Vanguard Research Initiative:

Survey 1 Documentation and Tabulations

John Ameriks, Vanguard Group

Andrew Caplin, New York University and NBER

Minjoon Lee, University of Michigan

Matthew D. Shapiro, University of Michigan and NBER

Christopher Tonetti, Stanford University

December 2014

Vanguard Research Initiative:

Survey 1 Documentation and Tabulations

ABSTRACT

This report provides documentation and analysis of the Vanguard Research Initiative (VRI)—a new panel study of clients of the Vanguard Group combining survey and administrative data—that is designed to yield high-quality measurements of a large sample of older Americans who face meaningful financial tradeoffs approaching and during retirement. It provides documentation relating to the design and sampling of the VRI panel, detailed sample characteristics, and detailed analysis of the survey and administrative measurement of financial assets

John Ameriks
The Vanguard Group, Inc.
PO Box 2600
Valley Forge PA 19482-2600
john_ameriks@vanguard.com

Andrew Caplin
Department of Economics
New York University
New York, NY 10012
 and NBER
andrew.caplin@nyu.edu

Minjoon Lee
Department of Economics
 and Survey Research Center
University of Michigan
Ann Arbor, MI 48109-1248
minjoon@umich.edu

Matthew D. Shapiro
Department of Economics
 and Survey Research Center
University of Michigan
Ann Arbor, MI 48109-1248
 and NBER
shapiro@umich.edu

Christopher Tonetti
Graduate School of Business
Stanford University
Stanford CA, 94305-7298
tonetti@stanford.edu

This report provides documentation and analysis of the Vanguard Research Initiative (VRI)—a new panel study of clients of the Vanguard Group combining survey and administrative data—that is designed to yield high-quality measurements of a large sample of older Americans who face meaningful financial tradeoffs approaching and during retirement. It provides documentation relating to the design and sampling of the VRI panel, detailed sample characteristics, and detailed analysis of the survey and administrative measurement of financial assets.

The sections of this report are as follows:

1. Survey Development and Testing

2. Sample Design

3. Invitations and Survey Mechanics

4. Response Analyses

5. Detailed Sample Characteristics

6. Survey versus Administrative Data: Analysis by Account-type and Regression Analysis

7. Use of the Correction Mechanism: Regression Analysis

8. Alternative Approaches to Measuring Stock Share: Account-by-account versus Overall

Further results and an application to estimating the relationship between financial wealth and expected time to retirement can be found in Ameriks et al. (2014).

REFERENCE

Ameriks, John, Andrew Caplin, Minjoon Lee, Matthew D. Shapiro, and Christopher Tonetti (2014) "The Wealth of Wealthholders," Vanguard Research Initiative Working Paper.

Vanguard Research Initiative
December 2014

1. Survey Development and Testing

We developed the survey instrument and sample using a number of deliberate steps to assure data quality and success of the survey design.

Cognitive interviews. We conducted two rounds of cognitive interviews in the early stages of the Vanguard Research Initiative.[1] Both sets of interviews were conducted by the Survey Research Center of the University of Michigan. The first set was based on a sample of Vanguard participants with over $100,000 in Vanguard assets. They were first given a small screening interview via the Internet. A subset was invited to do cognitive interviews via phone where the questions focused on retirement planning and interest in annuities. We used the results of this cognitive interview in designing multiple VRI surveys. The second set of cognitive interviews was conducted in person by the Survey Research Center on a small sample of individuals in the Ann Arbor area with similar characteristics to VRI. This round of interviews was used to test and evaluate the VRI Internet survey interface. Individuals were given pilot questions via a laptop and interviewed as they did the survey by SRC interviewers. For both sets of cognitive interviews, the research team analyzed and discussed detailed, de-identified summaries of the interviews. It used these results to improve wording and presentation of the surveys.

One important substantive finding of the cognitive interviews is that this group is familiar with annuities, has a good understanding of them, and has little interest in purchasing them.

---

[1] We gratefully acknowledge the collaboration of Wandi Bruine de Bruin, Alycia Chen, and Brooke Helppie McFall in the design, implementation, and interpretation of the cognitive interviews.

Subsequent VRI surveys are aimed at establishing whether this carries over to large-scale populations.

Pilot and Production Surveys and Samples. An initial round of survey was conducted using a Pilot Sample (747 completed surveys). The purpose of this pilot was to test our procedures, e.g., invitation, survey implementation, recording data, and whether the questions were working as designed. We used the pilot for the first VRI survey (the wealth survey discussed in this paper) to estimate response rates for the production survey as well as to estimate the oversampling rate of administrative singles (see Vanguard Research Initiative: Documentation and Supporting Analysis, "2. Sample Design"). The survey firm also conducted online chats with randomly-selected pilot respondents to assess their understanding and their attitudes about the survey (see below).

Based on the pilot, we made modest adjustments and correction to the survey instrument. We also adjusted our plans for the number of invitations based on the response rate and on the relationship between administrative- and survey-singles. Subsequent VRI surveys use the same structure, i.e., using the same pilot sample to test the design. Much of the pilot and production surveys are similar or identical, so we will be able to use the pilot data as well as the production data for some analyses.

This paper uses the Survey 1 production sample except where indicated.

iModerate. The survey firm contracted to have a separate firm conduct on-line chats (iModerate) with randomly-selected pilot respondents during the first two days that the pilot was in the field. These were conducted at the end of the pilot survey via a pop-up window. The chats were based on a loose script (similar to a cognitive interview) to assess which questions were difficult and whether the respondents found the survey understandable or difficult overall.

The typical chat lasted 15 to 20 minutes. The survey team made some adjustments based on de-identified summaries of these chats. Overall, the reaction of the respondents to the survey was quite positive. Even questions that we expected to be quite difficult were generally well-received.

2. Sample Design

1.   Screens using administrative data

Administrative data provide information about account holders including types and balances of accounts, age, gender, indicator of single status, and history of log-on to Web accounts. We used this information to create the sample frame for the study. In the following, we explain each layer of the sampling screen and discuss the effect of it.

