Add Health Weight Components

Kim Chantala
C. M. Suchindran

Carolina Population Center
UNC at Chapel Hill

Last Update: April 28, 2011

INTRODUCTION

This document describes the weight components available for the Add Health sample. A weight component is available for each level of sampling (schools and adolescents) at each wave of data and are useful for constructing sampling weights that differ in meaning from the sampling weights designed for estimating population-average (single-level) models that have been traditionally distributed with the Add Health data. In particular, these weight components are the basic building blocks needed for computing the multilevel weights with the methods given in the web document located at:
http://www.cpc.unc.edu/research/tools/data_analysis/ml_sampling_weights

A general description of the weight components available for each wave is presented in the next section followed by a section describing the details of the construction of the weight components and adjustments made for school and adolescent non-response.

AVAILABLE WEIGHT COMPONENTS

A weight component is available for each level of sampling (schools and adolescents) at each wave of data. The level-2 (school) weight component is the sampling weight computed for each school with adolescents participating in a specific round of Add Health data collection. This weight estimates the number of schools in the United States represented by the school contributing participating adolescents. The level-1 (adolescent) weight component is a sampling weight for each adolescent who participated in a specific round of Add Health data collection. The level-1 weight component for adolescent i interviewed at a specified wave represents the number of adolescents enrolled in school j for the 1994-1995 academic year who shared the same characteristics with adolescent i. Thus the sum of the level-one weights for all adolescents enrolled in a specific school is equal to the total enrolled in the sampled grades for that school.

In-School Weight Components

There were 128 of the 132 schools selected into the Add Health sample that allowed administration of the In-School Survey. All students from these 128 sampled schools who were in attendance on the day the survey was administered were asked to fill out the In-School Survey. The available weight components for the In-School data collection are listed in Table 1.

K:\ADHEALTH\SHARED\weights\wtcompon_doc.docx

Table 1. Available weight components designed for use with the In-School data.

Statistical Measures	Level-2 (School) weight component	Level-1 (Adolescent) weight component
Variable Name	schwt128	insch_wc
N	128	83,135
Minimum	36.53	0.89
Median	155.49	1.20
Maximum	4,244.76	7.38
Mean	251.42	1.33
Variance	182,682	0.253
Sum	32,181.22	110,524

In-Home Weight Components for Wave I, II, and III

Enrollment rosters from all 132 schools selected into the Add Health sample were used as the sampling frame for the Wave I In-Home sample. The sampling frame for the Wave II and Wave III sample was derived from the Wave I participants. The available weight components for the In-Home data collection are listed in Table 2.   Since all 132 schools had participants in each of the three sets of data, only one level-2 weight component for schools (schwt1) is needed for analyzing any combination of data from Wave I, Wave II or Wave III. However, a different level-1 (adolescent) weight component is needed for each Wave of data.

Table 2. Available weight components designed for use with the Wave I, II, or III In-Home Data.

Statistical Measures	Level-2 (School) weight component	Level-1 (Adolescent) weight component			
		Wave I In-Home	Wave II In-Home	Wave III In-Home[1]	Wave III In-Home[2]
Variable Name	schwt1	w1_wc	w2_wc	w3_2_wc	w3_wc
N	132	18,924	13,568	14,322	10,828
Minimum	35.89	1.00	1.00	1.00	1.00
Median	151.80	3.31	4.05	4.12	4.77
Maximum	4,170.13	119.98	119.98	119.98	119.98
Mean	243.80	6.17	7.18	8.12	8.96
Variance	171,803.92	54.88	70.86	102.20	119.40
Sum	32,181.22	116,706.97	97,361.25	116,266.04	97,018.23

[1] This weight component is constructed for the participants who filled out both Wave I and Wave III In-Home Surveys.
[2] This weight component is constructed for the participants who filled out Wave I, Wave II and Wave III In-Home Surveys.

K:\ADHEALTH\SHARED\weights\wtcompon_doc.docx

The Wave I In-Home survey selected adolescents from the enrollment rosters of the 132 schools with unequal probability of selection. Several special over-sampled groups of adolescents were also recruited for the Wave I interview. These include the core sample (roughly equal-sized samples), purposively selected schools (all students selected), non-genetic supplements (Black adolescents whose parents were college graduates, adolescents whose race was Cuban, Puerto Rican, or Chinese), the disabled sample, and the genetic supplement (biologically related adolescents, non-related adolescents living together). The Wave I level-1 weight component (w1_wc) was computed from grade-sex-race enrollment estimates for each school and the Wave I grand sample weight (gswgt1) that was designed for estimating population-average models.

