

CAN SURVEY PARTICIPATION ALTER HOUSEHOLD SAVING BEHAVIOUR?*

Thomas F. Crossley, Jochem de Bresser, Liam Delaney and Joachim Winter

We document an effect of survey participation on household saving. Identification comes from random assignment to modules within a population-representative Internet panel. The saving measure is based on linked administrative wealth data. Households that responded to a detailed questionnaire on needs in retirement reduced their non-housing saving rate by 3.5 percentage points, on a base of 1.5%. The survey may have acted as a salience shock, possibly with respect to reduced housing costs in retirement. Our findings present an important challenge to survey designers. They also add to the evidence of limited attention in household financial decision making.

Much empirical research in economics analyses 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 behaviour. 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 behaviour. 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 behaviour.

In this article, we test for survey effects in a central domain of economic research: household saving behaviour. In particular, we test whether being asked questions about retirement income needs leads to changes in household saving behaviour. Recent work in behavioural finance suggests possible mechanisms for survey effects on saving behaviour. Limited attention means that individuals tend to overlook some of the consequences of their decisions (DellaVigna, 2009). If those unnoticed consequences materialise 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 behaviour might be corrected by focusing individuals’ attention on the aspects they are missing. For example, recent literature suggests that sending out mailings can induce desired

* Corresponding author: Thomas F. Crossley, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK. Email: tcross@essex.ac.uk.

The authors thank the participants at the 2012 MESS workshop for helpful comments on an early version of this article. 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 The Federal Reserve Board, 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).

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

behaviour (such as tax compliance or personal care) cost effectively (Fellner *et al.*, 2013; Altmann and Traxler, 2014; Hallsworth *et al.*, 2014). 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 a 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 randomised 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 behaviour. 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, 2014), 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.¹ Our design, based on random invitations to a survey module, avoids concerns about attrition as well as age and time effects. In a recent review paper, Warren and Halpern-Manners (2012) note that the survey research literature on panel conditioning generally failed to employ randomised designs. They also point out that to date, there is little systematic evidence on panel conditioning in large-scale longitudinal social science surveys. Our article addresses this concern as well.

The psychology and marketing literatures have documented a number of related behavioural phenomena. The 'question-behaviour effect' refers to the observation that

¹ Das *et al.* (2011) consider a range of questions related to knowledge and behaviour, food choice, attitudes towards immigration, and pensions. The only questions that are related to the present article concern knowledge and expectations with respect to pensions. Generally, they do not find evidence of panel conditioning for those questions, with one exception, a knowledge question about a specific institutional feature of the Dutch pension system.

asking respondents to predict future behaviour results in more of the behaviour under consideration (Dholakia, 2010). For example, Spangenberg (1997) shows that asking people to predict their workout behaviour induces them to visit the gym more often. Where the behaviour 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 behaviours can increase the propensity to engage in those behaviours (Fitzsimons and Shiv, 2001; Fitzsimons and Moore, 2008; Dholakia, 2010). The public health literature reports some evidence that supports the use of self-recording of behaviour as a behavioural change intervention (Burke *et al.*, 2011; Michie *et al.*, 2011*a, b*), though recently Axinn *et al.* (2015) found little evidence of behavioural effects of keeping survey diaries of sexual behaviour 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 behaviour, or the recording of specific behaviour, to the propensity to engage in that behaviour. The survey effects that we study are less direct. We investigate the effect of being surveyed about topics related to the behaviour of interest, rather than direct questions about that behaviour. The question module that is our treatment has no specific questions about saving behaviour.

Two further relevant studies are Zwane *et al.* (2011) and Stango and Zinman (2014). Like us they consider survey effects that result from being surveyed on a topic related to the behaviour of interest, rather than from direct questions about that behaviour, or self-recording of that behaviour. Furthermore, they use administrative data to measure outcomes, in order to isolate genuine changes in behaviour 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 randomised 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 (2014) show that participants in a marketing research panel 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 article differs from Zwane *et al.* (2011) and Stango and Zinman (2014) 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 labour supply, than it is to induce small changes in the timing of transactions or portfolio composition, or to encourage the avoidance of fees. To our knowledge, there is no prior evidence on survey effects on core life-cycle choices like the level of savings or consumption. Our analysis further differs from Zwane *et al.* (2011), in that the data are drawn from a developed economy, and from Stango and Zinman (2014) in having a randomised design.

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

behaviour and retirement preparation in particular. Measurement is not innocuous. Even a random sample of a population, which has not been subject to non-random 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 randomised 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, 2014). 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 behaviour in different directions, depending on the context. While Karlan *et al.* (2012) find that deliberate and targeted reminders raise contributions to 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 (2014) 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 article is organised as follows. Section 1 describes the institutional context of our study. Section 2 describes the research design, data and methods. In Section 3, we present the main results as well as a number of robustness checks and falsification tests. Section 4 concludes.

