Center for Social Development

GEORGE WARREN BROWN SCHOOL OF SOCIAL WORK

Survey Response in a Statewide Social Experiment

Differences in Being Located and Collaborating by Race and Hispanic Origin

Yunju Nam
University at Buffalo, the State University of New York
School of Social Work

Lisa Reyes Mason
Center for Social Development

Youngmi Kim
Center for Social Development

Margaret Clancy
Center for Social Development

Michael Sherraden
Center for Social Development

Subsequent publication: Nam, Y., Mason, L. R., Kim, Y., Clancy, M., & Sherraden, M. (2013). Survey response in a statewide social experiment: Differences in being located and collaborating, by race and Hispanic origin. *Social Work Research, 37*(1), 64–74. doi:10.1093/swr/svs031

2011

CSD Working Papers
No. 11-01

Campus Box 1196 One Brookings Drive St. Louis, MO 63130-9906 • (314) 935.7433 • csd.wustl.edu

Acknowledgments

Support for this research comes from the Ford Foundation and Charles Stewart Mott Foundation. The authors thank Oklahoma State Treasurer Scott Meacham; Tim Allen, Deputy Treasurer for Policy and Administration; and James Wilbanks, former Director of Revenue and Fiscal Policy, who facilitated this research. We also appreciate contributions by staff at the Oklahoma State Department of Health—Kelly Baker and Sue Mallonee—and the survey research team at RTI International—Ellen Marks, Bryan Rhodes, and Jun Liu. We are grateful to Robert Zager for work with data management.

Survey Response in a Statewide Social Experiment: Differences in Being Located and Collaborating by Race and Hispanic Origin

This study examines whether and how survey response differs by race and Hispanic origin, using data from birth certificates and survey administrative data from a large-scale statewide experiment. The sample consists of mothers of infants selected from Oklahoma birth certificates using a stratified random sampling method (N=7,111). This study uses Heckman probit analysis to consider two stages of survey response: (1) being located by the survey team and (2) completing a questionnaire through collaboration with the survey team. Analysis results show that African Americans, American Indians, and Hispanics are significantly less likely to be located during the study recruitment than Whites, controlling for other demographic, socioeconomic, and environmental factors. Conditional on being located, however, the probability of collaboration does not differ among the four groups. Findings suggest that researchers should pay attention to separate stages of respondent recruitment and improve strategies to locate members of racial and ethnic minority groups during recruitment.

Key words: *Survey Research, Response Rate, Noncontact, Nonresponse, Race, Hispanic origin*

Racial and ethnic inequality in socioeconomic status and health is a serious challenge in the United States. Whites have, on average, advantaged positions over minority groups in educational achievement, employment status, income, wealth, and health. College-educated individuals represent 55% of non-Hispanic Whites but only 43% of African Americans, 42% of American Indians, and 30% of Hispanics (Bauman & Graf, 2003). Similarly, median household income is $54,920 among non-Hispanic Whites, $33,916 among African Americans, and $38,679 among Hispanics (DeNavas-Walt, Proctor, & Smith, 2008). The wealth gap is estimated to be even greater: median net worth of Whites is assessed at more than ten times that of non-Whites (Oliver & Shapiro, 1995; Shapiro, 2004). In terms of wellness indicators, racial and ethnic minority groups are estimated to have poorer health than Whites but to have more limited access to needed medical and preventive health care (Liao et al., 2004). Despite huge gaps in socioeconomic achievement, health, and other well-being measures, understanding of the causes and solutions to these disparities is limited (Yancey, Ortega, & Kumanyika, 2006).

One barrier to understanding racial and ethnic disparities is a lack of representative data for each racial and ethnic group. Knowledge and experience in achieving representative samples, recruiting survey respondents, and collecting accurate information from study target populations have expanded in recent decades (Baines et al., 2007; DeLeeuw et al., 2007; Groves, 2006; Groves & Couper, 1998), but strategies and tactics are still far from perfect and should be developed further to obtain more reliable information from samples of racial and ethnic minority groups (Yancey, Ortega, & Kumanyika, 2006). One potential obstacle to collecting reliable and valid data is unequal survey response rates among racial and ethnic groups. If response rates are consistently low among certain racial and ethnic groups, it may be hard to generate data representative of these groups (Yancey, Ortega, & Kumanyika, 2006).

Little is known about survey response among racial and ethnic groups. Existing empirical research provides inconsistent results on whether and how response rates differ across these groups (Galea & Tracy, 2007; Johnson et al., 2002). We also have limited knowledge on what affects minority groups' participation in scientific research. For example, we know little about whether response rates differ by race and ethnicity at certain stages of study recruitment: selection into a sampling frame, location by research team, or decision to participate (Buchwald et al., 2006). Furthermore, existing studies on racial and ethnic differences in survey response employ primarily bivariate analyses, and only a small number employ analytical approaches that control for covariates (Galea & Tracy, 2007; Groves & Couper, 1998). Also, the small number of existing studies focus mainly on difference in response rates between Whites and African Americans, while less attention has been paid to other groups, such as Hispanics, American Indians, and Asian Americans (Buchwald et al., 2006; Des Jarlais et al., 2005;Johnson et al., 2002).

To shed new light on survey response by race and ethnicity, this study takes advantage of data from a statewide social experiment, SEED for Oklahoma Kids (SEED OK hereafter). This study is unique in that the sample was drawn from a sampling frame with information on demographic and social status—2007 birth certificate data provided by the Oklahoma State Department of Health (OSDH). The design allows us to conduct statistical analysis with several possible predictors of survey response, including race or Hispanic origin. In other words, the sampling frame enables us to use analytical methods to compare characteristics between respondents and non-respondents based on information from birth certificates (Oropesa, 2002; Groves 2006).