The Wave II sample of adolescents was selected from the Wave I respondents. However, the following Wave I respondents were excluded from being eligible at Wave II:
- Wave I 12th graders who were not part of the genetic sample. Note this excludes most of the adolescents who were in 12th grade at Wave I
- Disabled adolescents who did not belong to any other over-sampled groups
- Adolescents found to be deceased at time of Wave II data collection

To compute the Wave II level-1 weight components (w2_wc), non-response adjustments were made to the Wave I level-1 weight components for sex by race (black/non-black) cells for adolescents attending each school. No attempt was made to inflate the level-1 weight components for the 12th graders to reflect the population represented by the 12th graders from Wave I.

The Wave III sample of adolescents who have weight components were selected from the 18,924 Wave I respondents who are part of the probability sample. Only the Wave I adolescents who were determined to be deceased at the Wave II interview were excluded from Wave III eligibility. Non-response adjustments were calculated separately by sex and race (black/non-black) for adolescents attending each school to compute the Wave III level-1 weight components from the Wave I level-1 weight components. Two level-2 weight components were constructed for the Wave III sample. The weight component variable w3_wc is designed for use with the 10,828 respondents who filled out the In-Home survey at Waves I, II, and III. It is also the appropriate weight component to use if analyzing data from Wave II and III. The level-2 weight component w3_2_wc is designed for use with the 14,322 respondents who filled out the survey at Wave I and Wave III. This weight component is also appropriate for cross-sectional analysis using only data from Wave III.

COMPUTATIONAL DETAILS

This section provides a description of the methods used to construct the Add Health weight components. First, we describe the computation of the level-2 weight (school) components for the In-School and In-Home data sets. Next, we describe the adjustments made to the population-average sampling weight available with the In-School data to create the level-1 In-School weight components. These weight components were used to define the grade-sex-race enrollment estimates for each school needed to compute the level-1 weight components for adolescents participating in Wave I. Because the Wave I respondents were used as the sampling frame for subsequent waves of data collection, the level-1 weight components for

Wave II and Wave III were computed by making non-response adjustments to the level-1 weight component from Wave I.

Level-2 (School) Weight Component

The level-2 sampling weight component is the final sampling weight computed for each school with adolescents participating in the Add Health Survey. It is an estimate of the number of schools in the United States that are similar to the recruited school.

There were 80 high schools and 52 feeder schools selected into the Add Health Sample. While all of these 132 schools agreed to provide enrollment rosters for recruiting students for the In-Home survey, only 128 of the schools permitted administration of the In-School survey.   Hence, a different level-2 weight component is needed for use with the In-School data versus data from any of the In-Home interviews.

Level-2 Component for Schools with Adolescents completing In-Home Interviews: schwt1

Adolescents from each of the 132 schools were recruited for the Wave I In-Home survey and then followed for the subsequent in-home surveys administered at Wave II and Wave III. The level-2 weight component for these 132 schools is available in a variable named schwt1.

Level-2 Component for Schools Allowing In-School Interviews: schwt128

The level-2 weight component for schools who allowed students to participate in the In-School Survey can be computed by making a non-response adjustment to the In-Home level-2 weight-component.   Let Q_j be an indicator variable with a value of 1 if school j allowed in-school data collection and a value of 0 otherwise. The level-2 weight component becomes:

$$schwt128_j = schwt1_j * \frac{\sum_{j=1}^{132} schwt1_j}{\sum_{j=1}^{132} schwt1_j Q_j} \tag{1}$$

Level-1 (Adolescent) Weight Components

Each wave of data has a different level-1 weight component. The weight component for adolescent i interviewed at a specified wave represents the number of adolescents enrolled in school j for the 1994-1995 academic year who shared the same characteristics with adolescent i. Appendix A provides the details of imputation of missing or out of range values of grade and sex for participants in the In-School and Wave I Surveys. If available, the grade assignments from the In-School survey were used for defining grade-sex-race cells for computing the level-1 weight components at Wave I. The same cell assignment was used at Waves II and III.   Subscripts used in the formulas that follow denote the following groups:

  i = adolescent
  j = school
  g = grade of adolescent for 1994-1995 academic year
  s = sex of adolescent
  r = race (black/non-black) of adolescent

In-School Weight Component: insch_wc

The In-School weight component is computed from the student in-school sampling weight that has been adjusted for non-response (variable $\text{kidwgt}_{i,j}$) and the number of students enrolled (E_j) in any grades at the school that were within the 7th to 12th grade range of eligible grades. For adolescent i attending school j the in-school weight component is computed as follows:

$$insch\_wc_{i|j(gs)} = \frac{E_j}{\sum\limits_{g}^{\#grades}\sum\limits_{s}^{2}\sum\limits_{i}^{m_{j(gs)}} kidwt_{i,j(gs)}} * kidwt_{i,j(gs)} \qquad (2)$$

