

Phone Surveys

How well do respondents represent average Americans?

By David Whitlark and Michael Geurts

Changes in communication technology and societal attitudes regarding unsolicited phone calls may be introducing personality bias into phone survey samples. A recent phone survey looking at the personality profile of a nationally representative sample of adults shows that the respondents are much more confident, outgoing, conscientious, and in the case of women, more agreeable than are average Americans.

Do national random phone surveys provide a representative sample of adult Americans? We tend to talk about them as if they did. There is no significant difference in the quality of data obtained from phone surveys and personal interviews according to Seymour Sudman and Edward Blair in their 1998 marketing research textbook *Marketing Research: A Problem-Solving Approach*. In *Marketing Research Essentials*, Carl McDaniel and Roger Gates state "if proper sampling and callback procedures are employed, the telephone approach probably produces a better sample than any other survey procedure." But with completion rates now amounting to only about 30% on phone surveys, we need to test the representativeness of today's samples. Are there factors other than the commonly-used demographics of age, gender, education, ethnicity, and household income that may also determine if a sample is representative?

Large random national studies with 1,000 or more adult respondents generally match U.S. demographic statistics reasonably well. A

typical sample consists of about 51% female and 49% male; 30% high school graduates, 25% with some college experience, 20% with at least an undergraduate college degree, and other demographic categories such as income, ethnicity, marital status, age, and home ownership. But due to growing changes in societal attitudes toward privacy and emerging communication technology, we may need to revise the way we think about representative samples. Some of the key changes are:

- The prevalence of telemarketing making people less willing to engage in phone conversations with people they do not know.
- Legislation directed at limiting telemarketing activities that validates public anxiety and annoyance felt by being called on the phone by strangers.
- Widespread availability of answering machines, phone mail, and caller identification devices that send up red flag warnings such as "unavailable," "anonymous," and/or "out of area."

Antonio M. Rosario/Image Bank

In today's society, people can choose not to answer phone calls that are from unknown or unwanted sources. An important question for survey researchers is if these changes in technology and attitudes are influencing the sample quality of telephone survey data. In our study, we find evidence that random phone surveys may have a personality bias. In our data sample, people with personality traits such as conscientiousness, agreeableness, and assertiveness are over-represented, while people with personality traits related to anxiety and worry are under-represented. This article describes the nature of personality bias found in our phone survey sample and discusses how this type of bias may affect the quality and interpretation of survey results.

PERSONALITY ASSESSMENT

We use the NEO Five-Factor Inventory (NEO-FFI) personality test to assess the personality profile of phone survey respondents and compare it to national norms. The assessment tool developed by Paul Costa and Robert McCrae is described in their 1992 NEO PI-R professional manual. It is a nationally recognized and validated assessment tool based on the "big five" personality factors. Its norms, also published in the same manual, are gender-specific and based on a representative, paper-pencil personal survey of 1,000 adult Americans age 21 and over. The traits measured by the NEO-FFI have high test-retest reliability and have been shown to be stable over long periods of time.

The NEO-FFI factors are neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. The factor definers are listed in Exhibit 1. However, to help people quickly grasp the nature of the "big five" personality model, L.R. Goldberg offers a simple description in a 1981 article "Language and Individual Differences: The Search For Universals in Personality Lexicons," published in the *Review of Personality and Social Psychology*. He explains, as we meet and get to know other people we implicitly ask ourselves questions:

- Is he crazy? (neuroticism)
- Will he bully me or can I bully him? (extraversion)
- Is he smart? (openness to experience)
- Is he sympathetic and kind? (agreeableness)
- Can I count on him? (conscientiousness)

Psychologists describe the NEO-FFI as being more inclusive than the Myers-Briggs and MMP1 personality tests. It is regarded as one of the best state-of-the-art tools currently available for assessing the fundamental dimensions of normal personality.

INDICATIONS OF PERSONALITY BIAS

Comparisons between the NEO-FFI phone survey results and its national norms are listed in Exhibit 2 (see pg 16). We include mean scores, standard deviations, and sample sizes. The differences we observe between survey respondents and NEO-FFI norms are clear cut and statistically significant at the 99% confidence level.

For males, phone survey respondents score significantly lower on the neuroticism factor and significantly higher on the extraversion and conscientiousness factors when compared to national norms. More specifically, neuroticism scores at the 42nd percentile, while extraversion and conscientiousness score at the 76th and 68th percentiles. Simply, male phone survey respondents tend to be somewhat less anxious and tense while being much more assertive and thorough than the national norm.

Similar to the males, female phone survey respondents score significantly lower on the neuroticism factor and significantly higher on the extraversion and conscientiousness factors when compared to national norms. Additionally these women score higher on the agreeableness factor than do average women or men. In terms of national percentiles, female score at the 36th percentile for neuroticism, 75th percentile for extraversion, while scoring at the 70th and 69th percentiles for agreeableness and conscientiousness.