Universe. We first form the universe by imposing a set of conditions on the entire pool of Vanguard account holders. Most importantly, to be in the universe, the account holder should be at least 55 years old. Several additional conditions include: excluding accounts with foreign address, those with an immediate annuity purchase from Vanguard (reserved for future use), and those with very large assets (approximately over $5M).

Web-survey eligible. The account holder must be registered for use of the Vanguard website, have a valid email address, and have logged on their Web accounts in the past six months. Additionally, the account holder cannot have requested not to receive a survey and not have been participated in a recent survey.

Balance limit. The account had to be either in the Vanguard individual client line of business or the employer sponsored line of business and have at least $10,000 in Vanguard assets in these lines.

Discussion of effect of sample screens. Appendix Table 2-1 shows the effect of these screens and how they interact. This account balance requirement binds only for a small fraction of accounts once the Web-eligibility requirements are imposed. Web-eligibility *per se* also does not have a big effect on respondent characteristics except for age.

Employer sponsored and individual client mutually exclusive. For convenience in administering the survey, we define the employer-sponsored account universe to be those account holders with only employer sponsored accounts. The individual account holders are the complement of this group, so they may also have an employer-sponsored account. There is also some fluidity across these groups, in particular, some of the individual account holders are former employee-sponsored account holds who rolled over their employer-sponsored accounts to Vanguard IRAs.

2.   Sampling by age and line of business.

We stratified the invitations to assure sufficient observations across all ages. The administrative data has a high-quality measure of age. We drew the sample from the universe in age bins so that 20% of the sample would fall in each of ages 55-59, 60-64, 65-69, 70-74, and 75 and above. For ages below 65, we drew evenly from the employer-sponsored and individual client lines of business. For older respondents, we made no distinction between line of business because most clients leave their employer-sponsored plan once they retire. See Table 2-2 for a summary of the sample design.

3.   Oversampling singles.

In order to have a larger sample of singles for the structural analyses we contemplated, we used the administrative indicator of single status to oversample singles. The administrative measure of marital status is constructed by Vanguard based on various indicators—sharing common registration and address. The measure is not perfect. Table 2-3 shows the distribution of survey reports of marital status by the administrative measure from the pilot survey (see

Vanguard Research Initiative: Documentation and Supporting Analysis, "1. Survey

Development and Testing" for the explanation on the pilot survey). In the production sample,

our target was to have approximately a third of the sample as single households. To hit this

target, based on the relationship between administrative single and survey single measure, we

chose the sampling rates shown in Panel B of Table 2-2.[1] For the younger ages, we oversample

administratively-single accounts by the ratio of 2 to 1. We reduce the oversampling for older

ages because singles occur at a greater natural rate as the account holder ages. For the sample

strata within employer-sponsored accounts, we are not able to effectively target singles because

employer-sponsored accounts are never joint. In the Appendix F, we show that we achieved

these targets across age groups. Note that in no case did we use quota sampling.

---

[1] We only oversampled singles in the individual client line of business. To be in the employer-sponsored sample, we required that they not have Vanguard individual accounts, so they are very unlikely to show up as non-singled in the administrative data.

Table 2-1. Effect of Sampling Screens in VRI

Screen	Fraction
Universe	100.0%
Web-survey eligible	34.1%
Asset cut-off	86.3%
Web-survey eligible and asset cut-off	29.6%
One sample from household id	28.1%

Note: Universe is as the set of account holders at least 55 years old. Additionally, the universe excludes accounts with foreign address, those with an immediate annuity purchase from Vanguard (reserved for future use), and those with very high assets (approximately over $5M).

Table 2-2. Sample Design

A. Sampling rates by age and client type

Age groups	Total	Client Type	
		Employer-sponsored	Individual client
55-59	20%	10%	10%
60-64	20%	10%	10%
65-69	20%	20%	
70-74	20%	20%	
75-	20%	20%	

B. Oversampling rates of administratively-single

Age groups	Client Type	
	Employer-sponsored	Individual client
55-59	1.0	2.0
60-64	1.0	2.0
65-69	1.5	
70-74	1.5	
75-	1.0	

Table 2-3. Marital Status: Administrative versus survey measures

		Survey measure				
		Married	Never married	Widowed	Divorced/ Separated	Partnered
Administrative	Non-single	85.72%	1.31%	9.14%	3.36%	0.47%
Measure	Single	44.64%	12.12%	11.72%	28.12%	3.40%

Note: Pilot Sample. Individual client sample only.

3. Invitations and Survey Mechanics

1.   Invitations and desired sample size.

Our goal in designing the VRI was to produce a minimum of 3,500 responses completing the

three panel surveys administered over the course of approximately one year. To hit this target,

we needed to estimate response rates for initial participation and attrition across surveys. In

designing the production sample, we had excellent information on response rates from the pilot

survey. The overall response rate was about 7.7 percent (ratio of completed surveys to non-

bounced invitations). This is a remarkably high response rate for such a Web survey.[1] To

decide on the overall number of invitations, we had to estimate projected attrition rates. We

assumed a one-third attrition rate between each survey.[2] We chose 130,000 accounts for email

invitations of which, at the time of invitation release, 121,201 had valid emails and did not

bounce.[3]   The actual response rate for Survey 1 production survey was 7.38%, with 8,950

completes. More than 4,600 respondents completed all the three surveys, which exceeds our

initial target.

2.   Invitation procedure.

Email accounts selected for the survey received a pre-alert email mailed by the Vanguard Group.

Roughly a week later, the selected individuals received a survey invitation, also emailed by

---

[1] Note also that this not a classic response rate, which is calculated conditional on the initial screen to create the frame. This response rate combines the success rate in survey to build the frame and the response to the survey given that a household is in the frame.
[2] The attrition rate from survey 1 to survey 2 was very close to what we estimated. The attrition rate from survey 2 to survey 3 (about a fifth) was quite lower than this estimate.
[3] An account is classified as to have bounced emails if not only the initial invitation but also all the following reminders are bounced (see below for detailed description for the reminder process).

Vanguard, containing a clickable survey link and instructions for completing the survey. See

Figure 3-1 for the text of these communications. Those who did not complete the survey within

a week received a reminder email, containing the same link and instructions. The Pilot had two

reminders, where the second one followed the first one after a week, if they had not completed

the survey by then. The production survey had one.

