DSDR | DATA SHARING FOR DEMOGRAPHIC RESEARCH

ICPSR 21600

National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008 [Public Use]

Kathleen Mullan Harris
University of North Carolina-Chapel Hill

J. Richard Udry
University of North Carolina-Chapel Hill

Wave IV: Public Use Weights User Guide

Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Terms of Use

The terms of use for this study can be found at:
http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/21600/terms

Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

National Longitudinal Study of Adolescent Health

Wave IV Weights

DEVELOPED BY:

Paul P. Biemer, PhD
Elvessa D. Aragon-Logan, MS

RTI International
3040 Cornwallis Road
P.O. Box 12194
Research Triangle Park, NC 27709-2194

Add Health Wave IV Weights

Introduction

In this document, we describe the process for weighting the Wave IV data to compensate for Wave IV nonresponse. The approach described here was used for Wave III, and is consistent with the approach that was used for Waves I and II. The weighting documentation for Wave III provides a brief review of the weighting approaches for Waves I and II.

Calculation of Weights for Wave IV

The same basic approach for adjusting the Wave I weights for Wave II nonresponse and adjusting the Wave I and Wave II weights for Wave III nonresponse will be used to adjust Wave I, Wave II and Wave III weights for Wave IV nonresponse. These weight adjustments were applied to form cross-sectional and longitudinal analysis weights for the grand sample and for the public-use sample. By following the same weight adjustment approach for all four waves, the impact of the weight adjustment methodologies at each wave on data analysis should be minimized.

Wave IV of the Add Health study consists of a follow-up interview with the 20,745 in-home sample members from Wave I. Excluded from the Wave IV sample were 687 cases that did not have sample flags or weights from Wave I, as well as 96 cases that were determined to be deceased at Wave III. In addition, 45 cases that were added in Wave II and included in Wave III (but not weighted) are excluded from Wave IV. Of the 19,962 cases available for Wave IV, 300 were selected for the pretest that was conducted in April and May of 2007. The remaining 19,662 cases, and 101 pretest cases who were not interviewed during the pretest period, comprise the sample for the main study that began in January 2008. The total number fielded for the main study was 19,763.

Eligibility for Wave IV

The 783 cases that were not fielded in Wave IV are ineligible and were not assigned Wave IV weights. Similarly, the 45 cases added in Wave II were never assigned weights in Wave II and Wave III. For purposes of weighting, these 45 cases were considered ineligible and will not be assigned any weights for Wave IV.

In Wave III, it was determined that “deceased” is the only group that had the same definition at Wave II and Wave III. Hence, only the 96 deceased cases found in Wave III were further deemed ineligible for purposes of Wave III weighting. To be consistent with previous waves, the only cases that will be further deemed ineligible for Wave IV weighting purposes will be the 131 deceased cases found in Wave IV. These cases were not assigned Wave III weights. Table A-1 of Appendix A shows the final disposition of the

Wave IV cases.

Weights for the Grand Sample

Because the Wave IV sample consists of Wave I in-home sample members, creating the Wave IV **cross-sectional weights** for the grand sample simply involved adjusting the Wave I grand sample weights for additional Wave IV nonresponse. The basic formulas for these adjustments appear in Appendix B.

Nonresponse adjustments were calculated separately for each school using the Wave I grand sample weights. Wave IV respondents who also have positive Wave I grand sample weights (*i.e.*, were Wave I respondents) were assigned cross-sectional weights. Weights were set to zero (for nonrespondents) or missing (for ineligible cases), otherwise.

The nonresponse-adjusted cross-sectional weights were poststratified to estimates of the grade-sex-race subpopulations derived from the Wave I grand sample weights, after adjusting for the deceased at Wave IV. These estimates reflect the portion of the 1995 population (represented by the Wave I sample) that would have been eligible at Wave IV. The estimates were calculated by summing the Wave I grand sample weights for all the sample members of each grade-sex-race domain that were alive at Wave IV. Table B-1 in Appendix B shows the grade-sex-race subpopulation estimates used to poststratify the cross-sectional weights to the Wave I population.

