Add Health
The National Longitudinal Study of Adolescent to Adult Health

Report prepared by

Robert A. Hummer

Sarah C. Dean

Dan Liao

Darryl Cooney

Caroline Jackson

Adrienne Griffiths

Allison E. Aiello

Wave VI Sampling and Mixed-Mode Survey Design

CAROLINA POPULATION CENTER | CAROLINA SQUARE - SUITE 210 | 123 WEST FRANKLIN STREET | CHAPEL HILL, NC 27516

https://doi.org/10.17615/ydjb-4305

Acknowledgement:

Data collection and dissemination for Wave VI of Add Health was funded by cooperative agreements U01 AG071448 (PI: Robert A. Hummer) and U01 AG071450 (MPIs: Robert A. Hummer and Allison E. Aiello) from the National Institute on Aging to the University of North Carolina at Chapel Hill, with cooperative funding from five other institutes and offices at the National Institutes of Health: the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institute on Minority Health and Health Disparities, the National Institute on Drug Abuse, the Office of Disease Prevention, and the Office of Behavioral and Social Science Research. Waves I-V of Add Health were funded by grant P01 HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Important Note: The above acknowledgement should be included in all presentations and publications using data from Wave VI of Add Health.

Citation for User Guide:

Hummer RA, Dean SC, Liao D, Cooney D, Jackson C, Griffiths A, Aiello AE. 2025. Add Health Wave VI Documentation: Wave VI Sampling and Mixed-Mode Survey Design. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Available from: https://doi.org/10.17615/ydjb-4305

Table of Contents

1. Introduction, Overview of Key Results, and Survey Data Files

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a nationally representative sample of U.S. adolescents who were in grades 7-12 during the 1994-1995 school year. Using a complex, school-based cluster-sampling frame, researchers selected high school and feeder school pairs from 80 communities across the United States. Wave I drew a sex- and grade-stratified random sample of roughly 12,000 adolescents. Special oversamples (e.g., racial/ethnic minority adolescents, genetic pairs) were added to that core sample to arrive at 20,745 adolescents for baseline inclusion in the study at Wave I. This sample has been followed from adolescence into early midlife across six waves of data collection to date, with the most recent wave of data collection (Wave VI) taking place between 2022 and 2025 when participants were ages 39 to 51, with an average age of 44.

Add Health Wave VI was approved by the University of North Carolina Institutional Review Board (IRB), which ensures the rights and welfare of human subjects, including data confidentiality, consent, and participant privacy. All Add Health participants provided informed consent for participation in Wave VI. Additionally, the amount of the incentive payment(s) provided to participants for their Wave VI participation was approved by the University's IRB.

Table 1 shows the overall results of Wave VI. A total of 11,979 Add Health sample members answered the Wave VI survey. Wave VI participants were part of either Sample 1 (largely web-based participants) or Sample 2 (largely in-person based participants). Together, Sample 1 (N=9,366) and Sample 2 (N=2,613) comprise the Wave VI total sample of 11,979.

Table 1. Key Wave VI Sample Size Results

	Wave VI Eligible and Fielded Cases*	N (completes)	Raw Response Rate (RRR)	Effective Response Rate (ERR)
Sample 1	15,103	9,366	62.0%	69.7%
Sample 1 (Phase 1)	15,103	8,200	54.3%	NA
Sample 1 (NRFU)	3,457	1,166	33.7%	NA
Sample 2	3,978	2,613	65.7%	65.7%
Total	19,081	11,979	62.8%	68.9%

*Ineligible cases and cases not fielded are excluded. Please see the design overview below for details.

The raw response rate (i.e., complete cases / eligible and fielded cases) for Wave VI was 62.8%, shown in the bottom row of Table 1. The effective response rate (ERR) for Wave VI was 68.9%, also in the bottom row of Table 1. The ERR takes into account that two phases of data collection were used in Sample 1. First, web-based attempts were made for all eligible cases in Sample 1, yielding 8,200 Phase 1 completes, or a rate of 54.3%. Second, strategic in-person contacts were made for approximately one-half of nonrespondents (i.e., non-response follow-up, or NRFU) in Sample 1, yielding an additional 1,166 completes in Sample 1. Together, the two phases of Sample 1 resulted in an 'effective response rate' of 69.7% for Sample 1. The term 'effective

response rate' refers to the idea that a two-phase strategy of data collection (for Sample 1 in this case) results in the same non-response bias risk compared with a single-phase data collection process with a 69.7% response rate (see Biemer et al. 2022 for more discussion of the ERR).