Moreover, we separate the survey response process into two stages: (1) being located by the survey team, and (2) completing the questionnaire through collaborating with the survey team. In this way, we distinguish the effects of race and Hispanic origin on being located by the survey team from those on collaborating with the survey team and completing the questionnaire (conditional on being located).

Background

Reliable data from representative samples of minority groups are necessary for expanding our understanding of socioeconomic and health disparities by race and ethnicity. However, knowledge is limited on how to generate representative samples from these under-represented groups (Yancey, Ortega, & Kumanyika, 2006). In fact, we do not have a definite answer even to the question of whether response rates differ across various racial and ethnic groups since existing empirical research provides inconsistent results (Galea & Tracy, 2007; Johnson et al., 2002). While some studies have found that minorities are less likely to participate in scientific research (Kim et al., 2008; Yancey, Ortega, & Kumanyika, 2006; Zaslavsky, Zaborski, & Cleary, 2002), others have found no statistical differences among groups (Groves, 2006), and the rest have shown that minority groups' response rates are higher than those of Whites (Groves & Couper, 1998).

We know little about why study participation rates differ between Whites and non-Whites and what affects minority group members' decisions about study participation, although previous studies have identified three factors that may reduce response rates among racial and ethnic minority groups. First, disparities in socioeconomic status may contribute to different response rates by race and ethnicity. It is well documented that those with higher socioeconomic status (e.g., highly-educated and high-income individuals) are more likely to participate in scientific research than those with

socioeconomic disadvantages (Galea & Tracy, 2007; Groves, Cialdini, & Couper, 1992; Groves & Couper, 1998). Because members of racial and ethnic minority groups tend to have lower levels of education and income than Whites, these differences may result in lower survey response rates.

Second, difficulty in locating or contacting members of minority groups is a challenge to recruitment. Some minority groups' geographic mobility rates are high, which results in frequent changes in addresses and telephone numbers (Johnson et al., 2002; Oropesa, 2002). For example, low-income renters and migrant farm workers, disproportionately members of minority groups, move more frequently than other populations. In addition, non-Whites are less likely to be contacted through traditional channels of recruitment. American Indians on reservations have a lower level of accessibility to telephone services than other groups (Kim et al., 2008). Low-income minority individuals are less likely to have telephones and more likely to work at odd hours, which reduce their chances of being selected into a sample. Kim and colleagues' (2008) descriptive analysis shows that American Indians' response rate is lower than non-Hispanic Whites because the former's chance of being located is much lower than that of the latter (70% versus 89%). Collaboration rates for those who were located are similar between the two groups (90% versus 92%) (Kim et al., 2008). Similarly, Zaslavsky and colleagues (2002) show that African Americans and Hispanics are less likely to be located than Whites. Their logistic regression also indicates that these two minority groups' rates of collaboration are significantly lower than that of Whites, among those who were located.

Third, mistrust and skepticism of research institutions may reduce minority groups' survey response rates, especially when government agencies and universities conduct research. Individuals are more willing to participate in a study when they perceive research institutions as a legitimate authority (Groves, Cialdini, & Couper, 1992). Members of minority groups are less likely to accept mainstream research institutions (e.g., government agencies, hospitals, and higher education institutions) as legitimate, however, because of historical and personal experiences of discriminatory treatment and misuse of research (e.g., Tuskegee Syphilis Study) (Buchwald et al., 2006; Fouad et al., 2000; Johnson et al., 2002; Oropesa, 2002; Yancey, Ortega, & Kumanyika, 2006).

Existing literature identifies one factor that may encourage members of minority groups to participate in scientific research: potential benefits to their family or community (Fouad et al., 2000). A study using vignettes that describe different types of research projects found that American Indians and Alaska Natives are more willing to participate in studies on diabetes and alcoholism, two major health concerns among these groups, than in other clinical studies. This finding suggests that these minority groups are more likely to participate in studies that address their communities' salient issues (Buchwald et al., 2006).

As valuable as they are, existing studies are not free from limitations. With the exception of a small but growing number of recent studies, most have examined bivariate associations (Groves & Couper, 1998;Yancey, Ortega, & Kumanyika, 2006). In addition, the majority of existing studies report only final response rates. Only a small number of studies (Groves & Couper, 1998; Kim et al., 2008; Zaslavsky, Zaborski, & Cleary, 2002) distinguish sample members who were not located from those who were located but did not collaborate with the survey team. For this reason, we cannot say when in the recruitment process differences among racial and ethnic groups occur. Furthermore, previous studies mostly focus on differences between Whites and African Americans, leaving other racial and ethnic minorities understudied (Buchwald et al., 2006; Des Jarlais et al., 2005; Johnson et al., 2002). It is likely that African Americans' decision process for survey response differs not only

from that of Whites, but also from those of other racial and ethnic minority groups. Accordingly, it is imperative to expand our understanding of survey response among Hispanics, American Indians, and Asians as well as various subgroups within these groups.

To fill gaps in existing knowledge, we examine survey response in the SEED OK baseline survey among African Americans, American Indians, Hispanics, and Whites. Using data from Oklahoma birth certificates and the SEED OK baseline survey, we first examine descriptive characteristics and response status by race and Hispanic origin. Next, we conduct Heckman probit analyses to see whether and how the likelihood of participating in SEED OK differs by race and ethnicity. These analyses separate sample members' probability of being located by the SEED OK survey team from their probability of collaborating with the baseline survey on the condition of being located.

Study Setting: SEED for Oklahoma Kids

SEED OK is a longitudinal social experiment that tests the policy concept of Child Development Accounts (CDAs) offered at birth.[1] The experiment used both random selection and random assignment into treatment and control groups. For infants in the treatment group only, an Oklahoma College Savings Plan (OCSP) account was automatically opened, and SEED OK made a $1,000 "seed" deposit into each of these accounts. Saving matches are also available to treatment infants from low- or moderate-income families for the first four years of the experiment (Zager et al., 2010).

