

APPENDIX A: SURVEY METHODOLOGY

The survey was conducted by telephone with a national sample of adults (18 years of age or older) living in all 50 U.S. states and the District of Columbia. The results reported here are based on 2,012 interviews (1,032 respondents were interviewed on a landline telephone and 980 were interviewed on a cell phone). Interviews were completed in English and Spanish by live, professionally trained interviewing staff under the direction of Princeton Survey Research Associates International from March 21 to April 8, 2013.

Survey Design

A combination of landline and cell random digit dial (RDD) samples was used to reach a representative sample of all adults in the United States who have access to either a landline or cellular telephone. Both samples were disproportionately stratified to increase the incidence of African-American and Hispanic respondents. Within each stratum, phone numbers were drawn with equal probabilities. The landline samples were list-assisted and drawn from active blocks containing three or more residential listings, while the cell samples were not list-assisted but were drawn through a systematic sampling from dedicated wireless 100-blocks and shared service 100-blocks with no directory-listed landline numbers. Both the landline and cell RDD samples were disproportionately stratified by county based on estimated incidences of African-American and Hispanic respondents.

The survey questionnaire included a split-form design whereby an additional 1,994 adults were asked a different set of questions. Results from those interviews are being held for future release. The total number of interviews conducted was 4,006. Thus, the data collection involved two simultaneous surveys, each of which is weighted separately to represent U.S. adults; where the same question was asked on each form, the results of the two forms can be combined to yield a representative survey of U.S. adults with the full 4,006 respondents.

Margin of Sampling Error

Statistical results are weighted to correct known demographic discrepancies, including disproportionate stratification of the sample. The table shows the unweighted sample sizes and the error attributable to sampling that would be expected at the 95% level of confidence for different groups in the survey.

The survey's *margin of error* is the largest 95% confidence interval for any estimated proportion based on the total sample – the one around 50%. For example, the margin of error for the entire sample is ± 2.9 percentage points. This means that in 95 out of every 100 samples drawn using the same methodology, estimated proportions based on the entire sample will be no more than 2.9 percentage points away from their true values in the population. Sampling errors and statistical tests of significance used in this report take into account the effect of weighting. In addition to sampling error, one should bear in mind that question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of opinion polls.

Interviewing Procedures

All interviews were conducted using a Computer Assisted Telephone Interviewing (CATI) system, which ensures that questions were asked in the proper sequence with appropriate skip patterns. CATI also allows certain questions and certain answer choices to be rotated, eliminating potential biases from the sequencing of questions or answers.

For the landline sample, half of the time interviewers asked to speak with the youngest adult male currently at home and the other half of the time asked to speak with the youngest adult female currently at home, based on a random rotation. If no respondent of the initially requested gender was available, interviewers asked to speak with the youngest adult of the opposite gender who was currently at home. For the cell phone sample, interviews were

Margins of Error

	Sample size	Plus or minus percentage points
All adults	2,012	2.9
White, not Hispanic	1,236	3.7
Black, not Hispanic	319	7.3
Hispanic	299	7.1
Protestant	985	4.2
White evangelical	353	6.9
White mainline	265	8.3
Black Protestant	255	8.2
Catholic	467	5.8
White Catholic	265	7.7
Hispanic Catholic	165	9.4
Unaffiliated	360	7.0

Note: The margins of error are reported at the 95% level of confidence and are calculated by taking into account the average design effect for each subgroup.

PEW RESEARCH CENTER

conducted with the person who answered the phone; interviewers verified that the person was an adult and could complete the call safely.

Both the landline and cell samples were released for interviewing in replicates, which are small random samples of each larger sample. Using replicates to control the release of the telephone numbers ensures that the complete call procedures are followed for all numbers dialed. As many as seven attempts were made to contact every sampled telephone number. The calls were staggered at varied times of day and days of the week (including at least one daytime call) to maximize the chances of making contact with a potential respondent.

Questionnaire Development and Testing

The Pew Research Center developed the questionnaire. The design of the questionnaire was informed by the results of additional pilot and pretesting on questions related to radical life extension.