Can we accept NEO-FFI norms as representing average American men and women? If so, the results of this initial study lead us to a profile of phone survey respondents that is more confident, assertive, and thorough than average Americans. And with respect to women respondents, we are also, on average, interviewing someone that is more appreciative, forgiving, generous, and kind than the national norm. At one time or another, all of us have either said or at least thought to ourselves that survey samples are populated with three kinds of people: complainers, contributors, and those too compliant to say no to a surveyor. The results of this study suggest that "contributors" and, in the case of women, "positive contributors" may be over-represented in national random phone surveys.

Stepping back a moment, one must consider several alternative explanations for the results reported above. First, there is a possibility that NEO-FFI norms are themselves biased by a less-than-perfect sample and that our phone survey averages better represent the national norm. In our opinion, however, phone survey results are less representative of average Americans than NEO-FFI norms. NEO-FFI norms are based on in-person surveying of a random sample with more than a 90% completion rate. With the typical phone survey, about 70 out of every 100 households called are eventually contact-

ed of which 35 potential respondents hang up during the survey introduction or after the first question with another four or five quitting at mid-survey or later for questionnaires lasting 20 minutes or longer. The possibility for non-response bias is huge.

Another explanation is phone survey respondents answer the NEO-FFI items so as to appear less neurotic and more extroverted, conscientious, and agreeable to the interviewers. We believe that it is just human nature to try and make ourselves look as good as possible when taking a psychological test regardless of the survey format. Comparison of test results to a national norm should wash out the effect of wanting to appear better than we really are. We cannot find evidence of extraordinary “demand effects” associated with the NEO-FFI results. For example, the survey sponsor’s target audience scored very high on anxiety (neuroticism). This is not socially desirable, but entirely consistent with the survey sponsor’s product category.

One may also question whether the length of the survey, which on average lasted about 30 minutes, affected the sample composition or quality of answers because the NEO-FFI items appear in the last half of the survey. As mentioned above, most survey respondents are lost because they cannot be contacted or they hang up during the introduction or after the first question. These statistics are the same for short as well as long surveys. In fact, even in longer surveys, very few respondents quit at mid-survey or later, usually 5% or less of the total sample. In our survey, most respondents enjoyed answering the NEO-FFI items and mid-survey attrition was low for a long survey, i.e., less than 5% of the total sample. In addition, we did not notice a degradation of answer quality from the first half to the second half of the survey. The NEO-FFI sets guidelines for answer quality. Based on the quality guidelines less than a handful of respondents were eliminated from the sample.

THE IMPACT OF PERSONALITY ON MARKETING SURVEYS

Marketing surveys often measure memory, attitudes, as well as a person’s willingness to try new behaviors. For example, we might see the following questions on a survey designed to measure the awareness and acceptance of a new consumer product:

- Do you recall recently seeing a television advertisement for Lynn Wilson’s Low Fat Burritos?
- In your own words, what did the advertisement say?

Exhibit 1

Traits defining neo-FFI personality inventory

Factors	Traits
Neuroticism	Worrier, Inferior, Go to Pieces, Lonely, Tense, Worthless, Fearful, Angry at Treatment, Get Discouraged, Depressed, Helpless, Ashamed
Extraversion	Like People, Laugh Easily, Light Hearted, Enjoy Talking, Enjoy Action, Do Things with Others, Bursting with Energy, High Spirited, Optimist, Fast-Paced Life, Very Active, Lead Others
Openness	Like to Daydream, Find New Ways, Intrigued by Nature, Listen to Controversial Speakers, Influenced by Poetry, Try New Foods, Aware of Moods and Feelings, Not Rely on Religion, Excited by Art, Speculate about Universe, Intellectual Curiosity, Play with Theories
Agreeableness	Courteous, No Arguments, Not Egotistical, Cooperate not Compete, Not Cynical of Others, Trust People, Please People, Not Cold, Not Hard-Hearted, Considerate, Try to Show I Like Others, Not Manipulative
Conscientiousness	Clean and Neat, Things Done on Time, Methodical, Conscientious, Clear Goals, Start Work Quickly, Work Hard, Follow Through on Commitments, Always Dependable, Productive, Organized

Source: 1992 NEO PI-R Professional Manual published by Psychological Assessment Resources, Inc.

- Would you say the advertisement made you feel more favorable, less favorable, or did not change your opinion about Lynn Wilson’s Low Fat Burritos?
- On a scale from 1 to 5 where 1 means “never” and 5 means “certain,” how likely are you to purchase a package of Lynn Wilson’s Low Fat Burritos during the next three months?
- Besides the Low Fat Burritos, are you familiar with other Lynn Wilson products?