There are 1,866 cases that have missing Wave I weights. Of these, 1,115 were eligible for Wave IV, and 901 responded. These 901 cases will have missing Wave IV cross-sectional weights, bringing the total of cases with positive Wave IV cross-sectional weights to 14,800 (instead of the 15,701 respondents indicated in Table A-1).

The Wave IV **longitudinal weights** were calculated similarly, except that the Wave III grand sample longitudinal weights were used instead of the Wave I grand sample weights (see Appendix B). That is, the Wave III grand sample longitudinal weights were adjusted for additional Wave IV nonresponse using weighted nonresponse adjustments to the Wave III grand sample longitudinal weights calculated separately for each school. For this analysis, a nonrespondent was defined as a unit that did not respond at either Waves I, II, III or IV. Sample members who were respondents in Wave II, Wave III and Wave IV were assigned positive weights. Nonrespondents were assigned zero weights. Weights were missing, otherwise. Of the 15,701 respondents in Wave IV, only 9,421 cases were also respondents in Waves II and III, and had nonmissing weights in Wave I.

Estimates of the grade-sex-race subpopulations derived from the Wave III grand sample longitudinal weights, excluding the deceased, were used to poststratify the nonresponse-adjusted longitudinal weights. The estimates were calculated by summing the Wave III grand sample longitudinal weights for all the sample members of each grade-sex-race domain that were still alive at Wave IV. Table B-2 in Appendix B shows the subpopulation estimates used to poststratify the longitudinal weights to the Wave III population.

Weights for the Public-Use Sample

The cross-sectional and longitudinal Wave IV weights for the public-use sample were calculated in a similar manner as the weights for the grand sample, except that the corresponding public-use sample weights from previous waves were used in the calculations instead of the grand sample weights. For reasons of confidentiality, nonresponse adjustments were calculated separately for each grade-sex cell instead of for each school. The basic formulas shown in Appendix B also apply to the public-use sample weights. Poststratification of the public-use sample weights to the respective Wave I and Wave III populations was done in the same manner as the grand sample weights. Tables B-3 and B-4 in Appendix B show the grade-sex-race subpopulation estimates used to poststratify the cross-sectional and longitudinal public-use sample weights, derived by summing the Wave I and Wave III public-use sample weights, respectively, for all the public-use sample cases still alive at Wave IV.

Weight Trimming

Since the nonresponse-adjusted weights were trimmed in the previous three waves, the question of whether to trim the weights in Wave IV was considered. Weight trimming can reduce the variances of estimates but may also increase their biases. The hope is that trimming will reduce the total mean squared error, but whether or not this is accomplished cannot be evaluated with the available data. Nevertheless, a good reason to trim the weights for Wave IV is to be consistent with the previous three waves.

To evaluate the effect of trimming the nonresponse-adjusted weights on the precision of the estimates, we calculated the unequal weighting effects (UWE) for the cross-sectional and longitudinal weights both with and without trimming. The UWE is defined as [1+ CV^2(weights)] where CV^2(weights) refers to the squared coefficient of variation of the weights. As was done in Wave I, the trimmed value was set to 6000; *i.e.*, all nonresponse-adjusted weights greater than 6000 were set equal to 6000. Poststratification adjustments were then applied to both the untrimmed and the trimmed nonresponse-adjusted weights. Table 1 below summarizes the results of our analysis of the trimming effects. Note that the public-use sample weights were not trimmed in previous waves, so they were not trimmed in Wave IV.

The Wave III grand sample non-response adjusted, trimmed and post-stratified longitudinal weights were used in order to be consistent with the use of the non-response adjusted, trimmed and post-stratified grand sample weights from Wave II to calculate the Wave III grand sample longitudinal weights.

From Table 1, note that the UWEs for the Wave IV longitudinal weights, trimmed and untrimmed, have decreased from Wave III, and are now lower than the UWEs for Wave I and Wave II weights. Thus, trimming does not appear to be necessary for Wave IV longitudinal weights to reduce the UWE to that of Wave I.

The UWEs for the Wave IV untrimmed cross-sectional weights are only slightly higher than the Wave I UWE and virtually identical to the Wave III UWEs. The UWEs for the Wave IV trimmed cross-sectional weights are lower than the Wave III UWEs. Thus, trimming does not appear to be necessary for Wave IV cross-sectional weights to reduce the UWE to that of Wave I.

