

The challenges of locating young adults for a longitudinal study: Improved tracing strategies implemented for the National Longitudinal Study of Adolescent Health, Wave IV

Amanda Meehan¹, Erica Saleska², Marjorie Hinsdale-Shouse³
Nick Kinsey⁴, Casey Tischner⁵

^{1, 2, 3, 4, 5} Research Triangle Institute, 3040 Cornwallis Road, P.O. Box 12194
Research Triangle Park, NC 27709

Abstract

Maintaining the integrity of the sample for a longitudinal study is key to its success. For a longitudinal study involving a young cohort with long periods of time between waves, like the National Longitudinal Study of Adolescent Health (Add Health), the need for effective tracking is underscored. Add Health is unique in that sample members are followed throughout the study regardless of past participation. Since young adults are traditionally difficult to locate because they move frequently, the five to six years between waves of Add Health has meant multiple moves for many sample members since our last contact.

Add Health is a study of individuals who were first interviewed in 1994-1995. Three waves of in-home interviews were conducted in 1995, 1996, and 2001-2002. Add Health Wave IV began in 2008 with a cohort ages 24 to 32 years old. Approximately 1,880 of the Wave IV sample members had not been interviewed since 1995.

During Wave III, 87% of the sample was located. To increase the locate rate in Wave IV, we developed tracing steps that enhanced traditional tracing methods and improved upon methods used previously. The Wave IV plan included the development of an electronic locating system, which served as a central location for all contact information in a secured system on field laptops.

We will compare the locating rates between Waves III and IV and analyze the impact of various factors such as the availability of SSNs, when the sample member last participated and the likelihood of participation based on previous waves of participation. We will also examine the success of specific tracing strategies and describe how enhanced tracing methods such as the creation of an electronic locating system helped maintain the sample for this large-scale national survey.

Key Words: Locating, Longitudinal, Tracing

1. Introduction

The success of longitudinal study requires a carefully developed plan that effectively and efficiently locates and maintains the integrity of the sample. To achieve this, detailed planning is needed prior to data collection with multiple strategies and approaches to locate sample members. Recent literature indicates that high locate rates on a longitudinal study are possible when a project uses a tailored locating approach involving multiple searches of databases, telephone calls, and in-person visits (Haggerty et al., 2008). Locating sample members becomes increasingly difficult in a young adult population who are traditionally a transient population. Cotter et al (2005) found that as sample members in a longitudinal study grow older they are more difficult to contact because life patterns change, making follow-up difficult. In order to retain sample members, it is important to use a combination of locating approaches. Strategies can include sending introductory or contact letters to determine the sample members most recent contact information, batch tracing efforts using specialized vendors and searches through the internet. Other strategies frequently used and described in the literature include phone contact and in-person locating as necessary. Cotter et al. (2005) also summarized that there is no agreement in the literature for how much time should be spent locating participants. The key is that interviewers need to remain determined.

Weinberger et al. (2001) described a multistep locating protocol to locate subjects who originally participated in clinical studies 27 years ago. Their locating efforts relied on personnel who aggressively used various locating databases and telephone calls to locate 84% of individuals. Literature suggests that high retention rates for a longitudinal sample are possible, but locating strategies cannot fully rely on technology as a means to locate their sample. Technology should be used to enhance the information and leads the interviewer will use to find the sample member. There is no one strategy to achieve high locate rates. Success is dependent on the project team working together and remaining persistent (Cotter et. al 2005).

Add Health differs from many longitudinal studies because original sample members are followed since the first wave of the study, regardless of participation in additional waves. A portion of the Wave IV sample consisted of people who had not been interviewed since the first wave of data collection in 1994-1995. To achieve the target locate rate of 92%, a comprehensive tracing plan, which relied on a variety of approaches was implemented. Additional innovative tracing methods were developed throughout Wave IV data collection to locate sample members for our most difficult to find cases.

This paper will describe the locating and tracing efforts used during The National Longitudinal Study of Adolescent Health (Add Health), Wave IV and present preliminary locating findings. Add Health Wave IV relied on lessons learned from the previous wave to develop innovative and enhanced tracing approaches. An extensive tracing and locating plan was developed that used available locating information from Wave III of the study. We will discuss the use of intensive tracing efforts prior to data collection and how the use of technology expanded our ability to locate sample members. We will also compare strategies used in Wave III of the study and how approaches were improved or modified to better assist in our efforts.