1. Institutional Context

Active saving will reflect the economic circumstances that households face and so to interpret the results that follow, some background is required. The Dutch pension system in 2008 was characterised by arrangements that were almost universal and provided extremely generous income replacement.

The Dutch system of income provision during retirement consists of four categories or 'pillars'. The first pillar is that of the public pension, which provides anyone 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 2% for each year spent abroad).² The level of the public pension is set with 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 organised 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, on average, 70% of the gross last earned income before retirement. This translates into replacement rates net of taxes above 80% (Kapteyn and De Vos, 1999; Bovenberg and Meijdam, 2001). The third pillar contains

² Technically, one accrues public pension rights 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.

all private savings vehicles that are aimed specifically at retirement, such as life annuities, and these accounted for 7% of retirement income on average in the relevant period. Such voluntary arrangements are especially important for individuals who cannot rely on occupational pensions, such as the self employed. Finally, the remainder of retirement income comes from the fourth pillar which contains all other forms of wealth that can be drawn down after retirement, such as savings accounts, investments, and real estate.

There were no major changes to the Dutch pension system between 2002 and 2013, except for the move away from tax exempt early retirement schemes towards actuarially fair adjustment of benefits which took place between 2000 and 2006; see Kapteyn and De Vos (1999, 2007) and Kapteyn *et al.* (2010) for further descriptions of the Dutch pension system during the period covered by our study.

De Bresser and Knoef (2015) use responses to the retirement needs module in the LISS panel (our treatment) combined with the administrative wealth data we used in this study and other data to analyse whether the Dutch pension system succeeded in providing an adequate retirement income to its contributors around the time of our study. 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 requirements, 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 the authors 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 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 of our study 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 1 January 2008.

2. Data and Research Design

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 core demographic, financial, health and social topics. The LISS has two features that are crucial to our study.

First, eligible respondents are selected randomly to be asked to complete additional, non-core survey modules. These modules are typically designed and submitted by

researchers and fielded only once. Researchers can specify a subset of the LISS panel as eligible for their module. Researchers' required sample size often does not exhaust the full subset of those eligible, so costs are reduced by randomly choosing eligible panel members to be asked to complete the module. 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 randomly assigned module in the LISS Panel.

Second, we can 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. Critically, an independent measure of saving is necessary to be sure that observed effects represent genuine behavioural change and not changes in survey reporting style.

Details of these data sources and our research design are as follows.

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 administered 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.

2.2. *Treatment*

We define treatment as participation in the first randomly assigned module in the LISS Panel – the retirement needs module. This module, formally titled 'What is an adequate old age income?' was designed 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). The module was then fielded in the LISS panel in January 2008. Binswanger *et al.* (2013) then compare responses to the module across the two panels to study differences in responses between experienced (CentERpanel) and novice (LISS) respondents. They find greater non-response 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 article. They compare novice and experienced panel members, all of whom receive the retirement needs module, to estimate the effect of past experience with an Internet panel on survey responses (particularly responses 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 of 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 previously been exposed to randomised survey content. Eligible panel members who were not offered the retirement needs module were not offered another module. Moreover random assignment to 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.

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 led to a total eligible sample of 5,435 individuals, 2,755 of which were selected at random and were offered the retirement needs module. Multiple members of a household can be members of the LISS panel, and so households may contain zero, one or two eligible individuals. The 5,435 eligible individuals were members of 3,125 households.

The basic unit of our analysis is the household since, as elaborated below, we measure both wealth and income at the household level. We classify a household as eligible if it contained at least one eligible individual, and we define a household as 'offered' if at least one household member that was eligible for the retirement needs module received the request to participate. Similarly, we define a household as treated if at least one household member was offered the module and actually completed the module. The completion rate among those who received the offer was 74% at the individual level and 79% at the household level.

The retirement needs module consists of around 60 items that concern desired expenditure levels in retirement, the trade-off 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.

The exact wording of the questions can be found in Binswanger and Schunk (2012).

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 randomised survey module did not constitute an information shock. Since the survey did not include any questions on predicted or intended savings either, the randomised survey module could not induce a question-behaviour effect; respondents were not asked to predict their own behaviour.

2.3. *Saving Measures*

We investigate the effect of survey participation on household saving. Though the LISS data include elaborate biannual surveys on assets and debt, we construct our saving measure from matched administrative data on wealth for two reasons. First, we want to isolate effects of survey participation on economic behaviour. If we found an effect of survey participation on self-reported savings, this could reflect altered survey-reporting styles rather than altered economic behaviour. 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).

The administrative wealth data we use 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 on 1 January of consecutive years. Note that the timing is ideal for measuring the effect of a module fielded in January of 2008. 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 behaviour are limited by having only data on specific assets or accounts.³ Thus, changes in contributions may represent portfolio

³ For example, in Duflo *et al.* (2006), Karlan *et al.* (2012), and in many studies of individual retirement accounts, the outcome variables are contributions to one account; see Crossley *et al.* (2012) for a discussion.

