Add Health
The National Longitudinal Study of Adolescent to Adult Health

**Report prepared by**

Dan Liao

Darryl Cooney

Allison E. Aiello

Robert A. Hummer

# Wave VI User Guide to Using Cross-Sectional and Longitudinal Weights

UNC CAROLINA POPULATION CENTER

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

https://doi.org/10.17615/8s2m-yk29

## 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:

Liao, D., Cooney, D. Aiello, A.A., Hummer, R.A. 2025. Add Health Wave VI Documentation: Guide to Using Cross-Sectional and Longitudinal Weights in Wave VI. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Available from: https://doi.org/10.17615/8s2m-yk29.

## Table of Contents

# 1. Introduction

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a longitudinal study of a nationally representative sample of over 20,000 adolescents who were in grades 7-12 during the 1994-95 school year and who have been followed for six waves to date, most recently in 2022-2025. Over the years, Add Health has collected rich demographic, social, familial, socioeconomic, behavioral, psychosocial, cognitive, and health survey data from participants and their parents; a vast array of contextual data from participants' schools, neighborhoods, and geographies of residence; and in-home physical and biological data from participants, including genetic markers, blood-based assays, anthropometric measures, and medications. Ancillary studies have added even more data over the years (Harris et al. 2019). The overall goal of Add Health Wave VI was to collect and disseminate the comprehensive data needed to best understand the social, economic, psychosocial, contextual, and biological determinants of health trajectories and disparities among this nationally representative cohort of Americans as they age into midlife.

Wave VI of Add Health used a multifaceted, mixed-mode design similar to that used in Wave V. The Wave VI sample was split into two random subsamples, Samples 1 and 2. Similar to the mixed mode subsamples in Wave V (Samples 1, 2a and 3 in Wave V), the data collection for Sample 1 employed a mixed-mode, two-phase survey design. In the first phase, sample members were contacted primarily by mail, email or text message and asked to complete a web questionnaire. In the second phase, a subsample of nonrespondents was followed up through nonresponse follow-up (NRFU). Most of these cases were completed online, and others were completed via in-person interviews when needed. Similar to Sample 2b in Wave V, the data collection for Sample 2 in Wave VI largely employed an in-person interview collection mode. In addition, in Wave VI, Sample 2 nonrespondents were provided a web questionnaire as follow-up (an approach that was not used in Wave V).

Table 1 shows the overall sample sizes and number of respondents for each sample component in 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). For an overview of the Wave VI sample design, refer to *Wave VI Sampling and Mixed-Mode Survey Design* (Hummer et al., 2025).

**Table 1. Key Wave VI Sample Size Results**

| | Wave VI Eligible and Fielded Cases* | N (respondents) | Raw Response Rate | Effective Response Rate |
|---|---|---|---|---|
| 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. See Hummer, et.al (2025) for more information about the overall sample design of Wave VI.

The data collected from survey respondents must be weighted to account for unequal probability sampling, as well as sample representativeness issues caused by nonresponse, ineligibility, and population under-coverage. As a panel survey, two types of weights were produced for Wave VI of Add Health: cross-sectional and longitudinal. Cross-sectional weights address representativeness issues for cross-sectional estimates, i.e., estimates describing the Add Health population at the time of the current data collection wave. Longitudinal weights are designed for analyses spanning two or more waves and are used to estimate wave-to-wave changes or to track trajectories of change across multiple waves. Since the design of Wave VI is similar to that of Wave V, we applied comparable weighting methods and procedures, as detailed in Biemer and Liao (2024), to ensure consistency. However, some modifications were made to reflect changes in the Wave VI design.

## 2. Weighting Methodologies for Wave VI

As mentioned earlier, Wave VI participants were divided into two subsamples: Sample 1, consisting primarily of web-based participants, and Sample 2, consisting primarily of in-person participants. Weights were created for Sample 1-only, Sample 2-only, and the Grand Sample (i.e., the combined Samples 1 and 2) using a methodology referred to as the *optimized weighting methodology*. In this context, "Sample 1-only" and "Sample 2-only" refer to the weighting of each individual sample to represent the entire target population of Add Health. In addition, weights were created for Sample 2-only using the *traditional weighting methodology.*

The traditional weighting methodology employs procedures that closely mirror those used in waves prior to Wave V. This approach is applied specifically to Sample 2-only, which followed an in-person data collection protocol similar to that used in earlier waves (Wave I through Wave IV), albeit with slight modifications—namely, that in-person survey nonrespondents were provided an opportunity to complete the survey online. Weighting Sample 2 independently using traditional weights ensures its representativeness for the entire Add Health sample population using the weighting method consistent with earlier waves. Some users may choose to use traditional weights to adjust for or control the effects of survey mode when conducting longitudinal analyses. Note that traditional weights were also created in Wave V for Sample 2b (the in-person subsample at Wave V) to support similar analyses.