2. Methods

2.1 Purpose of Add Health

The National Longitudinal Study of Adolescent Health (Add Health) is a longitudinal study designed and conducted under the direction of the Carolina Population Center (CPC) at the University of North Carolina at Chapel Hill (UNC), under a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health. The purpose of this congressionally-mandated study is to gather longitudinal data at a national level on health and health behaviors of adolescents and young adults. The study is designed to collect data that researchers and policy makers can use to identify ways to improve the health of the nation's youth and young adults. The fourth wave of the Add Health study included a pretest in 2007 and the main study conducted January 2008 through January 2009. Wave IV involved locating and interviewing as many of the original cohort of participants in 1995, now aged 24 to 32 to measure health attitudes, experiences, and behaviors of these same respondents as adults.

2.2 Sample

As a longitudinal study, Add Health follows a cohort of adolescents in 1994-95 through their transition to adulthood and now in adulthood at Wave IV. Add Health was first administered in 1994-1995 to around 90,000 students in 7th grade through 12th grade from 33 states to represent the nation as a whole. The sample for the in-school component was drawn from a nationally representative stratified probability sample of 80 high schools and scientifically selected middle schools that contributed adolescents to each high school's student body. These respondents were given a brief questionnaire in school. From this group of students and school rosters, a nationally representative sample of about 21,000 adolescents were sampled for an in-home interview. In 1995, these adolescents and one of their parents participated in the in-home interview. The in-home sample consisted of various subgroups including high education blacks, Cubans, Puerto Ricans, Chinese, and disabled persons; the sample also included twins, other siblings of twins, other full-siblings, half-siblings, and nonrelated household members. This first cycle of Add Health is considered Wave I, of this longitudinal study¹.

A follow-up interview, Wave II, was conducted in 1996 with the adolescents. The Wave II in-home sample was drawn primarily from the pool of participants in the Wave I in-home component, with the majority of the twelfth grade respondents not followed up in Wave II. Only the twelfth grade respondents who were part of the genetic pair sample were retained for Wave II. Wave II also included a small number of respondents who were not participants in the first wave, but were added in Wave II as supplements to the genetic pair sample. Table 1 provides an overview of the respondents' ages by data collection waves¹.

The data collected at Wave III in 2001-02 was designed to study the transition from adolescence to young adulthood. Specifically, questions were designed to determine how decisions made in adolescence influence the outcomes experienced in adulthood. The emphasis in Wave III was on the labor market, higher education, relationships, parenting,

¹ Aragon, et al (2003). The National Longitudinal Study, Wave III Final Report. Research Triangle Park, NC: RTI International.

and community involvement. Wave IV was a follow-up with 19,962 of the original Wave I in-home respondents. Wave IV respondents ranged from 24 to 32 years in age.

Table 1 Overview of Add Health Data Collection Waves and Age of Respondents

Year	Age of Respondents	Data Collection Waves
1994	11-18	Wave I – In-school interviews of adolescents in grades 7 through 12
1995	12-19	Wave I – In-home interviews of adolescents and parents
1996	13-20	Wave II – In-home follow-up interviews with adolescents
2001 – 2002	18-26	Wave III – In-home interviews with the same respondents from Wave I sample as young adults; subset of romantic partner.
2007 – 2009	24-32	Wave IV – In-home interviews with the same respondents from Wave I sample as adults

2.3 Wave IV Tracing Plan

A critical step in maintaining the integrity of the longitudinal cohort involved implementing an effective strategy to locate the sample members who had participated in past waves. At Wave IV, many of the sample members passed through one of the most transient times of their lives—the transition to young adulthood. Some, however, were still at an age of frequent moves due to the pursuit of education, new jobs, and families. In Wave III, we expected 25% of our population to have moved and in Wave IV the expected move rate was a little lower at 21% (US Census Bureau). The high mobility of young adults continued to pose a challenge for locating activities for Wave IV as it did during Wave III.

In Wave III, we faced locating difficulties because a third wave was not initially planned. Researchers determined after Wave II there was a great interest in a 3rd wave. Because the wave was not planned for, future contact information was not collected during Wave III. We had to rely on contact information from Wave I and Wave II which was at least 5-6 years old to begin our initial tracing efforts. During Wave III of Add Health (2001-2002), RTI located 37% of the main respondents through advanced tracing efforts. For the remaining cases, 26% were located through interactive tracing in RTI's Tracing Unit, and 12% were located in the field. In the end, about 13% were not located in Wave III and an additional 12% were considered ineligible, refusals or other non-interviews. The Wave IV tracing plan was devised to help increase the overall locate rate to help ensure response rate goals were met; at Wave IV, the study targeted an overall locate rate of 92%.