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 <500 euro or <15 euro 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. Such accounts are taxed only during the payout phase and are therefore invisible in tax records up to retirement. However, they were not very important during the period we study, since annuities contributed only 7% of household retirement income (Bovenberg and Meijdam, 2001). 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 is 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 labour income it includes income from entrepreneurship and from assets (interest payments and imputed rent for 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. *Matching Survey and Administrative Data and Sample Selection*

The construction of our estimation sample starts with 3,125 households that contain at least one member that was eligible for the retirement needs module according to the criteria mentioned in Section 2. Incomplete linkage is often an important concern when combining survey and administrative data. Out of the 3,125 households that contain at least one member that was eligible for the retirement needs module, we could find administrative data matches for only 1,602. Respondent refusal of consent for matching can be a reason for incomplete linkage (Sakshaug *et al.*, 2012) but in our case panel attrition is a bigger problem. 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, and by this time many of the households who were eligible for the retirements needs module had attrited from the panel. De Bresser and Knoef (2015) show that only 10% of the respondents to the retirement needs module that were still in the LISS Panel in 2011 refused consent to the administrative data linkage. 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 quantitatively similar results to those reported below.⁴

Table 1 compares our matched estimation sample to a 'full' eligible sample from the LISS that excludes only households for which all eligible members were retired in 2008 (from the 3,125 eligible LISS households this produces a sample of 2,816 households). We provide descriptive statistics for these two samples separately for couples and singles. Couples are households in which two partners live together irrespective of their marital status and, among couples, individual-specific attributes (such as gender and age) are reported for the head of the household. For both couples and singles, these samples are very similar in terms of observables.

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 important 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	2,167 (77.0%)		768 (76.9%)		649 (23.0%)		231 (23.1%)	

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

⁴ These additional estimates are available on request.

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
<i>N</i>		983			999			999	

Note. All assets and debt are reported in 2008 euros.

Table 2 summarises, for our matched estimation sample, administrative assets records for the years 2007, 2008 and 2009 (all in 2008 euro). 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.

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. Not surprisingly, 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.

2.5. Threats to Validity

Our analysis faces three threats to internal validity. Figure 1 provides an overview of the issues.

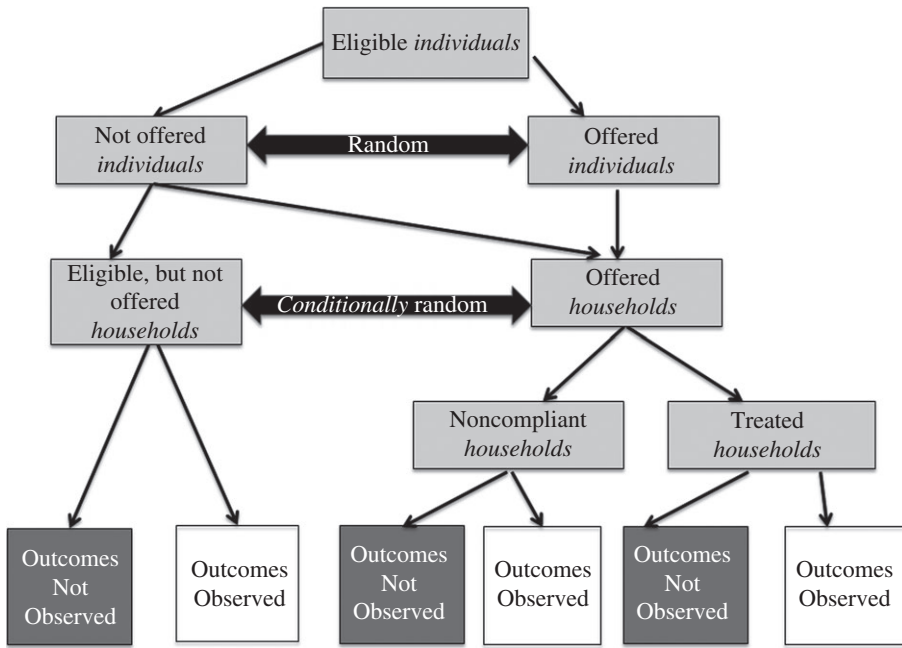


Fig. 1. Overview of the Research Design

The first threat is that the randomisation of the offer was performed at the level of the individual while the outcome variables we analyse 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. Thus we have conditionally random allocation to treatment and control: conditional on the number of eligible household members, randomisation across individuals ensures that the offer is random at the household level. Therefore in all of the regressions reported in the next Section, we control for the presence of multiple household members. As in practice, eligible households have either one or two eligible members, this amounts to including in our regression models a dummy variable equal to one if the household has a second eligible member.