The SEED OK study is unique in that its sampling frame was birth certificate data of all infants born in the state of Oklahoma in certain periods. The SEED OK survey team received from OSDH all birth certificate data of infants born in the state between April and June 2007 (stage 1). Due to a lower-than-expected response rate in the first stage of data collection, the SEED OK survey team added a second sample consisting of all infants born between August and October 2007 (stage 2). By using birth certificates, the SEED OK study obtained data for every infant born in the state of Oklahoma during these two time periods. Accordingly, SEED OK successfully acquired a sampling frame representative of its target population, the first step in constructing a representative sample for the study (Zager et al., 2010).

From the sampling frame of birth certificates, the SEED OK survey team used stratified random sampling to select infants for the study. Three minority groups were oversampled: African Americans, American Indians, and Hispanics. By oversampling, the SEED OK study aimed to create samples of these three minority groups sizeable enough for separate analyses by race and Hispanic origin (Marks, Rhodes, & Scheffler, 2008).

Recruitment and data collection occurred between August and December 2007 for the stage 1 sample and between January and April 2008 for the stage 2 sample. Mothers of the infants were recruited to join the SEED OK study.[2] SEED OK had more generous participation incentives than many studies, which may have motivated potential respondents to collaborate with the survey team: Mothers who joined the study and completed the baseline survey had a 50-50 chance of being

[1] SEED OK was designed by Michael Sherraden, other researchers at the Center for Social Development, and the SEED Research Advisory Council. Sampling, recruitment, and survey implementation were conducted by RTI International.
[2] In cases where the mother did not live with the infant in the sample, the main caretaker of the infant was asked to participate in the study.

assigned to the treatment group, and thus of receiving an OCSP account with a $1,000 deposit for their infants. In addition, SEED OK offered a $40 incentive to all individuals (both treatment and control group members) who completed the baseline survey. At the same time, the SEED OK study required the survey respondents to provide their infant's Social Security Number so that an OCSP account could be opened for the child. This requirement may have reduced survey collaboration rates because some individuals may have felt uncomfortable providing this confidential information (Marks, Rhodes, & Scheffler, 2008).

Since birth certificate data include names and addresses but not telephone numbers, the SEED OK researchers used various methods to contact participants and obtain phone numbers for interviews. First, an attempt was made to match OSDH birth certificate data to up-to-date telephone numbers and addresses, using automated databases. Second, an invitation letter, signed by Oklahoma State Treasurer Scott Meacham and sent by RTI International (the research firm implementing the baseline survey), briefly introduced the study to sample members and encouraged them to call the project's toll-free telephone number or visit the project's website.[3] Third, professional tracers used commercial sources and contacted possible relatives and neighbors of sample members. Fourth, a field representative visited the best known address of the sample members not located through other methods to explain the study and encourage participation (Marks, Rhodes, & Scheffler, 2008).

Methods

Data and Sample

This study uses data from three different sources: (1) Oklahoma birth certificates; (2) SEED OK administrative data on the baseline survey recruitment process; and (3) geographic data based on the U.S. Census Bureau's zip code classification system. Out of 7,297 cases in the sample originally selected from birth certificates, 182 ineligible cases (e.g., twin births and deaths of infants or mothers) are excluded. We also exclude four cases with information missing on child's birth weight, mother's marital status, and mother's birth place, three variables included in Heckman probit analysis. The final analysis sample consists of 7,111 cases.

Measures

The dependent variable is survey response status. We created this variable using administrative data that recorded the recruitment process of the SEED OK baseline survey. Since those who were not located may differ from those who were located but did not collaborate fully with the baseline survey interview, we created three categories for the dependent variable: (1) not located (i.e., unlocated cases); (2) located but did not complete the survey by refusing to participate or breaking off the interview (i.e., located non-collaborators) and (3) located and completed the survey questionnaire (i.e., respondents or located collaborators).

The independent variable is the race and Hispanic origin of infants from birth certificates. We use infant's race and Hispanic origin variable created based on the National Center for Health Statistics'

[3] Before SEED OK interviewers contacted sample members by phone for interviews, a letter and brochure invited them to participate in the study and detailed the study requirements. A professionally-designed "frequently asked questions" brochure addressed potential questions about the study and outlined participation requirements and confidentiality measures. The letter and brochure were also posted in English and Spanish on a simple study website.

Vital Statistics protocol (Buescher, Gizlice, & Jones-Vessey, 2005; Marks, Rhodes, & Scheffler, 2008). If the mother was reported as Hispanic, the infant was classified as Hispanic. If information regarding the mother's Hispanic origin was missing and the father was recorded as Hispanic, the infant was also categorized as Hispanic. The SEED OK sample, however, includes no child that was identified as Hispanic by using the father's information; among seven cases with missing information for the mother's Hispanic origin, no child had a Hispanic father. Infants not identified as Hispanic were assigned to African American, American Indian, or White categories based on the mother's race. As a result, the infant's race and Hispanic origin variable used in this study is identical to that of the mother, who was contacted for the survey.

The race and Hispanic origin variable consists of four categories: non-Hispanic White (White), non-Hispanic African American (African American), non-Hispanic American Indian (American Indian), and Hispanic. Although the White category is comprised primarily of non-Hispanic Whites, it also includes a small percentage of Asians (Chinese, Japanese, Hawaiians, or Filipinos). Asians comprise 1.34% of the total SEED OK sample, which is similar to the percentage reported in the American Community Survey (1.78%) (authors' calculation, using 2007 American Community Survey data from http://factfinder.census.gov/home/en/acs_pums_2007_1yr.html). Analysis results, after excluding Asians from the sample, do not differ substantively from those reported in this paper.