Our approach for Add Health Wave IV was tailored based on the quality of locating data and contact information from prior waves and our experiences during Wave III of the study. Based on the Wave III experience, strategies were implemented to maximize batch, interactive and field tracing for Wave IV. We also developed additional tracing methods over the course of data collection to further assist us in locating sample members.

The Wave IV tracing plan included the following components:

- Advance batch tracing
- Respondent mailing before data collection
- Interactive tracing through RTI's internal tracing unit
- Creation of Interactive Locating System (ILS) for use in field tracing
- Additional batch tracing during data collection
- Supplemental tracing strategies developed specifically for Wave IV cohort

Each of these components of the Wave IV tracing plan are documented below.

2.3.1 Advance Batch Tracing

Batch tracing is an automated service in which a person's name, address, and/or phone number are matched to information in multiple databases. The tracing vendors returned the most current address information for that person as determined by hard logic algorithms. By using these services, address confirmations or updates can be found quickly at a relatively low cost. Batch processes significantly reduce the percentage of cases requiring further tracing, thus lowering the overall cost of tracing efforts.

At Wave IV, we initially conducted batch tracing using the following sources: National Change of Address (NCOA), Transunion, Telematch, and Fast Data (FD) or Accurint (depending on the availability of Social Security Numbers). In Wave III, only NCOA and Telematch were used to search for updated respondent information prior to data collection. These advance batch tracing services were expanded from the services used in Wave III because we learned that address confirmation through NCOA follow up with Telematch was not sufficient for a highly mobile population. Another difference between the Wave III and Wave IV tracing protocols was that an initial postcard address update mailing was sent to all sample members and their parents before the batch tracing in Wave III. Only 7% of the postcards were sent back with address updates.

2.3.2. Respondent Mailing Before Data Collection

After completing batch tracing prior to data collection, RTI sent an update mailing to Add Health Wave IV sample members, using the best known address for the sample member and possibly the sample member's parent or other contact as well. The mailing occurred six weeks before beginning intensive tracing efforts and contained a letter reminding sample members about the study and asking them to update their contact information. Also, an update form was included that sample members were asked to complete by one of three methods: 1) filling out and returning the paper form using a postage-paid envelope, 2) calling a toll-free number, or 3) completing a web form online to provide updated contact information. This mailing also reminded all sample members of the study and included a return form for any updated address and telephone information. A toll-free number and a web update form were also included as a means to provide updated contact information. Any forms returned as undeliverable were subject to additional centralized tracing efforts by specialists in RTI's CCS Tracing Unit. The Wave IV approach to initial contacting was based on the previous wave of Add Health, but with several key adjustments to the implementation and timing. For Wave III, RTI mailed materials to all Add Health sample members and to some parents. The initial mailing was sent to the best address available from previous waves. The initial mailing included a detachable postage-paid postcard designed to capture updated contact information. This mailing was sent in March 1999 a year and a half before data collection started. The first mailing included a brief overview of the study and a request

to update the locating information for the sample members and family. Two additional mailings were sent before data collection began: 1) sample members were sent another address update letter and form requesting updated contact information and 2) about 83% of parents were sent a similar address update request for their child.. Sample members and parents were asked to update their information by returning the form via mail, calling TOPS to update over the phone, or sending an e-mail to RTI. The Wave IV approach, differed in that only one address update request was mailed prior to data collection, and sample members were given an option to update their contact information via the Add Health website rather than sending an e-mail. An e-mail address was also offered in Wave IV, but it was intended as a way for sample members to ask questions regarding participation or inform staff of website technical difficulties. Another key difference is that the Wave IV initial mailing was sent after the batch tracing for updated addresses was completed and letters were only mailed to parents' addresses if there was no viable mailing address for the sample member. In Wave III, the first mailing was sent before any batch tracing. The second and third mailings to sample members and parents were sent after two rounds of batch tracing updates using NCOA.

For the Wave IV pretest and main study, undeliverable address update letters were keyed into the Add Health website and coded as undeliverable. The data entry function of the website was accessible only to project staff. Undeliverable letters were sent weekly to RTI's internal tracing unit for additional tracing and forwarding addresses received were also used for additional tracing. This tactic provided an efficient method to send undeliverable cases for interactive tracing because we were informed of incorrect address prior to data collection. From our advance tracing mailings, we received approximately 20% (n=4,000) undeliverable letters. These cases were sent to the tracing unit for additional tracing before data collection. While the Wave III undeliverable returned mail rate was significantly lower than Wave IV, 11% (n=2,164) for the first mailing, this was not an indication that the Wave III mailing addresses were correct for the sample member. It is more likely that the address used was a valid address but it was for the parents. Returned mail for the initial mailing was tracked in Wave III but the cases were not sent directly to the tracing unit in Wave III. All cases were sent to NCOA after the results of the mailout were received. For the second and third mailings in Wave III, the return mail was not tracked. Instead, the field was instructed to contact the parents of any unconfirmed cases and try to reach the sample member through the parent or any other leads available.