The Sample 2 response rate was 65.7%, as shown in Table 1. As mentioned above, Sample 2 was conducted nearly all in-person, which was important for collecting measures of cognitive, physical, and sensory function added specifically for Wave VI.

For an overview of Wave VI results in comparison to earlier waves of Add Health, please see the cross-wave Add Health user guide, "Understanding Sample Sizes, Basic Sample Composition, and Cross-Sectional Sampling Weights, Add Health Survey Waves I-VI" (Griffiths et al., forthcoming).

Most analysts will likely use the total Wave VI sample size of 11,979. This is because the survey instrument was completely consistent across Sample 1 and Sample 2, with only the mode of data collection differing. Wave VI survey weights, both cross-sectional and longitudinal, are available in order to weight Wave VI so that it is nationally representative of the population from which it was drawn: adolescents enrolled in grades 7-12 in the United States in the 1994-95 school year. Please see the Wave VI weighting user guide for details (Liao et al. 2025). For more detailed Add Health weighting and analysis suggestions, please see "Guidelines for Analyzing Add Health Data" (Griffiths et al., forthcoming) for more guidance.

Some analysts will be solely interested in Sample 2 of Wave VI (N=2,613), largely collected in-person. In addition to survey data, Sample 2 included in-person assessments of cognitive, physical, and sensory function that are available in separate Wave VI data files, each with their own user guide (data forthcoming in fall 2025). Importantly, Sample 2 included oversamples of Black, Hispanic, and Asian American Add Health participants, making it useful for analysis of cognitive, physical, and sensory functioning disparities. Sample 2 weights, both cross-sectional and longitudinal, are also available so that results from Sample 2 are nationally representative of the population from which it was drawn: adolescents enrolled in grades 7-12 in the United States in the 1994-95 school year. Please see the updated Add Health weighting user guide, "Guidelines for Analyzing Add Health Data" (Griffiths et al., forthcoming) for guidance.

Corresponding to the above user needs, there are three core survey-based data sets disseminated for Wave VI:

1) Wave VI Mixed-Mode Survey (which is inclusive of both Sample 1 and Sample 2)
2) Wave VI Sample 2 Survey (for analysts who are interested in Sample 2 only)
3) Wave VI Survey Weights File

A fourth survey-based data set focuses solely on the medications that Wave VI participants reported using. More information on the Wave VI Survey Medications Inventory can be found in the associated user guide (Angel et al., 2025).

Finally, a fifth survey-based file includes final disposition data on the full set of cases that were initially deemed eligible for inclusion in Wave VI: Wave VI Final Disposition of Cases.

The remainder of this user guide delves more deeply into the Wave VI sample design and provides an overview of the mixed-mode survey.

2. Overview of the Wave VI Sample and Sample Design

2.1. Eligible and Fielded Cases

The vast majority (total of 19,446) of the 20,745 original Add Health Wave I participants were initially deemed to be eligible for participation in Wave VI and are included in the Wave VI Disposition File. The four criteria that were used to exclude Wave I Add Health sample members from initial eligibility included:

1. Sample members who were confirmed deceased at the *start* of Wave VI (n=578);
2. Sample members who had an unconfirmed deceased status at the *start* of Wave VI (n=31);
3. Sample members who have no Wave I grand sample weight and have been *excluded since Wave II* (n=687); and
4. Sample members who were *hostile refusal cases in the past* and who were excluded from the frame starting in Wave VI based on the decision of the UNC principal investigator (n=3).

As Wave VI data collection unfolded, an additional small number of Wave I participants (n=220) were deemed to be ineligible for Wave VI for the following reasons:

5. Additional sample members who were confirmed to be deceased *during* the Wave VI field period (n=136); and
6. Sample members who were *institutionalized, including those in long-term incarceration, during the entire Wave VI field period* (n=84).