We use demographic information as control variables. Information on infants includes gender (1 for male; 0 for female) and birth weight. Since the relationship between birth weight and probability of responding to the survey may not be linear, we use the logarithm of birth weight in multivariate analyses. Analysis results with the actual value of birth weight are substantively identical to those reported in this paper. Variables about mothers include education (no high school diploma; high school diploma; some college; college degree; and missing information); age; marital status (1 for married; 0 for others); nativity (1 for native-born; 0 for foreign-born); and indicator of being born in Oklahoma (1 for yes and 0 otherwise). Since a substantial proportion of birth certificates (18%) do not contain information on father's age or education, we have created an indicator for father's information (1 for those without information on father's age or education and 0 otherwise). Infants with fathers' information on their birth certificates may differ from those without in terms of parents' relationship and father's involvement with childbirth and rearing (Getahun, Ananth, & Vintzileos, 2006; Ma, 2008; Nicolaidis et al., 2004). These differences may affect the mothers' chance of being located by and collaborating with the SEED OK survey team.

We also use two variables related to the SEED OK baseline survey recruitment process: recruitment stage and the number of days from the child's birth to contact. As described above, the sample was selected from children born between April and June 2007 (stage 1) and between August and October 2007 (stage 2). For the recruitment stage variable, we assigned the value of "0" to those recruited at stage 1 and "1" to those recruited at stage 2. We suspect unobserved systematic differences between the two stages. For example, survey interviewers may have been more experienced in working with those in the SEED OK sample during the second stage compared to the first, and the length of the recruitment period differed between the two stages. The number of days between birth and contact variables was created using birthday and the first contact date of each recruitment stage. The first contact date was July 1, 2007 for stage 1 and November 1, 2007 for stage 2. We use a logarithm of this variable in the Heckman analysis because the relationship between this variable and survey response may not be linear. Analysis using the actual value of the

number of days between birth and contact produced substantively identical results from those from the main model.

In addition, we use a geographic residency variable, created based on the U.S. Census Bureau's classification of 5-digit Zip Code Tabulation Areas (ZCTAs). ZCTAs are the U.S. Census Bureau's generalized representation of postal zip codes (see http://www.census.gov/epcd/www/zipstats.html for additional information). First, we obtained a "Population by Zip Code Tabulation Area: 2000" table from the Oklahoma Deputy Treasurer for Policy and Administration. This table maps each 5-digit Oklahoma ZCTA to the standard U.S. Census Bureau geographical classification, based on 2000 U.S. Census Bureau population data. We then collapsed the ZCTAs to match the SEED OK baseline survey variable, which provides only the first three digits for confidentiality reasons. The geographic residency variable consists of three categories: (1) metropolitan only (3-digit ZCTAs that fall entirely in a metropolitan area); (2) non-metropolitan only (3-digit ZCTAs classified as entirely non-metropolitan areas); and (3) mixed (3-digit ZCTAs that are a combination of metropolitan and non-metropolitan areas). The 3-digit ZCTAs variable was then merged with the SEED OK data.

Statistical Approach

This study employs probit regression with sample selection (Heckman probit). The analytical method identifies factors that are associated with survey collaboration, conditional on one's probability of being located. Since only those who were located could decide whether or not to collaborate with the survey team and complete the survey questionnaire, analyses may produce a biased estimation if different (and unobserved) characteristics between those located and those not located (selection bias) are not considered. Heckman probit analysis deals with this selection bias by estimating one's probability of collaborating with the survey team, conditional on one's probability of being located (Greene, 2003; Sales et al., 2004).

The Heckman probit model requires at least one identifier that is supposed to affect selection (being located) but not the final outcome (collaboration) (Greene, 2003; Sales et al., 2004). We include two identifiers in the selection model: (1) whether the mother was born in Oklahoma, and (2) the number of days between the infant's birth and the first contact date. Mothers born in Oklahoma are more likely to have lived in Oklahoma longer and to have a stable residence than those born elsewhere. Furthermore, Oklahoma-born mothers are more likely to have relatives in the state, which may have facilitated tracking Oklahoma-born mothers, however, probably do not differ from others in their chance of collaboration if demographic, social, and economic factors are controlled. A shorter time gap between infant birth and the first day of contact is expected to improve the chance of being located because it decreases the probability of moving to a new place after the infant's birth, but likely does not affect the chance of collaborating conditional on being located and controlling for other factors.

Except for these two identifiers, both equations (location and collaboration) in the Heckman probit analysis include the same variables. We include dummy variables for race and Hispanic origin as independent variables. Control variables include infant's gender and birth weight (log form); mother's marital status, education, age, and nativity; an indicator of father's information missing in the birth certificates; residency (metropolitan, non-metropolitan, or mixed); and stage of recruitment.

In addition to the analysis model described above, we ran supplementary analyses to check the robustness of findings: (1) an analysis using bivariate probit, instead of Heckman probit; (2) a model using actual values of birth weight and the days between birth and the first contact, instead of their log forms; (3) a model using a categorical variable for mother's age (younger than 25; 25-34 years old; or 35 or older); and (4) an additional analysis with a sample that did not include Asians. These supplementary analyses produced substantively identical results to those reported in this paper. Results from supplementary analyses are available from the authors upon request.

Results

Table 1 summarizes demographic and other characteristics of the sample by race and Hispanic origin. As expected, Whites have, on average, more socioeconomic advantages than other groups. White infants are more likely to have mothers who are married and college-educated than minority infants. They are more likely to have the father's information recorded on their birth certificates than other groups, suggesting a more stable relationship between birth parents. Also as expected, mothers of Hispanic infants are less likely to be native born. The percentage living in metropolitan areas is much higher among African Americans and Hispanics.