Initial Pilot Tests. Two initial pilot studies were conducted Oct. 27-30, 2011, as part of an omnibus survey with a national sample of adults living in the continental U.S. The data collection involved a split-form design, which, in essence, collected data on two sets of questions, or pilot studies, simultaneously. Each form of the questionnaire, or pilot study, was based on both landline and cell phone telephone interviews; 262 adults completed Form A, and 265 adults completed Form B. The interviews were conducted in English under the direction of Princeton Survey Research Associates International. The purpose of the pilot tests was to gauge the feasibility of asking about radical life extension, a topic that involves reactions to a still-hypothetical development in medical science, and to compare two different variants in question wording. The questionnaire was brief. Form A included five closed-ended questions and two open-ended questions. Form B included two closed-ended questions and one open-ended question. The open-ended questions measured respondent's spontaneous reactions to the idea of radical life extension; these responses helped inform development of closed-ended questions on beliefs about the likely effects of radical life extension on society.

Full Pilot Test. An additional pilot study was conducted Feb. 27, 2013, with a sample of 103 adults living in the continental U.S. The sample was drawn from fresh RDD landline phone numbers and a sample of cell phone numbers from recontact cases in recent RDD omnibus studies. The interviews were conducted in English under the direction of Princeton Survey Research Associates International. The interviews tested the questions planned for the study questionnaire in the full survey context; pilot study interviews lasted about 24 to 26 minutes, on average.

Question Wording Tests. A question wording experiment was completed Feb. 21-26, 2013, through Google Consumer Surveys with a convenience sample of internet users. Those completing the online, two-question survey were randomly assigned to one of two forms of the questionnaire. A total of 301 respondents completed Form A; 318 respondents completed Form B. The purpose of the question wording experiment was to test the effect of including the specification that medical treatments would allow the average person to live decades longer “*to at least 120 years old*” compared with living decades longer when no specific age is mentioned. Both forms also included an open-ended question that measured respondents’ thoughts and associations related to radical life extension.

Pretest. As a final step, a traditional pretest was conducted March 12, 2013, with 22 adults living in the continental U.S. The sample was drawn from fresh RDD landline phone numbers and a sample of cell phone numbers from recontact cases in recent RDD omnibus studies. The interviews were conducted in English under the direction of Princeton Survey Research Associates International. The interviews tested the questions planned for the study questionnaire in the full survey context; pretest interviews lasted about 20 minutes, on average.

Weighting and Analysis

Several stages of statistical adjustment or weighting are used to account for the complex nature of the sample design. The weights account for numerous factors including (1) the different, disproportionate probabilities of selection in each strata, (2) the overlap of the landline and cell RDD sample frames and (3) differential nonresponse associated with sample demographics.

The first stage of weighting accounts for different probabilities of selection associated with the number of adults in each household and each respondent’s telephone status.¹⁰ This weighting also adjusts for the overlapping landline and cell RDD sample frames and the relative sizes of each frame and each sample. Due to the disproportionately stratified sample design, the first-stage weight was computed separately for each stratum in each sample frame.

After the first-stage weight adjustment, two rounds of poststratification were performed using an iterative technique known as raking. The raking matches the selected demographics to parameters from the U.S. Census Bureau’s 2011 American Community Survey data.¹¹ The

¹⁰ Telephone status refers to whether respondents have only a landline telephone, only a cell phone or both kinds of telephone.

¹¹ ACS analysis was based on all adults excluding those living in institutional group quarters.

population density parameter was derived from 2010 census data. The telephone usage parameter came from an analysis of the January-June 2012 National Health Interview Survey.¹² Raking was performed separately for those asked each form of the questionnaire using Sample Balancing, a special iterative sample weighting program that simultaneously balances the distributions of all variables using a statistical technique called the Deming Algorithm. The raking corrects for differential nonresponse that is related to particular demographic characteristics of the sample. This weight ensures that the demographic characteristics of the sample closely approximate the demographic characteristics of the population.

The first round of raking was done individually for the major race/ethnicity groups. The variables matched to population parameters for Hispanic respondents were gender by age, gender by education, age by education, census region and nativity (U.S. born versus foreign born). The variables for non-Hispanic blacks were gender by age, gender by education, age by education and census region. The variables for other non-Hispanic respondents (white, other or mixed race) were gender by age, gender by education, age by education, census region and race (white race versus other/mixed race).

A final round of poststratification raking was performed on the total sample. The total sample was raked to the following demographic variables: gender by age, gender by education, age by education, census region, race/ethnicity, population density and household telephone status.

¹² See Blumberg, S.J. and J.V. Luke. December 2012. Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, January-June, 2012. National Center for Health Statistics.