We recommend that the untrimmed final weights, cross-sectional and longitudinal, be used in analyses with Wave IV data.

Table 1. Effect of Weight Trimming on the UWE for Cross-Sectional and Longitudinal Weights

Type of Weight	Number of Positive Weights	Untrimmed Weights UWE	Trimmed Weights	
			Number Trimmed	UWE
Wave IV Cross-sectional				
Adjusted for NR	14800*	1.92437	208	1.80440
Final (Poststratified)	14800*	1.93466	208	1.81612
Wave IV Longitudinal				
Adjusted for NR	9421**	1.76658	267	1.69389
Final (Poststratified)	9421**	1.77218	267	1.70253
Wave III Cross-sectional				
Adjusted for NR	14322	1.92370	209	1.82769
Final (Poststratified)	14322	1.93530	209	1.84260
Wave III Longitudinal				
Adjusted for NR	10828	1.86821	209	1.74612
Final (Poststratified)	10828	1.87360	209	1.75417
Wave II Final	13568			1.82098
Wave I Final	18924			1.88615

** Of the 15,701 respondents in Wave IV, 901 had missing Wave I weights.*

*** In order to have a positive longitudinal weight, a case would have to be a respondent in all waves considered for the longitudinal weights. There were 9,421 cases that responded in all four waves.*

Appendix A
Final Disposition of Wave IV Sample

Table A-1. Final Disposition of Wave IV Cases

W4SAMP		Wave IV Final Disposition	Freq	Eligible E/I	Respondent R/NR
Excluded from Wave IV because these cases have no sample flags nor Wave I weights					
.	.		687	I	--
Excluded from Wave IV because these cases were found to be deceased at Wave III					
.	.		96	I	--
Cases added in Wave II to supplement the genetic sample; these have no Wave I, II or III weights					
.	.		45	I	
		Total	828		
Excluded from Wave IV weighting because these cases were found to be deceased					
1	459	Deceased	131	I	--
		Total	131		
Eligible, but non-interviews in Wave IV					
Cases considered eligible for weighting purposes					
1	482	Moved out of country	184	E	NR
1	486	Active Duty Military – Unavailable for Duration	87	E	NR
Other eligible nonrespondents					
1	419	Access denied	16	E	NR
1	420	No one home after repeated attempts	8	E	NR
1	422	SM Unavailable after repeated attempts	418	E	NR
1	423	SM Unavailable during interview period	3	E	NR
1	460	Final Refusal by SM	1587	E	NR

W4SAMP	Wave IV Final Disposition		Freq	Eligible E/I	Respondent R/NR
1	462	Final Refusal by Other	191	E	NR
1	470	Language Barrier - Spanish	4	E	NR
1	471	Language Barrier - Other	2	E	NR
1	475	Physically/Mentally Incapable	75	E	NR
1	477	Incarcerated	110	E	NR
1	478	Institutionalized	15	E	NR
1	480	Unlocatable	1348	E	NR
1	481	Moved out of Interviewing Area	3	E	NR
1	489	Other Non-Interview	39	E	NR
1	503	Interview Completed - Gender/DOB Discrepancies	24	E	NR
1	504	Interview Completed - Mentally Challenged Case - Mental Capacity Inadequate	5	E	NR
1	505	Interview Completed - Mentally Challenged Case - Unable to Determine Mental Capacity	10	E	NR
1	613	Interview Completed-Data Deleted	1	E	NR
		Total	4130		
Eligible interviews					
1	490	Breakoff/Partial Interview	7	E	R
1	491	Interview Completed	15694	E	R
		Total	15701		
		Grand Total	20790		

Appendix B
Basic Weighting Formulas

Basic Weighting Formulas for Wave IV

Define the following symbols:

D the domain of interest; for example, a particular grade-sex-race cell,
S the original Wave I sample,
R_{Dw} the set of respondents at Wave *w*, for *w* = 1,2,3,4
E_{Dw} the set of eligible persons at Wave *w*, for *w* = 1,2,3,4
I_{Dw} the set of ineligible persons at Wave *w*, for *w* = 1,2,3,4
ω_{1i} Wave I final weight (either grand sample or public-use sample, depending upon the context) defined for all *i* R_{D1}
ω_{3i} Wave III final weight (either grand sample or public-use sample, depending upon the context) defined for all *i* R_{D2}

