

Social Media for Recruitment Study

Supporting Statement – Section B

January 20, 2021

Program Official/Project Officer

Juan Rodriguez, MPH, MS
Epidemiologist
Division of Cancer Prevention and Control
National Center for Chronic Disease Prevention and Health Promotion
Centers for Disease Control and Prevention
4770 Buford Highway NE, MS F-76
Atlanta, GA 30341
770-488-3086
fph4@cdc.gov

Table of Contents

| | |
|--|----|
| Section B – Data Collection Procedures..... | 5 |
| B1. Respondent Universe and Sampling Methods..... | 5 |
| B2. Procedures for the Collection of Information..... | 7 |
| 2.1 Data Collection Procedures..... | 7 |
| 2.2 Estimation Procedure..... | 9 |
| B3. Methods to Maximize Response and Deal with Nonresponse..... | 12 |
| B4. Test of Procedures or Methods to be Undertaken..... | 13 |
| B5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data..... | 14 |

1LIST OF ATTACHMENTS

Attachment 1 – Section 301 of the Public Health Service Act

Attachment 2 – 60-Day FRN

Attachment 3 – Data Collection Forms

Attachment 3a - General Population Survey

Attachment 3b – Cancer Survivor Survey

Attachment 3c – High-Risk Survey

Attachment 3d – High-Risk Follow-Up Survey

Attachment 4 – Web Recruitment Ads

Attachment 5 – Informed Consent

Attachment 5a – Consent & Landing Page

Attachment 5b – Frequently Asked Questions

Attachment 5c – Follow Up Consent & Landing Page

Attachment 6 – Screeners

Attachment 6a – Baseline Screener

Attachment 6b – Follow Up Screener

Attachment 7 – Email Invitation

Attachment 8 – Data Collection Screenshots

Attachment 8a – Baseline and Follow-Up Screener Web Screenshots

Attachment 8b – General Population Survey Web Screenshots

Attachment 8c – Cancer Survivor Survey Web Screenshots

Attachment 8d – High-Risk Survey Web Screenshots

Attachment 8e – High-Risk Follow-Up Survey Web Screenshots

Attachment 9 – Human Subjects Approval

Attachment 9a – NORC IRB Approval Renewal

Attachment 9b – CDC IRB Reliance Letter

Attachment 9c – CDC IRB Authorization Agreement

Attachment 10 – 60-day Federal Register Notice and Comments

Attachment 10a – Published 60-day FRN

Attachment 10b – Comments Received from 60-day FRN

Attachment 10c – Response to Comments

Attachment 11 – Signed Privacy Impact Assessment

Section B – Data Collection Procedures

B1. Respondent Universe and Sampling Methods

Respondent Universe

Using social media, this research study will enroll adults living in the United States across three populations: women aged 40 and older, adults 18 and older who have been diagnosed with cancer, and adults 18 and older at high risk for cancer. Individuals will be recruited through paid ads on Facebook, Twitter, and Google. Targeted ads will use the sites' filtering abilities to show ads only to those who are most likely to be eligible; these filters may include factors such as age, gender, and social media profile.

Sampling Methods

The sample will be a non-probability based, targeted sample. Survey recruitment will continue for six months, or until 500-750 respondents (depending on the target population) for each survey are recruited, whichever comes first. Respondents will be recruited via social media and other internet sites, including Facebook, Twitter, and Google. Individuals meeting the filter criteria (age, gender, social media profile) who visit one of the aforementioned sites will be eligible to view the ads (**Attachment 4**). Ads will be restricted or targeted based on the profile of the viewer, using tools made available by the advertisement vendors.

Potential respondents who click on an ad will be routed to the survey landing page which will explain the purpose of the study and include consent language (**Attachments 5a & 5c**).

Respondents will consent by clicking on the "Continue" button on the survey landing page, where they will be taken to the web survey screener.

Table 1 illustrates our expected respondent sample size by race/ethnicity for Whites, Blacks and Hispanics. Recruitment vendor targeting may be used to display ads more frequently to minority groups if needed to permit sub-group analysis. If by the midpoint in data collection, a lower than expected proportion of respondents identify as Black or Hispanic, resources can be reallocated toward those ads and sites that are more productive at recruiting respondents that belong to a minority group.