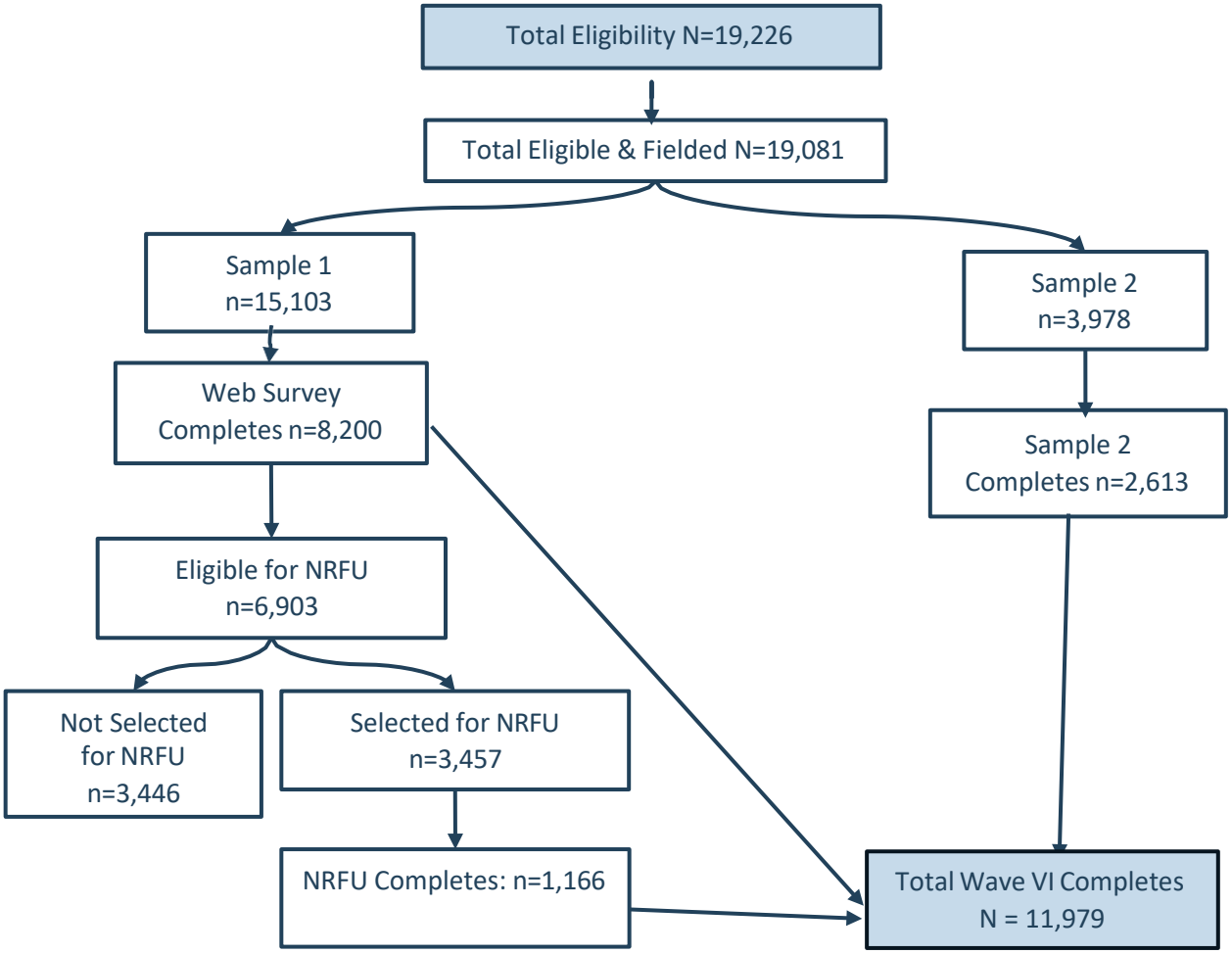
Subtracting out the six sets of ineligible sample members from the initial 20,745 participants yielded a final total of 19,226 eligible cases for Wave VI.

Although eligible for participation, an additional 145 Wave I participants were not fielded in Wave VI (either web-based or in-person) because, with certainty, there was no way to complete interviews with them. For example, they do not have name or identity information or they are permanently physically/mentally incapable. These 145 cases were considered to be in- scope (i.e., eligible) and thus are included in the Wave VI weighting process. All told, then, a total of 19,081 cases were eligible and fielded in Wave VI, as shown in Table 1 above.

2.2. Wave VI Two-Sample Design and Modes of Data Collection

As shown in Figure 1 below, the pool of eligible participants was split into two nationally representative subsamples. Sample 1 members received a primarily web-based, two-phase mixed-mode design. Sample 2 members received an in-person interview which included the Sample 1 survey content plus in-person cognitive, physical, and sensory functioning assessments. At the end of Sample 2 data collection, a small number of Sample 2 members who could not schedule an in-person interview were permitted to complete the web survey and thus did not receive the in-person assessments.