Grand Sample Cross-sectional Weights

For each sample school, define the school-specific weighted response rate for Wave IV given Wave I response as

$$r_{D4|1} = \frac{\sum_{i \in R_{D4}} \omega_{1i}}{\sum_{i \in E_{D4}} \omega_{1i}} \qquad \text{(B1)}$$

Then the Wave IV grand sample cross-sectional weight, before post-stratification adjustment, is

$$\omega^{C}_{D4i} = \frac{\omega_{1i}}{r_{D4|1}} \qquad \text{(B2)}$$

For the post-stratification adjustment, we divide an estimate of the total 1995 population in domain *D* who would be eligible in Wave IV by an estimate of the same population from Wave IV data as follows:

$$F^{C}_{D4} = \frac{\hat{X}_{E_{D4}}}{\sum_{i \in R_{D4}} \omega^{C}_{D4i}} \qquad \text{(B3)}$$

where

$$\hat{X}_{E_{D4}} = \sum_{i \in E_{D4}} \omega_{1i} \qquad \text{(B4)}$$

is the sum of the Wave I sample weights for the sample members who were eligible in Wave IV (*i.e.,* Wave I respondents who were alive at Wave IV). Table B-1 shows the subpopulation estimates (B4) for the grade-sex-race poststratification domains.

Then the final Wave IV grand sample weight, which is defined for all i R_{D4} is

$$\omega^{C}_{F,\,D4i} = F^{C}_{D4}\omega^{C}_{D4i} \qquad \text{(B5)}$$

The untrimmed cross-sectional weights were computed using Equations B1 to B5. The trimmed cross-sectional weights were computed in the same manner, except that the nonresponse-adjusted weight resulting from Equation B2 was trimmed to a value of 6000 before being substituted into Equations B3, B4 and B5.

Grand Sample Longitudinal Weights

For each sample school, define the school-specific weighted response rate for Wave IV given Wave II and Wave III response as

$$r_{D4|3} = \frac{\sum_{i \in R_{D4} \cap R_{D3} \cap R_{D2}} \omega_{3i}}{\sum_{i \in R_{D3} \cap E_{D4}} \omega_{3i}} \qquad \text{(B6)}$$

Then the Wave IV grand sample longitudinal weight, before post-stratification adjustment, is

$$\omega^{L}_{D4i} = \frac{\omega_{3i}}{r_{D4\,|\,3}} \qquad \text{(B7)}$$

which is defined for all $i \in R_{D2} \cap R_{D3} \cap R_{D4}$.

For the post-stratification adjustment, we divide an estimate of the total 1995 population in domain *D* who would be eligible for Waves II, III and IV by an estimate of the same population from Wave IV data as follows:

$$F_{D4}^{L} = \frac{\hat{X}_{E_{D2} \cap E_{D3} \cap E_{D4}}}{\sum_{i \in R_{D2} \cap R_{D3} \cap R_{D4}} \omega_{D4i}^{L}} \qquad \text{(B8)}$$

where

$$\hat{X}_{E_{D2} \cap E_{D3} \cap E_{D4}} = \sum_{i \in E_{D2} \cap E_{D3} \cap E_{D4}} \omega_{3i} \qquad \text{(B9)}$$

is the sum of the Wave III sample weights for the sample members who were eligible in Waves II, III and IV (*i.e.,* Waves I, II and III respondents who were alive at Wave IV). Table B-2 shows the subpopulation estimates (B9) for the grade-sex-race poststratification domains.

Then the final Wave IV grand sample longitudinal weight, which is defined for all $i \in R_{D2} \cap R_{D3} \cap R_{D4}$ is

$$\omega_{F,D4i}^{L} = F_{D4}^{L} \omega_{D4i}^{L} \qquad \text{(B10)}$$

The untrimmed longitudinal weights were computed using Equations B6 to B10. The trimmed longitudinal weights were computed in the same manner, except that the nonresponse-adjusted weight resulting from Equation B7 was trimmed to a value of 6000 before being substituted into Equations B8, B9 and B10.