Let’s assume that women are the primary sample for such a survey because they do most of the shopping for this type of product. Generalizing from our initial findings regarding personality bias, female phone survey respondents will be more confident, assertive, thorough, generous, and kind than average female consumers. These personality differences could create several effects on the survey data.

Respondents are confident and assertive.

Few people will argue against the supposition that confident and assertive people are more willing to take risks than others. In the “diffusion of innovation” model of product adoption, these are the people that are labeled innovators, early adopters, and the early majority. Said another

Exhibit 2

Comparing phone survey respondents to national norms

Averages for adult females

Factor	Phone survey ¹ n = 622	National norm ² n = 500	Difference score	P-Value
Neuroticism	17.88	20.54	-2.66	< .01
Extraversion	32.18	28.16	4.02	< .01
Openness	27.44	26.98	0.46	< .25
Agreeableness	35.56	33.76	1.80	< .01
Conscientiousness	37.58	35.04	2.54	< .01

Averages for adult males

Factor	Phone survey n = 386	National norm n = 500	Difference score	P-Value
Neuroticism	15.64	17.60	-1.96	< .01
Extraversion	31.13	27.22	3.91	< .01
Openness	26.45	27.09	-0.64	< .15
Agreeableness	31.32	31.93	-0.61	< .15
Conscientiousness	36.48	34.10	2.38	< .01

¹Source: 1995 Wirthlin Worldwide national telephone survey of adults 21 and older.

²Source: 1992 NEO PI-R Professional Manual, Table B-4, published by Psychological Assessment Resources, Inc.

way, the risk-taking nature of confident and assertive people make them more likely to purchase and use new products than others.

As a practical result of this personality bias, phone survey "purchase likelihood" studies taken at face value are likely to overstate the market acceptance of new products. In fact, some practitioners discount "purchase likelihood" data by multiplying top-box "certain to buy" responses by 75% and other "high chance to buy," "even chance to buy," response categories by 50% or less when forecasting sales for a new product from consumer surveys. There are several descriptive studies such as one we published in the Fall 1993 issue of the *Journal of Business Forecasting* or the August 1989 *Journal of Marketing Research* article by Linda Jamieson and Frank Bass that show the gap between purchase intentions and actual purchases. While discounting the face value of purchase intention scores is frequently done, personality bias in the phone survey sample provides a good rationale for consistently following the practice.

Respondents are thorough.

Thorough people are generally considered to be more careful, meticulous, and capable than others. It is also reasonable to assume that they are more likely to pay attention to and remember advertisements than less conscientious people. When evaluating a national mass media advertising campaign, clients usually track the percentage of the general

public that notice an advertisement, associate the advertisement with the appropriate product or advertiser, and recall the key selling points of the advertisement. Considering the heightened thoroughness of our phone survey respondents compared to average Americans, marketing research providers may be overstating all three measures of advertising effectiveness. Consequently, advertising agencies and advertisers may be underestimating the media dollars it currently takes to generate a given level of public awareness and message recall.

In addition to advertising recall, this type of personality bias can impact many different forms of marketplace recall that we test in marketing surveys. Being especially thorough, phone survey respondents may be more likely to notice new products, associated and complementary products, and changes in customer service policies, pricing and/or product quality, etc., than other consumers. In each case, phone surveys may create a false impression of how aware average consumers are of the changes being made to a company's marketing mix.

Female respondents are kind and generous.

Female phone survey respondents in our survey are more supportive than the "average" female. Their personal style suggests that they will try to sense the attitude of the interviewer, pinpoint the survey objective, and identify the survey sponsor so as to be as positive and supportive as possible. Consequently, personality bias may heighten the impact of "demand effects" in some survey situations. To counteract such effects, survey writers must take special care with the wording and ordering of questions so that the underlying objectives and survey sponsor stay hidden from the respondent as long as possible.

Sometimes even a well-written survey may not get around the bias introduced by the above-average supportiveness of female phone survey respondents. For example, consider an advertising awareness and attitude tracking study. In such a survey it is standard practice to ask a respondent demonstrating "proven ad awareness" to comment on whether the advertisement makes her feel more favorable, less favorable, or no change in her opinion about the advertiser's product. Personality bias guarantees that, on average, female respondents will overstate the favorable impact of advertising campaigns.

Other types of attitude tracking studies can be affected by above-average supportiveness of female respondents. For example, it is difficult to avoid personality bias in a customer satisfaction study. In such studies, the survey sponsor may be identified at the beginning of the survey to boost completion rates. In addition, these surveys cannot hide the objective of the study, i.e., how favorable a person feels about the sponsor's products and ser-

vices. In such situations, special care must be taken to ensure that customer satisfaction ratings do not overstate the true level of market approval.