3.   Incentive.

The incentive for completing all three surveys is a $20 payment by Vanguard. After each

survey, there is also a sweepstakes for prizes such as iPads, and a grand prize sweepstakes for

those who complete all three surveys. See Vanguard Research Initiative, "Survey 1

Questionnaire," pages 57-59, for details of incentive.

4.   Administration of survey.

The survey was programmed and administered for Vanguard by IPSOS, SA an internationally

recognized survey research organization. Neither the University of Michigan nor New York

University was engaged in collecting these data.

Figure 3-1. Pre-alert and Invitations Emails.

Vanguard Research Initiative
December 2014

4. Response Analysis

1.   <u>Response rates</u>

Table 4-1A gives the response rate by age and client type for the production survey.   The response rates are calculated as ratio between the number of completes and non-bounced emailed invitations. The overall response rate 7.38% is very close to 7.7% from the pilot survey which we used to determine the overall size of invitations for production survey. The response rate does not have a monotonic pattern in age. The response rate for the employer sample is noticeably lower than for the individual client sample.

Table 4-1B shows the response rate by assets in the Vanguard administrative data. Interestingly, the response rate increases with wealth for both client types with the effect much more powerful for the individual account respondents. This higher response rate for the high-wealth individuals is consistent with Vanguard's experience in earlier survey efforts. It runs against the conventional wisdom that high-wealth individuals are hard to survey both because they are hard to reach and because they have complex finances. Evidently, engagement with Vanguard overcomes these issues.

Table 4-2 shows the distribution of administrative assets conditional on criteria used for being invited to participate in the survey and conditional on responding. To be Web-survey eligible, the account holder must be Web registered, have logged on in the past six months, not opted out of being contacted, and not have been contacted recently (see Vanguard Research Initiative: Documentation and Supporting Analysis, "2. Sample Design"). Survey eligible account holders have higher wealth than the broader Vanguard population. Given Web-survey eligibility, most accounts satisfy our asset cut-off ($10K), so the additional effect of the

>$10,000 condition on mean assets is minimal. We stratified the sample in various ways, i.e., by line of business, age, and single status (see "2. Sample Design," Table 2-2). The consequence of this stratification—shown in the "invited" row—reduces assets, mainly in the upper tail. Administratively-single account holders also tend to have lower Vanguard assets.

The last line shows the asset distribution conditional on completing the survey. Higher wealth account holders are more likely to respond, so the completed sample has higher wealth both than the universe and the invited account holders. To further explore selectivity in response, Table 4-3 reports the estimates from the linear probability model where the dependent variable is one for completing the survey and zero otherwise. The independent variables include various demographic information for the account holders from the administrative data. The result reconfirms all the findings from the previous tables. The wealth variable has the largest quantitative effect: an account holder from the lowest quintile is 4.4 percentage points less likely to complete the survey compared to one from the highest quintile. The oldest group is about 1.5 percentage points less likely to complete the survey, while among the remaining age groups the likelihood is flat. Employer-sponsored sample is 1.5 percentage points less likely to complete the survey compared to individual client sample.

2.   Effect of reminder.

To improve response rates, we sent a reminder two weeks after the initial invitation. Figure 4-1 shows the response rate over time.[1] The vertical line indicates the date of the reminder. Before the reminder date the increase in the response rate slowed noticeably. The reminder leads to a

---

[1] To avoid having Vanguard's help desk overwhelmed with questions, the initial invitations were sent to two groups. Group A were invited on August 15 and Group B on August 20 and 22. The figure is for Group A.

very large jump in the response rate of about 2 percentage points (from 5.3% to 7.3%). In terms

of completed surveys, the reminder yields about 2,400 more observations.

3.   Break-off analysis.

Some of the non-completion comes from invitees who start the survey, but break off at some

point. Table 4-4 shows the number of respondents who completed the survey up to a selection of

questions. After the initial large drop prior to the first question (age), that is respondents who

clicked the survey link but did not participate at all, there are significant drops are observed in

the income section (between Q8 and Q45) and also in the financial wealth section (between Q56

and Q65). These sections are indeed the most challenging ones on the survey. Especially for the

financial wealth section, respondents face a battery of account-level questions. The wealth

battery is the heart of the survey. It takes the most time,[2] it has the most complex structure, and

is perhaps the most intrusive set of questions.   Within the wealth modules, we strove for a high

level of item response by asking again for any response left blank and only leaving implicit that

questions could be skipped after a second item-non-response. This strategy yielded very low

item non-response. For example, among those who said they have IRA accounts, only 0.2% of

them did not give any information on the balance or answered that the balance is $1, the

minimum value allowed to type in. Hence, for the final 8,950 sample, the quality of the data is

extremely high.[3] But the survey's insistence on responses for each item might well account for

the high level of breakoffs in this section.

        Table 4-5 shows the proportions of groups of respondents at various points in the survey.

By sample design, the shares of the age groups shown in the first row of Panel A are roughly

---

[2] The median length of the survey is approximately 40 minutes. The median time spent on the
financial wealth section is about a third of it.
[3] In the HRS or SCF, imputations are necessary to get better picture of overall wealth since the
fraction of non-response or bracketed answers is not negligible.

equal.[4] Panel A shows that older groups are more likely to accept the invitation, but have higher

with-in survey attrition. The net effect is that the completion rate for the survey declines with

age. Panel B does the same analysis across wealth groups. Wealthier households are more likely

to respond to the invitation and also more likely to complete the survey when they have started it.

Given the conventional wisdom that wealthier households are hard to survey, this finding may be

surprising, though it is consistent with what Vanguard has found in previous surveys. The

higher-asset individuals are likely more engaged with Vanguard and are potentially more

interested in the topic of the survey. Panel C shows the break-offs by for the employer-

sponsored versus individual client sectors. There are no noticeable patterns by sector.

---

[4] The shares of each aged group invited are exactly equal to 0.2. The shares of invitations
received are slightly different owing to different rates of bounced emails.