Wave IV sample members provided updated contact information via web updates, e-mail and the project hotline. At the end of February 2009, updates received totalled:

- Cumulative Web/E-mail Updates: 5% of all sample members (n=1,089)
- Cumulative Phone Updates: 2% of all sample members (n=429)

With the addition of the web update form, sample members were more willing to update their contact information electronically than over the phone.

Wave III cumulative updates of all three mailings to sample members and parents were as follows:

- Initial Mailing (a year and a half before data collection): 7% of all sample members (n=1,546)
- Second Mailing to Sample Members: 13% of all sample members (n=2,610)

- Third Mailing to Parents: 2.9% of all sample members (n=564)
- Nearly all Wave III address updates came from the returned paper form. Less than 1% of all sample member returns were from telephone or e-mail updates.

2.3.3 *Interactive Tracing Cases*

Cases that did not yield new or confirmed contact information through advance tracing (including address update mailing and batch database searches) were sent to interactive tracing. Interactive tracing differs from batch tracing in that a tracer has the ability to assess each case on an individual basis to determine which resources are most appropriate and the order in which they should be used. Additionally, interactive tracing is much more detailed due to the human review of information. During interactive tracing, tracers utilized all contact information collected during all waves of Add Health. Tracers made decisions about cases based on this contact information as well as information obtained from batch tracing and database searches. Our Wave IV tracing plan called for sending cases that did not generate new leads through batch tracing directly to interactive tracing prior to data collection starting. This was an assurance, that we were starting data collection with as many updated leads as possible.

In Wave III of Add Health, only 32% of the cases were sent to interactive tracing prior to data collection starting and not all cases were finalized in RTI's tracing unit prior to data collection. Only 2.8% of all cases were located interactively before data collection began. Unresolved cases were sent to the field to be confirmed. Any cases that could not generate viable leads in the field were returned to the centralized tracing unit for additional locating. The Wave IV approach was developed to reduce the need for field tracing effort that was required in Wave III. Additional time for advance tracing was built into the schedule for Wave IV to make this possible.

Cases in Wave IV sent to interactive tracing prior to data collection were given an average 85 minutes level of effort for tracing. This meant tracers had an average of 85 minutes to search for a confirmed lead for that case. Before data collection started cases that were not located within 85 minutes were given an additional 45 minutes of extra tracing work to try to find good leads before sending to the field. During the pre-data collection phase of interactive tracing, 44% of the cases (n=8,779) were traced and 36% of all cases (n=7,069) were located through interactive data collection prior to fielding.

Once in the field, cases that could not be located were sent back to the central office for interactive tracing. At the beginning of this phase of tracing, cases had a 60-minute level of effort for tracing, with the option to add additional thirty minutes if there were still viable leads to follow up. Toward the end of data collection, the parameters were adjusted so that cases that had never been to tracing before received a 75-minute level of effort, while cases that had been to tracing before received a 45-minute level of effort. During the second phase of interactive tracing, 21% of the cases (n=4,128) were traced and 14% of cases (n=2670) were located.

The level of effort for interactive tracing for Wave III was up to 1.5 hours per case. This was a cumulative level of effort that included any tracing work done before data collection as well as tracing for cases returned from the field during data collection.

2.3.4 Creation of Interactive Locating System (ILS)

The Interactive Locating System (ILS) was developed to house all the components of sample member information in a secured system on a field interviewer's laptop computer. Utilization of the ILS allowed Add Health to conform to rigorous data security standards by keeping sample members' personal information in the secure laptop instead of on paper and provided the interviewer with all tracing leads in one location.

The ILS was also created for Wave IV to better maintain, view and track the tracing leads. In Wave III of Add Health, tracing leads were tracked from paper records that were distributed with the initial case assignment and then updated and mailed to the interviewer when a case completed interactive tracing. Creating the ILS allowed the interviewer to see a history of all leads ever found for a case and allowed for better reporting of all of our tracing cases.

