

# Predicting Panel Attrition on a National Study: How We Can Optimize Locating Resources and Methods

Lauren M. McNamara<sup>1</sup>, Melissa Heim Viox<sup>1</sup>, Ned English<sup>1</sup>, Papia Paul<sup>1</sup>  
Katie Dekker<sup>1</sup>, Ron Hazen<sup>1</sup>

<sup>1</sup>NORC at the University of Chicago, 55 E. Monroe St, Chicago, IL, 60603

## Abstract

Longitudinal studies depend on the retention of sample members across an extended period of time. The propensity of sample members to attrite is not randomly distributed. Thus, longitudinal studies run the risk of introducing bias via differential attrition due to movement and nonresponse. To gain a better understanding of who is likely to become lost to follow-up, analysis and modeling comparing lost to retained participants were performed using data from the National Children's Study (NCS) Vanguard Study, a national longitudinal pilot study of children's health and development.

Once women enrolled in the Study, up to two participant interviews were scheduled to occur before childbirth. Women were then interviewed at childbirth and again at regularly-spaced intervals every three to six months for the first five years after birth. Women who were 'lost' were traced using best practice locating methods: contacting non-household contacts, accessing batch locating resources, interactive locating with free and proprietary databases, and in-person locating.

Our paper uses logistic regression to understand panel attrition and predict the propensity of a participant to become lost to follow-up. In so doing we pursue how attrition is dependent on recruitment strategy, demographics, geography, and number and distance of moves. Our results indicate that it may be most effective to focus resources in the form of extra locating effort or interim follow-ups on a subset of participants, while decreasing efforts spent on low risk or "stable" participants. Our findings shed light on the relationship between participant characteristics and panel attrition in longitudinal studies.

**Key Words:** attrition, modeling, longitudinal study

## 1. Introduction

It is generally understood that participant retention challenges all longitudinal panel surveys (Cotter et al. 2002). Longitudinal panel surveys are designed to measure change over time; for this type of study to be successful, a sufficient percentage of participants recruited at the start of the study need to remain in the study for its duration. However, participant attrition is expected in all longitudinal studies (Dodds et al. 1989). Participants may directly refuse or request to be dropped from the study, or participants may move or change phone numbers or e-mail addresses without informing the study they have done so. Those in the latter group will require locating wherein a researcher attempts to find new contact information and re-establish contact with the participant. If locating is unsuccessful

and yields no new accurate contact information, these participants may go on to be finalized as “unlocatable”.

Our analysis focuses on these unlocatable, or lost to follow-up, participants. Due to the considerable investment involved in recruiting and retaining participants for panel or cohort surveys it is of great value to be able to predict which participants are at risk of losing touch with these surveys. We hypothesize that attrition in this manner is non-random and that certain types of participants are more likely to become lost to follow-up.

### **1.1 Relevant Literature**

It is clear from the literature that the need to locate participants who have moved is pertinent to all longitudinal panel or cohort studies (Dodds et al. 1989; Cohen et al. 2000; Cotter et al. 2002). At issue is that the need for locating is not evenly distributed across the sample, generally as both a function of the propensity to move residence as well as to have “locatable” information (Graziotti et al. 2012).

Past studies support our hypothesis that being lost to follow-up is non-random, related to cooperation, movement, and the ability to be located (Cohen et al. 2000). Consequently, longitudinal studies that tend to lose particular types of households over time may be at risk of bias (Gustavson et al. 2012).

Researchers have found that marital status, socio-economic status, age, gender, race/ethnicity, education, and urbanicity impact panel retention due to locating and non-response (Klebanoff et al. 1998; Haunberger 2010; Olson and Witt 2011; Graziotti et al. 2012; Venevongsa et al. 2014). These studies have generally found that women, those with higher education and higher socio-economic status, homeowners, and non-urban residents were generally likely to remain on panels for longer than others, but the evidence across studies can be conflicting.

In addition to participant characteristics, interviewer attributes such as continuity, experience, and demographics have also been shown to influence nonresponse at the individual and household levels on panel surveys (Campanelli and O’Muircheartaigh 1999; West and Walston 2003). Analyses of operational variables have found that interim contact frequency, prior wave item-nonresponse or high numbers of contact attempts per data collection event were predictive of panel attrition (De Keulenaer 2005; Haunberger 2010).