Table 4-1. Response Rates

A. Age and Client Type

| Age | All | Client Type | |
		Employer-sponsored	Individual client
All	7.38%	6.64%	7.67%
55-59	6.93%	7.09%	6.76%
60-64	7.27%	6.49%	8.09%
65-69	7.85%	6.24%	8.34%
70-74	8.08%	6.30%	8.27%
75-	6.80%	5.37%	6.84%

B. Administrative Assets and Client Type

| Asset quintiles | Total | Client Type | |
		Employer-sponsored	Individual client
All	7.38%	6.64%	7.67%
1st (Lowest)	5.80%	5.96%	5.74%
2nd	6.24%	5.88%	6.36%
3rd	6.86%	6.11%	7.16%
4th	7.85%	6.93%	8.28%
5th (Highest)	10.23%	8.06%	11.14%

Note: Denominator of response rate is non-bounced invitations. 8950 observations (production sample).

Table 4-2. Administration Assets: Population to Sample

	Mean	Percentiles				
		10	25	50	75	90
Total Universe	272,264	6,892	24,570	86,611	261,995	635,017
Survey eligibility	309,715	10,366	35,672	117,501	319,447	688,349
Asset cut-off	287,873	19,972	42,727	114,455	307,339	700,283
Survey eligibility and Asset cut-off	310,967	23,311	53,470	142,518	352,837	727,007
Invited	277,172	22,868	52,150	138,327	331,887	655,287
Completed	355,329	27,151	68,174	188,191	445,497	826,482

Note: Age cut-off (higher than or equal to 55 and less than 100) is applied to all the rows

Table 4-3. Completed Survey (linear probability model)

Variable/Category	coefficient	standard error
Intercept	0.090	0.004
Asset quintile (1st, lowest)	-0.044	0.002
Asset quintile (2nd)	-0.040	0.002
Asset quintile (3rd)	-0.033	0.002
Asset quintile (4th)	-0.023	0.002
Sector - Employer-sponsored	-0.015	0.002
Age (55-59)	0.012	0.003
Age (60-64)	0.015	0.003
Age (65-69)	0.016	0.002
Age (70-74)	0.015	0.002
Admin non-single	0.002	0.002
Gender – F	0.000	0.003
Gender – M	0.006	0.003

Note: N = 121,201. Dependent variable is dummy variable for completion (1=completed). Omitted categories for independent variables are the highest asset quintile, individual client sector, the oldest, admin single and unknown gender.

Table 4-4. Progress through survey
(Respondents reaching various questions)

Question	Respondents
Start	19,477
Age (Q1)	14,091
Income (Q8)	13,829
Housing wealth (Q45)	11,495
Wealth (Q56)	11,438
Other assets (Q65)	9,163
Demographic (Q77)	9,090
End	8,950

Table 4-5. Break-off analysis

A. Age

	Age				
Question	55-59	60-64	65-69	70-74	75-
Invitation received	0.20	0.20	0.20	0.20	0.20
Start point	0.16	0.18	0.21	0.23	0.22
Age (Q1)	0.18	0.19	0.20	0.22	0.21
Income (Q8)	0.18	0.19	0.20	0.22	0.21
Housing wealth (Q45)	0.18	0.19	0.21	0.22	0.20
Wealth (Q56)	0.18	0.19	0.21	0.22	0.20
Other assets (Q65)	0.19	0.20	0.21	0.22	0.18
Demographic (Q77)	0.19	0.20	0.21	0.22	0.18
End	0.19	0.20	0.21	0.22	0.18

Note: Table show fraction of each age group (shares by row) at various points in the survey.

Table 4-5. Break-off Analysis (continued).

B. Administrative Assets

Question	Asset quintile				
	1st	2nd	3rd	4th	5th
Invitation received	0.20	0.20	0.20	0.20	0.19
Start point	0.17	0.18	0.20	0.22	0.24
Age (Q1)	0.16	0.17	0.19	0.22	0.25
Income (Q8)	0.16	0.17	0.19	0.22	0.25
Housing wealth (Q45)	0.16	0.17	0.19	0.22	0.26
Wealth (Q56)	0.16	0.17	0.19	0.22	0.26
Other assets (Q65)	0.16	0.17	0.19	0.22	0.27
Demographic (Q77)	0.16	0.17	0.19	0.22	0.27
End	0.15	0.17	0.19	0.22	0.27

Note: Table show fraction of each wealth group (shares by row) at various points in the survey.

C. Client type

Question	Employer-sponsored	Individual client
Invitation received	0.28	0.72
Start point	0.24	0.76
Age (Q1)	0.24	0.76
Income (Q8)	0.24	0.76
Housing wealth (Q45)	0.24	0.76
Wealth (Q56)	0.24	0.76
Other assets (Q65)	0.25	0.75
Demographic (Q77)	0.25	0.75
End	0.25	0.75

Note: Table show fraction of each client type (shares by row) at various points in the survey.

Figure 4-1. Response rate
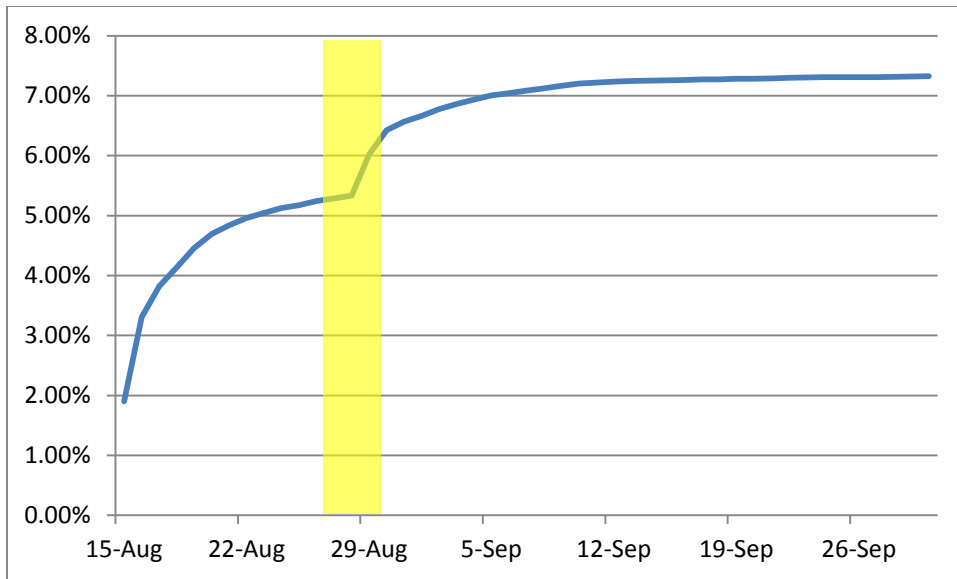
Note: Response rate of group A over time. The yellow vertical line indicates the date of the reminder.

