Supporting Statement for ALP Survey:
Model Benefits Statement for Participants in Employee Benefit Plans

Part B
Collections of Information Employing Statistical Methods

Version: June 2012

Table of contents

B.   Collection of Information Employing Statistical Methods

B.1. Respondent universe and sampling methods

This study will collect information through a household survey and a choice experiment. We will pretest the survey through four focus groups. The household survey will use the American Life Panel (ALP), an internet panel of individuals aged 18 and over. The ALP respondent universe and sampling methods are described as below.

Individuals who are currently members of the ALP have been recruited since 2002. Initially, they were recruited for a project that started in 2003, which compared internet interviewing with telephone interviewing (CATI). The ALP as it operates in its current form started in the beginning of 2006. At that point in time the first *household information survey* was conducted, asking the panel respondents a wide range of demographic questions on a quarterly basis (as is still the case today). This first household information survey was modeled after the demographic questions asked at that time within the Current Population Survey (which is conducted by the U.S. Census Bureau for the U.S. Bureau of Labor Statistics). This close alignment allows for extrapolation of analysis results of data collected within the American Life Panel to the U.S. population at large.

The focus groups are not meant to be statistically significant but rather will provide researchers qualitative information about how well the concepts and sample benefits statements in relation to employee sponsored benefit plans are understood. It allows us to probe why concepts may be misunderstood and whether or not the terminology used as well as the structure of the questions are appropriate for the audience.

B.1.a Household survey sampling approach

Since its start in 2003, the American Life Panel has expanded significantly and currently comprises approximately 4500 active individuals from U.S. households who have filled out the household information survey at least once during the past year (see section B.3.a. below for details). Participants in the ALP are recruited from survey programs that collect representative samples of U.S. consumers. Several cohorts can be distinguished based on their source and/or type of recruitment:

- University of Michigan internet panel cohort
- University of Michigan phone panel cohort
- Snowball cohort
- Stanford panel cohort
- Mailing cohort
- Phone cohort
- Vulnerable population cohort (active recruitment)
- Respondent driven sampling (RDS) cohort (active recruitment)

The cohort to which a respondent belongs is indicated by the **recruitment_type** variable, which is provided for each respondent. The specific category labels are (listed in order of appearance as above):

- 0 MS Internet
- 1 MS CATI
- 2 Snowballs

- 3 National Survey Project
- 6 Mailing Experiment
- 7 Phone Experiment
- 8 Vulnerable Population
- 9 RDS

In addition to the directly sampled panel members, the ALP also invites (adult) household members of the sampled panel members to join, thus allowing intra-household comparisons. These panel members are identifiable in the data as their identifier will end in a numeric value greater than 1 (e.g. identifier 10017494:2) in the data. However, this part of the ALP is currently relatively small, with less than 10% of the households having more than one panel member. For this reason the ALP cannot be used as a proper household survey panel and should be considered primarily a panel of individuals. The current study will only use original members from the large probability sample cohorts (MS Internet and MS CATI, National Survey Project, and Vulnerable Population) and not use the additional household members.

University of Michigan Internet Panel Cohort

The first cohort of participants in the ALP, the Michigan cohort, are respondents who were recruited from among individuals ages 18 years and older who had responded to the Monthly Survey (MS) of the University of Michigan's Survey Research Center (SRC). The MS is the leading consumer sentiments survey that incorporates the long-standing Survey of Consumer Attitudes (SCA) and produces, among others, the widely used Index of Consumer Expectations.

The sampling design of the MS is described in detail by Curtin (2002). The MS is conducted monthly. It uses a list of "hundred series" (i.e., the first eight digits of a phone number) that contains at least one listed residential landline phone number. This list is not older than six months. Phone numbers are then generated by randomly adding the last two digits, resulting in a random digit dialing (RDD) sample. Each month, approximately 300 households thus obtained are interviewed. Additionally, each month 200 individuals are re-interviewed from the RDD sample surveyed six months previously. The RDD procedure uses stratification by Census Division by MSA/non-MSA status and results in a stratified one-stage equal probability sample of telephone numbers in the 48 coterminous states and the District of Columbia. Within each household, one adult (18+) member is then selected using probability methods.

SRC screened MS respondents. At the end of the second interview, respondents were told that the University of Michigan was undertaking a joint project with RAND. They were asked if they would object to SRC sharing their information about them with RAND so that they could be contacted later and asked if they would be willing to actually participate in an Internet survey. Respondents who did not have Internet were told that RAND would provide them with free Internet. This attempt included the mention of the fact that participation in follow-up research carries a reward of $20 for each half-hour interview.

RAND received referrals from the University of Michigan each month from January 2002 through August 2008, with the exception of January through July 2004. RAND then contacted these individuals, asked them whether they would be willing to be part of the ALP, and entered them into the panel when they agreed. About 51 percent of the Michigan referrals agreed to be considered for the ALP, and about 58 percent of them actually participated in at least the household information survey. Thus, about 30

percent (51 percent × 58 percent) of the Michigan recruits became ALP participants. Originally, the ALP included only respondents 40 years of age and older, but since November 2006, the ALP has included respondents 18 years of age and older.

University of Michigan Phone Panel Cohort
The second cohort, the University of Michigan Phone Sample, comprises respondents who were recruited in the same MS surveys as the MS internet panel, but were first assigned to a phone panel, which was part of a study comparing Internet with CATI. They were invited to join the panel after that study had been completed. Thus, the sampling and recruitment of this cohort is the same as for the MS Internet cohort, but after recruitment they split for the duration of the initial project and then were recombined.

Stanford Panel Cohort
After August 2008, the ALP no longer received new respondents from the University of Michigan. Instead, in the fall of 2009 participants in the Face-to-Face Recruited Internet Survey Platform (FFRISP) were invited to join the ALP at the conclusion of the FFRISP panel. The FFRISP was an NSF-funded panel of Stanford University and Abt SRBI. Respondents were sampled from June to October 2008 in a multi-stage procedure based on address lists. More details of sampling are in Sakshaugh et al. (2009). The target population consisted of individuals 18 years or older who resided in a household in the 48 contiguous states or the District of Columbia and who were reportedly comfortable speaking and reading English. The sample was representative of this population.

Respondents were recruited in a face-to-face interview. They were offered a laptop (worth $500) and a broadband internet subscription, or $200 upfront and $25 per month (for 12 months) if they already had a computer and internet access. Additionally, they were paid $5 per monthly survey. The FFRISP recruited 1,000 respondents from a gross sample of 2,554 addresses that were not known to be ineligible. The panel was terminated after September 2009, but participants were offered to join the ALP under the same conditions (laptop, high speed internet, monetary compensation).

As noted, the recruitment strategy of FFRISP included a generous incentive, offering a free laptop to all households worth $500, high speed internet to households who did not have it (either by wireless card or DSL), and a cash incentive for participation in monthly surveys. Sometimes the generosity of the incentive was a source of refusal. Often, though, initial contact resulted in an unspecified refusal, without the interviewers having a chance to explain the study. Some examples of groups of people who refused include the elderly (7% of refusals), technophobes (14% of refusals), and skeptics (4% of refusals).

Respondents who did agree to join the American Life Panel form the Stanford cohort, or National Survey Project cohort. RAND renewed their laptop warranties and continues to pay for internet subscriptions as long as cohort members remain active in the ALP.

Snowball Cohort
A subset of respondents has been recruited through a so-called snowball sample, making up the Snowball cohort. ALP respondents were given the opportunity to suggest family members, friends, or acquaintances who might also want to participate. RAND then contacted those referrals and invited them to participate. Because this "snowball" sample is not randomly selected or representative of U.S.

residents, it is used mainly for testing surveys during piloting. Since May 2009 no new snowball respondents have been allowed to join the ALP. It is not possible to calculate a response rate for the Snowball Cohort, since this group does not represent a proper sample. **The Snowball Cohort is mainly used for pretests and small experiments and will not be part of the sample for the current study**.