The ILS contained updated contact information from RTI's internal tracing unit and from previous waves. Contact information came from the following sources:

- 1) The Case Management System (including preload information)
- 2) RTI's In-house Tracing Unit
- 3) Field Supervisors/Tracers
- 4) The field interviewer as they enter newly acquired leads

Field Interviewers accessed the ILS via the case management system loaded on their laptops. The initial grid in the ILS displayed a table format with each row representing a "lead" for the targeted sample member. Each lead contained specific information about the sample member including:

- Status of lead
- Date of lead
- Name (first, middle, and last)
- Street address
- City
- State
- Zip Code
- Home, Work, Cell phone number
- Comments from the tracing unit if applicable

In addition to the sample member's locating information, the ILS contained names, addresses, and phone numbers for additional contacts, previously provided by the sample member as someone who could help locate them in the future (such as a parent, relative, or friend). These other contacts were helpful for difficult-to-locate sample members. The ILS also provided interviewers with a script to contact and verify respondent information and to update new leads obtained in the field. Once field interviewers verified a sample member's contact information through the ILS, the Add Health interview was made available.

2.35 Additional Batch Tracing

In the fall of 2008, additional tracing strategies were implemented to improve sample member locate rates. Additional batch tracing services were used to generate potential

new leads for any difficult-to-locate cases. With the additional batch tracing, RTI developed 4,123 new leads for approximately 2,000 for our difficult to locate cases.

The additional batch tracing for Wave IV included the following vendors:

- Accurant Comprehensive People Batch
- FirstData address history batch with SSN lookup
- FirstData premium phone search
- Experian Residential Database Search Packages – MetroNet SSN Fall Thru
- Pension Benefits Information (PBI)

New tracing vendors not used for the advance tracing were used in the late fall 2007 for all pending unlocated cases. A similar approach was used for the difficult-to-locate cases in Wave III. In Wave III, 10,803 cases were sent to Lexis-Nexis for batch tracing. New leads were provided for 8,459 cases.

2.36 Supplemental Tracing Strategies

For Wave IV, RTI supplemented the more traditional approaches to tracing with additional locating efforts that have proven effective in locating a wide variety of populations. We experimented with all resources throughout data collection to identify what were the most cost-effective and fruitful tracing strategies to locate sample members.

Social Networking Websites. Toward the end of data collection, RTI decided to use Facebook and MySpace to generate new leads for the most difficult-to-locate sample members. MySpace and Facebook are both popular social networking sites that have roughly 100 million users each and target 18-35 year olds. Given the age range of Add Health sample members, we felt these websites could help us in our locating efforts to find as many sample members as possible before the end of data collection. Because Add Health is also in the unique position of having access to sample member age and high school information, we felt searching for sample members on MySpace and Facebook using these criteria would better ensure that the correct sample member was located.

When a sample member received a message, the sample member could either reply with another private message, respond by e-mailing us at the Add Health project e-mail account, or call the toll-free project number. If a sample member posted contact information such as an address, phone number, or e-mail address on their profile page, we also attempted to contact the respondent using that information. For Wave IV, we searched for 2,450 sample members and sent messages out to 359 sample members. In total, we received approximately 40 responses. This method did not provide a substantial boost to our locate rate in Wave IV, but did help to track down some of the more difficult to locate cases at the end of the data collection period.

The only social networking site utilized in Wave III for supplemental tracing was classmates.com. Because these searches were handled by the field supervisors, we do not have any data on the usefulness of the tracing efforts.

Respondent Assisted Tracing (RATS) We also attempted to improve the sample member locate rates for Wave IV utilizing a special type of tracing that was used successfully in Wave III, but was not part of the original tracing plan for Wave IV because the cohort was older and less likely to be in contact with friends from high

school. However, we did implement a modified version of the original respondent-assisted tracing (RATs) for Wave IV. This unique tracing activity involved re-contacting respondents to ask them if they would be willing to provide contact information for any of the unlocated sample members who went to the same school in their 1994/95 school year. This method was employed in January 2009. We only contacted sample members from schools where all students were a part of the original Add Health sample so as to protect the privacy of the sample members.

We identified cases to be contacted by grouping pending cases by school name electronically. Contacts were made by RTI project staff. For each school that had one or more pending cases, a similar list of completed interview cases was created and made available electronically.

Each Add Health respondent was re-contacted one time only to request address updates for any sample members who attended their school. Staff making the contacts used a scripted introduction that prompted the respondent to provide address, phone number, e-mail and other contact information. Staff never revealed the name of any of the sample members who were currently unlocatable; the sample member participating in RATs had to offer the names of classmates they knew how to contact. New leads were recorded in the database system and then transferred to the ILS as a new lead for the pending case.