Our paper synthesizes the above approaches to examine how participant characteristics, neighborhood-level characteristics, and operational factors predict panel attrition in a longitudinal study of young children and their caregivers.

### **1.2 Problem Statement**

Survey researchers will benefit from understanding the types of populations that tend to become lost to follow-up and considering this during the survey design phase, to mitigate the risk of bias entering a study. Understanding the characteristics of participants at risk of becoming lost will allow researchers to quantify any impact on panel composition, adjust recruitment methods, make predictions for panel loss over time, and consider differential locating methods for different participants.

This manuscript provides insight into the adult participants in the National Children’s Study (NCS) Vanguard Study who were lost to follow-up. First, we describe the characteristics of households who were lost to follow-up on the Study because they moved

residence and could not be located. Second, we present logistic regression models to predict the likelihood of a participant becoming lost to follow-up based on their characteristics. Such models have utility both for understanding participants lost to follow-up as well as to make predictions for types of participants that may become lost in the future.

## 2. Methods

### 2.1 National Children's Study

The National Children's Study (NCS) was designed as a large-scale, prospective longitudinal panel study with the goal of furthering our understanding of the roles various factors play in health and disease. The Study aimed to examine the impacts and interface of genetics and environment on the growth, development, and health of children across the United States, following them from before birth until age 21 years (Montaquila et al. 2009; Montaquila et al. 2010; Downs et al. 2010).

The NCS Vanguard Study was a pilot study of over 5,000 women and their newborn children, intended to test the proposed methods of the Main Study. Participation in the Vanguard Study was restricted to women who resided in a pre-selected geographic area, were within the ages of 18 to 49 years, and were pregnant or likely to become pregnant. The study population did not incorporate over-sampling of minority groups and this analysis is performed on raw, unweighted data.

### 2.2 Recruitment

The recruitment phase of the NCS Vanguard Study was comprised of five separate pilot groups, launched during three unique time periods. Seven Study Locations began the first wave of recruitment in 2009 using an Initial Household-Based protocol that included enumeration of selected households to confirm geographic eligibility and identify women to be screened for pregnancy status and age eligibility.

Three additional pilots were launched concurrently in 2011 across 30 Study Locations representing a mixture of urban and rural areas. The three new pilot groups used an identical sample design but varied in how potential participants were identified.

- The Enhanced Household locations utilized field workers to contact households in predetermined geographic areas.
- The Provider with Area Sample group locations worked with health care providers in the Study Location to identify women for eligibility screening and recruitment.
- The Direct Outreach locations relied on marketing, direct mail, and other referral techniques. Interested individuals were encouraged to contact the Study, at which time their eligibility could be determined.

Finally, in 2012, an additional Provider Sample pilot was launched in three Study Locations; this method compiled lists of providers and their associated patient volumes. Once complete, a sample of approximately 20 provider locations was selected from each study location. The final stage of sampling systematically sampled pregnant women from the selected provider locations.

At the end of recruitment in 2014, over 5,000 women had been enrolled, and the NCS Vanguard Study shifted focus toward participant retention.

### **2.3 Tracing and Locating**

The long-term quality of data in a longitudinal study is dependent upon the ability to locate participants for each data collection event. For participants whose contact information does not change, locating is as easy as calling them or visiting the last known address. However, as 29% of young adults aged 18-34 and 23% of those with children under 6 move addresses, locating efforts are especially important in a survey of children and their parents (Benetsky 2015).

The NCS Vanguard Study followed best practices for locating participants, utilizing a standard six-step procedure to trace participants. The first step was to ensure that the participant truly needed to be located, by checking phone numbers at multiple times of day or after a brief delay if it had been disconnected, by e-mail, and by use of signature-required letters to confirm whether a participant was at a physical address.

Once a participant was deemed “in locating”, data collectors began contacting participants’ “tracing contacts”. These individuals resided outside of the participant’s household (such as relatives and friends) and were provided by the participant during prior waves of data collection as a potential source of updated contact information. The third step was batch locating, in which multiple case files are sent to locating vendors to match a person’s previous contact information with their current contact information. The fourth step was interactive locating, in which cases are worked individually using a variety of approved databases, web searches, and contact attempts to obtain updated contact information.