Mailing and Phone Cohorts

As part of an experiment we recruited respondents from a random mail and telephone sample using the Dillman method. This experiment was initiated in 2010. **These are small cohorts (30 and 13 observations, respectively) and will not be part of the sample for the current study.**

Vulnerable Population Cohort

We are currently expanding the American Life Panel with panel members drawn from vulnerable groups and minorities. This addition will include a subsample for whom the interview language will be Spanish. We expect the expansion to be complete by the end of March 2012 ,The sample from which the panel members are recruited is address based; zip codes were chosen with a high percentage of Hispanics or with a high percentage of households with relatively low incomes. Potential panel members were first sent announcement letters. This was followed up with a paper survey in the mail (including a prepaid incentive). In this mail survey, households could express interest in becoming part of the ALP, or refuse to enter the ALP. We then attempted to contact households that did not respond to the paper survey within three weeks by phone. Again, during the phone interview, respondents were able to express interest in joining the ALP or refusing to join.

Respondent Driven Sampling (RDS) Cohort

Respondent driven sampling (RDS) is a technique to sample populations through social networks (Heckathorn, 1997, 2002, 2007). Each respondent recruits a fixed number of friends in the target population who in turn become the next generation of respondents. A crucial difference with snowball sampling is that the size of a respondent's social network must be known and that the individuals in the social network who are asked to join are a random selection from the network. Once sample equilibrium has been reached, sample proportions for a given variable of interest no longer change. Sample proportions in equilibrium are biased, however, because respondents with a greater social network are overrepresented. One can correct for this bias to derive unbiased population estimates (Heckathorn 2002). **This is a small experimental cohort and will be excluded from the sample for the current project.**

B.1.b Experiment sampling methods

The experimental population is the same as the survey population above.

B.2. Procedures for the Collection of Information

B.2.a. Statistical methodology for stratification and sample selection

B.2.a.1. Household survey sample selection

For this study, we will administer the survey to the respondents in the American Life Panel who belong to the following cohorts: University of Michigan internet panel cohort; University of Michigan phone panel

cohort; Stanford Panel cohort; Vulnerable Population cohort. We will only invite the original panel members and not the additional household members.

B.2.a.2. Experiment stratification and sample selection

For this study, we will treat all original panel members from the University of Michigan internet and phone panel cohorts, the Stanford Panel cohort, and the Vulnerable Population cohort in the American Life Panel as experimental subjects, which is the same selection as in the survey component in section B.2.a.1.

B.2.b. Data collection procedures

Respondents in the panel either use their own computer to log on to the Internet or they were provided a laptop with built in wireless, or a Web TV (http://www.webtv.com/pc/) that allows them to access the Internet, using their television and a telephone line.

About twice a month respondents receive an email with a request to visit the ALP URL and fill out questionnaires on the Internet. Typically an interview will not take more than 30 minutes. Respondents are paid an incentive of about $20 per thirty minutes of interviewing (and proportionately less if an interview is shorter). Most respondents respond within one week and the vast majority within three weeks. To further increase response rates, reminders are sent each week. For any given project, survey authors can receive data in real time during the field period so that one can start preliminary analysis before the field period has ended.

The incentives are in line with other main social science surveys, like the Panel Study of Income Dynamics ($60 per 77 minute interview) and the Health and Retirement Study ($50 per 121 minute interview). Providing incentives in panels is best practice and there is ample evidence that this helps in limiting attrition (see e.g., Göritz, 2006; Millar and Dillman, 2011). Moreover, the evidence suggests incentives have a positive effect on representativeness and data quality (see, e.g., Mack et al., 1998). Singer and Kulka (2002) present evidence suggesting the use of incentives in panel studies can be quite effective in reducing subsequent attrition. Limiting attrition both improves representativeness (e.g. Michaud et al., 2011) and reduces the added cost of sample recruitment. There is evidence that Internet response rates are increased by using incentives.

Focus group participants will be recruited through focus group vendors that maintain databases of names of individuals that have previously identified themselves as being willing to participate in future focus groups. Focus group participants will receive a $100 incentive fee to participate in the hour and a half discussion, which is a standard incentive fee for a focus group of this length in a metropolitan area.

B.3. Methods to Maximize Response Rates and Deal with Nonresponse

As noted above, participants are paid for their time. Once in the ALP, participants tend to remain indefinitely, which is reflected in relatively low attrition rates. It should be noted that panel members do not always give formal notification about their intent to leave the panel. Rather, they simply stop participating in surveys over a prolonged period of time. To avoid that such members remain in the panel indefinitely, at the beginning of each year RAND attempts to contact these members to ask them whether they are still interested in participating, and if contact attempts fail, removes them from the ALP. For

example, respondents who were not active in 2007, were removed from the panel in the beginning of 2008.

- Size of selected sample: Samples are drawn based on the selection criteria applicable to the survey, i.e. in relationship to the goal of the research. Only ALP panel members considered **active** at the time of the survey are selected. A member is considered to be active if s/he participated in the household information survey within one year prior to the fielding date of the survey. For example, if the survey was fielded on April 15, 2009, then a member is considered to be active if s/he responded to the household information survey in the period April 15, 2008 - April 14, 2009. If s/he did not respond to the household information survey in that period, s/he is considered to be **inactive** at the time of the survey and is not part of the selected sample.

Most respondents tend to respond within one week of the date the survey went into the field. Almost all respondents do so within three weeks. To further maximize response, the survey is kept in the field longer and/or additional reminders are sent.

B.3.a. Panel Attrition and Response Rates

To give an indication of typical response rates in the ALP, we computed the average response rate for all surveys conducted in 2011. For these computations, we follow AAPOR (2011) and Callegaro and DiSogra (2008). We compute cumulative response rates as the product of the following three factors:
- Recruitment rate (RECR), which is the fraction of individuals selected for inclusion in the panel who were recruited into the panel.
- Retention rate (RETR), which is the fraction of recruited individuals who were still panel members at any time during 2011.
- Completion rate (COMR), which is the fraction of invitations for participating in a survey in 2011 that resulted in a complete or partial interview.

Below, we provide more detailed definitions. AAPOR and Callegaro and DiSogra also define a *profile rate* (PROR), which is the fraction of recruits who complete an initial profile survey, consisting typically of basic demographics (the household information survey mentioned above). We have not computed separate rates for this, so this is included in the retention rate in our computations.

For this study, we will only select respondents from the large cohorts that were obtained from well-defined probability samples: the University of Michigan's Survey of Consumer Attitudes (SCA), Stanford University's National Survey Project (FFRISP), and Vulnerable Populations. We will only use original panel members and not added household members. The latter are a fairly small group and omitting them avoids complications in the computations of response rates, standard errors, and design effects. Furthermore, because the MS Internet and MS CATI subsamples were drawn from the same source, we combine them for the purpose of computing response rates.

Table B.1 summarizes the results. It shows that overall we have a recruitment rate of 22%, a retention rate of 72%, and a completion rate of 79%, resulting in a cumulative response rate of 12.5%. This is twice as high as the example given in Callegaro & DiSogra (2008). Note that the retention rate differs greatly between cohorts, which is due to the variation in the time when the samples were initially recruited. For the Michigan samples, the earliest members were recruited in 2002, whereas the Vulnerable Populations have been recruited very recently, and in fact recruitment for these is still ongoing.

Table B.1: Summary of response rates in 2011 (%).

	recruitment rate (RECR)	retention rate (RETR)	completion rate (COMR)	cumulative response rate (CUMRR2)
MS Internet + MS CATI	18.4	45.8	78.3	6.6
National Survey Project	23.9	98.9	80.8	19.1
Vulnerable Population	28.9	100.0	78.6	22.7
Total	22.1	72.1	78.6	12.5

Note. As explained in the text, some of these numbers are based on assumptions, and not only straightforward computations.