CONTINUING CHALLENGES OF REPRESENTATIVE SAMPLING

Assembling representative samples from year to year is a lot like trying to hit a moving target. As survey researchers, we must look for changes in technology as well as in societal attitudes and practices that may influence the representativeness of our samples. Currently, the challenge centers on the ease with which less confident, less outgoing, less conscientious and often less agreeable people can avoid phone surveys. As of today, it appears as if we are losing the battle. On average, for every hour an interviewer spends on the phone, 30 minutes is spent just trying to find a person that will agree to be surveyed. Consequently, even if our phone survey samples approximately match the nation on demographics, they may not match on personality and personality bias can have a significant impact on survey results.

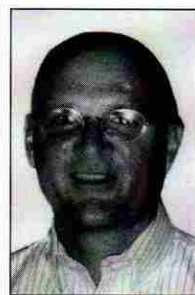
To address the issue of personality bias we recommend several options for survey research done by phone. First, inducement questions placed at the beginning of phone surveys improve response

rates. An article we published in the Summer 1994 issue of *Marketing Research* provides examples of successful inducement questions. Knowing the nature of the personality bias we are likely to find in a phone survey, researchers can write inducement questions to draw out the types of people we most often miss. For example, we could include an inducement question dealing with an issue of personal safety or privacy on which an anxious or fearful person would want to comment.

As an alternative, recruiting panels that have the right demographic and personality mix may be an effective way to deal with personality bias. When constructing a panel, researchers can assemble an appropriate demographic and personality profile by providing greater incentives to "hard-to-get" panel members. In our experience, paid respondents will be honest evaluators when asked to look at all the aspects of a product or service, both positive and negative, and when reminded that overly positive evaluations harm rather than help the survey sponsors. ■

ACKNOWLEDGMENTS

The authors want to thank the New York office of Wirthlin Worldwide and particularly Cathy Shores and Elizabeth Kehler for their contributions to this research.



David B. Whitlark is Associate Professor of Business Management, Marriott School of Management at Brigham Young University.



Michael D. Geurts is J. Earl Garrett Professor of Business Management, Marriott School of Management at Brigham Young University.

RESEARCH PROCEDURE

In 1995 Wirthlin Worldwide conducted a national random phone survey consisting of 1,008 completed interviews with men and women age 21 and over for a client interested in contrasting the general attitudes, interests, and personality profiles of adult consumers in their target market with other adult consumers. On average, the total survey sample scored significantly lower than the national norm on neuroticism and higher than the norm on extraversion and conscientiousness. In contrast, the sample respondents representing the survey sponsor's target market scored significantly higher than the national norm on neuroticism and at about the norm for each of the other personality factors.

To assess personality profiles, the 60-question NEO Five-Factor Inventory (NEO-FFI) was included as part of the survey instrument. In the NEO-FFI, respondents use a five-point scale ranging from "strongly agree" to "strongly disagree" to answer inventory questions. In a phone interview, the inventory takes approximately 15 minutes to complete. The NEO-FFI section was located in the last half of the survey instrument. Similar to most national telephone studies, respondents were not paid for participating in the survey.

To meet the guidelines set by Costa and McCrae, respondents needed to answer all 60 questions in the section to be included in the study. Also, respondents could not answer "strongly disagree" to more than 6 consecutive items, "disagree" to more than 9 consecutive items, "neutral" to more than 10 consecutive items, "agree" to more than 14 consecutive items, or "strongly agree" to more than 9 consecutive items and still be included in the study. People willing to complete the survey but saying they were either running short on time or were caught at an inconvenient time were rescheduled and called back by the Wirthlin Worldwide interviewers.

The sample for the survey was drawn to be representative of adult males and females age 21 and over. However, females were over-sampled because they were more likely than males to fall into the client's target market. Because females comprise approximately 62% of the sample rather than the national average of 51% in the study, we compare personality profiles and draw conclusions for each gender separately rather than for the adult population as a whole.

We use two-tail t-tests to evaluate the statistical significance of the survey results because prior to the study we could not pre-

dict whether the phone respondent NEO-FFI scores would be higher or lower than national norms. In the study, differences in mean scores need to be significant at the 90% confidence level or higher in order to rule out sampling error as the reason for observing differences between phone respondents and NEO-FFI national norms.

For males, differences are statistically significant in excess of the 99% confidence level for neuroticism, extraversion, and conscientiousness. Differences for openness and agreeableness are insignificant. For females, differences are statistically significant in excess of the 99% confidence level for neuroticism, extraversion, conscientiousness, and agreeableness. The difference in mean scores for openness is insignificant.

The NEO-FFI is a self-administered instrument only requiring a sixth-grade reading level to complete. According to Costa and McCrae, individuals not having formal training in clinical psychology, personality, or related fields may administer and score the test. Professionals wishing to include the NEO-FFI in a survey instrument must first obtain a formal, written licensing agreement from Psychological Assessment Resources, Inc., publisher of the manual.