The optimized weighting methodology, developed in Wave V, leverages the two-phased structure of Sample 1, comprising a web survey in Phase I and a NRFU process in Phase II, to maximize the use of available information. This methodology is designed to compensate for nonresponse in both Samples 1 and 2, providing a more refined solution for addressing representativeness issues across the Add Health sample.
The following subsections provide a detailed description of both weighting methods. For additional mathematical details on these methodologies, please refer to Biemer and Liao (2024).

## 2.1. Optimized Weighting Methodology

For Wave VI, optimized cross-sectional weights were created for Sample 1-only, Sample 2-only, and the Grand Sample (i.e., the combined data from Samples 1 and 2).

***Cross-sectional Grand Sample Weighting***

The process for constructing optimized cross-sectional Grand Sample weights involved three key steps:

**Step 1: Creation of Phase I Weights**

Phase I weights were created for respondents in Sample 1-Phase I and Sample 2. These weights were derived through a two-step process:

- First, nonresponse adjustment was applied to the Wave I Grand Sample weights to account for differential participation among respondents. This adjustment involved estimating response probabilities using variables available for all sample members (e.g., age, race/ethnicity, response indicators, and the number of contacts made in the previous waves). The inverse of these response probabilities was then applied to the Wave I Grand Sample weights for respondents, while nonrespondents were assigned missing weights.

- Second, calibration was performed to align the adjusted weights with known control totals based on the full set of 19,081 eligible and fielded cases from Wave VI. Control totals were computed using cross-classifications of race (4 groups), sex (2 groups), and age (5 groups). This procedure accounted for ineligibility and addresses coverage issues, such as the underrepresentation or overrepresentation of certain subpopulations, ensuring that the weighted sample accurately reflects the target population.

**Step 2: Joint Weighting of Phases I and II Respondents**

In this step, weights were calculated for all Wave VI respondents, including those from Phase II (i.e., the NRFU). This involved jointly weighting Phase II respondents and Phase I respondents with adjustment to account for selection probabilities associated with Phase II inclusion (in Sample 1). Subsequently, nonresponse adjustments and calibration were applied.

**Step 3: Combining Weights from Steps 1 and 2**

In this final step, the two sets of weights, those generated in Step 1 for Phase I respondents and Step 2 for Phases I and II respondents, were combined to produce optimized Grand Sample weights. This combination was performed by minimizing total error, balancing two key trade-offs:

- **Variance reduction:** Phase I weights generally exhibit lower variability because Phase II sampling can introduce weight inflation due to differential probabilities and smaller subsample sizes. As a result, Phase I weights often have more desirable variance properties.
- **Bias reduction:** Combined Phase I-II weights mitigate nonresponse bias as they incorporate information from respondents captured during the NRFU phase. Nonresponse bias is reduced through the inclusion of data from NRFU participants who would otherwise remain unrepresented.

By systematically optimizing the mix of these estimators, Step 3 achieved a balance between bias reduction and variance control, producing final Grand Sample weights designed to minimize mean squared error (a combination of bias and variance).

***Cross-Sectional Sample 1-Only Weighting***

The process for creating the optimized cross-sectional Sample 1-only weights followed a methodology similar to that used for constructing the Grand Sample weights, with one key distinction: the weights were adjusted exclusively for Sample 1 respondents (excluding Sample 2) and included an additional adjustment to account for their selection probabilities into Sample 1.

***Cross-Sectional Sample 2-Only Weighting***

The optimized cross-sectional weights for Sample 2-only were produced through two additional steps after making the cross-sectional Grand Sample weights:

**Step 4:** The Wave I weights for Sample 2 respondents were adjusted for their selection probabilities into Sample 2 and Wave VI nonresponse.