We checked the conditional random allocation of households to treatment and control by regressing our instrument (an indicator for being offered the retirement needs module), on all the socio-demographic variables listed in Table 1 while controlling for the presence of multiple eligibles in the household. As expected, the dummy for multiple eligibles does predict receipt of offer at the household level. However, conditional on this control, the other covariates are jointly insignificant ($p = 0.901$). One-by-one balance tests also confirm that treatment and control groups are similar in terms of demographics, with no statistically or economically significant differences. Hence, the covariates are (conditionally) balanced, as one would expect given the conditional randomisation.

The second issue is incomplete compliance with the offer of treatment: not every LISS Panel member who was offered the retirement needs module completed this module. 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 (conditionally) 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, since they rely on exogenous variation in the module offer.⁵

Our intention to treat analysis, exploiting the conditionally random allocation of eligible households to the offer of the retirement needs module, is implemented by estimating the following regression:

$$S_{h,t} = \beta_0 + \beta_1 IME_h + \beta_2 Z_h + e_{h,t}, \quad (1)$$

where $S_{h,t}$ is the saving (level or rate) of household h at time t ; IME_h is an indicator variable for the presence of multiple eligible individuals in the household and Z_h is our instrument, which is equal to one if one of those eligible in the household received the offer of the retirement-needs-module. The parameter β_2 is the ITT effect.

Our estimate of the treatment effect is obtained by estimating the equation:

$$S_{h,t} = \alpha_0 + \alpha_1 IME_h + \alpha_2 T_h + u_{h,t}, \quad (2)$$

where T_h is the treatment dummy, equal to one if at least one of those eligible in the household received the offer of the retirement needs module and completed the module. We estimate this equation by instrumental variables with Z_h as the instrument for T_h . The parameter α_2 is the treatment effect. Importantly, the research design we use is characterised 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.

In addition to IV regressions for the conditional mean of the savings distribution, we also estimate decile treatment effects in order to establish the robustness of our results. We use the estimator proposed in Frölich and Melly (2013).⁶

The final threat to internal validity is the incomplete matching to the administrative data, described above. This would invalidate our design if, among those eligible, selection into our estimation sample were related to the offer of the retirements needs survey (i.e. if selection were related to our instrument). It is important to check whether this is the case.

The descriptive statistics in Table 1 show that the estimated sample is quite like the full sample of those eligible in terms of observables characteristics. However, one difference between the full sample and the estimation sample that is not reported in Table 1 is the degree of compliance with the survey offer. As noted above, in the full sample, 74% of individuals who were offered the survey participated. Household-level compliance in the full sample is 5 percentage points higher at 79%.⁷ In the estimation sample, the

⁵ This randomised assignment to a treatment and control group at a given point in time allows us to identify the effect of the survey on behaviour. Other research questions could be answered using different randomisations, such as randomised inclusion into a panel at different times in order to measure panel conditioning.

⁶ The Stata code we used for implementing this estimator is discussed in Frölich and Melly (2010).

⁷ Again, this is because some households have two eligible individuals who both received an offer and if either completes the survey we consider the household to be treated.

corresponding compliance rates are 82% for individuals and 87% for households. It is not surprising that compliance with the offer of the retirement needs module is related to inclusion in the estimation sample. Non-compliers with the survey offer are also more likely to leave the LISS Panel altogether. As a result, non-compliers were less likely to be present in 2011 to give consent for the match to administrative records. However, this does not compromise our research design, so long as the instrument is orthogonal to this selection process.

In the next Section, we formally test whether selection into the estimation sample is (mean) independent of the instrument. We do this by estimating the following regression on the full sample of eligible households:

$$IES_h = \gamma_0 + \gamma_1 IME_h + \gamma_2 Z_h + v_h, \quad (3)$$

where IES_h is a dummy variable which is equal to one if the household is included in the estimation sample. The hypothesis that inclusion in the sample is independent of the instrument implies $\gamma_2 = 0$, and this is our test.

3. Results

3.1. Outcomes: Descriptive Statistics

Table 3 presents summary statistics for net income and for the outcome variables (savings and saving rates) used in the analysis. Statistics are presented for 2007 and 2008, and the matched estimation sample of eligible households is split into those offered and not offered the module which is our treatment (i.e. values of our binary instrumental variable). In 2008 mean household income is 36,479 euro for the not-offered group and 38,979 for the offered group. Both values are slightly higher than the average of 33,100 euro for the Dutch population at large (Centraal Bureau voor de Statistiek (CBS), 2012). This is unsurprising given that conditions for eligibility for the module included a minimum income. The difference in mean incomes between the not-offered and offered groups reflects the fact that random assignment of the module to those eligible occurred at the level of the individual respondent, as described in the previous Section. This means that households with multiple eligible members were more likely to be offered and so offered households are more likely to have multiple eligible members. Table 3 reveals that in 2008, 82% of offered households contained multiple eligible individuals while only 62% of not-offered households contained multiple eligible individuals. These differences, and the resulting income differences, emphasise again that, at the household level, we have conditional random assignment to the offer of the module (conditional on the number of eligible individuals in the household). As explained in the previous Section, our empirical strategy accounts for this.