Table 2 presents response status by race and Hispanic origin. In addition to the number of individuals in each category of status, Table 2 shows three summary measures of response: location rate (the percentage of located individuals in the entire sample), collaboration rate (the percentage of located individuals who completed the survey questionnaire), and response rate (the percentage of complete cases in the entire sample). As shown in Table 2, different measures produce distinct results by race and Hispanic origin. Whites have the highest location rate (89%), while African Americans the lowest rate (82%). In contrast, Whites show the second lowest collaboration rate while African Americans have the highest rate. Hispanics have low rates for both location and collaboration, and American Indians have rates in the middle for both measures. As a result, the response rate is highest among African Americans and lowest among Hispanics, while Whites' and American Indians' response rates fall in the middle.

Table 3 summarizes the Heckman probit analysis. Results for control variables are as expected. Those with a college degree are more likely than those without a high school diploma to be located and to collaborate with the survey team, consistent with findings in previous studies (Galea & Tracy, 2007; Groves, Cialdini, & Couper, 1992). Native-born mothers are more likely to collaborate with the survey team once located by SEED OK survey team. Foreign-born mothers may have language and cultural barriers, which could discourage them from participating in the survey. In addition, these mothers may feel uncomfortable revealing personal information if they are undocumented. Consistent with existing studies (Groves, 2006; Keeter et al., 2006), those living in non-metropolitan areas are more likely to be located than those living in metropolitan areas. SEED OK sample members recruited at stage 2 are more likely to be located, probably because the survey team acquired skills through experience in stage 1. However, those recruited at stage 2 are less likely to collaborate with the survey team, conditional on being located.

Analysis of race and Hispanic origin variables reveal different patterns for location and collaboration. All three minority groups are significantly less likely to be located than Whites (the reference group), even when demographic, socioeconomic, and geographic factors are considered. Conditional on being located, however, the probability of collaboration does not differ significantly by race or Hispanic origin.

Table 1. Sample Characteristics by Race and Hispanic Origin (N=7,111)

	Whites	African Americans	American Indians	Hispanics	Total
Infant is Male, %	52.02	52.16	51.22	54.47	52.34
Birth-weight (gm), Mean	3277	3054	3305	3297	3249
Mother is Married,%	68.47	27.81	47.01	52.02	54.59
Mother's Education, %					
No High School Diploma	15.01	22.49	25.02	49.73	24.55
High School Diploma	35.45	40.15	41.25	32.93	36.87
Some College	24.26	24.34	21.55	10.62	21.25
College Degree	24.82	12.26	11.88	5.50	16.71
Missing	0.46	0.76	0.30	1.22	0.62
Mother's Age, Mean	26.11	24.37	24.46	25.19	25.34
Mother is a Native Citizen, %	95.44	95.35	99.26	37.51	85.49
Mother Born in Oklahoma, %	61.21	67.62	83.84	16.58	58.37
Father's Information Missing, %	11.21	37.19	17.42	18.11	17.99
Residence, %					
Metropolitan	26.62	66.86	16.01	59.66	37.38
Mixed	65.44	29.08	74.91	31.93	55.03
Non-Metropolitan	7.94	4.06	9.08	8.40	7.59
Recruited at Stage 2, %	49.08	48.52	56.16	49.35	50.39
Days from Child Birth to Contact, Mean	50.41	49.51	50.93	51.08	50.48
Number of Observations	3,264	1,183	1,355	1,309	7,111
(%)	(45.90)	(16.64)	(19.05)	(18.41)	(100.00)

Table 2. Response Status by Race and Hispanic Origin (N=7,111)

	Number of Unlocated Cases (A)	*Number of Located Non-collaborators (B)*	*Number of Respondents (C)*	*Location Rate (B+C)/(A+B+C)*	*Collaboration Rate C/(B+C)*	*Response Rate [C/(A+B+C)]*
White	364	1,651	1,249	88.85%	43.07%	38.27%
African American	209	503	471	82.33%	48.36%	39.81%
American Indian	192	644	519	85.83%	44.63%	38.30%
Hispanic	216	629	464	83.50%	42.45%	35.45%
Full Sample	981	3,427	2,703	86.20%	44.09%	38.01%

Table 3. Heckman Probit Results: Collaboration Conditional on Being Located (N=7,111)

Variables	Located	Collaborated
Race: African American	-0.166***	0.084
	(0.058)	(0.052)
American Indian	-0.150***	0.046
	(0.055)	(0.045)
Hispanic	-0.129*	0.066
	(0.069)	(0.058)
Male	0.037	0.031
	(0.039)	(0.031)
Birth-Weight (logarithm)	0.140*	0.023
	(0.084)	(0.074)
Mother is Married	0.139***	0.027
	(0.049)	(0.041)
Mother's Education: High School Diploma	0.013	-0.023
	(0.050)	(0.043)
Some College	0.064	-0.000
	(0.061)	(0.051)
College Degree	0.296***	0.209***
	(0.076)	(0.058)
Missing	0.340	0.063
	(0.259)	(0.199)
Mother's Age	0.007*	0.002
	(0.004)	(0.003)
Mother is a Native Citizen	0.053	0.167***
	(0.077)	(0.060)
Father's Information Missing	-0.041	0.063
	(0.055)	(0.048)
Residence: Mixed	0.068	0.020
	(0.045)	(0.037)
Non-Metro	0.149*	-0.011
	(0.079)	(0.064)
Recruited at Stage 2	0.685***	-0.174***
	(0.041)	(0.060)
Mother Born in Oklahoma	0.038	
	(0.045)	
Log (Birth to Contact)	-0.042	
	(0.028)	
Constant	-1.200*	-0.520
	(0.700)	(0.637)
Athrho	0.953* (0.539)	
Rho	0.741 (0.243)	