As a final step for the most difficult-to-locate cases, in-person strategies for tracing and locating were utilized. Local interviewers or field managers began with the best known address for the participant to try to find the participant or ask individuals at the address or neighbors to acquire information on the participant’s new contact information, acting on any leads obtained. While these steps were followed in order for most cases, tracing is a very iterative process and the procedures were modified on a case-level basis to best track lost participants. If after all steps were followed, the participant was still unable to be located, s/he was coded as “final unlocatable” and no further efforts were made to locate the participant.

### **2.4 Model Building**

Raw, unweighted data from the NCS Vanguard Study were downloaded from a central repository on November 21, 2014. Lost to follow-up (LTFU) status was determined based on a report listing the participants that had become unlocatable as of July 31, 2014. All time-sensitive data was limited to that which had been collected prior to July 31, 2014. Most demographic variables were sourced from a Study analytic dataset developed by a Eunice Kennedy Shriver National Institute of Child Health and Human Development contractor. Household language, home ownership, tracing contacts, and address-related variables were derived using raw instrument datasets. Census tract variables associated with a participant’s location were pulled from the 2009-2013 American Communities Survey (ACS) 5-year estimates based on address.

For this longitudinal study, the last-known data point for a participant was retained for time-varying variables. For example, for household income, if a participant had only one study visit, the income would reflect this single visit; if another participant had five study visits, the data would reflect income as of the fifth visit. Demographic data reflects the NCS Vanguard Study child’s primary caregiver on the assumption that the primary

caregiver will be the primary contact rather than the minor children. Where primary caregiver was unknown, the biological mother was assumed to be the primary caregiver.

Education and home ownership data were collected via NCS Vanguard Study surveys but these variables had a high rate of missing data. The collected data was replaced with ACS data describing the percentage of residents 18 years and older with a high school diploma or less education and the percentage of homeowners in the participant's last known census tract in an effort to improve the predictive ability of our regression model. The Akaike Information Criterion (AIC) and Schwarz Criterion (SC) model-building statistics supported this choice. Although this data is not directly reported by participants, we were able to obtain education and homeownership information for more participants using this method.

Using participant address, we calculated the total distance a participant moved by calculating the distance from first to second reported address, plus second to third reported address, and so on. If a participant did not move, the distance moved was set to zero.

Percentages and odds ratios were computed for categorical variables by LTFU status and significance testing was performed using chi-square tests of independence. All continuous variables were found to be non-normally distributed using the Kolmogorov-Smirnov Goodness-of-Fit Test and thus the Wilcoxon signed-rank test was used to test for significant differences between the LTFU and non-LTFU groups. Binary logistic regression was used to examine the relative strength of risk factors associated with unlocatable status. A full model was first fit to the data, and then a reduced model was selected using the SAS stepwise model selection procedure. All computations were completed using SAS 9.3.

### 3. Results

We completed both univariate and multivariate analyses to assess the effect of various participant characteristics on lost to follow-up status.