In 2007, the mean level of savings among eligible households who were subsequently offered the module was 1,207 euros, with a median of 1.6 euro. This is non-housing saving, computed as the difference between the non-housing wealth stocks of 1 January 2007 and 2008. For eligible households that were subsequently offered the module, mean savings were 1,247 euro, with a median of 184 euro. Note that the savings levels are very dispersed, with the standard deviation in both groups about 9,000 euro. In the

Table 3
Outcomes: Descriptive Statistics

	Mean	SD	P25	P50	P75
Panel (a) 2007					
Not offered					
HH income*	35,916	19,130	22,810	33,486	43,201
Level of savings†	1,271	8,854	-1,627	1,594	3,770
Savings rate	0.023	0.176	-0.054	0.000	0.110
Hholds w. multiple eligibles			0.58		
N			330		
Offered					
HH income*	39,624	18,619	28,780	36,761	47,020
Level of savings†	1,247	9,325	-2,295	184	4,270
Savings rate	0.017	0.178	-0.067	0.005	0.114
Hholds w. multiple eligibles			0.82		
N			684		
Panel (b) 2008					
Not offered					
HH income*	36,479	19,320	24,241	33,940	44,169
Level of savings†	1,105	9,513	-2,044	352	3,422
Savings rate	0.016	0.175	-0.062	0.010	0.108
Hholds w. multiple eligibles			0.62		
N			325		
Offered					
HH income*	38,978	16,739	28,223	36,787	46,661
Level of savings†	-305	9,334	-3,857	0.000	2,823
Savings rate	-0.016	0.199	-0.112	0.000	0.082
Hholds w. multiple eligibles			0.82		
Module participation			0.79		
N			674		

Notes. *HH income net of taxes. †Savings corrected for inflation and net of property value and mortgages. All monetary variables in 2008 euros.

same year, the mean savings rates were 2.3% for the not-offered and 1.7% for the offered. We compute savings rates as the level of savings divided by after-tax income. The medians are respectively 0% and 1%. As with saving levels, savings rates are very dispersed with a standard deviation of 18 percentage points in both groups.

In 2008, the mean saving rate of the not-offered group fell somewhat, to 1.6%. The mean saving rate of the offered group fell much more, from 1.7% to (-1.6)% (a fall of 3.1 percentage points). This suggests that the participation in the module may have had an effect on the saving behaviour of the 79% of the offered group that did participate. Note, however, that the saving rate of the offered group in 2007 does not provide a credible counterfactual for the saving rate of that group in 2008 (because of time effects). The saving rate of the non-offered group in 2008 also does not provide a credible counterfactual for the saving rate of the offered in 2008 (because the groups differ in the frequency of households with multiple eligible individuals). We now turn to the ITT and IV estimators that provide a credible counterfactual and therefore credible estimates of the causal effect of participation in the module.

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

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

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 ²	0.688
F(1, 996)	4,818.37***
N	999

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

3.2. Validity and Relevance of the Instrument

As explained in Section 4, 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 (3). This a linear regression that uses our instrument, called ‘offer’, and a dummy for households with multiple eligible individuals, 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 research design.

Turning to instrument relevance, the first-stage regression reported in Table 5 demonstrates that the instrument is highly relevant: the F-statistic for the coefficient of the instrument in a model that controls for the presence of multiple eligible individuals is 4,818.37.

Table 6
The Effect of Survey Participation on Savings (ITT Effects)

	Mean	Deciles†						
		0.20	0.30	0.40	0.50	0.60	0.70	0.80
Dependent variable: 2008 non-housing savings (thousands of euro)								
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				

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

Together these two tables confirm that our empirical strategy overcomes both incomplete matching of households to the administrative wealth data and non-compliance with the offer of module participation.

3.3. Main Results

Table 6 presents ITT estimates of the effect of (the offer 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 reduction in the savings rate of 3.1 percentage points, or about 1,500 euro 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.

These ITT estimates, in conjunction with the summary statistics in Table 3, allow us to calculate counterfactual savings levels and rates for the offered group in 2008. In particular, subtracting the ITT estimate of (-3.1) percentage points from the observed saving rate of (-0.016) gives a counterfactual saving rate of 1.5% for the offered group in 2008. An analogous calculation gives a counterfactual saving level of 1,173.

Table 7 presents our main results, obtained from IV regressions of the two savings measures (levels and rates) on participation in the retirement needs module, where we instrument survey participation with the random offer of the survey and control for multiple eligibles in the household. The leftmost column shows the estimated coefficients and accompanying standard errors for the treatment dummy in 2SLS models, which capture the mean effect. Participation in the survey caused households to save 1,683 euro less on average during 2008. This is a large effect, and can be compared to the counterfactual saving of the offered group of 1,173 euro.