Note: Standard errors are in parentheses.
* p <0.1 ** p <0.05 *** p < 0.01

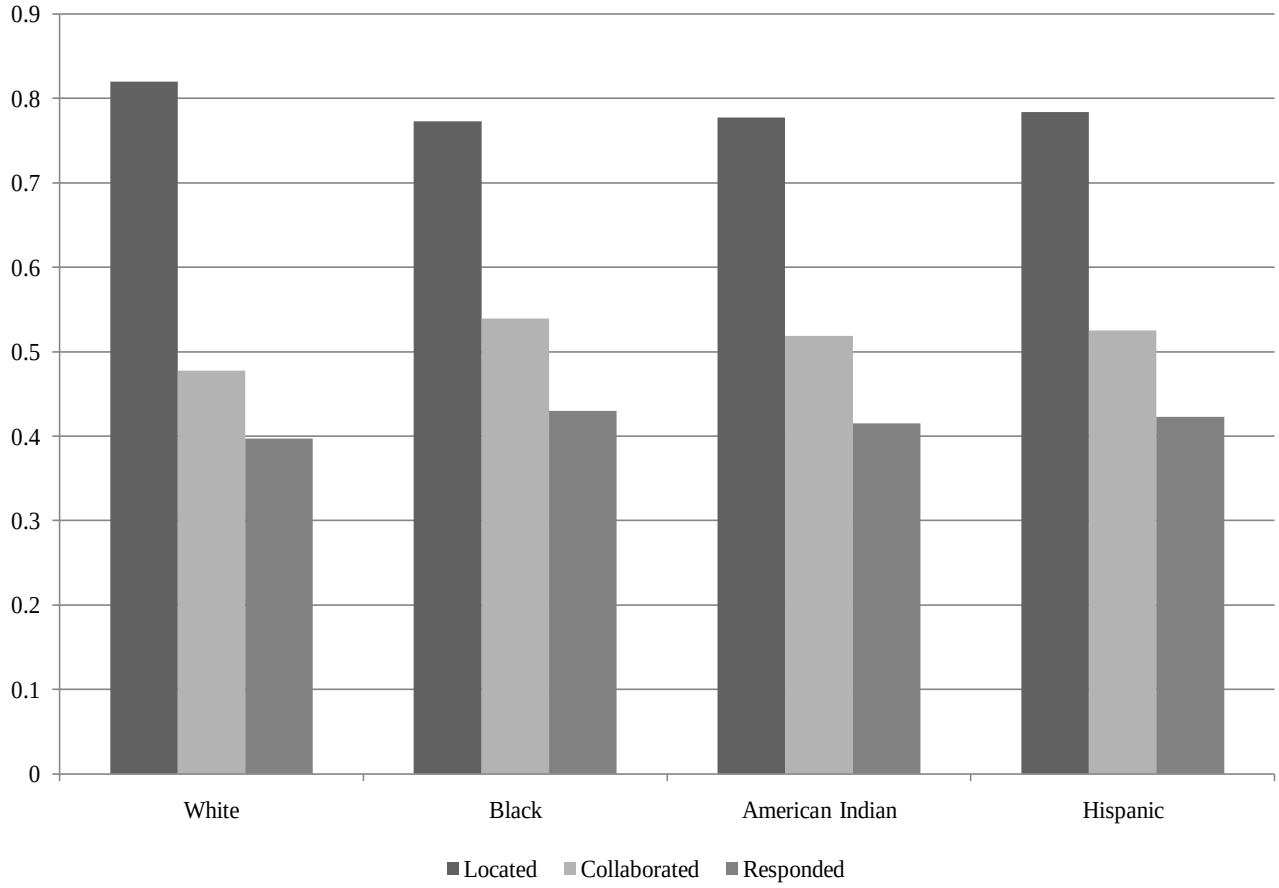
Figure 1 presents predicted probabilities of survey participation based on the Heckman probit results reported above. We estimate predicted probabilities for a typical case in the sample: a male infant with a birth weight of 3,250 grams; whose mother is a 25-year-old, Oklahoma-born, married woman with a high school diploma, living in a metropolitan area; whose father's information is not missing; and who was recruited at stage 1 and born 50 days before the first contact was made by the SEED OK survey team.

The first bars in Figure 1 present predicted probabilities of being located for the four racial and ethnic groups in our study. Consistent with descriptive statistics, Whites are more likely to be located than minority groups (82% versus 77-78%). Predicted probabilities produce a somewhat different result from the descriptive statistics for American Indians: the likelihood of being located for American Indians is comparable to the other two minority groups. This result contrasts with the descriptive finding where the location rate for American Indians is 2 or 4 percentage points higher than those of African Americans and Hispanics. These results suggest that differences in location rates between American Indians and the other two minority groups found in the descriptive statistics may be caused by distinct characteristics. American Indians are less likely to live in metropolitan areas (Table 1), where the location rate is typically lower than in other areas (Table 3). In addition, the percentage of married mothers is much higher among American Indians in comparison with African Americans (47% versus 28%). According to the Heckman probit analysis, married mothers are more likely to be located than non-married mothers.

The second bars show predicted probabilities of collaboration by race and Hispanic origin once located. Predicted probability of collaboration is lowest among Whites (48%) and highest among African Americans (54%), which is consistent with descriptive statistics in Table 2. The other two minority groups' predicted probabilities of survey collaboration lie between those of Whites and African Americans: 52% for American Indians and 53% for Hispanics.

The last bars show predicted probabilities of responding to the survey, combining probabilities of being located by and collaborating with the survey team. Predicted probability of survey response is lowest among Whites (40%), due to their low level of collaboration. Predicted probability of response is highest among African Americans (43%), followed by Hispanics and American Indians (both around 42%). These results are somewhat different from descriptive statistics that show the lowest response rate among Hispanics and a relatively high response rate among Whites. These discrepancies suggest that different response rates among racial and ethnic groups may be explained at least partially by distinct characteristics among these groups. Hispanic infants have a lower percentage of mothers with a college degree and a higher percentage of foreign-born mothers; these variables are estimated to decrease one's chance of being located and/or of collaborating with the survey team. At the same time, White infants have the highest percentage of mothers with a college degree, at least twice as high as other groups (Table 1); this variable is associated with higher chances of location and survey collaboration.