Table 1: Projected Sample Size by Survey and Race/Ethnicity (n=4,000)

| Survey | Eligible Respondents (Planned) ^a | Estimated number of eligible respondents who complete the survey | | | |
|-------------------------------|---|--|------------|------------|------------|
| | | White | Black | Hispanic | Other |
| Women aged 40 and older | 500 | 290 | 70 | 90 | 50 |
| Cancer Survivors ^c | 750 | 613 | 65 | 45 | 27 |
| High-Risk | 750 | 613 | 65 | 45 | 27 |
| High-Risk Follow-Up | 750 | 613 | 65 | 45 | 27 |
| Total | 2,750 | 2,557 | 301 | 261 | 131 |

^a Total projected sample size is based on expected response size of 500-750 respondents per survey. Respondents may be eligible for both the high-risk and high-risk follow-up survey, though new respondents may also be recruited for the follow-up survey using social media for recruitment.

^b Distribution by race/ethnicity was based on a weighted average of the adult population in the US according to the V2015 Census Bureau data (<https://www.census.gov/quickfacts/table/PST045216/00#headnote-js-a>) and social media user demographics (<http://www.pewinternet.org/fact-sheet/social-media/>).

^c Distribution by race/ethnicity was based on a weighted average of adult cancer survivors (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3697294/>) and social media user demographics (<http://www.pewinternet.org/fact-sheet/social-media/>).

Each advertisement will be closely monitored to determine which ads are performing best in terms of respondent clicks and survey completion rates. Based on monitoring results, ads that are not performing well can be shut down to transfer resources to the content that is performing best. In addition, the project will analyze the effectiveness of the landing page and survey instrument to determine if there are any points where potential or engaged respondents tend to drop out of the survey. Comparing ‘clicks’ to survey completes will show if respondents are frequently exiting at the landing page before even beginning the survey or if they are breaking off mid-way through survey completion. This will help to identify if changes to survey length or landing page

content would be more successful at improving completion rates, either overall or to improve representativeness across certain groups.

B2. Procedures for the Collection of Information

The *women aged 40 and older* survey covers topics related to mammography use and beliefs about breast cancer. The *survivorship* survey covers topics related to general health and well-being post treatment. The *high risk* survey covers topics related to risk communication among family members. The *high-risk follow-up* survey covers topics related to respondent perspectives on tools and resources for communication of family genetic cancer risk. The data collection instruments are included with this submission as **Attachments 3a-3d**.

Specifically, the women aged 40 and over survey covers: (1) utilization and practices around mammography, (2) information seeking behavior, (3) access to health care, and (4) beliefs and attitudes associated with breast cancer and cancer prevention. The survivorship survey includes questions on: (1) cancer diagnosis and treatment, (2) physical, mental, and emotional health post-cancer, (3) the impact of cancer on employment, and (4) financial well-being related to cancer and cancer treatment. The high risk survey covers: (1) reasons for genetic testing, (2) reasons for sharing or not sharing test results with family, (3) mechanisms for sharing results with family, (4) sources of cancer information, (5) patterns of communication with family, and (6) resources for sharing information with family. The high-risk follow-up survey asks respondents about (1) factors influencing their decision to get tested, (2) satisfaction with the testing process, (3) how they were informed of test results, (4) reasons for sharing test results with family members, and (5) ways to improve process for receiving test results.

2.1 Data Collection Procedures

Respondents to the **women aged 40 and over, survivor, and high-risk surveys** will be recruited via ads posted on Facebook, Twitter, and Google. Various ads will be posted using different strategic targeting options on the listed social media and search engine sites in order to assess the effectiveness of each site, ad, and recruitment strategy.

Adults 18 and older who visit one of the three sites (Facebook, Twitter, Google) will be eligible to view the sample ads (**Attachment 4**). Ads will also be targeted to women aged 40 and older, and those who have liked or follow pages associated with BRCA mutations or cancer survivorship. Interested individuals who click on an ad will be routed to the survey landing page which will explain the purpose of the study and include consent language (**Attachments 5a & 5c**). To determine eligibility of interested respondents, the web survey includes a brief screener to assess eligibility as well as determine skip patterns for the questionnaire (**Attachment 6a**). After screening for eligibility, the respondent will continue to the full web survey appropriate for the individual based on their responses to the screener instrument. Respondents to the women aged 40 and over survey and the high-risk surveys (**Attachments 3a and 3c**) will be asked varying questions depending on cancer screening recommendations for their age and sex. The survivorship survey (**Attachment 3b**) also depends upon screener-based skip logic.