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

Table 7
The Effect of Survey Participation on Savings (IV Treatment Effects)

Mean	Deciles†							
	0.20	0.30	0.40	0.50	0.60	0.70	0.80	
Dependent variable: 2008 non-housing savings (thousands of euro)								
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				

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

less on average, compared with counterfactual mean saving rate of 1.5% and a standard deviation of 18 percentage points (both for the offered group). On average, participation in the survey moved households from modest saving to modest dis-saving.

The remaining columns of Table 7 report quantile IV effects. We report estimates for the second to eighth deciles in order to show that participation in the retirement needs module affected subsequent savings throughout the distribution of savings rates. 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 saving 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 participation in the survey, with similar effect sizes below and above the median. The mean effects reported above are not driven by responses in the tails of the saving distribution.

We investigated the robustness of these results in a number of other ways as well. In particular, alternative trimming rules yield quantitatively similar results, as do narrower definitions of wealth that include only risky assets and bank accounts.⁸

Our identification is based on the randomised offer of the retirement needs module to a subset of the eligible panel members. This implies conditional randomisation to households (conditional on the number of eligibles in a household) and allows us to cleanly measure the causal effect of interest. Nevertheless, we added to the models

⁸ Assets held in bank accounts and risky investments are provided directly by banks to the tax authority and as a result are probably measured most accurately. Therefore, we also tried yearly savings in those categories as alternative outcome variables. Estimates are available on request.

Table 8
Falsification Tests (IV Treatment Effects)

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

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

reported in Table 7 all the covariates listed in Table 1 and found that all effects are robust to including these additional controls.⁹

3.4. Falsification Test

As a falsification test, we estimate the same IV model from (2) with 2007 savings as the dependent variable, that is on behaviour realised before exposure to the retirement needs module. The results of this exercise are reported in Table 8. We find no evidence of any systematic difference in savings behaviour, neither in terms of the mean level or rate of saving, nor at any of the deciles. Note that, in contrast to Table 7, the coefficients of the various deciles vary in sign. This provides further assurance that our research design is valid and that our main results are not driven by incomplete data linkage (as any differential selection into the match sample would also affect the 2007 results).

3.5. 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 pooled IV approach, in which we interact the treatment (survey participation) dummy with household characteristics, and then generate additional instruments by interacting our instrument (the offer dummy) with household characteristics. In addition to the interaction between selected covariates and the treatment dummy, the estimated model includes all the covariates from Table 1 as exogenous independent variables. We investigate heterogeneity along the lines of income, education and age.

⁹ Estimates available on request.

Table 9
Heterogeneous IV Treatment Effects – Savings Rate

Dependent variable: 2008 savings rate (1 = 100%)						
Treated	0.0283 (0.0347)					
Treated × HH inc. high	0.0526* (0.0301)					
Treated × educ. middle	−0.0172 (0.0363)					
Treated × educ. high	−0.115*** (0.0355)					
Treated × age 40–54	−0.0249 (0.0337)					
Treated × age 55+	−0.0838** (0.0380)					
Controls	Yes					
R ²	0.0586					
N	999					
Heterogeneous effects						
	Income below median			Income above median		
	Education			Education		
	Low	Middle	High	Low	Middle	High
Age <40	0.0283 (0.0347)	0.0112 (0.0305)	−0.0862*** (0.0323)	0.0809* (0.0427)	0.0637* (0.0355)	−0.0336 (0.0337)
Age 40–54	0.00348 (0.0337)	−0.0137 (0.0349)	−0.111*** (0.0367)	0.0560 (0.0351)	0.0389 (0.0319)	−0.0585* (0.0301)
Age 55+	−0.0555* (0.0308)	−0.0727* (0.0395)	−0.170*** (0.0385)	−0.00293 (0.0387)	−0.0201 (0.0426)	−0.117*** (0.0388)

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

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 treatment 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 survey did not affect the average savings rate of the base group: 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 evidence for effect heterogeneity by current income is weaker. Once we allow the treatment effect to interact with education, the interaction with income is positive (indicating that, conditional on education, households with higher current income cut saving by less). But the income interaction is only weakly statistically significant ($p < 0.1$) and much smaller in magnitude than the interaction with higher education. It may be that education is a better measure of long-run economic position or retirement preparedness.

We then use the coefficients of this model to calculate treatment effects for households with different combinations of characteristics. These are reported in the lower panel of Table 9. Participation actually raised the saving rate of households with poorly educated heads who are younger than 40 and have a disposable income above the sample median by 8 percentage points (though this effect is not particularly precise). 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 mean treatment effect is $-3/-9$ percentage points for households below age 40; $-5/-11$ percentage points for age 40–54; and $-12/-17$ percentage points for households aged 55 or older. We find similar heterogeneity in the survey effect on the level of savings.¹⁰

Overall, 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). In addition, we also investigated effect heterogeneity by gender and family type, and whether the number of household members participating matters. In particular, we checked whether the effect 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 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 the 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 standardised 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 two 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.