**Step 5:** The nonresponse-adjusted weights from Step 4 were calibrated to control totals derived from the Grand Sample weights. This calibration process adjusted the Sample 2-only weights so that their weighted totals match key control totals for demographic variables (e.g., age-sex-race cross-classifications) and key survey questionnaire variables from Wave VI based on the Wave VI Grand Sample weights (derived from Steps 1 to 3). This calibration ensured that the Sample 2 respondents adequately represent the entire target population. Because the Grand Sample relies on a larger combined dataset from both Sample 1 and Sample 2, its weighted estimates were assumed to provide a more accurate reflection of the population and thus were used as control totals in this process.

## 2.2. Traditional Weighting Methodology

For Wave VI, traditional weights were created specifically for Sample 2-only. In waves prior to Wave V, the weighting process included only a simple nonresponse propensity adjustment,

without employing more advanced calibration techniques. At the time, calibration weighting approaches were neither used nor deemed necessary. Instead, the respondents in these earlier waves were post-stratified to match the weighted grade-sex-race subpopulation totals from the Wave I Grand Sample (using the Wave I Grand Sample weights), after excluding sample members who had passed away since Wave I. The same general methodology was used to create the Wave VI traditional weights.

## 3. Cross-sectional Weights and Longitudinal Weights for Wave VI

In Wave VI, both cross-sectional and longitudinal weights were created.

**Cross-sectional weights should be used by most analysts for cross-sectional analyses**. The optimized Wave VI Grand Sample weight (GSW6) is expected to be the primary weight utilized for these analyses. Additionally, to perform cross-sectional estimations for the cognitive, physical, and sensory functioning assessments in Sample 2, optimized weights have been created specifically for Sample 2 (GSW6_2). The optimized weights for Sample 1 only (GSW6_1), as well as the traditional weights for Sample 2 only (GSW6_2T), were generated to evaluate how longitudinal changes may be influenced by differences in design and weighting methodologies. This allows for a comparison between estimates based on the new mixed-mode and optimized weighting methodology (i.e. Sample 1-only and its optimized weights) and those based on the traditional in-person mode and weighting methodology (i.e. Sample 2-only with traditional weights), highlighting how estimates can differ based on data collection mode and weighting methods.

**Longitudinal weights, designed for different combinations of waves, should be used to analyze individuals who responded during specific waves.** The optimized longitudinal weights for the Grand Sample will serve as the primary weights for longitudinal analyses involving specific waves. Optimized weights were created for three sets of longitudinal series for the Grand Sample: I-II-III-IV-V-VI, I-III-IV-V-VI, and I-IV-V-VI (named as GSW123456, GSW13456, GSW1456, respectively). The basic optimized weighting methodology for creating the longitudinal weights was essentially the same as the method used to create the cross-sectional weights as described in Section 2. For each set of longitudinal weights, we defined a response set, R, consisting of individuals from Add Health Wave I who responded in all waves included in the specified longitudinal series. To be more specific,

- for weight set I-II-III-IV-V-VI, R consisted of all sample members who responded at Waves I, II, III, IV, V, and VI;
- for weight set I-III-IV-V-VI, R consisted of all sample members who responded at Waves I, III, IV, V, and VI;
- for weight set I-IV-V-VI, R consisted of all sample members who responded at Waves I, IV, V, and VI.

What primarily differentiates longitudinal weighting from cross-sectional weighting is the nonresponse adjustment applied to the base weights (i.e., Wave I weights). This distinction arises because the responding sample for longitudinal weighting is a proper subset of the responding sample for cross-sectional weighting. For example, for weight set I-IV-V-VI, the respondents are defined as those who responded at Waves I, IV, V and VI, a subset of all Wave VI respondents. Since each longitudinal weight set corresponds to a different responding sample, the nonresponse and calibration adjustments were performed separately for each set using the Wave I Grand Sample weights as the base. As a result, these adjustments (and thus the final weights) will differ from the cross-sectional weights and will also vary across the different types of longitudinal weight sets.

For users who wish to conduct longitudinal analyses that fall outside of the three sets of weights that have been created (e.g., an analysis spanning Waves I, II, V, and VI), additional considerations are required. Users will either need to construct custom weights themselves or select a weight series that spans a broader range of waves. For example, to analyze data across Waves I-II-V-VI, users should utilize weights created for Waves I-II-III-IV-V-VI. While this approach may result in some loss of cases (i.e., respondents who did not participate in Waves III and IV but responded in other waves), it ensures consistency in cross-wave representativeness.