#### 3.1 Descriptive Statistics

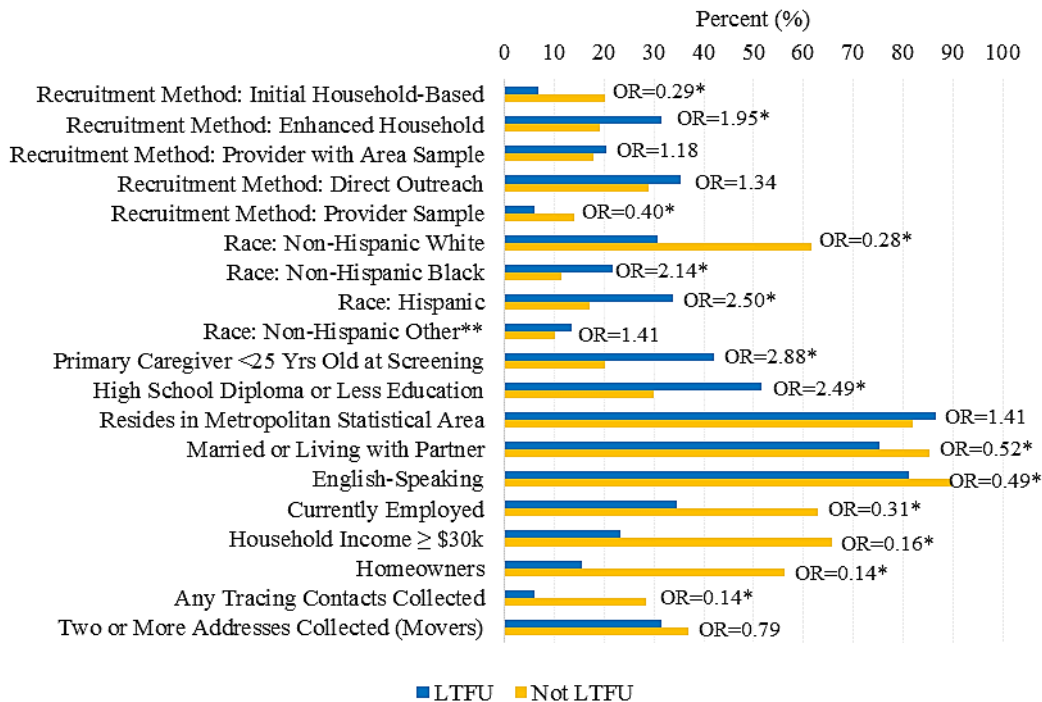
##### 3.1.1 Odds ratios and means

Figure 1 presents univariate descriptive characteristics for participants lost to follow-up (LTFU) and active participants. Participants who were recruited using the Enhanced Household method were more likely to be LTFU (OR = 1.95), whereas participants recruited through the Provider Sample (OR = 0.40) or Initial Household-Based (OR = 0.29) methods were less likely to become LTFU. Hispanic (OR = 2.50) and non-Hispanic Black participants (OR = 2.14) were more likely to be lost to follow-up than non-Hispanic White participants (OR = 0.28). Primary caregivers younger than 25 years old at screening (OR = 2.88) and participants with a high school education or less (OR = 2.49) also tended to have higher rates of attrition.

Other characteristics appear to shield against attrition. Participants that were less likely to become LTFU were married or living with a partner (OR = 0.52), English-speaking (OR = 0.49), employed (OR = 0.31), earning a higher income (OR = 0.16), homeowners (OR = 0.14), or had provided tracing contacts (OR = 0.14). No statistically significant univariate

impact was found for participants living in a metropolitan statistical area or those who had provided two or more addresses to the study.

**Figure 1: Descriptive Statistics (Percentages) for LTFU and Not LTFU Participants**



\*Statistically significant, p<0.05

\*\*Includes American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, Multiple Race, and Other Race

Table 1 shows the means and standard errors for continuous measures of participant characteristics. The average total distance moved was greater for those who were lost to follow-up than those who were not, but this difference was not statistically significant. ACS-derived education and homeownership results mirrored the participant-reported results, with participants living in census tracts with lower rates of higher education and homeownership being more likely to become LTFU.

**Table 1:** Descriptive Statistics (Means) for LTFU and Not LTFU Participants

	LTFU	Not LTFU
Distance Moved in Miles	128.6 (46.94)	68.0 (5.49)
Percent 18+ with High School Diploma or Less Education in Census Tract	0.5 (0.02)*	0.4 (0.00)
Percent Homeowners in Census Tract	0.5 (0.02)*	0.6 (0.00)

Data are presented as mean (standard error)

\*Statistically significant,  $p < 0.05$

### 3.1.2 No address collected

A number of participants ( $N = 300$ ) did not provide any household address while enrolled in the Study, comprising 5.2% of those not lost to follow-up, and 18.5% of the lost to follow-up group. Participants who did not give an address were significantly more likely to attrite from the study ( $OR = 4.1$ ,  $p < 0.05$ ), but this was not included as a factor in the regression modeling and these participants were otherwise excluded from the analysis due to missing data on address-related variables.

## 3.2 Modeling

In the binary logistic regression analysis, odds ratios indicate the odds of being in the lost to follow-up group compared to not being in this group. The analysis was performed on  $N = 4,700$  participants with no missing data on any of the variables of interest since regression models require complete data. Results from the logistic regression models are presented below.