Figure 1: Overview of Wave VI Sample Design and Completes



All sample members selected for Sample 1 were asked to complete a web-based survey during Phase 1. Phase 1 was split into two subsamples, a pilot and main sample. The Sample 1 pilot was administered to test a set of new web-based cognitive assessments from TestMyBrain (TMB) (The Many Brains Project. URL: https://www.manybrains.net/; see Singh et al. 2021). The aim was to evaluate whether early or late survey placement of TMB impacted the participants' survey performance and/or increased survey breakoffs. An experiment was conducted which randomly assigned participants to the differing TMB placements and the best performing placement was selected for the main Sample 1 data collection. The TMB data set will be released in fall 2025 and will have its own user guide (Aiello et al., forthcoming 2025).

Multiple recruitment contacts of varying kinds (i.e., email, letter, and text) were sent to Sample 1 fielded cases to elicit response. After approximately 12 months of recruitment, Sample 1 nonrespondents were then eligible to be sampled for the Phase 2 Non-Response Follow-Up (NRFU). A subsample of approximately 50% of Sample 1 nonrespondents were selected to receive the NRFU data collection protocol. NRFU data collection procedures involved field interviewers contacting the nonrespondents and asking them to complete, via self-administration, the full web-based survey on a laptop provided by the interviewer in the participant's home (or another

location, at the request of the participant) or to go online and complete the web-based survey at their convenience. After several months of interviewer recruitment effort, the nonrespondents in the NRFU sample were then sent a series of final reminders to complete the web-based survey without the aid of an interviewer.

Sample 2 eligible and fielded participants were likewise split into a pilot and main sample. The Sample 2 pilot was necessary to test the in-person cognitive, physical, and sensory functioning assessments new to this wave (i.e., grip strength, hearing test, animal fluency, NIH toolbox assessments) and make necessary protocol adjustments. Unlike Wave V, the survey content for the bulk of Sample 2 was self-administered. This decision was made to reduce participant burden after the Sample 2 pilot, which demonstrated that the duration of the full Sample 2 protocol (survey plus interviewer- administered assessments) was more than 2.5 hours. Switching to a self-administered survey significantly reduced the time to complete the full Sample 2 protocol. While this approach differs from prior waves, there was precedent for the participant completing computer-assisted self- interview (CASI) sections containing sensitive questions during prior waves; therefore, CASI survey completion was familiar to our participants.

3. Systematic Scheme to Select the Two Samples

The sampling frame for Wave VI consisted of 19,226 eligible Wave I sample members, after excluding all of the ineligible cases as described above. To ensure that Samples 1 and 2 were random samples of the entire Wave VI sample and each was representative of the Add Health target population, a stratified systematic sampling scheme was implemented to draw the two samples.

The Sample 2 starting sample consisted of 1,488 eligible cases from Sample 2b of Wave V, which was collected entirely in-person. Given that Sample 2b of Wave V was relatively small and not very racially/ethnically heterogenous, we selected 2,490 additional cases for Sample 2, referred to as the Sample 2b Supplement. Together, this resulted in a total eligible and to be fielded sample size of 3,978 for Sample 2 of Wave VI, shown in the upper right of Figure 1. Because the cases from Sample 2b of Wave V were fixed, the Wave VI sample selection concentrated on selecting the Sample 2b Supplement from Wave V Samples 1, 2a, and 3 via stratified random sampling, while the remaining cases in the Wave VI frame (n=15,103, shown in the upper left of Figure 1) were included as eligible for and fielded in Sample 1. The target sample sizes of the Sample 2b Supplement by eight domains of interest were determined based on power analysis to ensure an adequate sample size for detecting small to medium effect sizes across all domains.

Table 2. Sample Sizes by Domain for Sample 2b of Wave V, the Sample 2b Supplement, and Sample 2 of Wave VI

Domain	Wave V Sample 2b	Wave VI Sample 2b Supplement	Wave VI Sample 2 Total	Wave VI Sample 2 Total Participants	Wave VI Response Rate (%)
Male/Non-Hispanic Black	145	381	526	331	62.9%
Male/Non-Hispanic White & Other (excluding Asian)	407	136	543	353	65.0%
Male/Hispanic	117	462	579	347	59.9%
Male/Asian	57	399	456	274	60.1%
Female/Non-Hispanic Black	176	257	433	325	75.1%
Female/Non-Hispanic White & Other (excluding Asian)	420	82	502	364	72.5%
Female/Hispanic	114	379	493	340	69.0%
Female/Asian	52	394	446	279	62.6%
Total	1,488	2,490	3,978	2,613	65.7%

For the stratified systematic sampling, all the cases from Wave V Sample 2b were first excluded from the sampling frame. Then, the remaining sampling frame was stratified across the eight sex-race domains listed in Table 2. Within each stratum, all the cases in the frame were sorted by key sorting variables so that the two samples, the Sample 2b Supplement and Sample 1, were approximately balanced with respect to the sorting variables and the proportions of the sample in each implicit stratum were approximately the same for each sample. The sorting variables from Wave V were used in the following order: region of residence, state of residence, age, sexual orientation, and Wave I grand sampling weights.