It is important to put this response rate in perspective. Some studies have been carried out to compare the quality of results based on probability internet panels with similar response rates as the ALP to population representative telephone surveys and convenience Internet surveys (which typically have much larger numbers of respondents, but unknown response probabilities). Chang and Krosnick (2009) simultaneously administered the same questionnaire (on politics) to an RDD (random digit dialing) telephone sample, an Internet probability sample, and a non-probability sample of volunteers who do Internet surveys for money. They found that the telephone sample had most random measurement error, while the non-probability sample had the least. At the same time, the latter sample exhibited most bias (also after reweighting), so that it produced the most accurate self-reports from the most biased sample. The probability Internet sample exhibited more random measurement error than the non-probability sample (but less than the telephone sample) and less bias than the non-probability Internet sample. On balance, the probability Internet sample produced the most accurate results. Yeager et al. (2009) conducted a follow-up study comparing one probability Internet sample, one RDD telephone sample, and seven non-probability Internet samples and a wider array of outcomes. Their conclusions are the same: Both the telephone sample and the probability Internet sample showed the least bias; reweighting the non-probability samples did not help (for some outcomes, the bias gets worse; for others, better). They also found that response rates do not appear critical for bias. Even with relatively low response rates, the probability samples yielded unbiased estimates. It is not clear a priori why non-probability samples do so much worse. As they note, it appears that there are some fundamental differences between Internet users and non-Internet users that cannot be redressed by reweighting. Indeed, Couper et al. (2007) and Schonlau et al. (2009) show weighting and matching do not eliminate differences between estimates based on samples of respondents with and without Internet access. Several other studies point at equally mixed results, including Vehovar et al. (1999); Duffy et al., (2005); Malhotra and Krosnick (2007), Taylor (2000), Loosveldt and Sonck (2008). Of course a distinguishing characteristic of the ALP is that it provides Internet access to respondents who did not have it before joining the study.

The computation of the recruitment rates in the ALP varies greatly between the various cohorts, because of their different backgrounds. Therefore, we discuss each of these separately, and then discuss retention rates and completion rates.

University of Michigan (MS Internet and MS CATI)

Recruitment of panel members through the University of Michigan's Survey of Consumer Attitudes (SCA) proceeded through several stages. Stage 1 is the initial (first round) SCA interview. Stage 2 is the

6-month follow-up interview in the SCA. In this follow-up interview, age-eligible respondents were asked permission to refer them to RAND, so stage 3 is referral to RAND. Referred individuals were then entered into RAND's database and given an individual identifier. We define recruitment as the result of these three stages. Once recruited, individuals could ask to be removed from the panel at any time (including immediately after recruitment), or they could be removed from the panel by RAND due to inactivity for more than a (calendar) year. If not removed before 2011, they are considered to be part of the panel in 2011, which defines the retention rate.

We have been unable to obtain detailed information from the University of Michigan about response rates, age distributions, and permission rates, so for this document we have used a combination of the available information and assumptions (appropriately noted) to compute recruitment rates for the Michigan subsample. We base our computations on the AAPOR RR2 definition:

$$RR2 = \frac{(I+P)}{(I+P)+(R+NC+O)+(UH+UO)}$$

where I is the number of complete interviews, P is the number of partial interviews, R is the number of refusals, NC is the number of non-contacts, O is "other", and UH and UO denote cases in which it is unknown whether the selected unit (telephone number for the SCA) is an eligible unit, that is, whether it is a residential landline telephone number.

Let N_1 be the number of telephone numbers selected for a first-round interview in a given month. The SCA contains 300 first-round interviews per month, and thus the AAPOR RR2 first-round response rate $p_1 = 300/N_1$. The SCA contains 200 second-round interviews per month, and thus the AAPOR RR2 cumulative response rate for the second-round interviews is $p_1 p_2$, where $p_2 = 200/300 = 66.7\%$ is the conditional second-round response rate, given a first-round response. The combined response rate of the SCA is then the number of interviews in a month (500) divided by the number of originally selected phone numbers for this sample (first-round plus second round; 2 N_1). Thus, 2 N_1 = 500/RR2. From Dixon and Tucker (2010, Figure 19.1), we obtain the SCA RR2 = 57%, 54%, 52% in 2002, 2004, and 2006, respectively. We interpolate and extrapolate this to obtain RR2 = 55.5%, 53%, 51%, 50% in 2003, 2005, 2007, and 2008, respectively, from which we obtain N_1 for each month the ALP received referrals from the University of Michigan.

Let p_3 be the fraction of second-round respondents who are age-eligible for the ALP. Since November 2006, the ALP includes respondents 18 and over, which is the same age restriction as the SCA, so $p_3 = 100\%$. Before November 2006, the ALP only recruited individuals 40 and over. We obtain the number of individuals 40 and over as a fraction of the number of individuals 18 and over from population estimates from the US Census Bureau (US Census Bureau, 2011), and use this as our estimate of p_3 before November 2006. Let p_4 be the referral rate, that is, the fraction of individuals asked for permission who were referred to RAND. Then the number of referrals received by RAND (number of recruits) in a given month is $N_4 = N_1 p_1 p_2 p_3 p_4$. The number of recruits N_4 is observed in our records and with the estimated $N_1, p_1, p_2,$ and p_3 as above, we can solve for p_4. Because p_3 refers to a selection based on eligibility and not nonresponse, the recruitment rate is RECR = $p_1 p_2 p_4$. We assume that the various probabilities are constant within a year and the same for individuals aged 18-39 and age 40 and over We compute these numbers per year, rather than on a monthly basis, because the monthly numbers are more subject to random variation. In 2006, the age 40 restriction was imposed for 10 months, and lifted in November, and thus we compute p_3 as (10/12) * F + (2/12) * 1, where F is the fraction of individuals age 40 and older among the number of individuals 18 and over.

9

Summarizing, and aggregating to the annual level, we compute the various components as follows:

$$RR2 = \begin{cases} 57\% \ (2002), 54\% \ (2004), 52\% \ (2006)(\text{Dixon \& Tucker, 2010}) \\ 55.5\% \ (2003), 53\% \ (2005)(\text{interpolated}) \\ 51\% \ (2007), 50\% \ (2008)(\text{extrapolated}) \end{cases}$$

$$p_3 = \begin{cases} F \ (2002\text{-}2005)(\text{derived from Census Bureau data}) \\ (10/12)F + (2/12)(2006)(F \ \text{derived from Census Bureau data}) \\ 1 \ (2007\text{-}2008) \end{cases}$$

$N_4 = $ number of recruits according to our records

$M = $ number of months we received UMich recruits

$N_2 = $ number of 1st round interviews in SCA $= 300 M$

$N_3 = $ number of 2nd round interviews in SCA $= 200 M$

$N_1 = (N_2 + N_3)/(2 \ RR2)$

$p_1 = N_2 / N_1$

$p_2 = N_3 / N_2 = 2/3$

$p_4 = N_4 /(N_1 p_1 p_2 p_3)$

$RECR = p_1 p_2 p_4$

The results are given in Table B.2. An overall recruitment rate can be computed by weighting the recruitment rates for the separate years by the corresponding numbers of selected households. The number of selected households, accounting for age-ineligibility in earlier years, is $W = N_4 / RECR$. This is naturally equal to N_1 for the years in which all SCA respondents were eligible for the ALP (2007 and 2008), but it is less by the eligibility-adjustment factor p_3 in 2002-2006. Attaching a t subscript for the year, the combined recruitment rate is

$$RECR = \frac{\sum_{t=2002}^{2008} W_t ´ RECR_t}{\sum_{t=2002}^{2008} W_t} .$$

The resulting combined recruitment rate for the MS cohorts is 18.4%.

Table B.2: Computation of recruitment rates for the Michigan sample.

year	SCA RR2 (%)	F	N4	No. months 40+ elig.	No. months 18+ elig	SCA no. interviews 1st rnd	2nd rnd	N1	p1 (%)	p2 (%)	p3 (%)	p4 (%)	RECR (%)
2002	57.0	58.3	484	12	0	3600	2400	5263	68.4	66.7	58.3	34.6	15.8
2003	55.5	58.8	603	12	0	3600	2400	5405	66.6	66.7	58.8	42.7	19.0
2004	54.0	59.3	275	5	0	1500	1000	2315	64.8	66.7	59.3	46.4	20.0
2005	53.0	59.7	642	12	0	3600	2400	5660	63.6	66.7	59.7	44.8	19.0
2006	52.0	60.0	669	10	2	3600	2400	5769	62.4	66.7	66.6	41.8	17.4

| 2007 | 51.0 | 60.2 | 1120 | 0 | 12 | 3600 | 2400 | 5882 | 61.2 | 66.7 | 100.0 | 46.7 | 19.0 |
| 2008 | 50.0 | 60.3 | 762 | 0 | 8 | 2400 | 1600 | 4000 | 60.0 | 66.7 | 100.0 | 47.6 | 19.1 |

Note. See text for definition of N1, p1-p4, N4, and F. RECR = Recruitment rate.