Figure 1. Predicted Probabilities of Survey Response by Race and Hispanic Origin



Discussion

In this study, we investigate differences in survey response among four racial and Hispanic origin groups, using data from OSDH birth certificates and the SEED OK baseline survey data. Heckman probit analysis finds that African Americans, American Indians, and Hispanics are significantly less likely than Whites to be located during study recruitment. Since this study controls for several demographic, socioeconomic, and geographic factors, differences among the four groups cannot be attributed to discrepancies in these observed characteristics. Findings of this study are consistent with other research that finds lower location rates for American Indians, African Americans, and Hispanics in comparison with Whites (Azar et al., 1984; Kim et al., 2008).

Heckman probit analysis also demonstrates that, conditional on being located, the probability of collaboration does not differ between Whites and minority groups. This result differs from two existing studies that use a different analytical approach: logistic regression with a sample consisting only of located cases (Groves & Couper, 1998; Zaslavsky, Zaborski, & Cleary, 2002). Zaslavsky et al. (2002) find that African Americans and Hispanics are less likely to collaborate with research than Whites. Groves and Couper's (1998) regression results indicate that Hispanics are more likely to collaborate than Whites, while African Americans' collaboration probability is not significantly

different from that of Whites. To test whether different analytical approaches (Heckman probit versus logistic regression with located cases) produced distinct analysis results, we ran a logistic regression with a sample composed only of located cases, following the two previous studies. With this analytic approach, the three minority groups are significantly more likely to collaborate with the survey team than Whites, in contrast with Zaslavsky and colleagues' study, but somewhat similar to Groves and Couper's (1998).

It is plausible that the generous incentives available to SEED OK study participants may have affected located individuals' decision to collaborate, especially among members of minority groups. Before the survey was conducted, SEED OK informed sample members that their child would have a 50% chance of receiving a $1,000 deposit into a college savings plan account and (if income-eligible) saving matches if they participated in the study. The possibility of receiving financial incentives may have motivated some located sample members to collaborate with the survey team. It is also possible that requesting the infant's Social Security Number may have discouraged some sample members from collaborating with the survey. In addition, the study population of SEED OK— mothers of infants in SEED OK—also differs from the other two studies. The study population is Medicare beneficiaries in Zaslavsky et al. (2002) and general households in Groves and Couper (1998). These factors may explain discrepancies in findings between this study and the other two studies. Further research is warranted on whether survey collaboration rates are the same or different among racial and ethnic groups when their chances of being located are taken into account, whether and what financial incentives associated with study participation affect collaboration, and whether the relationship between race/ethnicity and chances of collaboration differs by age, gender, and other characteristics.

This research addresses key methodological deficiencies found in previous studies to offer important new insights on survey response by race and ethnicity. First, our analytical method uses several possible predictors of response—such as education, and race or Hispanic origin—from the Oklahoma birth certificate sampling frame to compare characteristics between survey respondents and non-respondents. Second, this study addresses whether survey response rates differ by race and ethnicity at two distinct stages in the recruitment process: (1) being located by the survey team and (2) collaborating with the survey team by completing the survey questionnaire. Thus, this analysis distinguishes the effects of race and Hispanic origin on survey responses by separating three distinct groups of potential respondents: those who were not located, those who were located but did not complete the questionnaire, and those who were located and collaborated with the survey team by completing the questionnaire. Third, this research includes American Indians, a racial group that is rarely studied. In this way, this study expands understanding of survey response by race and ethnicity.

Despite these methodological strengths, this study has some limitations. First, this study uses data from only one state, so that findings of this study cannot be generalized to other states or the nation. Second, we are unable to create a separate category for Asian Americans because of the small number included in our sample. Third, birth certificate data do not provide a couple of critical variables, such as family income and homeownership. The omission of these variables may have produced biased analysis results. Fourth, a multi-racial category was not available for this study. This category is desirable since population compositions in the United States are changing and multi-racial populations are growing (Hirschman, Alba, & Farley, 2000). Last, some hypotheses described in the background section are not able to be tested directly. For example, we cannot test the

hypothesis that mistrust and suspicion toward research institutions reduces response rates among minority groups.

Turning to key results and implications, we find that the three minority groups were significantly less likely than Whites to be located by the survey team. These results suggest that researchers should improve strategies to locate African American, American Indian, and Hispanic study subjects to increase response rates of these groups. This finding also calls for further investigation of why it is more difficult to locate members of minority groups than Whites. The causes of low location rates for members of minority groups (e.g., high mobility rates, lack of access to a phone, or mistrust and fear) should be explored. At the same time, researchers should develop strategies to locate members of minority groups.

Furthermore, results show that minority groups' chances of collaborating with the survey do not differ significantly from that of Whites, when we take into account these groups' probabilities of being located and several demographic and socioeconomic factors. If these results are confirmed by future studies, it would contradict the often-mentioned hypothesis that members of minority groups refuse to collaborate with survey researchers because they mistrust and fear mainstream institutions and scientific research (Buchwald et al., 2006; Fouad et al., 2000; Johnson et al., 2002; Oropesa, 2002; Yancey, Ortega, & Kumanyika, 2006). However, it is also possible that significantly lower probabilities of being located may reflect minority group members' mistrust and suspicion. Those who were suspicious and mistrusting may not have responded to mail requesting their phone numbers nor answered the phone when the SEED OK survey team called. In general, more empirical evidence and in-depth understanding of fear and mistrust in survey collaboration are needed before making conclusions about how these factors affect survey response among minority groups.

Results of this study call for more detailed attention to the recruitment process of survey response. This study demonstrates that the association of race and Hispanic origin with survey response differs at two distinct stages: 1) being located and 2) collaborating with the survey team. Thus, researchers may not be able to fully understand the survey response process if they examine only the final response rate. Further, members of minority groups may not be fully represented in research studies if improved location techniques are not given a higher priority.