To compensate for item missingness in the geographic and demographic characteristics in Wave V, the information reported in the most recent wave from each frame member was used. As an example, if the state of residence was known for a frame member based on information collected for Wave V, then that information was used in sorting. Otherwise, the most recent information available on state of residence was used.

After sorting within each stratum, the determined number of cases (as shown in Table 2) was selected from the sorted frame via a systematic sampling scheme. The selected cases were combined with the cases in Wave V Sample 2b to form the entire eligible for and fielded Sample 2 cases. Then, all the remaining cases in the frame were included as eligible for and fielded Sample 1 cases.

4. Subsampling for Secondary Objectives

4.1. Selecting Sample 1 and Sample 2 Pilot Samples

To ensure quality control and test the complex instrumentation, procedures, and protocols for Samples 1 and 2, a pilot study was conducted prior to the main data collection for each sample. It is important to note that completed pilot cases are included in each respective Wave VI sample (Sample 1 and Sample 2, respectively); pilot cases are indeed Add Health participants. Moreover, there is no way to distinguish between pilot and main sample cases in the datasets.

Pilot selection was based on geographic location for convenience, as representative samples were not necessary for the pilot tests. The Sample 1 pilot consisted of 553 eligible and fielded cases, while the Sample 2 pilot consisted of 261 eligible and fielded cases.

The sex and education level distributions in each of these two pilot study samples were similar to their counterparts in the main samples. Nevertheless, because the selections of the pilot study samples were based on particular geographic locations, most of the sample members in the pilot study were White (over 45%) or Black (over 40%), with small proportions of cases Hispanic, Asian, and other races. Therefore, the racial/ethnic distributions in the pilot samples were slightly different than their counterparts in the main samples. However, as the purposes of the pilot studies were to test the data collection instruments, procedures, and protocols (rather than generate estimates for the target population), those differences were deemed to be acceptable.

4.2. Selecting Nonresponse Follow-up Cases for Sample 1 Phase 2 NRFU

The Sample 1 NRFU sample consisted of a main sample and a reserve sample to be fielded as necessary. The combined main NRFU sample and reserve sample, with approximately 5,000 cases in total, was selected as one sample initially, and then the final NRFU sample with 3,457 cases was randomly subsampled. Before sampling, the sample list was sorted by region, state, sex, race, sexual orientation, and age to implement *implicit stratification*. This ensured that the selected sample was more balanced across these characteristics.

The NRFU sample was then selected systematically with probability proportionate to size measure. The size measure for i^{th} sample case was calculated as:

$$s_i = \begin{cases} 500{,}000 & \text{when } ith \text{ case has less than high school education} \\ 10 * \dfrac{\omega_i}{p_{NRFUi}} & \text{when } ith \text{ case is non-White with high school education or above} \\ \dfrac{\omega_i}{p_{NRFUi}} & \text{Otherwise} \end{cases} \quad (1)$$

where ω_i was the Wave I grand sample weight and p_{NRFUi} was the estimated NRFU response propensity indicating the likelihood for i^{th} sample case to respond in NRFU.

Setting the size measure according to equation (1) served several important objectives:

1. By setting the size measure as 500,000, an extremely large value, we ensured the inclusion of all the cases with less than a high school education in the NRFU sample. This was particularly crucial for increasing the number of survey completes within this subgroup, which historically tends to have low participation.

2. The size measure was proportional to the Wave I grand sample weight and inversely proportional to the estimated NRFU response propensity. This adjustment not only ensured more interviews were completed for types of sample cases that have larger base weights or tend to have low response rates, but also minimized variability of the final weights.

3. We implemented an oversampling strategy for minority populations to ensure sufficient sample sizes for accurate point estimation within each racial/ethnic group. Specifically, the size measures for non-White groups were inflated by a factor of 10. Our extensive simulation studies explored various inflation values (10, 20, 40, and 100), with 10 being identified as the optimal value. This choice effectively balanced the objectives of oversampling and addressed potential issues such as large mean squared errors for key estimates, and the unequal weighting effects that can be caused by large weight variation due to oversampling.

