

**Vanguard Research Initiative**  
December 2014

**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 holders 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.<sup>1</sup> 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.

---

<sup>1</sup> 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

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

Note: Pilot Sample. Individual client sample only.