Respondents to the **follow-up survey** will be recruited via email invitations to eligible participants who completed the high-risk survey and agreed to be contacted for a second survey, as well as new individuals recruited via social media (Facebook, Twitter) and Google, if needed. Prior respondents will be sent a personalized email inviting them to participate in the survey, with a unique link to the web survey (**Attachment 7**). If necessary to meet sample size goals, additional respondents will be recruited via advertisements placed on social media and other internet sites such as Facebook, Twitter, and Google, following the same procedure as described above for the original survey. Targeting may be used so that individuals who “like” or “follow” certain pages are more likely to be shown the ads (**Attachment 4**). Ads that were productive in recruiting eligible respondents for the high-risk survey will be used again in recruiting the follow-up sample. Interested respondents who click on an ad will be routed to the survey landing page which will explain the purpose of the study and include assent/consent language (**Attachments 5a & 5c**). To determine eligibility of interested respondents, the web survey includes a brief screener to assess eligibility (**Attachment 6b**). Upon completion of the screener, eligible individuals will continue to complete the high-risk follow-up survey (**Attachment 3d**). Individuals recruited via email (i.e., who participated in the original high risk survey) will not be required to complete the screener and will continue directly to the high-risk follow-up survey.

2.2 Estimation Procedure

Survey Weights

Given that this is formative research and respondents are being recruited from the Internet and social media platforms to develop a quota sample, no survey weights will be used. Respondent data will be analyzed without survey weights and will be interpreted with caution. This data is not intended to be representative of any of the populations of interest or to be used to draw inferences about any of these groups more broadly. The results will only be representative of the opinions and experiences of those recruited and not generalizable beyond the study sample.

Power Analysis

Table 1 presents ranges of acceptable error rates and corresponding sample sizes for 90% and 95% confidence intervals. In order to arrive at the model based on 500 to 750 completes, the acceptable amount of error was set at 90% confidence ($\pm .03/3\%$). As seen highlighted in Table B2-1, this calculation means that a minimum of 752 completed surveys are needed for a 90% confidence level. Therefore, with 752 or more completed surveys, statistical power and confidence can be increased, and it may be possible to analyze sub-groups. While the proposed sample size of the survey targeting women over 40 is below this number (500 vs. 752), the estimate of 752 is based on 10% of the population, which also includes men. In addition, given that we do not expect survey results to be used to make inferences or generalizable to the population, we are not expressly concerned about being underpowered. In addition, we do not expect to conduct any significant sub-group analyses. The sample size for the high risk and cancer survivor surveys are set at 750 in order to approximate the 90% confidence estimated sample size and to ensure that there are sufficient individuals recruited that had varying experiences with cancer treatment and genetic testing so as to capture details on an array of experiences. This will allow us to better inform future studies and intervention materials.

The estimates in Table B2-1 are based on the following assumptions:

1. The statistic of interest is the number of respondents who provide a response to a yet-to-be-determined variable of interest.

2. Non-probability sampling will be used to recruit respondents through the placement of advertisements on social media sites (e.g., Facebook, Twitter, or other site) and search engines (e.g., Google).
3. Based on extant research, respondents who enter the survey will respond and we anticipate ending with 752 completes (Stern, Bilgen, McClain, & Hunscher, 2017).
4. Probability statements refer to the entire population to the extent they are known.
5. The proportion of individuals selecting a given response will be .5 (50%). This assumption provides the most conservative (largest) sample size estimates.