After the total NRFU sample was selected, the final NRFU sample of 3,457 cases was selected using the same probability proportion to size sampling scheme as used to select the total sample.

5. Mixed-Mode Survey

Over the years, Add Health has collected a wealth of information from participants and their parents about demographic and socioeconomic characteristics, family structure, social relationships, health behaviors, cognition, physical and mental health status, medication usage, health care access, and more. Add Health also has collected anthropometric, cardiovascular, metabolic, renal, hepatic, inflammatory/immune, infectious, neurodegenerative, and multi-omic biomarkers from participants. In addition, Add Health has merged multilevel contextual data characterizing the economic, school, neighborhood, policy, and environmental contexts in which the participants are embedded to the core survey at each wave. The full Add Health data archive thereby provides researchers with rich opportunities to explore the causes and consequences of health across multiple contextual domains as individuals age across the life course.

Wave VI continued the mixed-mode survey design first utilized during Wave V, featuring both web-based and in-person survey administration. This mixed-mode survey design required the development of two different data collection protocols, as the in-person administration involved particular cognitive, physical, and sensory assessments that can only be collected in-person

The questionnaire content in the web-based survey was identical to the self-administered survey content in Sample 2 and Sample 1 Phase 2/NRFU with two exceptions:

(1) The web survey utilized a Google Maps application which asked participants to enter their residential address into a field, confirm on a map whether the Google pinpoint of the entered address was accurate, and - if the pinpoint was not accurate - drag the pin and drop it on the correct address. This application required an internet connection and therefore could not be replicated for the offline in-person Sample 2 interview.

(2) The Sample 1 web survey utilized the online Single Page Application (SPA) version of the TMB cognitive assessments, while the in-person instrument utilized a Progressive Web Application (PWA) loaded on a laptop with the exact same TMB cognitive assessments.

The data set contents – which is nearly identical for the full Mixed-Mode file (N=11,979) as it is for the Sample 2 file (N=2,613) – is summarized in Figure 2 below.

Figure 2: Add Health Survey Content as Organized in the Data Files

Section A: Interview Variables	Survey sample, survey mode, survey date MM/YY, platform used by web-based participants, consents, assisted computer use, Census region
Section 1: Background	DOB, biological sex, race/ethnicity, education
Section 2: Household Roster	HH roster and members' DOB, biological sex, race/ethnicity, education
Section 3: Military/Employment and Work-Life Balance	Military service dates/duty, current or most recent job(s) characteristics (hours, shifts, autonomy, physical activity, satisfaction, commute, job security, effect on health, benefits), Census Industry and Occupation Coding questions, work-life balance and impacts, income, federal assistance programs, gifts/inheritances, debts, residence ownership (value, mortgage)
TestMyBrain: Please note that these data are included in a separate Wave VI file with a separate user guide	Cognitive assessments
Section 4: Health Care and Illness	Self-rated health, ADLs, health conditions, menopause, health insurance, foregone care, dental care, counseling, sleep, hearing, vision, smell, prescription medications, physical activity, injury, memory, health care discrimination
Section 5: Feelings, Personality, Social Support	5-item CES-D, suicide ideation, GAD-7, LOT-R, Grit, risk, purpose, subjective age, social support, social stress, friends, institutional trust, neighborhood characteristics, isolation, chronic pain
Section 6: Discrimination and Feelings	Major experiences of discrimination, everyday discrimination, heightened vigilance, chronic workplace discrimination
Section 7: Turner Stress Measures	Turner Stress measures
Section 8: Anticipatory Stress	Anticipatory stress measures

Section 9: Tobacco, Alcohol, Substance Abuse, Religion	Tobacco use & cessation, electronic cigarette use & cessation, alcohol use & cessation, marijuana use & cessation, prescription medication misuse, pain killer misuse/frequency/cessation, illegal substance use, religious beliefs & practices
Section 10: Relationships, Parents, Siblings, Family Member Deaths	Biological parent deaths, biological parent closeness, sibling deaths, weapons ownership
Section 11: Caregiving & Romantic/Sexual Relationships	Intergenerational caregiving (to and from parents of participant and/or spouse/partner; to and from children of participant and/or spouse/partner), intergenerational monetary gifts, anticipatory parental caregiving, caregiving stress, caregiving benefits, marriage & cohabitation (current and total number of partners), length of relationship, current partner's demographics (age, race/ethnicity), relationship satisfaction, sexual behavior, reproductive plans, partner violence & sexual assault
Section 12: Criminal Justice System Involvement & Civic Activities	Arrests, incarceration, law enforcement, volunteering, voting
Section 13: Sexual Experiences, Pregnancy, Live Births, and Parenting	Sexual orientation, sexual behavior, lifetime sexual assault, pregnancies (total number, current status & due date), infertility, total number of biological children, live birth characteristics, total number of step/foster/adopted children, child deaths