5. Detailed Sample Characteristics

1.   Demographics.

Table 5-1 shows the distribution of education, health, gender, and marital status of the sample.

Overall, the sample is very well educated and very healthy. More than 70% of the sample has a

college degree. Post-graduate degrees are very common in the individual client sample. More

than 70% of total sample say that they are either in excellent or very good health. The sample

also has a high marriage rate despite our having oversampled singles. The sample is also

disproportionately white (Table 5-2). These distinctive features of VRI can be accounted for by

the fact that this sample is very affluent, since education level, health, being married, and being

white are positively correlated with wealth. In Section 4 of Ameriks et al. (2014), we take a

closer look at this issue by comparing VRI to the HRS and the SCF samples.

        Table 5-3 shows the marital status by age bins to check whether the oversampling by

administrative single status was successful in hitting our target of having one-third of the sample

as single (see Vanguard Research Initiative: Documentation and Supporting Analysis, "2.

Sample Design"). With the exception of the 70-74 age bin, our stratification procedure resulted

in hitting the target almost exactly.

2.   Income.

Before we describe the income data, we briefly explain how income is measured. In the

beginning of the income section, we ask "what was the approximate total combined income for

you [and your spouse] during 2012? You may respond by listing your household's "Adjusted

Gross Income" (AGI) as computed for your 2012 tax return, or provide an estimate of your

household's total income for 2012." The survey included a check box to indicate whether the

reply was AGI. About 58% checked this box. The medians and interquartile ranges are almost identical for those who checked the box versus those who did not. After this question we ask questions by income source – Social Security, pension, immediate annuity, wage, and disability benefits. For each source, respondents are asked to answer their own and their spouses' income. These incomes are asked in monthly terms, except for wages that are asked on annual basis. At the end of the income section, a summary table containing annualized income from these sources is presented and respondents are asked whether all the information is correct. If they say "no" to some items, they are brought back to the corresponding parts and allowed to change their answers. This correction mechanism is very similar to what we designed for the wealth section (see Section 2 of Ameriks et al. (2014)).

Table 5-4 reports the distribution of annual income. The first row tabulates the distribution of the total household income. Mean annual income is about $121,000 while median is about $82,000. Remaining rows show the distribution of annualized income from different sources.

3.   Financial Assets by line of business

Tables 5-5A and 5-5B report the distribution of the financial assets separately for the employer-sponsored and the individual client samples. The individual client sample has higher level of assets and especially they tend to have more wealth in IRA, mutual fund and brokerage accounts. By construction, almost all of the employer-sponsored respondents have employer-sponsored plans, but for other types of wealth, both the likelihood of having that type of account and the mean value conditional on having that type are lower compared to the individual client sample.

4.   Total Net Wealth by line of business

Table 5-6 tabulates the distribution of total net wealth by line of business. Individual client sample not only has more financial wealth but also has more real estate and business. As a result, total survey net wealth is higher for the individual client sample.

Table 5-1. Education, Health, Gender, and Marital Status.

| | | | Client Type | |
		Total	Employer-sponsored	Individual client
Education	College graduate	32.18%	33.69%	31.67%
	Post-college	38.45%	26.25%	42.53%
Health	Poor	0.84%	0.53%	0.94%
	Fair	4.77%	3.48%	5.20%
	Good	21.77%	22.33%	21.58%
	Very Good	41.84%	42.25%	41.71%
	Excellent	30.78%	31.42%	30.57%
Gender	Male	64.26%	65.29%	64.91%
	Female	35.74%	34.71%	36.09%
Marital status	Married	65.21%	72.19%	62.87%
	Never married	6.92%	5.70%	7.32%
	Widowed	8.93%	4.01%	10.57%
	Divorced	16.41%	15.42%	16.75%
	Separated	0.54%	0.98%	0.39%
	Having partner	2.00%	1.69%	2.10%

Table 5-2. Race and Ethnicity

| | | Client Type | |
	Total	Employer-sponsored	Individual client
White	93.65%	89.90%	94.90%
Black	1.62%	4.01%	0.82%
Asian	2.66%	3.07%	2.52%
N. Hawaiian / Pacific islander	0.13%	0.27%	0.09%
American Indian / Alaska Native	0.18%	0.31%	0.13%
Multi-racial	0.38%	0.36%	0.39%
Other	1.37%	2.05%	1.15%

Table 5-3. Marital Status by Age

| | Age | | | | |
	55-59	60-64	65-69	70-74	75-
Non-single	65.72%	66.33%	67.58%	69.74%	66.25%
Single	34.28%	33.67%	32.42%	30.26%	33.75%

Table 5-4. Annual Income

Income category	Mean	N	Conditional on having positive amount					
				Percentiles				
			Mean	10	25	50	75	90
Total income (Q8)	121,481	8,950	121,481	27,004	50,000	82,017	125,000	191,616
Social security	20,283	5,628	32,237	12,912	18,372	25,830	33,510	40,380
Pension	20,080	4,803	37,530	4,428	11,376	26,316	48,000	74,808
Immediate Annuity	945	523	16,269	1,200	3,576	7,800	15,708	31,200
Wage	68,174	5,136	119,184	6,000	28,000	72,000	125,000	192,000
Disability benefit	902	416	19,423	1,548	4,740	12,054	23,706	37,200

Note: Total income (Q8) is total household income. The other categories are surveys as monthly for respondent and spouse, and then annualized. Wage is asked on annual basis.