Table B2-1. Sample Size Requirements for SMFR

| <u>Sample Sizes for Single Proportions</u> | | <u>Confidence Interval for Proportions</u> | |
|--|----------------|--|-----------|
| Population Size: | 1,222,438 * | Population Size: | 1,222,439 |
| Acceptable Error (+/-): | 0.03 | Sample Size: | 1067 |

Specific Table for Pop. = 1222438.86 and Error of +/- 0.03

Level of Confidence

| Acceptable Error +/- | 90% | 95% |
|-----------------------------|------------------|------------------|
| | Sample n= | Sample n= |
| 0.03 | 752 | 1067 |

General Table for Population Size = 1222438.86

Level of Confidence

| Acceptable Error +/- | 90% Sample n= | 95% Sample n= |
|-----------------------------|----------------------|----------------------|
| 0.01 | 6727 | 9530 |
| 0.02 | 1690 | 2397 |
| 0.03 | 752 | 1067 |
| 0.04 | 424 | 601 |
| 0.05 | 271 | 385 |
| 0.06 | 189 | 268 |
| 0.07 | 139 | 197 |
| 0.08 | 107 | 151 |
| 0.09 | 84 | 120 |
| 0.10 | 69 | 97 |
| 0.12 | 48 | 68 |
| 0.15 | 31 | 44 |

*Based on 10% of the population, thus, conservative to those on social media

B3. Methods to Maximize Response and Deal with Nonresponse

The following procedures will be used to maximize cooperation and achieve the desired participation rates:

- Respondents who break off during the survey will be prompted to complete the survey via an e-mail reminder sent one to two weeks after survey initiation.
- A \$5 Amazon gift code will be offered to participants who complete the web survey. A review of 49 experiments has shown that incentives significantly increase the odds of survey completion (Edwards et al., 2007) and this has lately been proven to be critical

when seeking to recruit for a survey from social media sites and search engines (Stern, Bilgen, McClain, & Hunscher, 2017).

NORC will provide a toll-free telephone number for the CDC COR to address specific questions or obtain clarifications on the study purpose, as well as for the NORC IRB hotline should participants have any questions about their rights as study participants. In addition, an email address and 800-line will be provided should participants encounter technical issues with the web survey.

All participants will be routed to federal web resources related to cancer risk, where they may find additional information to answer questions or seek help. This text displayed below in Exhibit B3-1 is well-established as a best practice in the field of web surveys (Dillman, Smyth, & Christian, 2014).

Exhibit B3-1

If you would like more information about genetic testing for cancer risk, please visit the following resources:

Bring Your Brave (BRCA testing)

https://www.cdc.gov/cancer/breast/young_women/bringyourbrave/

Know: BRCA <https://www.knowbrca.org/>

Talking to family members <http://kintalk.org/>

NCI Cancer Genetics Services Directory <https://www.cancer.gov/about-cancer/causes-prevention/genetics/directory>

B4. Test of Procedures or Methods to be Undertaken

The high-risk questionnaire was cognitively tested with nine English-speaking men and women at high risk for cancer, all of whom were living and/or working in the Chicago metropolitan area. The cognitive testing assessed clarity, quality, and usability of the study materials (**Attachments 3a-3d, 5a-5c, 6a-6b**) and was used as a tool to estimate time burden associated with completing the survey and other study materials including the screener and consent. Feedback from cognitive testing was incorporated into the final version of the study materials. The estimates of time burden presented in Part A of the Supporting Statement were generated from these cognitive testing results.

Some of the items included in the women aged 40 and over and survivorship surveys were taken from validated, national surveys such as NHIS and HINTS (Nelson et al., 2004; Willson, 2014). Some were developed specifically for this study or are modifications of existing scales or questions. However, in the high-risk surveys the questions are 1) drawn directly from the two source surveys above; 2) adaptations of questions from those surveys; or 3) new measure developed for this work tested through cognitive interviewing.

As stated above, respondents will be recruited via social media and other internet sites, including Facebook, Twitter, and Google. Individuals meeting the determined criteria (age, gender, social media profile) who visit one of the aforementioned sites will be shown the ads and have the option of entering the survey. These methods are established (Stern, Bilgen, McClain, & Hunscher, 2017) but this will be the one of the few studies assessing their use in the area of cancer prevention and control.

In terms of survey design, we are using appropriate visual design in order to reduce measurement error to the degree possible through appropriate spacing and groups of survey items and response options (Dillman, Smyth, & Christian, 2014; Stern, Bilgen, & Sterrett, 2016). In addition to design, we have used the best science available that addresses the placement of sensitive questions at appropriate points in the survey and in proper context. (Bradburn, Stern, & Johnson, forthcoming).