Traditional weights were created for two sets of longitudinal series for Sample 2-only: I-III-IV-V-VI and I-IV-V-VI. These traditional longitudinal weights for Sample 2 were primarily designed to examine longitudinal changes as measured using the traditional in-person survey mode and weighting methodology. The basic traditional approach for creating these longitudinal weights is essentially the same as the method used to create the traditional weights for Sample 2, as described in Section 2. As with the optimized longitudinal weights, each traditional longitudinal weight set began by defining a response set, R, consisting of individuals selected into Wave VI Sample 2 who responded in all waves included in the specified longitudinal series. To be more specific,

- for weight set I-III-IV-V-VI, R consisted of all sample members who were selected into Wave VI Sample 2 and responded at Waves I, III, IV, V, and VI;
- for weight set I-IV-V-VI, R consisted of all sample members who were selected into Wave VI Sample 2 and responded at Waves I, IV, V, and VI.

The traditional weights for each response set were then created following the same traditional methodology described in Section 2.

## 4. Weights for Biomarker Sample and Add Health Wave VI Public Use File

Additionally, both cross-sectional and longitudinal weights were provided for the biomarker sample and the Add Health Wave VI Public Use File (PUF), available through UNC Dataverse and the Inter-university Consortium for Political and Social Research and National Archive of Computerized Data on Aging (ICPSR/NACDA).

## 4.1. Weights for Biomarker Sample

***Cross-sectional Weights for Biomarker Sample***

The full Wave VI respondent sample includes 11,979 individuals, of whom 6,073 completed the biomarker data collection. To ensure that this biomarker subsample is representative of the target population reflected in the full Wave VI sample, additional nonresponse and calibration adjustments were applied to the Wave VI optimized cross-sectional weights for the Grand Sample to create the optimized cross-sectional weights for the biomarker sample.

In the nonresponse adjustment process, Wave VI respondents were assigned a binary response indicator, with 1 representing respondents who completed the biomarker data collection and 0 representing those who did not complete. This response indicator was treated as the dependent variable, while other variables known for all sample members (such as demographic characteristics and information about the level of effort required to contact respondents in previous waves) were used to estimate response propensities for biomarker data collection participation. The estimated response propensities were then inverted and applied as the nonresponse adjustment factor, refining the Wave VI Grand Sample weights to compensate for differential participation across respondent groups.

Following the nonresponse adjustment, a calibration step was performed using control totals derived from Wave I data, after excluding all ineligible cases at Wave VI. Calibration aligned the weighted sums of the adjusted weights with known control totals for race-sex-age groups.

***Longitudinal Weights for Biomarker Sample***

Two sets of optimized longitudinal weights were created for the biomarker sample: one for Waves I-IV-V-VI and another for Waves I-V-VI. The methodology for developing these longitudinal weights closely follows the approach used to create the optimized cross-sectional weights for the biomarker sample. For each series, additional nonresponse and calibration adjustments were applied to the Wave VI optimized cross-sectional weights for the Grand Sample to create the optimized longitudinal weights for the biomarker sample. In the nonresponse adjustment process, respondents were defined as those who completed biomarker data collection in all specified waves.

## 4.2. Weights for Add Health Wave VI Public Use File (PUF)

There are 3,937 Add Health respondents included in the Wave VI PUF. To ensure that this subsample is representative of the target population reflected in the full Wave VI sample, additional nonresponse and calibration adjustments were applied to the Wave VI optimized cross-sectional weights for the Grand Sample to create the optimized cross-sectional (or longitudinal) weights for the PUF. Similar to the weighting process for the biomarker sample,

the nonresponse adjustment accounts for instances where sample members are included in the Grand Sample but not in the PUF. Two sets of optimized longitudinal weights were created for the PUF: one for Waves I-III-IV-V-Vi and another for Waves I-IV-V-VI.