¹⁰ Estimates available on request.

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 the treated sample, we cannot test for treatment effect heterogeneity in this specific dimension. Nevertheless, this difference raises the possibility that the survey made salient to high education households likely decreases in housing expenditure after retirement.

3.6. *Evidence on Portfolio Effects*

Finally, we consider whether the observed effect of survey participation on household saving might reflect portfolio reshuffling rather changes in active saving. There are two concerns. First, as explained earlier, the administrative wealth data do not contain information on a relatively small category of tax-favoured savings accounts. 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-favoured 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.¹¹ While the survey data may be 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.¹² 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. These checks point to the effect documented above being an effect on active saving, and not an effect on portfolio allocation.

4. Discussion and Conclusion

In this article, 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

¹¹ In all these regressions, we find $p > 0.3$ for the effect of the treatment dummy; estimates are available on request.

¹² Estimates are available on request.

have been affected by the intervention. Participation in the survey is instrumented by randomised 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 (on a base saving rate of 1.5%). 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 become available.

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 heterogeneous 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 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. They only need to consume more.

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 (2015) show that many Dutch households were over-saving, relative to their self-reported desired expenditure level in retirement. They also show that 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 onwards. 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. It is not clear how households would react to the salience shock of the survey if they were largely under-saving (especially if adjustment costs are asymmetric), or if information on retirement preparedness was costly to obtain.

While effects which are likely due to limited attention have been documented previously (Stango and Zinman, 2014), 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 (2014) 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. These salience effects might be exploited for purposes of pension policy. But it is only through the continued accumulation of evidence on how salience shocks manifest in widely-varied contexts, and with respect to different outcomes, that general models of these effects can be formulated and convincingly tested.

The validity of empirical research that uses survey data relies on the representativeness of the sample.¹³ Survey effects such as the one documented in this article 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 external validity of any study based on such data will be compromised. Moreover, if the behaviour of survey participants is altered by participation, new ethical issues arise in survey research. This issue is already being debated in the public health literature (Fitzsimons and Moore, 2008; Schwarz, 2008).

Such considerations may be a further argument for greater use of administrative data whenever available (Einav and Levin, 2014).¹⁴ 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 in the social sciences. Our analysis shows one way in which this can be done: by exploiting the randomisation of content that is becoming more common in such surveys. We strongly support their call for further research.

Essex University and Institute for Fiscal Studies, London

University of Groningen

University of Stirling

University of Munich

Accepted: 15 December 2015

Additional Supporting Information may be found in the online version of this article:

Data S1.

References

- Altmann, S. and Traxler, C. (2014). 'Nudges at the dentist', *European Economic Review*, vol. 72, pp. 19–38.
- Angrist, J., Imbens, G. and Rubin, D. (1996). 'Identification of causal effects using instrumental variables', *Journal of the American Statistical Association*, vol. 91(434), pp. 444–55.
- Axinn, W.G., Jennings, E.A. and Couper, M.P. (2015). 'Response of sensitive behaviors to frequent measurement', *Social Science Research*, vol. 49, pp. 1–15.
- Binswanger, J. and Schunk, D. (2012). 'What is an adequate standard of living during retirement?', *Journal of Pension Economics and Finance*, vol. 11(2), pp. 203–22.
- Binswanger, J., Schunk, D. and Toepoel, V. (2013). 'Panel conditioning in difficult attitudinal questions', *Public Opinion Quarterly*, vol. 77(3), pp. 783–97.

¹³ A recent example is Heffetz and Rabin (2013) who use survey measures of subjective well-being to illustrate the selection problems that arise from survey non-response.

¹⁴ However, we note that administrative data are subject to various methodological challenges as well. The perhaps most important problem of many administrative data sets is sample selectivity; see Browning *et al.* (2014) for a discussion.