Public-use Sample Weights

The calculation of the public-use sample cross-sectional and longitudinal weights follow the approach used above for the grand sample cross-sectional and longitudinal weights. The only distinctions are that the public-use sample weights from Waves I and III replace the grand sample weights in the formulas, and the nonresponse adjustments are calculated separately for each grade-sex cell. Tables B-3 and B-4 show the subpopulation estimates (B4 and B9, respectively) for the grade-sex-race poststratification domains, derived using the Wave I and Wave III public-use sample weights, respectively, for all public-use sample cases that were alive at Wave IV.

Table B-5 summarizes the total population estimates derived from all the sample weights.

Table B-1. Wave I Subpopulation Estimates for Wave IV Poststratification Domains
(Based on Wave I grand sample weights for all cases alive at Wave IV)

Gender	Grade	All Races	Black	Non-Black
Male	6	81	.	81
	7	1,918,767	272,464	1,646,303
	8	1,875,833	303,477	1,572,356
	9	1,920,770	308,641	1,612,129
	10	1,710,623	310,486	1,400,137
	11	1,762,643	264,022	1,498,621
	12	1,952,873	334,848	1,618,025
	Total	11,141,590	1,793,938	9,347,652
Female	.	2,168	.	2,168
	7	1,918,940	295,601	1,623,339
	8	1,715,004	294,068	1,420,936
	9	1,818,086	274,891	1,543,195
	10	1,852,448	339,427	1,513,021
	11	1,649,940	288,115	1,361,825
	12	1,871,578	324,080	1,547,498
	Total	10,828,164	1,816,182	9,011,982
Total		21,969,754	3,610,120	18,359,634

Table B-2. Wave III Subpopulation Estimates for Wave IV Poststratification Domains

(Based on Wave III grand sample longitudinal weights for all cases alive at Wave IV)

Gender	Grade	All Races	Black	Non-Black
Male	7	1,837,171	252,007	1,585,164
	8	1,837,493	292,612	1,544,881
	9	1,879,016	291,147	1,587,869
	10	1,679,980	299,262	1,380,718
	11	1,668,105	241,726	1,426,379
	12	380,818	60,450	320,368
	Total	9,282,583	1,437,204	7,845,379
Female	7	1,902,008	282,017	1,619,991
	8	1,708,678	287,536	1,421,142
	9	1,815,611	268,997	1,546,614
	10	1,856,602	342,710	1,513,892
	11	1,608,220	274,311	1,333,909
	12	384,660	62,030	322,630
	Total	9,275,779	1,517,601	7,758,178
	Total	18,558,362	2,954,805	15,603,557

Table B-3. Wave I Subpopulation Estimates for Wave IV Poststratification Domains
(Based on Wave I public-use sample weights for all cases alive at Wave IV)

Gender	Grade	All Races	Black	Non-Black
Male	7	1,895,423	258,522	1,636,901
	8	1,827,597	303,759	1,523,838
	9	1,988,973	324,216	1,664,757
	10	1,759,754	336,793	1,422,961
	11	1,726,694	252,039	1,474,655
	12	1,933,868	329,796	1,604,072
	Total	11,132,309	1,805,125	9,327,184
Female	7	1,904,478	302,111	1,602,367
	8	1,736,932	282,166	1,454,766
	9	1,862,084	289,814	1,572,270
	10	1,869,993	319,861	1,550,132
	11	1,657,389	302,717	1,354,672
	12	1,850,851	320,988	1,529,863
	Total	10,881,727	1,817,657	9,064,070
Total		22,014,036	3,622,782	18,391,254

Table B-4. Wave III Subpopulation Estimates for Wave IV Poststratification Domains

(Based on Wave III public-use sample longitudinal weights for all cases alive at Wave IV)