## 5. Summary

Table 2 provides an overview of all weights created for Add Health Wave VI. Each weight is designed to ensure that its corresponding sample accurately represents the Wave VI target population. Below are general instructions on which weights should be used for different types of analysis:

- **Cross-sectional analysis with the Wave VI Grand Sample:** Use the optimized cross-sectional Wave VI Grand Sample weight (GSW6).
- **Longitudinal analysis with the Wave VI Grand Sample:** Use the optimized longitudinal Wave VI Grand Sample weight (GSW123456 for I-II-III-IV-V-VI; GSW13456 for I-III-IV-V-VI; GSW1456 for I-IV-V-VI). For longitudinal analyses outside the three weight sets (e.g., Waves I, II, V, and VI), users must either create custom weights or use a broader weight series, for example, the one spanning Waves I-II-III-IV-V-VI for analyzing data across Waves I-II-V-VI.
- **Cross-sectional analysis with the Wave VI biomarker sample**: Use the optimized cross-sectional weights for biomarker sample (W6BIOWGT).
- **Longitudinal analysis with the Wave VI biomarker sample**: Use the optimized longitudinal weights for biomarker sample (W6BIOWGT_1456 for I-IV-V-VI and W6BIOWGT_156 for I-V-VI). For longitudinal analyses outside the two weight sets, users must either create custom weights or use a broader weight series, for example, the one spanning Waves I-IV-V-VI for analyzing data across Waves I-IV-VI.
- **Cross-sectional analysis with the PUF**: Use the optimized cross-sectional weights for PUF (GSW6).
- **Longitudinal analysis with the PUF**: Use the optimized longitudinal weights for the PUF (GSW13456 for I-III-IV-V-VI and GSW1456 for I-IV-V-VI). For longitudinal analyses outside the two weight sets, users must either create custom weights or use a broader weight series, for example, the one spanning Waves I-III-IV-V-VI for analyzing data across Waves I-III-V-VI.
- **Evaluating the Effects of Mode and Weighting Methods on Longitudinal Changes:** For cross-sectional comparison, use the traditional Sample 2-only weights (GSW6_2T) with Sample 2-only data to obtain estimates based on the traditional in-person data collection mode and weighting method. Then, compare these estimates with those derived using optimized Sample 1-only weights (GSW6_1) and Sample 1-only data, or with those derived using optimized Grand Sample weights (GSW6) and the Wave VI Grand Sample. For longitudinal comparison, use the traditional Sample 2-only weights (GSW13456_2T for I-III-IV-V-VI and GSW1456_2T for I-IV-V-VI).

**Table 2. Weights Created for Add Health Wave VI**

| | Weights | Weighting Method | Number of Respondents with Positive Weights | Variable Name |
|---|---|---|---|---|
| **Cross-Sectional Weights for Wave VI** | | | | |
| 1 | Grand Sample (Combined Samples 1 and 2) | Optimized | 11,979 | GSW6 |
| 2 | Sample 1 only | Optimized | 9,366 | GSW6_1 |
| 3 | Sample 2 only | Optimized | 2,613 | GSW6_2 |
| 4 | Sample 2 only | Traditional | 2,613 | GSW6_2T |
| **Longitudinal Weights for Wave VI** | | | | |
| 5 | I-IV-V-VI | Optimized | 8,846 | GSW1456 |
| 6 | I-II-III-IV-V-VI | Optimized | 6,028 | GSW123456 |
| 7 | I-III-IV-V-VI | Optimized | 7,673 | GSW13456 |
| 8 | I-IV-V-VI (Sample 2 only) | Traditional | 1,748 | GSW1456_2T |
| 9 | I-III-IV-V-VI (Sample 2 only) | Traditional | 1,508 | GSW13456_2T |
| **Weights for Wave VI Biomarker Sample** | | | | |
| 10 | Cross-Sectional Weight | Optimized | 6,073 | W6BIOWGT |
| 11 | Longitudinal Weight I-IV-V-VI | Optimized | 3,395 | W6BIOWGT_1456 |
| 12 | Longitudinal Weight I-V-VI | Optimized | 3,633 | W6BIOWGT_156 |
| **Weights for Wave VI Public Use File** | | | | |
| 13 | Cross-Sectional Weight | Optimized | 3,937 | GSW6 |
| 14 | Longitudinal Weight I-III-IV-V-VI | Optimized | 2,571 | GSW13456 |
| 15 | Longitudinal Weight I-IV-V-VI | Optimized | 2,996 | GSW1456 |

**References**

Biemer and Liao (2024). Add Health-Wave V: Full Sample Cross-Sectional Weighting Specifications. Available at: https://addhealth.cpc.unc.edu/wp-content/uploads/2024/09/Full-Add-Health-Sample-Cross-Sectional-Weighting-Specifications.pdf

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