Please note that users should be careful when considering the skip patterns involved in the data. **Indeed, not all questions were asked of all participants**. A simple example is some of the health questions in Section 4: some questions were specific to females while others were specific to males. Another is with the household roster in section 2: questions regarding the composition of large households were (legitimately) skipped by most participants. A final example is live births in Section 13: participants provided updates since their last Add Health interview, so users will need to link data across waves to piece together participant birth histories.

Once the questionnaire content was completed, Sample 2 participants then completed a series of cognitive, physical, and sensory functioning assessments. The in-person Sample 1 NRFU participants did not complete these assessments. The complete battery of cognitive, physical, and sensory functioning assessments completed by Wave VI participants are referred to as Add CAPS and described in detail in a series of Add CAPS Wave VI user guides (forthcoming 2025).

As noted previously, most prior Add Health interviews were administered by a field interviewer who could key in special codes of 'Don't Know' or 'Refuse' if the participant indicated that was their response. However, since most Wave VI participants completed the survey on their own, they were not provided with 'Don't Know' and 'Refuse' options. This decision was made after consulting survey literature, which indicated that participants would be less likely to answer a question if they could see and select the 'Don't Know' or 'Refuse' options on the screen. Participants could choose not to answer a question and click 'Next' to advance to the next question if desired. As a result, Wave VI data have more (.) missing codes than previous waves.

5.1 Sample 1 Survey Modes (Phase 1 and 2)

The web survey was programmed and administered using Blaise 5.8 software platform (Blaise Software. URL: https://www.blaise.com). Blaise software is widely used for a variety of research studies and offers multi-mode and multi-device support. The Blaise software was ideal due to its responsive nature. It was expected that a significant percentage of participants would choose to complete the web survey on a mobile device – whether tablet or mobile phone – since they did so during Wave V data collection. It was critical that the survey questions appeared in a suitably designed format for such devices.

The web survey took an average duration of 90 minutes to complete. All participants were offered the opportunity to complete the web survey on a device of their choice (there were no restrictions as to what type of device could be used to complete the web survey). Participants were provided with a unique password in order to access and complete the survey; identity checks were run based on certain identifiers provided within the survey to make sure the participant was the actual Add Health sample member. Participants were able to pause the survey and log in at a later time to continue/complete the survey if they chose. Once a participant completed and submitted the web questionnaire, their password was locked, and they were not allowed to re-enter. All participants were provided with a monetary incentive for their time.

A primary aim of Wave VI data collection was to provide the comprehensive data needed to understand the life course trajectories, determinants, and consequences of critical dimensions of health and health behavior among the Add Health cohort as they age into midlife. The survey content maintained the longitudinal integrity of the project by including questions from prior waves while also enhancing content in domains most important to the cohort as they entered early midlife. Survey questions were added to assess cumulative stress, discrimination, despair, work-life balance, memory, physical limitations, and caregiving. When possible, survey questions were harmonized with the Health and Retirement Study (HRS) and other aging studies. Due to content additions relevant to early midlife, the Wave VI survey was significantly longer than the 50-minute Wave V instrument.

As discussed in Section 2, Phase 1 of data collection involved recruiting participants to complete the web survey. The Phase 2/NRFU effort for Sample 1 also utilized a self-administered survey, in some cases with a field interviewer present. This self-administered in-person instrument was administered using Blaise 5.12 software. For these cases, the web survey was transferred to static laptops, which field interviewers took into the participants' homes. Interviewers went to participant homes and asked them to complete the interview, but each participant did so on their own without the interviewer administering any questions. Thus, the Sample 1 Phase 2/NRFU participants essentially completed the web survey with the two exceptions noted above (Google Maps Application and PWA version of TMB). An identical but offline version of the TMB web assessments used in the web survey was utilized during in-person fieldwork.