The RATs approach used in Wave IV differed from how RATs was used in Wave III. In Wave III, RATs was incorporated into the Add Health interview. At the end of the interview, respondents, were asked for assistance in finding classmates that had not yet been interviewed. When a respondent could provide updated locating information about another respondent, the interviewer recorded the lead and reported the new locating information to the supervisor (if the information was for a case not in his/her own assignment). We modified this approach in Wave IV as one of our final attempts to try and locate our most difficult cases.

Using the modified (RATs) approach, we attempted to reach 169 respondents. Out of those respondents, we spoke to 46 people, and 3 of them provided some kind of information for unlocatable sample members. We generated 3 new phone numbers, 1 case generated an e-mail only, and 1 case generated MySpace info (which we had already obtained from our own MySpace searches.) While the results of RATs at Wave IV did not make a significant impact on our locate rate, the approach did help us to identify the location of a few cases we would not otherwise have been able to locate.

Detailed Case Review. Another key component of the tracing process was careful case review of information on difficult-to-locate respondents; Supervisory staff would use information available in the ILS leads and electronic record of calls to determine if all leads had been attempted and whether or not leads needed to be recontacted; specific times and strategies were then developed for recontacting leads.

Supervisory staff also would use approved web searches to develop potential new leads for field interviewers, if available ILS leads did not pan out. Case reviews also identified cases that had an e-mail address associated with the case; field supervisors were asked to e-mail these respondents in the hopes of reaching them via e-mail to further discuss the

study and potentially schedule an appointment. Another option was to send the sample member a “trying to reach you” letter, if case review revealed that the field interviewer was having trouble reaching the respondent.

Case reviews were also essential to identify cases that were ready to move back to the interactive tracing step with the RTI tracing unit. If a case did need to return for additional interactive tracing, case review helped identify known dead ends so that our tracing unit could focus on new leads that the interviewer had not already ruled out during the field tracing process. The case review protocol for Wave IV was more deliberate and organized than Wave III in that specialized staff were assigned to review cases in great detail. In Wave III, case review was done primarily by field management staff who had limited time to devote to case review. The level of review for Wave IV helped ensure that all tracing leads were being followed up as expeditiously as possible.

3. Results

3.1 Overall Results

At the completion of the Wave IV main study data collection, we had located 18,289 of the 19,763 Wave IV sample members – the Wave IV locate rate of 92.5% was an increase over the Wave III locate rate of 87%. Located cases are defined as cases where an interview was complete and then cases where we were able to confirm the sample member but did not complete an interview. These located non-interview cases include: unavailable, deceased, refusals, physically/mentally disabled, incarcerated, institutionalized, out of country, and active duty military cases. The comparison of percent locates for Waves III and IV by each tracing step is shown in Table 2; these data include both main study and pre-test cases. The results presented are preliminary, pending the confirmation of deceased cases from the National Death Index. Our results also do not include Wave IV pretest cases, because the specific tracing strategies discussed in this paper were not implemented until the Wave IV main study. In reviewing the data across waves, we were able to locate more cases at Wave IV in our first tracing step that included batch tracing and interactive tracing prior to the start of data collection. This then reduced the number of cases located at later stages, allowing for more efficient use of interviewer time. It is important to note that the locate percentages from Wave III only account for located interview cases, while at Wave IV, locate percentages account for both located interview and non-interview cases. To account for this difference, there is a category for Wave III data only that shows located non-interviews because we cannot identify the source that lead to the location of those respondents.

Table 2 Wave III and Wave IV Comparison -- Tracing Step Locates

Tracing Step at which Sample Member Located	Wave III	Wave IV
Advance tracing (batch tracing or interactive tracing prior to data collection start)	37%	66%
Interactive tracing (after start of data collection)	26%	10%
Located in the field	12%	14%
Located via other tracing step	0%	3%

Located noninterviews, unknown source ²	12%	0%
Not Located	13%	8%

3.3 Impact of tracing factors

To understand our improved locate rate result at Wave IV, we evaluated certain key locating criteria to determine their impact on our locating efforts; specifically, we reviewed the following data from the Wave IV main study: age, the last wave that the sample member participated in, whether or not the sample member was a refusal at Wave III, and the availability of sample member Social Security Number (SSN).