### 3.2.1 Full model

Results for the full regression model are presented in Table 2. Characteristics that significantly increased a participant's likelihood of being lost to follow-up were Hispanic ( $OR = 2.6$ ) or Non-Hispanic Other ( $OR = 2.4$ ) race, residing in a Metropolitan Statistical Area ( $OR = 1.7$ ), having a primary caregiver less than 25 years old at screening ( $OR = 1.5$ ), and being recruited via the Direct Outreach method ( $OR = 2.3$ ). Characteristics that significantly decreased a participant's likelihood of being lost to follow-up were having two or more addresses collected ( $OR = 0.5$ ), being currently employed ( $OR = 0.5$ ), living in a census tract with a greater percentage of homeowners ( $OR = 0.3$ ), having a household income over \$30,000 ( $OR = 0.2$ ), and having provided one or more tracing contacts ( $OR = 0.2$ ). The Enhanced Household, Provider with Area Sample, and Provider Sample recruitment methods, Non-Hispanic Black race, participant's spoken language, being married or living with a partner, and education level were not significant predictors of lost to follow-up status in the full multivariate model.

**Table 2:** Logistic Regression Full Model Results

Variable	Odds Ratio	95% Confidence Interval	Wald $\chi^2$	p-Value
Recruitment Method: Initial Household-Based (REFERENCE GROUP)	---	---	---	
Recruitment Method: Enhanced Household	2.1	(1.0, 4.5)	3.4	0.065
Recruitment Method: Provider with Area Sample	1.7	(0.8, 3.9)	1.6	0.203
Recruitment Method: Direct Outreach	2.3	(1.1, 4.9)	4.5	0.033
Recruitment Method: Provider Sample	0.4	(0.1, 1.0)	3.5	0.060
Race: Non-Hispanic White (REFERENCE GROUP)	---	---	---	
Race: Non-Hispanic Black	1.7	(0.9, 3.0)	2.7	0.098
Race: Hispanic	2.6	(1.5, 4.6)	11.6	0.001
Race: Non-Hispanic Other*	2.4	(1.3, 4.5)	8.6	0.003
Resides in Metropolitan Statistical Area	1.7	(1.0, 3.0)	4.0	0.045
Age: <25 Years Old at Screening	1.5	(1.0, 2.3)	4.4	0.036
English-Speaking	1.5	(0.8, 2.6)	1.5	0.216
Married or Living with Partner	1.3	(0.8, 2.1)	1.1	0.304
Percent with High School Diploma or Less Education in Census Tract	1.0	(0.3, 3.3)	0.0	0.996
Two or More Addresses Collected (Movers)	0.5	(0.3, 0.7)	12.8	0.000
Currently Employed	0.5	(0.3, 0.7)	16.3	0.000
Percent Homeowners in Census Tract	0.3	(0.1, 0.6)	9.2	0.002
HH Income $\geq$ \$30k	0.2	(0.2, 0.4)	30.2	0.000
Any Tracing Contacts Collected	0.2	(0.1, 0.5)	15.4	0.000

\*Includes American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, Multiple, and Other Race

### 3.2.2 Reduced model

Table 3 provides results for the reduced regression model. In this model, factors that significantly increased a participant's likelihood of being lost to follow-up were Hispanic (OR = 1.8) or Non-Hispanic Other (OR = 1.9) race, residing in a Metropolitan Statistical Area (OR = 1.8), and primary caregiver being less than 25 years old at screening (OR = 1.7). Factors that significantly decreased a participant's likelihood of being lost to follow-up were having two or more addresses collected (OR = 0.5), being currently employed (OR = 0.5), being currently employed (OR = 0.4), a greater percentage of homeowners in the census tract (OR = 0.3), being recruited via the Provider Sample recruitment method (OR = 0.2), having a household income over \$30,000 (OR = 0.2), and providing one or more tracing contacts (OR = 0.2).



**Table 3:** Logistic Regression Reduced Model Results

Variable	Odds Ratio	95% Confidence Interval	Wald $\chi^2$	<i>p</i> -Value
Recruitment Method: Provider Sample	0.2	(0.1, 0.4)	17.6	0.000
Race: Hispanic	1.8	(1.2, 2.7)	7.8	0.005
Race: Non-Hispanic Other*	1.9	(1.1, 3.2)	4.8	0.029
HH Income $\geq$ \$30k	0.2	(0.1, 0.4)	37.7	0.000
Age: <25 Years Old at Screening	1.7	(1.1, 2.5)	6.9	0.009
Currently Employed	0.4	(0.3, 0.7)	17.1	0.000
Percent Homeowners in Census Tract	0.3	(0.1, 0.6)	10.9	0.001
Any Tracing Contacts Collected	0.2	(0.1, 0.4)	17.7	0.000
Two or More Addresses Collected (Movers)	0.5	(0.3, 0.7)	12.0	0.001
Resides in Metropolitan Statistical Area	1.8	(1.1, 3.1)	5.4	0.021

\*Includes American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, Multiple, and Other Race

The full and reduced models were quite similar in terms of the variables that significantly predicted lost to follow-up status and their scale, with the only exception being recruitment method. In each of the full and reduced models, one recruitment method was the only significant predictor of lost to follow-up status, and both the method and the direction of the association changed across the two models.