5.2 Sample 2 Modes

The in-person interview was administered via Blaise so that the questions included within both web and in-person instruments would appear identically across the modes, reducing any potential mode effects due to appearance on the screen. In a departure from prior waves, the main Sample 2 full survey was self-administered. As mentioned above, this reduced the length of the survey and participant burden after concerns were raised during the Sample 2 pilot. Reducing participant burden was critical due to the importance of the cognitive, physical, and sensory tests that followed the Sample 2 survey. Thus, the Sample 2 participants essentially completed the web survey with the two exceptions noted above (Google Maps Application and PWA version of TMB).

The Waves IV and V cognitive measures (word recall, digits backwards), which could not be included in a self-administered web survey, were included in the Wave VI Sample 2 in-person interview. An identical but offline version of the TMB web assessments used in the web survey was utilized during in-person fieldwork. New assessments administered to Sample 2 participants this wave included grip strength, hearing text, animal fluency, and NIH Toolbox assessments.

At the end of data collection, we offered the web mode to Sample 2 nonrespondents; thus, there were a small number of web survey completes in Sample 2 that are flagged as such in the Sample 2 data file.

5.3 A Note on Wave VI Survey Data Quality Control

The Wave VI survey instruments were programmed so that participant keystroke mistakes or the entering of illogical values (e.g., heights in excess of 7 feet) were minimized. The Add Health team has also performed light data cleaning procedures to try and clean up any obvious mistakes on the part of participants. Nevertheless, as in any collection of survey data, there are values listed by small numbers of participants for some variables that may be inaccurate and/or illogical (e.g., a very small number of individuals who report working more than 120 hours per week), but perhaps not completely impossible. We have decided, for the most part, to let the participant data “speak for themselves.” As such, researchers may wish to carefully review the variables they are working with and make judgments about the values that have been reported by participants; it is possible that some values may be best categorized with others and/or declared missing.

6. Wave VI Final Disposition Codes

Each Add Health Sample Member was assigned a code when the Wave VI eligibility sample was initially drawn. Throughout Wave VI data collection, these codes were continually updated whenever new information about sample members was obtained from tracing or fieldwork (e.g., sample member was found to be deceased). These codes were monitored regularly to remove ineligible sample members from recruitment and to perform record searches for anyone reported as deceased. At the conclusion of data collection, the codes were reviewed and finalized as the Wave VI Disposition Codes. These codes mirror those which were disseminated with prior waves of Add Health data. The Wave VI Disposition Codes are available in the Wave VI Final Disposition of Cases data file.

References

Aiello AE, Stebbins RC, Parker C, Rob R, Hummer RA. 2025. Add Health Wave VI Documentation: TestMyBrain User Guide. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Available at: forthcoming.

Angel RA, Grago J, Dean SC, Qu L, Carrier KS, Hummer RA, Whitsel EA. 2025. Add Health Wave VI Documentation: Medication Use – Add Health Survey. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Available at: https://doi.org/10.17615/kvxf-5r56

Biemer PB, Harris KM, Burke BJ, Liao D, Halpern CT. 2022. Transitioning a Panel Survey from in-person to Predominantly Web Data Collection: Results and Lessons Learned. *Journal of the Royal Statistical Society Series A: Statistics in Society*, Volume 185(3): 798-821. https://doi.org/10.1111/rssa.12750

Blaise Software. Available at: https://www.blaise.com.

Griffiths A, Dean SC, Harris KM, RA. 2025. Understanding Sample Sizes, Basic Sample Composition, and Cross-Sectional Sampling Weights, Add Health Survey Waves I-VI. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Available at: forthcoming.

Griffiths A, Harris KM, Liao D, Hummer RA. 2025. Guidelines for Analyzing Add Health Data. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Available at: forthcoming.

Liao D, Cooney D, Aiello AE, Hummer RA. 2025. Guide to Using Cross-Sectional and Longitudinal Weights in Add Health Wave VI. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Available at: forthcoming.

Many Brains Project. TestMyBrain Cognitive Tests. Available at: https://www.manybrains.net

Singh S, Strong RW, Jung L, Li FH, Grinspoon L, Scheuer LS, Passell EJ, Martini P, Chaytor N, Soble JR, Germine L. 2021. The TestMyBrain Digital Neuropsychology Toolkit: Development and Psychometric Characteristics. *J Clin Exp Neuropsychol*, Volume 43(8): 786-795. doi: 10.1080/13803395.2021.2002269. Epub 2021 Dec 15. PMID: 34907842; PMCID: PMC892997.