Age. In reviewing the age of our sample members, we divided the population into two groups; 46% of our Wave IV sample members were ages 24 to 28 years, and the remaining 54% of our Wave IV sample members were 29 years of age or older. Out of all located cases, the younger age group accounted for 47% of our located sample members, while the older age group accounted for 53% of our located sample members. Table 3 shows the percentage of sample members who were located vs. not located among these two different age groups. Comparing the significantly different locate rates between the two age groups, our tracing effort proved to be more effective for our younger population, aged 24-28 ($\chi^2=103.578$, $p<.05$). It may be that locating our younger population was easier due to more cooperative parent contacts; the younger population may also have been more motivated by the interview incentive and were more likely to respond to our contacting efforts.

Table 3 Wave IV Located Sample Members - Age Comparison

Age Group* p<.05	Total Number of Sample Members in Group	Percentage Located	Percentage Not Located
24-28 years old	9,016	95%	5%
29 – 32+ years old	10,742	91%	9%

*Table 3 does not include five cases that do not have ages.

Last Wave of Participation. We reviewed the last wave of participation for each of our Wave IV sample members; 9% of our Wave IV sample members last participated at Wave I, 15% last participated at Wave II, and 76% of our sample members last participated at Wave III. Out of all located cases at Wave IV, the sample members who last participated at Wave I accounted for 8% of our located sample members; the sample members who last participated at Wave II accounted for 13% of our located sample members, and the remaining 79% of our sample members were last interviewed at Wave III. Table 4 shows the percentage of sample members who were located vs. not located among these three groups, differences that were found to be significantly different ($\chi^2=1012.79$, $p<.05$). As expected, we were able to locate a higher percentage of respondents at Wave IV, based on the last time the sample member participated. Sample

² In Wave III, located non-interviews were not accounted for in the locate percentages; thus there are 12% of Wave III cases where we cannot identify the tracing step that lead to locating the respondents of located non-interview cases.

members who participated in Wave III were easier to locate than those who had only participated in Wave I or Wave I and II. These findings were anticipated as Add Health had more recent locating information for Wave III respondents over the other groups; Wave III respondents were also better able to remember their participation from the previous wave, thus making contact with them an easier process.

Table 4 Wave IV Located Sample Members - Last Wave of Participation Comparison

Last Wave of Participation* p<.05	Total Number of Sample Members in Group	Percentage Located	Percentage Not Located
Wave I	1,882	79%	21%
Wave II	2,909	85%	15%
Wave III	14,972	96%	4%

Refusal Status at Wave III. In reviewing the refusal status from Wave III, 5% of our total Wave IV sample members were refusals at Wave III, with the remaining 95% of the Wave IV sample being a non-refusal at Wave III. Out of all located cases at Wave IV, the sample members who were Wave III refusals accounted for 5% of our located sample members, and sample members who were non-refusals accounted for 95% of our located sample members. Table 5 shows the percentage of sample members who were located vs. not located across the Wave III refusal status. There was a small difference in locate rates between the Wave III refusal status groups, which was found to be significantly different ($\chi^2=4.907$, $p=.026$).

Table 5 Wave IV Located Sample Members – Wave III Refusal Status Comparison

Wave III Refusal Status* p=.026	Total Number of Sample Members in Group	Percentage Located	Percentage Not Located
Wave III Refusal	1,042	91%	9%
Wave III Non-refusal	18,721	93%	7%

Social Security Number availability at Wave IV. We also reviewed the impact of having a Social Security Number (SSN) that could be used for tracing purposes in regards to our Wave IV tracing plan. We had SSNs that could be used for tracing for 60% of our total Wave IV sample members, with the remaining 40% of the Wave IV sample not having an SSN or not permitting us to use the SSN for tracing purposes. Out of all located cases at Wave IV, the sample members for whom we had an SSN we could use for tracing accounted for 62% of our located sample members, and sample members for whom we did not have an SSN to be used for tracing accounted for 38% of our located sample members. Table 6 shows the percentage of sample members who were located vs. not located across SSN status. Not surprisingly, when SSN was available for tracing, we located a significantly higher percentage of cases ($\chi^2=177.537$, $p<.05$).

Table 6 Wave IV Located Sample Members – SSN Status Comparison

SSN Status* p<.05	Total Number of Sample Members in Group	Percentage Located	Percentage Not Located
SSN available for tracing	11,904	95%	5%
SSN not available for tracing	7,859	89%	11%

4. Conclusion

The locating improvements made from Add Health Wave III to Wave IV made a difference in the number of cases we were able to find to achieve our 92% locate rate. Our improvements started with the advance planning and improvements to the tracing plan based on Wave III experience. Leveraging the lessons learned from Wave III by strategically implementing necessary modifications made an impact on the locating rates overall. The most significant change was starting the tracing process well in advance of data collection and implementing multiple rounds of batch and interactive tracing rather than multiple rounds of address update mailings, as was done in Wave III. The result was that we were able to get many more updated cases to the field at the beginning of data collection.