Gender	Grade	All Races	Black	Non-Black
Male	7	1,860,549	237,089	1,623,460
	8	1,814,151	293,325	1,520,826
	9	1,951,814	308,508	1,643,306
	10	1,754,860	321,152	1,433,708
	11	1,662,334	240,478	1,421,856
	12	283,818	65,733	218,085
	Total	9,327,526	1,466,285	7,861,241
Female	7	1,888,831	282,139	1,606,692
	8	1,714,545	272,313	1,442,232
	9	1,824,688	275,746	1,548,942
	10	1,865,095	335,928	1,529,167
	11	1,648,269	302,560	1,345,709
	12	258,365	45,130	213,235
	Total	9,199,793	1,513,816	7,685,977
	Total	18,527,319	2,980,101	15,547,218

Table B-5. Weight Totals

Weights	Sample Members with Sample Weights		Sample Members Eligible (Alive) at Wave IV	
	n	Sum of Weights	n	Sum of Weights
Wave IV Grand Sample Nonresponse-Adjusted Untrimmed Cross-sectional Weight	14,800*	21,969,754	14,800*	21,969,754
Wave IV Grand Sample Nonresponse-Adjusted Trimmed Cross-sectional Weight	14,800*	21,614,717	14,800*	21,614,717
Wave IV Grand Sample Poststratified Untrimmed Cross-sectional Weight	14,800*	21,967,505	14,800*	21,967,505
Wave IV Grand Sample Poststratified Trimmed Cross-sectional Weight	14,800*	21,967,505	14,800*	21,967,505
Wave IV Grand Sample Nonresponse-Adjusted Untrimmed Longitudinal Weight	9,421**	18,558,364	9,421**	18,558,364
Wave IV Grand Sample Nonresponse-Adjusted Trimmed Longitudinal Weight	9,421**	18,248,782	9,421**	18,248,782
Wave IV Grand Sample Poststratified Untrimmed Longitudinal Weight	9,421**	18,558,362	9,421**	18,558,362
Wave IV Grand Sample Poststratified Trimmed Longitudinal Weight	9,421**	18,558,362	9,421**	18,558,362
Wave IV Public-Use Sample Nonresponse-Adjusted Cross-sectional Weight	5,114	22,014,034	5,114	22,014,034
Wave IV Public-Use Sample Poststratified Cross-sectional Weight	5,114	22,014,038	5,114	22,014,038
Wave IV Public-Use Sample Nonresponse-Adjusted Longitudinal Weight	3,342	18,527,318	3,342	18,527,318
Wave IV Public-Use Sample Poststratified Longitudinal Weight	3,342	18,527,319	3,342	18,527,319

** Of the 15,701 respondents in Wave IV, 901 had missing Wave I weights.*

*** In order to have a positive longitudinal weight, a case would have to be a respondent in all waves considered for the longitudinal weights. There were 9,421 cases that responded in all four waves.*

Table B-6. Correspondence between Equations in Appendix B and Variable Names

Equation		Grand Sample		Public-use Sample
		Untrimmed	Trimmed	
Cross-sectional Weights				
B1 Nonresponse adjustment	Numerator	GSCRB1NU		PSCRB1NU
	Denominator	GSCRB1DE		PSCRB1DE
	Result	GS4NRCR		PGS4NRCR
B2 NR-adjusted weights		GSW4CR1U	GSW4CR1T	PGSW4CR1
B3 Poststratification adjustment	Denominator	GSCRB3DU	GSCRB3DT	PSCRB3DE
	Result	GS4PSCRU	GS4PSCRT	PGS4PSCR
B4		GSCRB4	GSCRB4	PSCRB4
B5 Final weight		GSW4CR2U	GSW4CR2T	PGSW4CR2*
Longitudinal Weights				
B6 Nonresponse adjustment	Numerator	GSLOB6NU		PSLOB6NU
	Denominator	GSLOB6DE		PSLOB6DE
	Result	GS4NRLO		PGS4NRLO
B7 NR-adjusted weights		GSW4LO1U	GSW4LO1T	PGSW4LO1
B8 Poststratification adjustment	Denominator	GSLOB8DU	GSLOB8DT	PSLOB8DE
	Result	GS4PSLOU	GS4PSLOT	PGS4PSLO
B9		GSLOB9	GSLOB9	PSLOB9
B10 Final weight		GSW4LO2U	GSW4LO2T	PGSW4LO2**

*Add Health public-use cross-sectional variable name is GSWGT4_2

** Add Health public-use longitudinal variable name is GSWGT4