National Survey Project

The National Survey Project cohort consists of individuals who were previously part of the Face-to-Face Recruited Internet Survey Platform (FFRISP) at Stanford University. The FFRISP had an AAPOR RR4 recruitment rate of 43% (Yeager et al., 2011). RR4 is similar to the RR2 formula given above, except that the "unknown" term (UH + UO) in the denominator is multiplied by a factor e that is an estimate of the fraction eligible among the unknown. The resulting panel size of the FFRISP was 1,000. RAND received 556 of these as recruits (488 in 2009, 67 in 2010, and 1 in 2011), so our recruitment rate for this cohort is RECR = 43% * (556/1,000) = 23.9%.

Vulnerable Populations

Unlike the MS and National Survey Project cohorts, the Vulnerable Populations cohort is not inherited from a previous survey or panel but specifically sampled for the ALP. The number of addresses selected was 15,000, divided into three batches of 5,000 addresses each. Table B.3 shows the results of the initial brief survey that was sent to all 15,000 individuals in the sample.

Table B.3: ALP Vulnerable Populations supplement survey: recruiting survey results

Sample Batch	Mail Survey	Undeliverable	Refusal by Mail	Phone follow-up sample	Refusal by Phone	Completed by Mail	Completed by Phone	Total # of completes
Batch 1	5000	497 (9%)	217 (4%)	1270	245 (19%)	1310 (26%)	229 (18%)	1539 (31%)
Batch 2	5000	763 (15%)	205 (4%)	1337	326 (24%)	1194 (24%)	172 (13%)	1366 (27%)
Batch 3	5000	790 (16%)	183 (4%)	1675	174 (10%)	1226 (25%)	200 (12%)	1426 (29%)

The numbers in parentheses in the final column are the response rates obtained as of December 15, 2011, according to the AAPOR RR1 definition. The overall response rate across the three batches is 4,331/15,000=28.9%. Analogous to our definition of recruitment rate for the other cohorts, we define this as our recruitment rate (RECR). Among respondents, 86% respond positively to the question whether they are interested in joining the panel. We are still in the process of signing up the respondents who expressed a willingness to participate in the panel; currently we have signed up 11% of the original sample, but we expect this to increase significantly. If indeed ultimately 86% join the panel, this would mean a retention rate of 86% for 2012 according to our definition of recruitment and retention. These response rates compare favorably with similar recruitment strategies elsewhere. For example, DiSogra (2010) reports a "successful" recruitment for Knowledge Networks' KnowledgePanel with an initial response rate of 14% and 75% of those becoming panel members.

Attrition and retention rates

The ALP records indicate whether a previously recruited individual has been removed from the panel ("dropped") and an indication of the reason why the individual was dropped. Additionally, individuals can be classified as (temporary) nonparticipants, for example, if they are traveling or moving overseas for a prolonged but (expected) finite period of time. For the current computations, we define an individual as being in the panel in a given year if the individual was recruited during or before this year, was not

dropped before the start of this year, and was not a nonparticipant during this whole year. Only five individuals ever returned to the panel after being temporary nonparticipants, and these all did so within the same year or the next year after becoming nonparticipants. Thus, there are no respondents for whom we observe discontinuous panel membership as defined on an annual basis. Therefore, we can unambiguously define attrition between two consecutive panel years as being in the panel in one year and not being in the panel in the next year. This treats nonparticipants who did not return in the same or the next year as attritors. The (year to year) attrition rate is the number of individuals thus classified as attritors as a percentage of the panel size in the former year, and the cumulative attrition rate after year *t* is the fraction of individuals who were recruited in year *t* or earlier who were not in the panel after year *t*. The (cumulative) retention rate RETR is 100% minus the cumulative attrition rate.

Table B.4 shows the number of panel members in the Michigan cohorts, by panel year and recruitment year. From this table, it follows that prior to the start of the ALP in its current form, no individuals had been dropped from the panel. There were two waves in which relatively large numbers of inactive members were removed from the panel: in 2007 for members who were recruited in 2002-2006 and in 2010 for members who were recruited in 2006-2008. Cumulative attrition rates after 2011 vary from 40% to 74%. We can compute the combined cumulative retention rate analogously to the combined recruitment rate as a weighted average of the recruitment year-specific retention rates:

$$\text{RETR} = \frac{\sum_{t=2002}^{2008} W_t \, \text{RETR}_t}{\sum_{t=2002}^{2008} W_t},$$

where for the computation of retention rates the weights are equal to the number of recruits. The result is equal to the panel size in the year following the panel year divided by the number of individuals recruited up to the panel year. Thus, the retention rate after 2010 was RETR = 2087/4555 = 45.8% and the cumulative attrition rate was 100% - RETR = 54.2%.

Table B.4: Panel size and attrition in the Michigan cohort.

| Recruitment year | No. recruits | Panel year | | | | | | |
		2006	2007	2008	2009	2010	2011	2012
		Panel size						
2002	484	484	393	229	223	214	209	206
2003	603	603	529	223	221	212	206	203
2004	275	275	259	74	73	73	72	72
2005	642	642	630	180	178	177	175	174
2006	669	669	662	506	490	452	306	304
2007	1120	0	1120	1095	1070	1029	665	657
2008	762	0	0	762	733	712	454	452
Total	4555	2673	3593	3069	2988	2869	2087	2068
		Attrition rate (after the panel year, %)						
2002		18.8	41.7	2.6	4.0	2.3	1.4	
2003		12.3	57.8	0.9	4.1	2.8	1.5	

2004	5.8	71.4	1.4	0.0	1.4	0.0
2005	1.9	71.4	1.1	0.6	1.1	0.6
2006	1.0	23.6	3.2	7.8	32.3	0.7
2007		2.2	2.3	3.8	35.4	1.2
2008			3.8	2.9	36.2	0.4
Cumulative attrition rate (after the panel year, %)						
2002	18.8	52.7	53.9	55.8	56.8	57.4
2003	12.3	63.0	63.3	64.8	65.8	66.3
2004	5.8	73.1	73.5	73.5	73.8	73.8
2005	1.9	72.0	72.3	72.4	72.7	72.9
2006	1.0	24.4	26.8	32.4	54.3	54.6
2007		2.2	4.5	8.1	40.6	41.3
2008			3.8	6.6	40.4	40.7
Total	7.5	39.2	34.4	37.0	54.2	54.6

The National Survey Project cohort was recruited in 2009-2011. There was no attrition between 2009 and 2010. After 2010, 3 of the 488 individuals recruited in 2009 and 3 of the 67 individuals recruited in 2010 were removed from the panel, resulting in attrition rates of 0.6% and 4.5%, respectively, and a combined attrition rate of 1.1% and a RETR of 98.9%. Another 15 individuals from the 2009 recruits and none from 2010 and 2011 left the panel after 2011, resulting in attrition rates of 3.1%, 0%, and 0%, respectively. The cumulative attrition rates after 2011 were 3.7%, 4.5%, and 0%. The combined attrition rate after 2011 is 3.8% and the RETR is 96.2%.

Recruitment for the Vulnerable Populations cohort started in 2011, so there has not been any attrition before 2011. Of the 3759 individuals who were entered into the panel in 2011, 73 (1.9%) were removed during or after 2011. Note, however, that entering these individuals had not been completed yet by the end of 2011 (e.g., 265 additional individuals have been entered between January 1 and January 25, 2012). As indicated in the section on recruitment, 4331 individuals were recruited and 86% of these expressed interest in joining the panel, which would imply an attrition rate (as defined here) of 14% after 2011.

Table B.5 shows the reasons for attrition. It follows that the vast majority of "attritors" are individuals who signed up for the panel but never actually participated and once active, individuals are much less likely to leave the panel. Note that "ineligible" and "died" should generally not be considered attrition. Rather, these individuals should be removed from both the numerator and the denominator of the attrition rates. Given their small numbers, the computed rates are not noticeably affected by their inclusion.

Table B.5: Reasons for attrition.