We were also able to improve our batch tracing prior to data collection by expanding batch tracing vendors used in Wave III. The additional batch tracing searches completed during data collection also allowed us to find leads for our very difficult to locate cases. As shown in Table 2, our expanded advance tracing steps achieved in Wave IV decreased the amount of tracing work needed during data collection, which saved the project time and money. Advance tracing steps included the batch tracing searches we conducted and the interactive steps taken before cases were released to the field.

The creation of the Interactive Locating System (ILS) allowed the project to better track all case leads and communicate in a standardized manner to ensure that all leads were accounted for. Also, by eliminating the need to rely on paper to view leads provided a greater efficiency in distributing leads to the field from RTI's tracing department. In Wave IV, we also implemented more systematic and detailed case reviews by supervisory staff. This allowed us to manage leads and confirm the field was following up thoroughly.

Our use of Respondent Assisted Tracing (RATS), differed from Wave III and was not something we implemented until the end of data collection. In reviewing leads received from this technique, it did not provide a significant help to improving our locate rate. If this approach were to be used in a future wave of the study, it would need to be implemented during the interview and consistently updated throughout data collection to be effective in locating sample members.

The use of social networking sites also did not have a significant impact on our locating rates because the approach was not taken until later in data collection and only with the most difficult-to-locate cases. In any future waves, social networking websites could be used as part of our advanced tracing steps when possible and over the course of data collection. We feel this approach would allow us to locate more of our sample members, especially considering the popularity of social networking with our target population. In

the future, tracing through social networking sites could be another means to improve locate rates and trace sample members more cost effectively. The remaining challenge to using these sites is the level of effort required to conduct the searches. It can also be difficult to confirm that the correct sample member has been identified.

While it is important to use a variety of strategies to locate sample members, the type and age of information available to find a sample member obviously is a factor. As indicated in Table 4, we were able to locate 96% of our Wave III sample members at Wave IV because we had more recent locating information. These sample members remembered participating in the last round of the study, making them more likely to participate in Wave IV. Also, the availability of tracing with a SSN increases your chances of finding the most recent contact information for an individual. Using an SSN is extremely helpful during batch tracing and interactive tracing when searching the various locating databases we used.

Before Wave IV began, we understood that tracing a young adult population would be a challenge. We allotted tracing resources to help us overcome this challenge and adapted our tracing strategies along the way to achieve our necessary locate rate. As the literature suggested, there is not a clear tracing method you can use to locate a longitudinal sample population. Numerous approaches are needed and an understanding of challenges faced during previous waves can allow tracing methods of improvement to be implemented. On Add Health Wave IV, we looked at the locating obstacles we faced during Wave III and improved on a variety of strategies to achieve the project goals. For future waves of Add Health, it will be important to continue the tracing methods used in Wave IV, but also use expanded batch tracing searches prior to data collection. The use of social networking sites prior to and throughout data collection to locate more of our sample members should be explored. Locating for future waves may not be as challenging because the population will be older and less transient. The key to success will be to weigh all these factors and determine what activities are most likely to improve the advanced tracing efforts as this will ultimately conserve the resources needed to locate the most difficult sample members.

References

- Aragon, A., Biemer, P., Causey, E., Hinsdale, M., Thornburg, V (2003). The National Longitudinal Study of Adolescent Health, Wave III Final Report (prepared for the Carolina Population Center under contract number UNC Subcontract No. 5-51757 Prime Grant No. 5 P01 HD31921-09 CFDA No. 93.864). Research Triangle Park, NC: RTI International.
- Cotter, R., Burke, J., Stouthamer-Loeber, M., & Loeber, R. (2005). Contacting participants for follow-up: how much effort is required to retain participants in longitudinal studies? *Evaluation and Program Planning*, 28, 15-21.
- Haggerty, K., Fleming, C., Catalano, R., Petrie, R., Rubin, R., & Grassley, M. (2007) Ten years later: Locating and interviewing children of drug abusers. *Evaluation and Program Planning*, 31, 1-9.
- U.S. Census Bureau. Data from 1980, 1990, and 2000 Decennial Census. www.census.gov
- Weinberger, M., Weinberger, M., Fineberg, N., Fineberg, E., & Wagner, U. (2002). *Journal of Clinical Epidemiology*, 55, 230-234.