Table 5-5A. Survey Financial Assets: Employer-sponsored sample

Account type	Mean	N	Conditional on having positive amount					
			Mean	Percentiles				
				10	25	50	75	90
Total financial assets	840,592	2,244	840,592	82,251	203,400	507,653	1,023,500	1,822,200
IRA	185,385	1,220	340,986	10,000	30,000	100,000	310,000	692,282
Employer sponsored	349,943	2,113	371,638	35,000	96,568	230,000	480,000	825,000
Pension	61,754	562	246,578	19,000	36,000	100,000	300,000	678,367
Other retirement asset	11,412	136	188,304	10,000	23,711	75,000	168,500	450,000
Checking	13,664	2,135	14,362	740	1,800	4,500	12,000	32,000
Saving	21,556	1,771	27,314	500	2,000	8,100	26,000	70,000
Money market	15,702	580	60,750	1,327	5,000	20,000	65,968	146,420
Mutual fund	18,340	353	116,587	3,000	10,000	40,000	125,000	300,000
Certificate of deposit	7,798	302	57,944	3,000	8,167	24,000	62,000	145,000
Brokerage	86,147	749	258,097	5,000	20,000	89,000	275,600	677,000
Directly held securities	10,756	257	93,914	1,000	4,700	17,000	70,000	180,000
Annuity	19,468	284	153,824	10,457	27,500	83,450	179,442	335,000
Life insurance	31,183	714	98,004	5,000	10,350	32,000	100,000	250,000
Educational related	1,712	101	38,026	1,705	7,500	19,250	40,000	90,000
Other accounts	5,772	82	157,942	1,200	7,000	35,000	135,000	350,000

Note: Pension, annuity, and life-insurance are current cash values.

Table 5-5B. Survey Financial Assets: Individual client sample

Account type	Mean	N	Conditional on having positive amount					
			Mean	Percentiles				
				10	25	50	75	90
Total financial assets	1,306,099	6,704	1,306,099	145,000	330,702	709,750	1,354,209	2,384,500
IRA	417,337	6,083	460,079	38,000	103,000	265,743	596,000	1,081,000
Employer sponsored	170,672	2,517	454,719	19,000	70,000	216,700	471,000	850,000
Pension	13,188	454	194,792	5,941	25,000	96,678	200,000	500,000
Other retirement asset	13,848	466	199,281	8,500	28,000	86,000	240,000	452,554
Checking	17,967	6,502	18,530	1,050	2,500	6,014	16,300	41,500
Saving	23,510	4,391	35,905	500	2,300	10,300	35,000	88,000
Money market	32,526	3,496	62,392	1,200	5,500	23,000	70,000	153,000
Mutual fund	302,931	3,589	566,023	10,000	35,000	125,000	332,000	729,000
Certificate of deposit	19,514	1,332	98,242	4,341	12,000	37,240	101,000	247,250
Brokerage	213,904	3,435	417,595	7,000	30,000	115,000	362,600	882,100
Directly held securities	26,608	1,544	115,567	2,600	10,000	32,850	103,350	240,000
Annuity	21,260	879	162,193	13,400	36,000	96,000	202,000	367,000
Life insurance	17,663	1,982	59,763	4,870	10,000	25,000	64,061	131,900
Educational related	3,460	512	45,321	4,000	8,450	21,000	50,000	104,000
Other accounts	11,322	347	218,796	2,000	11,000	50,000	220,000	499,300

Note: Pension, annuity, and life-insurance are current cash values.

Table 5-6. Total Net Wealth

A. Employer-sponsored sample

| | | Mean | N | Conditional on having positive amount | | | | | |
| | | | | Mean | Percentiles | | | | |
					10	25	50	75	90
Wealth category	Total financial wealth	840,592	2,244	840,592	82,251	203,400	507,653	1,023,500	1,822,200
	Primary residence	292,995	2,044	321,664	100,000	160,000	250,000	400,000	600,000
	Mortgage in primary residence	94,930	1,383	154,030	30,000	62,500	121,500	203,000	305,000
	Other real asset	93,078	626	333,652	35,000	80,000	180,000	350,000	800,000
	Other Mortgages	20,320	246	185,360	25,000	50,000	112,000	231,000	400,000
	Other financial wealth	21,295	269	177,645	7,000	12,000	25,000	100,000	300,000
	Business	12,115	101	269,168	15,000	25,000	100,000	250,000	650,000
	Debt	14,538	1,211	26,940	3,178	9,000	18,000	30,000	52,000
	Survey net wealth	1,159,363	2,244	1,159,363	159,000	346,700	738,091	1,395,274	2,355,180

B. Individual client sample

| | | Mean | N | Conditional on having positive amount | | | | | |
| | | | | Mean | Percentiles | | | | |
					10	25	50	75	90
Wealth category	Total financial wealth	1,306,099	6,704	1,306,099	145,000	330,702	709,750	1,354,209	2,384,500
	Primary residence	324,764	6,138	354,817	120,000	175,000	275,000	420,000	675,000
	Mortgage in primary residence	49,178	2,405	137,127	20,000	50,000	104,000	180,000	294,000
	Other real asset	115,642	1,916	404,748	35,000	100,000	220,000	450,000	825,000
	Other Mortgages	13,433	492	183,087	28,000	70,000	125,000	230,000	378,000
	Other financial wealth	26,474	934	190,083	7,000	15,000	42,722	150,000	400,000
	Business	16,858	309	365,851	15,000	30,000	100,000	350,000	775,000
	Debt	5,377	1,687	21,375	2,000	5,000	12,000	22,000	40,000
	Survey net wealth	1,732,214	6,706	1,732,214	293,500	551,817	1,042,915	1,869,000	3,148,000

Vanguard Research Initiative
December 2014

6. Survey versus Administrative Data: Analysis by Account-type and Regression Analysis

1.   Comparison by asset types

In this document we compare survey and administrative data for subsets of wealth–IRA, employer sponsored plan and taxable assets.[1] All comparisons are made conditioning on survey-singles. Table 6-1 shows results for IRAs. Survey and administrative data line up very well. The median difference is very small, while the interquartile range does not go over ±5%.

One source of error is misreports of whether the asset is at Vanguard. Note that the comparison in Panel A of Table 6-1 is done conditioning on that the respondent has IRA either in survey or administrative data. In the following two panels, we show the cross-tabulations of indicators of having Vanguard IRA wealth in the survey and the administrative data. The correlation between the two indicators is very high – if they do (not) have IRA wealth in administrative data, with more than 90% likelihood, in the survey they answered they do (not) have IRA wealth.   But the correlation is not perfect, so some of the discrepancy arises from misclassification of assets at Vanguard or not.

