions were generous, replacing 80% of final income after tax on average, and covered nearly the entire population. De Bresser and Knoef (2013) show that many Dutch households were over-saving, relative to their self-reported desired expenditure level in retirement. They also show that older and more educated households could look forward to higher occupational pensions. Moreover, financial institutions were obliged to provide all pension holders with Universal Pensions Overviews (UPOs) from 2008 onward. This may have meant that households whose attention was drawn to their retirement needs and plans could obtain information on current entitlements and projections for age 65 at very low cost.

While effects likely due to limited attention have been documented previously (Stango and Zinman, 2013), our results highlight that such effects can operate in surprising directions depending on the context. We also show that the patterns of heterogeneous effects are context specific. We find the largest effects for the most educated while Stango and Zinman (2013) find the opposite. Perhaps most importantly, we show that such effects can operate on the most central of life-cycle choices, such as the level of consumption and saving. It is only through the continued accumulation of evidence on how survey effects manifest in widely-varied contexts, and with respect to different outcomes, that general models of these effects can be formulated and convincingly tested.

Survey effects such as the one documented in this paper imply that panels may not be representative of the underlying population even if the initial sampling was representative and the representativeness of the panel has not been degraded by attrition or non-response. If survey participation alters the behaviour of respondents the the external validity of any study based on such data will be compromised. Such considerations may be a further argument for greater use of administrative data whenever available (Einav and Levin, 2014).[11] Also, as noted by Zwane et al. (2011), survey effects may mean that it is better to achieve statistical power with large panels and infrequent measurement rather than with smaller, but more frequently measured samples.

In their recent survey, Warren and Halpern-Manners (2012) call for more research on panel conditioning and survey effects that employs experimental or quasi-experimental designs, and which documents effects in the kinds of large-scale longitudinal surveys that underpin much research

[11]However, we note that administrative data are subject to various methodological challenges as well. The perhaps most important problem of many administrative datasets is sample selectivity; see Browning et al. (2014) for a discussion.

in the social sciences. Our analysis shows one way in which this can be done: by exploiting the randomization of content that is becoming more common in such surveys. We strongly support their call for further research.

References

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Axinn, W. G., Jennings, E. A., and Couper, M. P. (2014). Response of sensitive behaviors to frequent measurement. Mimeo, Survey Research Center, University of Michigan.

Binswanger, J. and Schunk, D. (2012). What is an adequate standard of living during retirement? *Journal of Pension Economics and Finance*, 11(2):203–222.

Binswanger, J., Schunk, D., and Toepoel, V. (2013). Panel conditioning in difficult attitudinal questions. *Public Opinion Quarterly*, 77(3):783–797.

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5, pages 3705–3843. Elsevier, North-Holland, Amsterdam.

Bovenberg, A. and Meijdam, L. (2001). The Dutch pension system. In Börsch-Supan, A. H. and Miegel, M., editors, *Pension Reform in Six Countries*. Springer, New York.

Browning, M., Crossley, T. F., and Winter, J. K. (2014). The measurement of household consumption expenditures. *Annual Review of Economics*, 6:475–501.

Browning, M., Gørtz, M., and Leth-Petersen, S. (2013). Housing wealth and consumption: a micro panel study. *Economic Journal*, 123(568):401–428.

Burke, L. E., Wang, J., and Sevick, M. A. (2011). Self-monitoring in weight loss: a systematic review of the literature. *Journal of the American Dietetic Association*, 111(1):92–102.

Centraal Bureau voor de Statistiek (CBS), D. (2012). Welvaart in nederland – inkomen, vermogen en bestedingen van huishoudens en personen.

Crossley, T. F., Emmerson, C., and Leicester, A. (2012). Raising household saving. The British Academy, London.

Das, M., Toepoel, V., and Van Soest, A. (2011). Nonparametric tests of panel conditioning and attrition bias in panel surveys. *Sociological Methods & Research*, 40(1):32–56.

De Bresser, J. and Knoef, M. (2013). Can the Dutch meet their own retirement expenditure goals? Mimeo, Tilburg University.

DellaVigna, S. (2009). Psychology and economics: evidence from the field. *Journal of Economic Literature*, 47(2):315–372.

Dholakia, U. (2010). A critical review of question-behavior effect research. *Review of Marketing Research*, 7(7):145–197.

Duflo, E., Gale, W., Liebman, J., Orszag, P., and Saez, E. (2006). Saving incentives for low- and middle-income families: Evidence from a field experiment with H&R block. *Quarterly Journal of Economics*, 121(4):1311–1346.

Einav, L. and Levin, J. D. (2014). The data revolution and economic analysis. In Lerner, J. and Stern, S., editors, *Innovation Policy and the Economy*, pages 1–24.

Fitzsimmons, G. and Moore, S. (2008). Should we ask our children about sex, drugs and rock & roll? Potentially harmful effects of asking questions about risky behaviors. *Journal of Consumer Psychology*, 18(2):82–95.

Fitzsimmons, G. and Shiv, B. (2001). Nonconscious and contaminative effects of hypothetical questions on subsequent decision making. *Journal of Consumer Research*, 28(2):224–238.

Frölich, M. and Melly, B. (2010). Estimation of quantile treatment effects with Stata. *The Stata Journal*, 10(3):423–457.

Frölich, M. and Melly, B. (2013). Unconditional quantile treatment effects under endogeneity. *Journal of Business and Economic Statistics*, just-accepted.

Glenn, N. (1998). The course of marital success and failure in five American 10-year marriage cohorts. *Journal of Marriage and Family*, 60(3):569–576.

Kapteyn, A. and De Vos, K. (2008). Social security and retirement in the Netherlands. In Gruber, J. and Wise, D. A., editors, *Social Security and Retirement Around the World*, pages 269–304. University of Chicago Press.

Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2012). Getting to the top of mind: How reminders increase saving. Working Paper No. 16205, National Bureau of Economic Research (NBER).

Michie, S., Churchill, S., and West, R. (2011a). Identifying evidence-based competences required to deliver behavioural support for smoking cessation. *Annals of Behavioral Medicine*, 41(1):59–70.

Michie, S., Hyder, N., Walia, A., and West, R. (2011b). Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. *Addictive Behaviors*, 36(4):315–319.

Morwitz, V. G., Johnson, E., and Schmittlein, D. (1993). Does measuring intent change behavior? *Journal of Consumer Research*, 20(1):46–61.

Sakshaug, J. W., Couper, M. P., Ofstedal, A. B., and Weir, D. R. (2012). Linking survey and administrative records: Mechanisms of consent. *Sociological Methods & Research*, 41(4):535–569.

Scherpenzeel, A. (2011). Data collection in a probability-based internet panel: how the LISS panel was built and how it can be used. *Bulletin of Sociological Methodology*, 109(1):56–61.

Sherman, S. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology*, 39(2):211–222.

Spangenberg, E. (1997). Increasing health club attendance through self-prophecy. *Marketing Letters*, 8(1):23–31.

Stango, V. and Zinman, J. (2013). Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees. *Review of Financial Studies*, forthcoming.

Van Landeghem, B. (2012). Panel conditioning and self-reported satisfaction: evidence from international panel data and repeated cross-sections. SOEPpapers on Multidisciplinary Panel Data Research.

Warren, R. and Halpern-Manners, A. (2012). Panel conditioning in longitudinal social science surveys. *Sociological Methods & Research*, 41(4):491–534.

Zwane, A., Zinman, J., Dusen, E. V., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D., Hornbeck, R., Gine, X., Duflo, E., Devoto, F., Crepon, B., and Banerjee, A. (2011). Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences*, 108(1):1821–1826.