B5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

Juan Rodriguez, MPH, MS, of the Division of Cancer Prevention and Control, is the Principal Investigator and Technical Monitor for the study, and has overall responsibility for overseeing the design, conduct, and analysis of the study. He will approve and receive all contract deliverables. In addition, he, along with Lucy Peipins, an epidemiologist at CDC, will be analyzing data from this collection effort. Telephone: 770-488-3086.

The survey instrument, sampling and data collection procedures, and analysis plan were designed in collaboration with researchers at NORC at the University of Chicago. NORC will oversee the data collection.

Michael Stern, PhD, has overall technical and financial responsibility for the study at NORC and led the NORC effort to design this protocol. Dr. Stern will direct the overall data collection and participate with CDC investigators in analysis efforts. Telephone: 312-357-3891.

Other personnel involved in design of the protocol and data collection instruments are:

CDC

Lucy Peipins, PhD
Centers for Disease Control and Prevention
Epidemiology and Applied Research Branch
Division of Cancer Prevention and Control
Atlanta, GA 30341
lbp6@cdc.gov
770-488-3034

NORC

Melissa Heim Viox, MPH
3520 Piedmont Rd NE, Suite 225
Atlanta, GA 30305
heimviox-melissa@norc.org
404-240-8412

Erin Fordyce, MA
55 E Monroe, Floor 30
Chicago, IL 60603
fordyce-erin@norc.org
312-357-7011

Ipek Bilgen, PhD
55 E Monroe, Floor 30
Chicago, IL 60603
bilgen-ipek@norc.org
312-357-3874

Mayo Clinic College of Medicine

Lila Rutten, PhD
200 First Street SW
Rochester, MN 55905
rutten-lila@mayo.org
507-293-2341

REFERENCES

- Andridge, R. R., & Little, R. J. A. (2010). A Review of Hot Deck Imputation for Survey Non-response. *International Statistical Review = Revue Internationale de Statistique*, 78(1), 40–64. doi:10.1111/j.1751-5823.2010.00103.x
- Bradburn, N., Stern, M. J., & Johnson, T. (Forthcoming) *Asking Questions: The Definitive Guide to Questionnaire Design—For Market Research, Political Polls and Social and Health Questionnaires 3rd Edition*. San Francisco: Josey-Bass.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72. doi: 10.1111/j.1467-6419.2007.00527.x.
- Dillman, D., Smyth, J., & Christian, M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. Hoboken, NJ: Wiley.
- Edwards, P., Roberts, I., Clarke, M., DiGuseppi, C., Pratap, S., Wentz, R., Cooper, R. (2007). Methods to increase response rates to postal questionnaires. *Cochrane Database of Systematic Reviews*, 2, MR000008.
- Leyden, W. A., Manos, M. M., Geiger, A. M., Weinmann, S., Mouchawar, J., Bischoff, K., Yood, M. U., Gilbert, J., Taplin, S. H. (2005). Cervical cancer in women with comprehensive health care access: attributable factors in the screening process. *Journal of the National Cancer Institute*, 97(9), 675-683.
- Nelson, D. E., Kreps, G. L., Hesse, B. W., Croyle, R. T., Willis, G., Arora, N. K., Rimer, B. K., Viswanath, K. V., Weinstein, N., Alden, S. (2004). The Health Information National Trend Survey (HINTS): development, design, and dissemination. *Journal of Health Communication* 9, 443-460.
- SAS Institute. (2012). *SAS/STAT 9.3 User's Guide*. SAS Institute: Cary, NC.
- Stern, M., Bilgen, I., McClain, C., & Hunscher, B. (2017). "Can We Effectively Sample From Social Media Sites? Results from Two Sampling Experiments." *Social Science and Computer Review* 35(6), 713-732.
- Stern, M., Bilgen, I., & Sterrett, D. (2016). "The Effects of Grids on Web Surveys Completed with Mobile Devices." *Social Currents*, 3(3): 217-233.
- Willson, S. (2014). *Cognitive Interview Evaluation of the 2015 NHIS Cancer Control Supplement*. National Center for Health Statistics. Online: https://wwwn.cdc.gov/qbank/report/Willson_2014_NCHS_NHIS.pdf.