Table 6-2 shows the same comparisons for employer-sponsored plan, while Table 6-3 is on taxable assets. All the patterns are fairly similar to what we found with IRA, though these items turn out to be somewhat noisier.

2.   Comparison conditioning on other variables.

---

[1] Since some of the administrative data were pulled after the survey, we do not have administrative data on respondents who died or otherwise left Vanguard between the administration of survey 1 and when this analysis commenced.   As a result, we do not have administrative data on 58 respondents who took Survey 1. Going forward, the VRI will have administrative data up to the point that a client leaves Vanguard for any reason.

The administrative data we used so far in the comparison is the snapshot of the account

holder's portfolio at the beginning of months, in addition to basic demographic information

available in the administrative records. The administrative data contains other information

useful for analyzing survey responses. We highlight several measures—summarized in Table 6-

4— that are useful for understanding corrections as well as having independent interest.

Vanguard tracks usage of its Web site. From these data, we can determine about 6% of

respondents logged in while taking the survey.

The data also contain the start date of their web registration (potentially earlier than

2003), which can be used as a proxy for their length of relationship with Vanguard. The median

length of relationship is 10 years.

The administrative data also allow us to track volatility of balances. The interquartile

range of mean monthly percentage change is from 2.6% to 5.1%.[2]

These variables might be related to discrepancies between the survey and administrative

data. Table 6-5 reports the conditional distributions of the difference between the wealth

measures of the survey and the administrative data (singles only). Logging into the account

during the survey reduces the interquartile range for the individual client sample, but the number

of respondents who logged in is too small for a meaningful comparison. The interquartile range

is smaller when they have been web-registered for more than 10 years. Finally, the difference

tends to be larger for the respondents whose historical balance has been more volatile. The

interquartile range is about 50% wider for the group whose historical volatility has been larger

---

[2] The volatility of historical data is calculated in the following way. First, the percentage
difference of total balances between any two consecutive months is calculated using a formula
very similar to the one explained in footnote 8 of the main paper. The only difference is that the
numerator is now calculated as absolute value difference, since we are interested in the
magnitude of changes in balance not the direction here. Then the historical average percentage
difference is calculated as the mean of these monthly percentage differences.

than the median. There are two possible explanations for this. On the one hand, respondents with volatile balances might find it harder to keep track of their financial records, and as a result their survey response error can be larger. On the other hand, when the volatility measured as monthly changes in the balance is high, it is also likely that their intra-month volatility is also high. If that is the case, the difference between the survey measure, which is measured at the moment of the survey, and the administrative measure, which is a snap shot at the end of months, can be accounted for by the timing issue.

3.   Regression analysis of deviation of survey and administrative data

Table 6-6 shows the results from the regression analyses where the dependent variable is the absolute percentage value of discrepancy between the survey and the administrative wealth measures and the explanatory variables capture all the variables we considered so far. The discrepancies are smaller if they are younger than 75, if they checked for the record for majority (more than 80%) of their accounts, if they logged in during the survey, if their history of balance at Vanguard has been less volatile, and if they had longer relationship with Vanguard.

Table 6-1. Survey versus Administrative Data: IRA

A. Balances (Singles)

	Mean	Percentiles				
		10	25	50	75	90
Survey	202,498	3,000	32,078	101,272	269,000	519,000
Administrative	214,047	16,422	40,989	114,408	277,162	519,706
Difference	-11,549	-51,640	-4,918	-116	1,866	17,809
% Difference	-14.60%	-171.66%	-4.41%	-0.13%	1.44%	18.03%

Note: Figures on this table condition on having positive holding in either survey or administrative data. 1,914 observations.

B. Have or Not (Singles, Employer-sponsored)

		Survey measure		N
		Have	Do not have	
Administrative	Have	NA	NA	0
Measure	Do not have	4.62%	95.38%	585

C. Have or Not (Singles, Individual client)

		Survey measure		N
		Have	Do not have	
Administrative	Have	90.70%	9.30%	1,860
Measure	Do not have	5.32%	94.68%	489

Table 6-2. Survey versus Administrative Data: Employer-sponsored plan

A. Balances (Singles)

| | Mean | Percentiles | | | | |
		10	25	50	75	90
Survey	214,663	12,000	35,000	111,000	275,000	518,581
Administrative	195,204	0	24,402	91,425	247,532	500,999
Difference	19,460	-41,710	-5,276	476	11,018	98,652
% Difference	15.17%	-63.77%	-3.70%	0.44%	12.19%	200.00%

Note: Figures on this table condition on having positive holding in either survey or administrative data. 760 observations.

B. Have or Not (Singles, Employer-sponsored)

| | | Survey measure | | N |
		have	do not have	
Administrative	Have	91.42%	8.58%	583
Measure	do not have	0%	100%	3

C. Have or Not (Singles, Individual client)

| | | Survey measure | | N |
		have	do not have	
Administrative	Have	88.52%	11.48%	61
Measure	do not have	5.07%	94.93%	2,288

Table 6-3. Survey versus Administrative Data: Taxable assets

A. Balances (Singles)

	Mean	Percentiles				
		10	25	50	75	90
Survey	189,852	1,900	20,000	74,000	229,770	500,000
Administrative	147,860	0	11,971	55,333	179,005	417,028
Difference	41,991	-27,334	-2,623	94	13,800	130,508
% Difference	16.31%	-96.17%	-3.70%	0.23%	25.74%	200.00%

Note: Figures on this table condition on having positive holding in either survey or administrative data. 1,629 observations.