Reason	Number	Percent
Ineligible for the panel	13	0.5
Died	44	1.7
Unable to (re)contact	167	6.4
Never participated	1997	76.6
Problems with computer/internet/WebTV	51	2.0
Health/cognition/claims too old	29	1.1
No time/busy	33	1.3

Other			272	10.4
Total			2606	100.0

Completion rates

For every survey, a number of panel members are invited by email to participate in the survey. Generally, due to survey-specific eligibility criteria or budget reasons, not all panel members are invited for a survey. Callegaro and DiSogra (2008) and AAPOR (2011) define the completion rate (COMR) of a survey as the number of panel members who delivered a complete (I) or partial (P) interview, divided by the number of panel members who were invited to participate in the survey. AAPOR (2011) also mentions break-offs and includes a discussion of rules of thumb of how one might define complete, partial, and break-off. In ALP surveys, almost all panel members who start a survey also finish it, and item nonresponse is very low, so the computation of completion rates is insensitive to alternative definitions. For the current computations, we define a partial interview as one that was started (variable tsstart is a valid date/time stamp) but not finished (tsend is missing) and a complete interview as one that was finished (tsend is a valid date/time stamp), and a nonresponse as an individual who was invited to participate in the survey but did not start the survey (tsstart is missing). We do not make a distinction between break-offs and partial interviews.

We computed completion rates for all surveys that were both opened and closed in 2011 (the field dates were completely within 2011). The surveys included are 162, 164, 165, 167, 169, 171, 173, 174, 175, 176, 177, 178, 179, 180, 182, 185, 188, 189, 190, 194, 197, 198, 199, 200, 211, 213, 217, 219, 220, 221, 225, and 231. See https://mmicdata.rand.org/alp/index.php/Data for the list of ALP surveys. Table B.6 gives a summary of the results. The number of partial interviews is 1% of the number of invites, which confirms the minor role played by partial interviews. The completion rates are approximately 80% taken over all surveys. (There is considerable variation across surveys, though.)

Table B.6: Completion rates in 2011.

		invited	nonresp	partial (P)	complete (I)	completion rate (COMR)
MS Internet + MS CATI	N	36,529	7,945	353	28,231	
	%	100.0	21.7	1.0	77.3	78.3
National Survey Project	N	5,236	1,007	53	4,176	
	%	100.0	19.2	1.0	79.8	80.8
Total	N	41,765	8,952	406	32,407	
	%	100.0	21.4	1.0	77.6	78.6

The Vulnerable Population cohort was signed up in 2011 (and 2012) and thus many of these members have not completed their first survey yet. Some have been invited already, and their completion rates are about 30%, but we believe that this low number is due to them not being used to the system yet and not a reliable indication of likely response rates in 2012 and later. We will spend considerable efforts in activating these panel members and we expect their completion rates to be similar to the completion rates of the other cohorts. Thus, in the computation of the cumulative response rates, we assume the average 78.6% completion rate of the other cohorts applies to the Vulnerable Population cohort.

Combining the rates of the different cohorts

Recruitment rates, retention rates, and completion rates of the different cohorts are combined analogously to the way the different recruitment years were combined, that is, by weighting the cohort-specific rates by their sizes. Here, the "size" is the denominator for the relevant rate. For the recruitment rates, this is the number of addresses or telephone numbers selected; for the retention rates, it is the number of individuals who were recruited; and for the completion rates, it is the number of invitations to participate in a survey. The results were given in Table B.1 above.

Sample Size

As mentioned above, the ALP pool contains roughly 4,500 respondents. After excluding the Snowball cohort, the Mailing and Phone Experiment cohorts, and the RDS cohort, we are left with about 3,600 panel members. Based on the previous discussion, we assume a completion rate of 80%. Accordingly, the expected sample size for the project will be 2,900 individuals.

B.4. Tests of Procedures or Methods to be undertaken

Sample Weights

As for most microeconomic surveys based on random sampling, the composition of the ALP sample does not necessarily match the one of the reference population. Hence, sample weights are needed to derive population estimates. RAND constructs such weights to allow for generalization to the population it intends to represent.

The reference population for the entire ALP consists of individuals aged 18 and older, who are not in the Armed Forces or institutionalized. The weighting procedure adopted by RAND, however, allows targeting specific sub-populations depending on the sample selection criteria of distinct surveys (e.g. the population of persons in a certain age bracket).

The benchmark distributions against which ALP surveys are weighted are derived from the Current Population Survey (CPS) Annual Social and Economic Supplement (administered in March of each year). This choice follows common practice in other social science surveys, such as for instance, the Health and Retirement Study (HRS).

Three weighting methods have been implemented for the ALP: cell-based post stratification, logistic regression, and raking. After experimentation over time, raking has been found to give the best results among these different methods. It allows finer categorizations of variables of interest (in particular, age and income) than cell-based post-stratification does, while still matching benchmark distributions of such variables exactly.

The weighting procedure consists of two steps. In the first one, individual demographics from the CPS are mapped onto those available in the ALP and selected weighting variables are recoded into strata (or categories). Such re-categorization, which applies to both CPS and ALP variables, is required particularly when weighting variables are continuous (e.g., income) or take values in a finite, but relatively large set (e.g., age).. In the second step, the raking algorithm is implemented and sample weights are generated by matching the proportions of pre-defined strata in the ALP to those in the CPS. More precisely, the weighting algorithm is performed on the following set of two-way marginals:

- **Gender ×Age**

with twelve categories: (1) male, 18-30; (2) male, 31-40; (3) male, 41-50; (4) male, 51-60; (5) male, 61-74; (6) male, 75+; (7) female, 18-30; (8) female, 31-40; (9) female, 41-50; (10) female, 51-60; (11) female, 61-74; (12) female, 75+.

- **Gender × Ethnicity**

  with six categories: (1) male, non-Hispanic White; (2) male, non-Hispanic African American; (3) male, Hispanic or Other; (4) female, non-Hispanic White; (5) female, non-Hispanic African American; (6) female, Hispanic or Other.

- **Gender × Education**

  with six categories: (1) male, high school or less; (2) male, some college or a bachelor's degree; (3) male, more than a bachelor's degree; (4) female, high school or less; (5) female, some college or a bachelor's degree; (6) female, more than a bachelor's degree.

- **Gender × Household Income**

  with eight categories: (1) male, <$35,000; (2) male, $35,000-$59,999; (3) male, $60,000-$99,999; (4) male, $100,000+; (5) female, <$35,000; (6) female, $35,000-$59,999; (7) female, $60,000-$99,999; (8) female, $100,000+.

- **Household Income × Number of Household Members**

  with six categories: (1) single, <$60,000; (2) single, $60,000+; (3) couple, <$60,000; (4) couple, $60,000+; (5) 3+ members, <$60,000; (6) 3+ members, $60,000+.

The above strata are defined such that none of them contains less than 5% of the ALP sample. This rule of thumb is commonly adopted (DeBell and Krosnick, 2009) in poststratification weighting. It aims at preventing very small cell sizes and, therefore, extremely high weights.

While sample weights are necessary to correctly infer population parameters, their use requires estimation techniques that take design effects into account and adjust variance estimates accordingly. This is a crucial step to enable valid statistical inference. The design effect (*deff*) measures the extent to which the sampling design (as described by the sample weights) influences the computation of any statistic of interest. It is defined as the ratio of the variance of the statistic from the weighted sample (under complex survey design) to the variance of the statistic from an equally weighted sample (under simple random sample) with the same number of observations. A design effect greater than 1 indicates that the use of sample weights decreases the precision of estimates compared to an equally weighted sample. In this case, standard formulas for the variance of estimates should be appropriately amended. Suppose, for instance, that we are interested in the population mean of a variable *x*. After computing the weighted sample average, $\overline{x}$, a 95% confidence interval for the population mean of *x* will be given by:

$$\overline{x} \pm 1.96 \times SE(\overline{x}) \times \sqrt{deff},$$

where $SE(\overline{x})$ is the standard error of $\overline{x}$. Analogous formulas apply to other estimators, such as regression coefficients.

We will run our statistical analyses using Stata, software capable of correcting variance estimates for design effects. In Table B.7, we present mean estimates and corresponding design effects for a number of individual demographics. We compare estimated quantities across different samples: the 2011 CPS sample of individuals aged 18 and older, the entire ALP sample, and the reference sample for this project (excluding Snowball, Mail and Phone, and RDS cohorts from the ALP sample as well as Added Members).

The results show that weighted sample means in the ALP are in line with their CPS counterparts. Moreover, the increase in the variance introduced by sample weights, as measured by the design effect, appear modest.