## 4. Discussion and Conclusions

### 4.1 Application

Our research has direct application for other longitudinal studies that need to maintain cohorts of children, pregnant women, or caregivers of young children over an extended period of time. Using these models, one can calculate the probability a given participant may be lost to follow-up, and adjust the level of interaction accordingly. Moreover, projects could consider incenting participants to provide more locating information, as well as to notify data collectors of moves or changes to contact information. Retention methods could be tailored to each participant, with those prone to attrition receiving more frequent check-ins. Information provided by this model could be used to direct efforts at participants with higher risk of being lost to follow-up, optimizing limited study resources.

### 4.2 Limitations

It is important to note that our data and analysis does carry specific limitations. Most importantly, the data are unweighted and thus not representative of the general population.

In addition, there is no formal, industry-standard definition of being “lost to follow-up”. This analysis did not include withdrawn, deceased, or refused participants in the lost to follow-up group, but further consideration of whether these groups represent participants that have attrited may be useful. It is also possible that some participants considered lost could return to the study in the future.

Future analysis should also take interviewer characteristics into consideration. Although these are likely to play a role in participants becoming lost to follow-up, we limited this analysis to individual characteristics, household characteristics, and operational indicators.

Lastly, logistic regression does not account for time-varying data, and we were restricted to using the most recent non-missing value for variables (such as income) that can change over time. NCS Vanguard Study questionnaires did not collect the time-varying variables included in our model at every time point, and combined with item non-response, we faced missing data at the study event level which would have made a more detailed model difficult to work with. In future analyses, a multi-level, mixed-effect regression model would be more appropriate.

Nonetheless, we consider the results to be useful for the purposes of understanding the types of participants likely to become lost to follow-up on longitudinal studies similar to the NCS Vanguard Study.

### **4.3 Conclusion**

We considered individual characteristics, household characteristics, and operational indicators in describing participants likely to become lost to follow-up on the NCS Vanguard Study. Consistent with the literature, participants lost to follow-up tended to be non-White and Hispanic. Additionally, young primary caregivers (25 years or younger at the time of screening) and individuals with a high school diploma or less education showed higher rates of attrition. In contrast, caregivers who were married or living with a partner, English-speaking, currently employed, earning \$30,000 or more per year, homeowners, and providers of multiple addresses and tracing contact information were lost to follow-up at lower rates. Over 18% of lost to follow-up participants did not provide any address information, and these individuals were much more likely to attrite than participants for whom we have address information.

We also developed regression models to predict whether a participant would be lost to follow-up or not. These models found that participants with increased risk of attrition were Hispanic or non-Hispanic other race, resided in urban areas, and were younger than 25 years old at the time of screening. Conversely, those with decreased risk of attrition were currently employed, lived in areas with higher concentrations of homeowners, and provided multiple addresses and non-household tracing contact information. Recruitment method can also influence whether a participant was lost to follow-up, and further research is needed into the relationship between recruitment method and retention status. It is likely that the protective nature of the Provider Sample recruitment method, which is significant in the reduced model, is due to its later rollout and thus shorter timeframe for a participant to become LTFU.

These findings can be useful in designing future longitudinal studies of young women, parents, and children. Identification of populations that are likely to be at high risk for loss to follow-up can allow for adjustments to the sample design as well as direction of increased retention efforts toward these participants to minimize bias from attrition.

### **Acknowledgements**

*Sources of funding:* This analysis was conducted as part of the National Children’s Study, supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and funded, through its appropriation, by the Office of the Director of the National Institutes of Health.

*Acknowledgements:* The National Children’s Study is indebted to the many communities, families and their health care providers, whose participation made this study possible. The authors of this manuscript wish to thank Elise Wisnieski and Shannon McEwen for their assistance in data manipulation and Lindsay Miller for her systematic review of the literature.

*Disclaimer:* The views expressed in this article are the responsibility of the authors and do not necessarily represent the position of the National Children’s Study, the National Institutes of Health, or the U.S. Department of Health and Human Services.