B. Have or Not (Singles, Employer-sponsored)

		Survey measure		N
		Have	Do not have	
Administrative	Have	NA	NA	0
Measure	Do not have	2.90%	97.10%	586

C. Have or Not (Singles, Individual client)

		Survey measure		N
		Have	Do not have	
Administrative	Have	91.44%	8.56%	1,414
measure	Do not have	21.07%	78.93%	935

Table 6-4. Other Administrative Variables: Summary Statistics

Variable	Fraction	25 percentile	median	75 percentile
Logged in during survey	5.77%			
Length of relationship (years)		5	10	13
Mean volatility of balance history		2.59%	3.53%	5.08%

Table 6-5. Effects of logging in during survey, length of relationship and volatility of historical balances: Percentage deviation, Survey versus administrative wealth. (Singles only)

A. Logged in during survey

Sample	Logged in	N	Mean	Percentiles				
				10	25	50	75	90
Employer-sponsored	Yes	4	33.79%	0.27%	0.51%	2.53%	67.07%	129.84%
	No	581	1.83%	-22.06%	-3.07%	0.33%	6.21%	29.56%
Individual Client	Yes	177	4.91%	-9.75%	-0.64%	0.08%	1.52%	30.08%
	No	2,172	-1.10%	-23.93%	-3.26%	-0.04%	2.20%	23.49%

B. Length of relationship

Sample	Length of relationship	N	Mean	Percentiles				
				10	25	50	75	90
Employer-sponsored	< 5 years	188	0.73%	-37.64%	-5.60%	-0.10%	5.97%	66.33%
	5-10 years	230	2.96%	-22.80%	-3.51%	0.41%	7.49%	31.71%
	> 10 years	167	2.28%	-6.64%	-0.82%	0.79%	5.25%	18.75%
Individual Client	< 5 years	524	-0.86%	-27.40%	-3.50%	-0.08%	2.36%	32.06%
	5-10 years	724	-2.47%	-26.09%	-3.73%	-0.05%	2.10%	20.45%
	> 10 years	1,101	2.60%	-16.09%	-2.30%	0.01%	2.02%	24.38%

C. Volatility of historical balances

Sample	Volatility	N	Mean	Percentiles				
				10	25	50	75	90
Employer-sponsored	< 3.5%	250	1.61%	-14.61%	-2.00%	0.28%	4.70%	26.91%
	> 3.5%	335	2.38%	-27.26%	-3.69%	0.44%	7.49%	37.56%
Individual Client	< 3.5%	1,188	-2.76%	-20.39%	-2.50%	-0.09%	1.52%	14.62%
	> 3.5%	1,161	1.52%	-23.87%	-3.52%	0.05%	2.87%	31.36%

Table 6-6. Regression analysis of survey – administrative wealth measure discrepancy (Singles only)

Variable/Category	coefficient	Standard error
Intercept	0.215	0.045
Male	-0.005	0.017
Employer-sponsored sample	0.009	0.023
Age (55-64)	-0.067	0.024
Age (65-74)	-0.082	0.023
Logged in during survey	-0.052	0.035
Checked records (> 80% of accounts)	-0.032	0.017
Number of accounts	0.004	0.002
Length of relationship (<5yrs)	0.062	0.021
Length of relationship (5-10yrs)	0.016	0.020
Volatility of historical balance (>3.5%)	0.050	0.017

Note: N = 8,948. Dependent variable is absolute value of percentage difference defined in footnote 10. Omitted categories for independent variables are individual client sector, the oldest, married, female, checked for record for less than 80% of accounts, did not log in during the survey, and having relationship with Vanguard for more than 10 years.

7. Use of Correction Mechanism: Regression Analysis

To get insight into what type of respondents make corrections, we estimate a linear probability model for making a correction to balances. Table 7-1 reports the estimation result of the linear probability model where the dependent variable is a dummy variable which is 1 when the respondent corrected their answers after seeing the summary table containing balances, 0 otherwise. The explanatory variables include demographics, wealth group by quartiles, number of accounts, ratio of accounts referred to record. We also include variables indicating whether they logged into their account or not during the survey and the length of their relationship with Vanguard.

- Higher wealth respondents are much more likely to make a correction. When they belong to the lowest wealth quartile group, it is about 10.9% less likely that they will use this correction mechanism, compared to the highest quartile group.

- Those who referred to records for more than 80% of their accounts are 6.3% more likely to make a correction.

- Having one more account increases the chance of a correction by 0.7%.

- Logging onto the Vanguard account Web page during the survey increase the correction rate by 6.9%.

- Males and younger respondents are less likely to make corrections.

Of course, these figures conflate the need to make a correction and willingness to make one. For example, we cannot tell from them whether males had more accurate initial entries or were less willing to own mistakes. Similarly, having more accounts increases the opportunities for making entries that need corrections, but also perhaps signals a willingness to share information.

Table 7-1. Propensity to Make a Correction (Linear Probability Model)

Variable/Category	coefficient	standard error
Intercept	0.296	0.027
Wealth quartile (1st - poorest)	-0.109	0.014
Wealth quartile (2nd)	-0.061	0.013
Wealth quartile (3rd)	-0.043	0.013
Male	-0.036	0.010
Employer-sponsored sample	-0.032	0.011
Age (55-64)	-0.027	0.013
Age (65-74)	-0.021	0.012
Single	-0.013	0.010
Logged in during survey	0.069	0.019
Checked record (> 80% of acccounts)	0.063	0.009
Number of accounts	0.007	0.001
Length of relationship (<5yrs)	-0.006	0.012
Length of relationship (5-10yrs)	-0.013	0.010

Note: N = 8,948. Dependent variable is dummy variable for making corrections (1=making corrections). Omitted categories for independent variables are the highest wealth quintile, individual client sector, the oldest, married, female, checked for record for less than 80% of accounts, did not log in during the survey, and having relationship with Vanguard for more than 10 years.

8. Alternative Approaches to Measuring Stock Share:
Account-by-account versus Overall

In VRI, stock share is also measured account-by-account. For accounts other than saving/checking/MMMF, the respondents are shown the table with balance and asked to enter the share of stock held in each account. When the respondents did not answer at least one of the account-level stock share questions, the survey asks the overall stock share of their entire portfolio.

Table 8-1 shows the stock share distribution by type of responses. First, we can see that the vast majority of respondents answered all the account-by-account stock share questions. Second, the two distributions look almost identical. The numbers from the overall question look like rounded-up version of the numbers from the account-specific questions.

Table 8-1. Stock Share: Account-by-account versus overall responses

	N	Percentiles		
		25	median	75
Account specific (Q63)	8,696	35.18%	54.71%	74.58%
Overall (Q63a)	209	35.00%	60.00%	75.00%