Sample means will be, together with regression coefficients, the type of estimates we will generalize to the national level. The results in Table B.7 can be used to determine the extent to which the margin of error will be affected by the design effect. For our statistical analysis, we will follow the common practice of setting the reference level of significance for hypothesis tests at 5%. This represents the probability of rejecting the null hypothesis when it is indeed true. Suppose we are interested in the average age of the population. Using our reference sample, we obtain an estimate of 46.72 and a standard error of 0.453 (see Table B.7). These numbers give the following 95% confidence interval [45.83, 47.60]. The 95% confidence interval appropriately adjusted to account for the design effect is [45.33, 48.10], which is only marginally wider. As it can be seen from Table B.7, the estimate of average age has a relatively high standard error and design effect. Even in this case, which we chose deliberately for illustrative purposes, the impact of the sample design on the margin of error is rather modest. We expect the impact of the sample design on the margin of error to be around or below this order of magnitude when conducting statistical inference for the proposed investigation.

Table B.7: Estimated Means and Design Effects

Gender: Male (1), Female (0)					
	Min	Max	Mean	Std. Err.	Deff
CPS	0	1	0.48	0.0018	1.28
ALP Sample	0	1	0.48	0.0094	1.60
Ref. Sample	0	1	0.48	0.0123	1.99
Ethnicity: White (1), Non-White (0)					
	Min	Max	Mean	Std. Err.	Deff
CPS	0	1	0.68	0.0017	1.34
ALP Sample	0	1	0.68	0.0089	1.62
Ref. Sample	0	1	0.68	0.0115	1.99
Age (years)*					
	Min	Max	Mean	Std. Err.	Deff
CPS	18	85	46.33	0.063	1.29
ALP Sample	18	85	46.63	0.348	1.91
Ref. Sample	18	85	46.72	0.453	2.42
Education (categories)**					

	Min	Max	Mean	Std. Err.	Deff
CPS	1	16	10.17	0.0098	1.29
ALP Sample	1	16	10.42	0.0440	1.51
Ref. Sample	1	16	10.47	0.0564	1.78
Household Income (categories)*					
	Min	Max	Mean	Std. Err.	Deff
CPS	1	15	10.54	0.014	1.29
ALP Sample	1	15	10.39	0.080	1.69
Ref. Sample	1	15	10.37	0.107	2.14

Note:

*Top-coded to match the CPS. Specifically, 80 means 80-84 and 85 means 85+.

**Education categories Less than 1st grade, 1;1st,2nd,3rd,or 4th grade, 2; 5th or 6th grade, 3; 7th or 8th grade, 4; 9th grade, 5; 10th grade, 6; 11th grade, 7; 12th grade NO DIPLOMA, 8; High school DIPLOMA or the equivalent (For example: GED), 9; Some college but no degree, 10; Associate degree in college Occupational/vocational program, 11; Associate degree in college Academic program, 12 ; Bachelor's degree (For example: BA,AB,BS), 13; Master's degree (For example: MA,MS,MEng,MEd,MSW,MBA), 14; Professional School Degree (For example: MD,DDS,DVM,LLB,JD), 15 Doctorate degree (For example: PhD,EdD), 16.

***Income categories: Less than $5,000, 1; $5,000 to $7,499, 2; $7,500 to $9,999, 3; $10,000 to $12,499, 4; $12,500 to $14,999, 5;$15,000 to $19,999, 6;$20,000 to $24,999, 7;$25,000 to $29,999, 8;$30,000 to $34,999, 9;$35,000 to $39,999, 10;$40,000 to $49,999, 11;$50,000 to $59,999, 12;$60,000 to $74,999, 13;$75,000 to $99,999, 14;$100,000 or more, 15

In Tables B.8 and B.9, we assess the extent to which sample weights correct for over- or under-representation of strata by comparing weighted distributions in the reference sample with those in the CPS. For this purpose, we form strata by interacting gender, age, working status, and income. The chosen combinations are different from those used to generate sample weights because they feature 1) working status (which is not used in the weighting procedure) and 2) age and income categories based on the quartiles of the 2011 CPS age and income distributions, respectively. This approach should increase the power of the test. The results show a satisfactory alignment of the proportions across strata, once sample weights are applied.

Table B.8: Gender x Working Status x Income Distribution: CPS vs. Weighted Reference Sample

Stratum	Gender	Working Status	Income Bracket	CPS	Ref. Sample (unweighted)	Ref. Sample (weighted)
1	Male	Working	Less than $30,000	5.79	3.92	6.11
2	Female	Working	Less than $30,000	5.68	7.91	6.03
3	Male	Not Working	Less than $30,000	7.67	6.81	8.21
4	Female	Not Working	Less than	10.70	11.65	10.5

			$30,000			
5	Male	Working	$30,000-$59,999	8.96	7.42	8.56
6	Female	Working	$30,000-$59,999	8.24	11.53	8.13
7	Male	Not Working	$30,000-$59,999	5.31	4.54	4.85
8	Female	Not Working	$30,000-$59,999	6.78	6.71	6.74
9	Male	Working	$60,000-$99,999	8.35	7.39	9.12
10	Female	Working	$60,000-$99,999	7.35	8.74	7.74
11	Male	Not Working	$60,000-$99,999	2.85	2.58	2.08
12	Female	Not Working	$60,000-$99,999	3.72	3.46	3.33
13	Male	Working	$100,000+	7.66	6.68	8.16
14	Female	Working	$100,000+	6.25	6.68	6.30
15	Male	Not Working	$100,000+	1.82	1.59	1.33
16	Female	Not Working	$100,000+	2.87	2.39	2.83

In columns 5-7 we report the fraction of individuals in each stratum.

Table B.9: Gender x Working Status x Age Distribution: CPS vs. Weighted Reference Sample

Stratum	Gender	Working Status	Age Group	CPS	Ref. Sample (unweighted)	Ref. Sample (weighted)
1	Male	Working	18-32	8.95	3.49	10.05
2	Female	Working	33-44	8.09	6.68	7.71
3	Male	Not Working	45-57	4.78	1.13	3.38
4	Female	Not Working	58+	5.51	4.05	5.42
5	Male	Working	18-32	8.24	5.43	7.68
6	Female	Working	33-44	7.05	8.74	6.86
7	Male	Not Working	45-57	1.89	1.62	2.13
8	Female	Not Working	58+	3.32	3.95	3.29
9	Male	Working	18-32	9.05	10.94	10.29
10	Female	Working	33-44	8.42	13.24	9.20

11	Male	Not Working	45-57	2.85	2.94	2.43
12	Female	Not Working	58+	4.04	5.76	3.99
13	Male	Working	18-32	4.51	5.55	3.94
14	Female	Working	33-44	3.96	6.19	4.43
15	Male	Not Working	45-57	8.13	9.81	8.52
16	Female	Not Working	58+	11.19	10.45	10.69

In columns 5-7 we report the fraction of individuals in each stratum.

B.5. Pre-Testing

Once the survey questionnaire has been developed, we will pre-test it using ALP respondents excluded from the reference sample. Specifically, we plan to pilot the questionnaire by administering it to 50 respondents from the Snowball cohort. There are around 550 ALP respondents belonging to this cohort. Compared to those in the reference sample, their characteristics are somewhat different. In particular, the Snowball cohort features more female respondents (63% versus 52% in the reference sample), younger (average age of 43.73 versus 46.72 in the reference sample), and slightly higher educated individuals (average education is 10.81 versus 10.37). We will appropriately choose the 50 respondents for pre-testing so as to mimic the composition of the reference sample as closely as possible. Data from the pretest will be used for quality checks, which may lead to changes in the questionnaire. Once the final questionnaire is ready, it will be fielded to the respondents in the reference sample.

B.6. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

This study is being conducted by the contractor, The RAND Corporation, under contract to the U.S. Department of Labor. The RAND Principal Investigator is Dr. Angela Hung. Ms. Noreen Clancy will oversee the collection of focus group data. Dr. Lauren Fleishman and Dr. Jeremy Burke will oversee the design and analysis of the survey and experiments.