## References

- Benetsky, M.J., Burd, C., and Rapino, M. (2015), “Young Adult Migration: 2007–2009 to 2010–2012,” American Community Survey Reports ACS-31, U.S. Census Bureau.
- Campanelli, P., and O’Muircheartaigh, C. (1999), “Interviewers, Interviewer Continuity, and Panel Survey Nonresponse,” *Quality and Quantity*, 33(1), 59-76.
- Cohen, S.B., Machlin, S.R., and Branscome, J. (2000), “Patterns of Survey Attrition and Reluctant Response in the 1996 Medical Expenditure Panel Survey,” *Health Services and Outcomes Research Methodology*, 1(2), 131-148.
- Cotter, R.B., Burke, J.D., Loeber, R., and Navratil, J.L. (2002), “Innovative Retention Methods in Longitudinal Research: A Case Study of the Developmental Trends Study,” *Journal of Child and Family Studies*, 11(4), 485-498.
- De Keulenaer, F. (2005), “Using Process Data to Predict Attrition from a Panel Survey: a Case Study,” in *American Statistical Association Proceedings of the Joint Statistical Meetings*, pp. 3844-3851.
- Dodds, S., Furlong, A., and Croxford, L. (1989), “Quality and Quantity: Tackling Non-Contact Attrition in a Longitudinal Study,” *Sociology-The Journal of the British Sociological Association*, 23(2), 275-284.
- Downs TJ, Ogneva-Himmelberger Y, Aupont O, Wang Y, Raj A, et al. (2010), “Vulnerability-Based Spatial Sampling Stratification for the National Children's Study, Worcester County, Massachusetts: Capturing Health-Relevant Environmental and Sociodemographic Variability,” *Environmental Health Perspectives*, 118(9), 1318-1325.
- Graziotti, A.L., Hammond, J., Messinger, D.S., Bann, C.M., Miller-Loncar, C., Twomey, J.E., Bursi, C., Woldt, E., Nelson, J.A., Fleischmann, D., and Alexander, B. (2012), “Maintaining Participation and Momentum in Longitudinal Research Involving High-Risk Families,” *Journal of Nursing Scholarship*, 44(2), 120-6.
- Gustavson, K., von Soest, T., Karevold, E., and Roysamb, Ed. (2012), “Attrition and Generalizability in Longitudinal Studies: Findings from a 15-year Population-Based Study and a Monte Carlo Simulation Study,” *BMC Public Health*, 12, 918.
- Haunberger, S. (2010), “The Effects of Interviewer, Respondent and Area Characteristics on Co-operation in Panel Surveys: A Multilevel Approach,” *Quality and Quantity*, 44(5), 957.

- Klebanoff, M.A., Zemel, B.S., Buka, S., and Zierler, S. (1998), "Long-Term Follow-Up of Participants in the Collaborative Perinatal Project: Tracking the Next Generation," *Paediatric and Perinatal Epidemiology*, 12(3), 334-346.
- Montaquila, J. M., Brick, J.M., and Curtin, L.R. (2010), "Statistical and Practical Issues in the Design of a National Probability Sample of Births for the Vanguard Study of the National Children's Study," *Statistics in Medicine*, 29(13), 1399-90.
- Montaquila, J.M., Hsu, V., Brick, J.M., English, N., and O'Muircheartaigh, C. (2009), "A Comparative Evaluation of Traditional Listing vs. Address-Based Sampling Frames: Matching with Field Investigation of Discrepancies," in *American Statistical Association Proceedings of the Survey Research Methods Section*, pp. 4855-4862.
- Olson, K., and Witt, L. (2011), "Are We Keeping the People Who Used to Stay? Changes in Correlates of Panel Survey Attrition Over Time," *Social Science Research*, 40(4), 1037-1050.
- Venevongsa, J.C., Chooniedass, R., Kozyrskyj, A.L., Ramsey, C.D., and Becker, A.B. (2014), "Attrition in the Canadian Healthy Infant Longitudinal Development (CHILD) study," *Allergy, Asthma and Clinical Immunology*, 10(Suppl 1), A16.
- West, J., and Walston, J. (2003), "Staying the Course: Patterns of Nonresponse in a National Longitudinal Study of Young Children," in *American Statistical Association Proceedings of the Joint Statistical Meetings*, pp. 4500-4507.