- Bound, J., Brown, C. and Mathiowetz, N. (2001). 'Measurement error in survey data', in (J. Heckman and E. Leamer, eds.), *Handbook of Econometrics*, vol. 5, pp. 3705–843, Amsterdam: Elsevier.
- Bovenberg, A. and Meijdam, L. (2001). 'The Dutch pension system', in (A.H. Börsch-Supan and M. Miegel, eds.), *Pension Reform in Six Countries*, pp. 39–67, New York: Springer.
- Browning, M., Crossley, T.F. and Winter, J.K. (2014). 'The measurement of household consumption expenditures', *Annual Review of Economics*, vol. 6, pp. 475–501.
- Browning, M., Gørtz, M. and Leth-Petersen, S. (2013). 'Housing wealth and consumption: a micro panel study', *ECONOMIC JOURNAL*, vol. 123(568), pp. 401–28.
- 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*, vol. 111(1), pp. 92–102.
- Centraal Bureau voor de Statistiek (CBS). (2012). 'Welvaart in nederland – inkomen, vermogen en bestedingen van huishoudens en personen'. Den Haag/Heerlen.
- Crossley, T.F., Emmerson, C. and Leicester, A. (2012). *Raising Household Saving*, London: The British Academy.
- Das, M., Toepoel, V. and Van Soest, A. (2011). 'Nonparametric tests of panel conditioning and attrition bias in panel surveys', *Sociological Methods & Research*, vol. 40(1), pp. 32–56.
- De Bresser, J. and Knoef, M. (2015). 'Can the Dutch meet their own retirement expenditure goals?', *Labour Economics*, vol. 34, pp. 100–17.
- DellaVigna, S. (2009). 'Psychology and economics: evidence from the field', *Journal of Economic Literature*, vol. 47(2), pp. 315–72.
- Dholakia, U. (2010). 'A critical review of question-behavior effect research', *Review of Marketing Research*, vol. 7(7), pp. 145–97.
- 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*, vol. 121(4), pp. 1311–46.
- Einav, L. and Levin, J.D. (2014). 'The data revolution and economic analysis', in (J. Lerner and S. Stern, eds.), *Innovation Policy and the Economy*, pp. 1–24, Chicago, IL: The University of Chicago Press.
- Fellner, G., Sausgruber, R. and Traxler, C. (2013). 'Testing enforcement strategies in the field: threat, moral appeal and social information', *Journal of the European Economic Association*, vol. 11(3), pp. 634–60.
- Fitzsimons, 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*, vol. 18(2), pp. 82–95.
- Fitzsimons, G. and Shiv, B. (2001). 'Nonconscious and contaminative effects of hypothetical questions on subsequent decision making', *Journal of Consumer Research*, vol. 28(2), pp. 224–38.
- Frölich, M. and Melly, B. (2010). 'Estimation of quantile treatment effects with Stata', *Stata Journal*, vol. 10(3), pp. 423–57.
- Frölich, M. and Melly, B. (2013). 'Unconditional quantile treatment effects under endogeneity', *Journal of Business & Economic Statistics*, vol. 31(3), pp. 346–57.
- Glenn, N. (1998). 'The course of marital success and failure in five American 10-year marriage cohorts', *Journal of Marriage and Family*, vol. 60(3), pp. 569–76.
- Hallsworth, M., List, J., Metcalfe, R. and Vlaev, I. (2014). 'The behavioralist as tax collector: using natural field experiments to enhance tax compliance', Working Paper No. 20007, National Bureau of Economic Research.
- Heffetz, O. and Rabin, M. (2013). 'Conclusions regarding cross-group differences in happiness depend on difficulty of reaching respondents', *American Economic Review*, vol. 103(7), pp. 3001–21.
- Kapteyn, A. and De Vos, K. (1999). 'Social security and retirement in the Netherlands', in (J. Gruber and D. A. Wise, eds.), *Social Security and Retirement Around the World*, pp. 269–303, Chicago, IL: University of Chicago Press.
- Kapteyn, A. and De Vos, K. (2007). 'Simulation of pension reforms in the Netherlands', in (J. Gruber and D. A. Wise, eds.), *Social Security Programs and Retirement Around the World: Fiscal Implications of Reform*, pp. 327–50, Chicago, IL: University of Chicago Press.
- Kapteyn, A., De Vos, K. and Kalwij, A. (2010). 'Early retirement and employment of the young in the Netherlands', in (J. Gruber and D. A. Wise, eds.), *Social Security Programs and Retirement Around the World: the Relationship to Youth Employment*, pp. 243–59, Chicago, IL: 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*, vol. 41(1), pp. 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*, vol. 36(4), pp. 315–9.
- Morwitz, V.G., Johnson, E. and Schmittlein, D. (1993). 'Does measuring intent change behavior?', *Journal of Consumer Research*, vol. 20(1), pp. 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*, vol. 41(4), pp. 535–69.
- 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*, vol. 109(1), pp. 56–61.
- Schwarz, N. (2008). 'Introduction to research dialogue', *Journal of Consumer Psychology*, vol. 18(4), pp. 242–3.
- Sherman, S. (1980). 'On the self-erasing nature of errors of prediction', *Journal of Personality and Social Psychology*, vol. 39(2), pp. 211–22.
- Spangenberg, E. (1997). 'Increasing health club attendance through self-prophecy', *Marketing Letters*, vol. 8(1), pp. 23–31.
- Stango, V. and Zinman, J. (2014). 'Limited and varying consumer attention: evidence from shocks to the salience of bank overdraft fees', *Review of Financial Studies*, vol. 27(4), pp. 990–1030.
- Van Landeghem, B. (2014). 'A test based on panel refreshments for panel conditioning in stated utility measures', *Economics Letters*, vol. 124(2), pp. 236–8.
- Warren, R. and Halpern-Manners, A. (2012). 'Panel conditioning in longitudinal social science surveys', *Sociological Methods & Research*, vol. 41(4), pp. 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*, vol. 108(1), pp. 1821–6.