Individuals consulted on Statistical Aspects:

Dr. Erik Meijer, Econometrician, RAND
Dr. Arie Kapteyn, Econometrician, RAND
Dr. Marco Angrisani, Quantitative Economist, RAND

Contact information:

The RAND Corporation:

Principal Investigator: Angela Hung, PhD
Senior Economist and Director, Center for Financial and Economic Decision Making
RAND Corporation
1776 Main Street
P.O. Box 2138
Santa Monica, CA 90407-2138

Office: 310-393-0411 x6081
Fax: 310-393-4818
Email: ahung@rand.org

Department of Labor:
Anja Decressin, Ph.D.
Keith Bergstresser, Ph.D
Department of Labor
Employee Benefits Security Administration, N5718
200 Constitution Ave., NW
Washington, DC 20210
Phone: (202) 693-8417
Decressin.Anja@dol.gov

References

AAPOR (2011). *Standard definitions: Final depositions of case codes and outcome rates for surveys* (7th ed.). The American Association for Public Opinion Research.

Alwin, D.F. (2007), *Margins of Error. A study of Reliability in Survey Measurement*, New Jersey: Wiley.

Andersen, S., G. Harrison, M. Lau, and E. Rutstroem (2008), Eliciting risk and time preferences, *Econometrica*, 76(3), pp. 583-618.

Avendano, M., A. Scherpenzeel, and J.P. Mackenbach (2011), Can Biomarkers be Collected in an Internet Survey? A Pilot Study in the LISS Panel, In: Das, M., P. Ester, and L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*. Boca Raton: Taylor & Francis.

Bellemare, C. and S. Kröger (2007), On representative social capital, *European Economic Review*, 51, pp. 183–202.

Bellemare, C., S. Kröger, and A. van Soest (2008), Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities, *Econometrica*, 76(4), pp. 815-839.

Bennink, M., G. Moors, and J. Gelissen (2011), Exploring Response Differences between Face-to-face and Web surveys: A Qualitative Comparative Analysis of the Dutch 'European Values Survey 2008', Working paper, Department of Methodology and Statistics, School of Social and Behavioral Sciences, Tilburg University.

Berthoud, R. and J. Gershuny (2000), *Seven Years in the Lives of British Families. Evidence on the Dynamics of Social Change from the British Household Panel Survey*, Bristol: The Policy Press.

Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly*, *72*, 1008-1032.

Carmines, E.G. and R.A. Zeller (1994), Reliability and validity assessment, In: Lewis-Beck, E. (Ed.), *Basic measurement*, London: Sage.

Chang, L. and J.A. Krosnick (2009), National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality, *Public Opinion Quarterly*, 73, pp. 641-678.

Couper, M.P. (2008), *Designing Effective Web Surveys*. Cambridge: Cambridge University Press.

Couper, M. P. 2000, Web Surveys: A Review of Issues and Approaches, *Public Opinion Quarterly*, 64, pp. 464–94.

Couper, M. P., F.G. Conrad, and R. Tourangeau (2007), Visual context effects in Web surveys, *Public Opinion Quarterly*, 71, pp. 623–634.

Couper, M.P., A. Kapteyn, M. Schonlau, and J. Winter (2007), Noncoverage and Nonresponse in an Internet Survey, *Social Science Research*, 36(1), pp. 131-148.

Couper, M.P. and M.B. Ofstedal (2009), Keeping in Contact with Mobile Sample Members. In: Lynn, P. (Ed.) *Methodology of Longitudinal Surveys*, Chichester: John Wiley & Sons.

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

Couper, M. P., R. Tourangeau, and K. Kenyon (2004), Picture this! Exploring visual effects in Web surveys, *Public Opinion Quarterly*, 68, pp. 255–266.

Curtin, R. T. (2002). Surveys of consumers. Retrieved January 25, 2012, from http://www.sca.isr.umich.edu/

Das, M. and A. van Soest (1999), A Panel Data Model for Subjective Information on Household Income Growth, *Journal of Economic Behavior and Organization*, 40, pp. 409–426.

Das, M., J. Dominitz, and A. van Soest (1999), Comparing Predictions and Outcomes: Theory and Application to Income Changes, *Journal of the American Statistical Association*, 94, pp. 75–85.

De Leeuw, E. (2005), To Mix or Not to Mix Data Collection Modes in Surveys, *Journal of Official Statistics*, 21 (2), pp. 233-255.

De Leeuw, E. and J. Hox (2011), Internet Surveys as Part of a Mixed-Mode Design, In: Das, M., P. Ester, and L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*. Boca Raton: Taylor & Francis.

De Leeuw, E., J. Hox, and A.C. Scherpenzeel (2010a), Emulating Interviewers in an Online Survey: Experimental Manipulation of 'Do-Not-Know' over the Phone and on the Web, *JSM Proceedings, Survey Research Methods Section*, pp. 6305-6314. Alexandria, VA: American Statistical Association.

De Leeuw, E., J. Hox, and A.C. Scherpenzeel (2010b), Mode Effect or Question Wording? Measurement Error in Mixed Mode Surveys, *JSM Proceedings, Survey Research Methods Section, pp. 5959-5967*. Alexandria, VA: American Statistical Association.

Dillman, D.A., G. Phelps, R. Tortora, K. Swift, J. Kohrell, J. Berck, and B.L. Messer (2009), Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet, *Social Science Research*, 38(1), pp. 1-18.

DiSogra, C. (2010). Update: Address-based sampling nets success for KnowledgePanel® recruitment and sample representation. Retrieved January 25, 2012, from http://knowledgenetworks.com/accuracy/spring2010/disogra-spring10.html

Dixon, J. & Tucker, C. (2010). Survey nonresponse. In P. Marsden & J. Wright, *Handbook of survey research* (2nd ed.; pp. 593-630). Bingley, UK: Emerald Group Publishing Limited.

Duffy, B., K. Smith, G. Terhanian, and J. Bremer (2005), Comparing data from online and face-to-face surveys, *International Journal of Market Research*, 47, pp. 615–639.

Duncan, G.J. and G. Kalton (1987), Issues of Design and Analysis of Surveys Across Time, *International Statistical Review*, 55, pp. 97-117.

Göritz, A.S. (2006), Incentives in Web Studies: Methodological Issues and a Review, *International Journal of Internet Science*, 1, pp. 58-70.

Groves, R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau (2004), *Survey Methodology*, New York: John Wiley & Sons.

Harris, T. J., Owen, C. G., Victor, C. R., Adams, R., Cook, D. G., 2009. What factors are associated with physical activity in older people, assessed objectively by accelerometry? *Br J Sports Med*, *43*, 442-450.

Harrison, G., M. Lau and M. Williams (2002), Estimating individual discount rates in Denmark: A field experiment, *American Economic Review*, 92(5), pp. 1606–1617.

Hill, D.H. (2002), Wealth Dynamics: Reducing Noise in Panel Data, Working paper, University of Michigan.

Hill, D., M. Perry, and R.J. Willis (2004), Do Internet Surveys Alter Estimates of Uncertainty and Optimism about Survival Chances? Working paper, University of Michigan.

Hoogendoorn, A.W. (2004), A Questionnaire Design for Dependent Interviewing that Addresses the Problem of Cognitive Satisficing, *Journal of Official Statistics*, 20, pp. 219-232.

Hoogendoorn, A.W. and J. Daalmans (2009), Nonresponse in the Recruitment of an Internet Panel Based on Probability Sampling, *Survey Research Methods*, 3, pp. 59-72.

Hurd, M. and A. Kapteyn (2004), The Quality of Subjective Probability Data: Telephone versus Internet, Working paper, RAND.

Jäckle, A. (2009), Dependent Interviewing: A Framework and Application to Current Research, In: Lynn, P. (Ed.), *Methodology of Longitudinal Surveys*, Chichester: John Wiley & Sons.

Jäckle, A. (2008), Dependent Interviewing: Effects on Respondent Burden and Efficiency, *Journal of Official Statistics*, 24, pp. 411-430.

Jäckle, A. and P. Lynn (2007), Dependent Interviewing and Seam Effects in Work History Data, *Journal of Official Statistics*, 23, pp. 529-551.

Kahneman, D., A.B. Krueger, D.A. Schkade, N. Schwarz, and A.A. Stone (2004), A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method, *Science*, 306, pp. 1776-1780.

Kalton, G. and C.F. Citro (1993), Panel Surveys: Adding the Fourth Dimension, *Survey Methodology*, 19, pp. 205-215.