References

Azar, S. T., Robinson,D. R., Hekimian, E., & Twentyman, C. T. (1984). Unrealistic expectations and problem-solving ability in maltreating and comparison mothers. *Journal of Consulting & Clinical Psychology, 52*, 687-691.

Baines, A. D., Partin, M. R., Davern, M., & Rockwood, T. H. (2007). Mixed-mode administration reduced bias and enhanced poststratification adjustments in a health behavior survey. *Journal of Clinical Epidemiology, 60*, 1246-1255.

Bauman, K. J., & Graf, N. L. (2003). *Educational Attainment:2000* [Census 2000 Brief]. Washington, DC: U.S. Census Bureau.

Buchwald, D., Mendoza-Jenkins, V., Calvin, C., McGough, H., Bezdek, M., & Spicer, P. (2006). Attitudes of urban American Indians and Alaska Natives regarding participation in research. *Journal of General Internal Medicine, 21*, 648-651.

Buescher, P. A., Gizlice, Z., & Jones-Vessey, K. A. (2005). Discrepancies between published data on racial classification and self-reported race: Evidence from the 2002 North Carolina live birth records. *Public Health Reports, 120*, 393-398.

DeLeeuw, E., Callegaro, M., Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys. *Public Opinion Quarterly, 71*, 413-443.

DeNavas-Walt, C., Proctor, B. D., & Smith, J. (2008). *Income, poverty, and health insurance coverage in the United States: 2007. Current population reports*. Washington, DC: U.S. Census Bureau.

Des Jarlais, G., Kaplan, C. P., Haas, J., Gregorich, S. E., Perez-Stable, E., & Kerlikowske, K. (2005). Factors affecting participation in a breast cancer risk reduction telephone survey among women from four racial/ethnic groups. *Preventive Medicine, 41*, 720-727.

Fouad, M. N., Edward, P., Lee, G. B., Kohler, C., Wynn, T., Nagy, S., & Churchill, S. (2000). Minority recruitment in clinical trials: A conference at Tuskegee, researchers and the community. *Annals of Epidemiology, 10*, S35-S40.

Galea, S., & Tracy, M. (2007). Participation rates in epidemiologic studies. *Annals of Epidemiology, 17*, 643-654.

Getahun, D., Ananth, C. V., & Vintzileos, A. M. (2006). Uteroplacental bleeding disorders during pregnancy: Do missing paternal characteristics influence risk? *BMC Pregnancy and Childbirth 6*,doi:10.1186/1471-2393-6-2.

Greene, W. H. (2003). *Econometric analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*, 646-675.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*, 475-495.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: John Wiley.

Hirschman, C., Alba, R., & Farley, R. (2000). Meaning and measurement of race in the U.S. census: Glimpses into the future. *Demography, 37*, 381-393.

Johnson, T., O'Rourke, D., Burris, J., & Owens, L. (2002). Culture and survey nonresponse. In R. M. Groves, D. A. Dillman, J. L. Etlinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 55-70). New York: John Wiley & Sons, Inc.

Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly, 70*, 759-779.

Kim, S. Y., Tucker, M., Danielson, M., Johnson, C., Snesrud, P., & Shulman, H. (2008). How can PRAMS survey response rates be improved among American Indian mothers? Data from 10 States. *Maternal and Child Health Journal, 12*, S119-S125.

Liao, Y., Tucker, P., Okoro, C. A., Giles, W. H., Mokdad, A. H., & Harris, V. B. (2004). *REACH 2010 - Surveillance for health status in minority communities, United States, 2001-2002*. Vol. 53. Atlanta, GA: Center for Disease Control and Prevention.

Ma, S. (2008). Paternal race/ethnicity and birth outcomes. *American Journal of Public Health, 98*, 2285-2292.

Marks, E.n L., Rhodes, B. B., & Scheffler, S. (2008). *SEED for Oklahoma Kids: Baseline analysis*. Research Triangle Park, NC: RTI International.

Nicolaidis, C., Ko, C. W., Saha, S., & Koepsell, T. D. (2004). Racial discrepancies in the association between paternal vs. maternal educational level and risk of low birthweight in Washington state. *BMC Pregnancy and Childbirth, 4*(10), doi:10.1186/1471-2393-4-10.

Oliver, M. L., & Shapiro, T. M. (1995). *Black wealth/White wealth: A new perspective on racial inequality*. New York: Routledge.

Oropesa, R. S., & Landale, N. S. (2002). Nonresponse in follow-back surveys of ethnic minority groups: An analysis of the Puerto Rican maternal and infant health study. *Maternal and Child Health Journal, 6*, 49-58.

Sales, A. E., Plomondon, M. E., Magid, D. J., Spertus, J. A., & Rumsfeld, J. S. (2004). Assessing response bias from missing quality of life data: The Heckman Method. *Health and Quality of Life Outcomes, 2*, 49.

Shapiro, T. M. (2004). *The hidden cost of being African-American: How wealth perpetuates inequality*. Oxford; New York: Oxford University Press.

Yancey, A. K., Ortega, A. N., & Kumanyika, S. K. (2006). Effective recruitment and retention of minority research participants. *Annual Review of Public Health, 27*, 1-28.

Zager, R., Kim, Y., Nam, Y., Clancy, M., & Sherraden, M. (2010). *The SEED for Oklahoma Kids Experiment: Initial account opening and savings* (CSD Research Report,No. 10-14). St. Louis, MO: Washington University in St. Louis, Center for Social Development.

Zaslavsky, A. M., Zaborski, L, & Cleary, P. D. (2002). Factors affecting response rates to the Consumer Assessment of Health Plans study survey. *Medical Care, 40*, 485-499.