Kapteyn, A. and J.Y. Ypma (2007), Measurement Error and Misclassification: A Comparison of Survey and Administrative Data, *Journal of Labor Economics*, 25 (3), pp. 513-551.

Kapteyn, A., J.P. Smith, and A. van Soest (2007), Vignettes and self-reports of work disability in the US and The Netherlands, *American Economic Review,* 97 (1), pp. 461-473.

Knauper, B. (1999), The Impact of Age and Education on Response Order Effects in Attitude Measurement, *Public Opinion Quarterly*, 63, pp. 347-370.

Laurie, H., R. Smith, and L. Scott (1999), Strategies for Reducing Nonresponse in a Longitudinal Panel Survey, *Journal of Official Statistics*, 15, pp. 269-282.

Loosveldt, G. and N. Sonck (2008), An evaluation of the weighting procedures for an online access panel survey, *Survey Research Methods*, 2, pp. 93–105.

Lynn, P. (2009), Methods for Longitudinal Surveys, In: Lynn, P (Ed.), *Methodology of Longitudinal Surveys*, Chichester: John Wiley & Sons.

Lynn, P., A. Jäckle, S.P. Jenkins, and E. Sala, E. (2006), The Effects of Dependent Interviewing on Responses to Questions on Income Sources, *Journal of Official Statistics*, 22, pp. 357-384.

Lynn, P. and E. Sala (2006), Measuring Change in Employment Characteristics: The Effects of Dependent Interviewing, *International Journal of Public Opinion Research*, 18, pp. 500-509.

Manski, C. F. (1990), The Use of Intentions Data to Predict Behavior: A Best-Case Analysis, *Journal of the American Statistical Association*, 85, pp. 934–940.

Mathiowetz, N.A. and T.J. Lair (1994), Getting Better? Changes or Errors in the Measurement of Functional Limitations, *Journal of Economic & Social Measurement*, 20, pp. 237-262.

Mathiowetz, N.A. and K.A. McGonagle (2000), An Assessment of the Current State of Dependent Interviewing in Household Surveys, *Journal of Official Statistics*, 16, pp. 401-418.

Meurs, H., L. van Wissen, and J. Visser (1989), Measurement Biases in Panel Data, *Transportation*, 16, pp. 175-194.

Mack, S., V. Huggins, D. Keathley, and M. Sunduckchi (1998), Do Monetary Incentives Improve Response Rates in the Survey of Income and Program Participation? *Proceedings of the Section of Survey Research Methods*, American Statistical Association.

Malhotra, N. and J.A. Krosnick (2007), The effect of survey mode and sampling on inferences about political attitudes and behavior: Comparing the 2000 and 2004 ANES to Internet surveys with non-probability samples, *Political Analysis*, 15, pp. 286–323.

Millar, M.M. and D.A. Dillman (2011), Improving Response to Web and Mixed-Mode Surveys, *Public Opinion Quarterly*, 75 (2), pp. 249-269.

Oosterveld, P. and P. Willems (2003), Two modalities, one answer? Combining Internet and CATI surveys effectively in market research, In: Fellows, D.S. (Ed.), *Technovate*, Amsterdam: ESOMAR.

Revilla, M.A. and W.E. Saris (2010), Comparison of Surveys Using Different Modes of Data Collection: European Social Survey versus LISS Panel, Working Paper, Research and Expertise Centre for Survey Methodology (RECSM), Universitat Pompeu Fabra, Spain.

Sakshaug, J., Tourangeau, R., Krosnick, J.A., Ackermann, A., Marka, A., DeBell, M., et al. (2009). Dispositions and outcome rates in the "Face-to-Face Recruited Internet Survey Platform" (the FFRISP). Paper presented at the 64th Annual conference of the American Association for Public Opinion Research (AAPOR), Hollywood, FL.

Saris W.E. and I. Gallhofer (2007), Design, Evaluation, and Analysis of Questionnaires for Survey Research, New York: Wiley.

Scherpenzeel, A.C. (2009a), Recruiting a probability sample for an online panel: effects of contact mode, incentives and information, Working paper, CentERdata, Tilburg University (resubmitted to Public Opinion Quarterly).

Scherpenzeel, A.C. (2009b), Online interviews and data quality: A multitrait-multimethod study, Working paper, CentERdata, Tilburg University.

Scherpenzeel, A.C. and M. Das (2011), True Longitudinal and Probability-Based Internet Panels: Evidence from the Netherlands, In: Das, M., P. Ester, and L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*. Boca Raton: Taylor & Francis.

Schonlau, M., A. van Soest, A. Kapteyn, and M. Couper (2009), Selection Bias in Web Surveys and the Use of Propensity Scores, *Sociological Methods and Research*, 37, pp. 291-318.

Sikkel, D. and A.W. Hoogendoorn (2008), Panel surveys. In: De Leeuw, E.D., J.J. Hox, and D.A. Dillman (Eds.), *International handbook of survey methodology*, European Association of Methodology Series. New York: Lawrence Erlbaum Associates.

Singer, E. and R.A. Kulka (2002), Paying Respondents for Survey participation. In: M. Ver Ploeg, R.A. Moffitt, and C.F. Citro (eds.), *Studies of Welfare Populations: Data Collection and Research Issues*, Committee on National Statistics, National Research Council.

Stiglitz, J., A. Sen, and J.-P. Fitoussi (2009), *Report of the Commission on the Measurement of Economic Performance and Social Progress*, Commission on the Measurement of Economic Performance and Social Progress, Paris.

Taylor, H. (2000). Does Internet research work? Comparing online survey results with telephone surveys, *International Journal of Market Research,* 42(1), pp. 51–63.

Toepoel, V. and D.A. Dillman (2011), How visual design affects the interpretability of survey questions, In: Das, M., P. Ester, and L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*. Boca Raton: Taylor & Francis.

Toepoel, V., M. Das, and A. van Soest (2008). Effects of design in web surveys: Comparing trained and fresh respondents, *Public Opinion Quarterly*, 72(5), pp. 985-1007.

Toepoel, V., M. Das, and A. van Soest (2009), Relating Question Type to Panel Conditioning: A Comparison between Trained and Fresh Respondents, *Survey Research Methods*, 3(2), pp. 73-80.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., McDowell, M. 2008. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc, 40*, 181-188

US Census Bureau (2011). *Intercensal estimates of the resident population by single year of age, sex, race, and hispanic origin for the United States: April 1, 2000 to July 1, 2010* (Release date September 2011). Retrieved January 25, 2012, from

http://www.census.gov/popest/data/intercensal/national/nat2010.html

Van der Sar, R., E.P.M. Brouwers, I.A.M. van de Goor, H.F.L. Garretsen (2010), The opinion of adolescents and adults on Dutch restrictive and educational alcohol policy measures, *Health Policy*, 99 (1), pp. 10-16.

Van Soest, A. and A. Kapteyn (2011), Mode and Context Effects in Measuring Household Assets, In: Das, M., P. Ester, and L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*. Boca Raton: Taylor & Francis.

Van Wilsem, J. and M. Heemskerk (2011). Socially desirable crime rates? A Dutch study on socially desirable responding and its impact on the self-report of violent victimization, Working paper, Leiden University.

Vehovar, V., Z. Batagelj, and K. Lozar Manfreda (1999), Web surveys: Can the weighting solve the problem? *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 962–967.

Von Gaudecker, H-M., A. van Soest, and E. Wengström (2011), Heterogeneity in Risky Choice Behavior in a Broad Population, *American Economic Review*, 101(2), pp. 664–694.

Voortman, H. (2009), *Political Trust; a Matter of Personality Factors or Satisfaction with Government Performance? A study on the influence of personality traits, moods and satisfaction with government performance on political trust*, PhD Thesis, University of Twente, Enschede.

Yeager, D.S., J.A. Krosnick, L. Chang, H.S. Javitz, M.S. Levindusky, A. Simpser, and R. Wang (2009), Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples, Working paper, Stanford University.

Yeager, D. S., Larson, S. B., Krosnick, J. A., & Tompson, T. (2011). Measuring Americans' issue priorities: A new version of the most important problem question reveals more concern about global warming and the environment. *Public Opinion Quarterly*, *75*, 125-138.