This equation reduces to the inverse of the response rate for each grade-sex (g,s) cell in the school. This can be shown as follows: The in-school sampling weight for adolescent i has been computed as the School weight (S_j) multiplied by the inverse of the response rate for the grade-sex cell to which adolescent i belongs:

$$kidwt_{i,j(gs)} = S_j \frac{n_{j(gs)}}{m_{j(gs)}}$$

The response rate is the number of adolescents of sex s in grade g who filled out the in-school survey (designated by $m_{j(gs)}$) divided by the number of adolescents of sex s enrolled in grade g (designated by $n_{j(gs)}$). The total number of adolescents enrolled in the eligible grades is just the sum of the number of students enrolled in each grade-sex combination. By substituting these quantities into equation 2 we can show that the in-school weight component is just the inverse of the response rate for the grade-sex cell to which an adolescent belongs:

$$insch\_wc_{i|j} = \frac{\sum\limits_{g}^{\#grades}\sum\limits_{s}^{2} n_{j(gs)}}{\sum\limits_{g}^{\#grades}\sum\limits_{s}^{2}\sum\limits_{i}^{m_{j(gs)}} S_j \frac{n_{j(gs)}}{m_{j(gs)}}} * S_j \frac{n_{j(gs)}}{m_{j(gs)}}$$

The School weight (S_j) cancels out of the equation while the inner-most summation in the denominator (indexed by i) has the effect of multiplying by the number of adolescents who completed the questionnaire ($m_{j(gs)}$).

$$insch\_wc_{i|j(gs)} = \frac{\sum\limits_{g}^{\#grades}\sum\limits_{s}^{2} n_{j(gs)}}{\sum\limits_{g}^{\#grades}\sum\limits_{s}^{2} m_{j(gs)} \frac{n_{j(gs)}}{m_{j(gs)}}} * \frac{n_{j(gs)}}{m_{j(gs)}}$$

which reduces to:

$$insch\_wc_{i|j} = \frac{n_{j(gs)}}{m_{j(gs)}}$$

Wave I In-Home Level-1 Weight Component: w1_wc.

The level-1 weight components for Wave I were constructed from the final sampling weight (Add Health variable named gswgt1) designed for estimating population average models and enrollment estimates derived from the In-School level-1 weight components.   The Wave I final sampling weight was computed using a multiplicity approach. First a sampling weight was computed for each of the over-sample groups for which the participant was eligible. Then the Wave I final sampling weight was computed by taking an average of these weights and then applying a post-stratification adjustment. Non-response adjustments and post-stratification adjustments were automatically incorporated in this weight.   The Wave I sampling weights for each adolescent i attending school j have the form:

$$gswgt1_{i,j} = X_{g,s,r}W_{i,j}^{*} = X_{g,s,r}S_{j}W_{i|j}^{*}$$

where:

$gswgt1_{i,j}$ = the number of adolescents in the U. S. that are represented by adolescent i
sampled from school j.

$X_{s,g,r}$ = sex-grade-race post-stratification factor made to ensure that the Add Health weights
would agree with the Census Bureau estimates of the size of each grade-sex-race
(black vs. non-black) subpopulation in the US for the year the schools were sampled.

$W_{i,j}^{*}$ = the average of the sampling weights across eligible groups for adolescent i sampled
from school j.

S_{j} = the sampling weight for school j

$W_{i|j}^{*}$ = the average of the level 1 weight components from each eligible sample for adolescent
i sampled from school j.

Because $X_{s,g,r}$ is not available, the level 1 weight component can be estimated as follows. First an estimate the number of adolescents ($M_{j(gsr)}$) in each grade-sex-race cell at school j is computed from the level 1 in-school weight components:

$$M_{j(gsr)} = \sum_{i} insch\_wc_{i|j(gsr)}$$

The general formula for estimating the level 1 weight component for adolescent i who was interviewed from school j at Wave I was computed as:

K:\ADHEALTH\SHARED\weights\wtcompon_doc.docx

$$wt\_comp_{i|j(gsr)} = \frac{M_{j(gsr)}}{\sum_i gswgt1_{i,j(gsr)}} gswgt1_{i,j(gsr)} \qquad (3)$$

Note that because the post-stratification factor ($X_{g,s,r}$) and school weight (S_j) are common factors in the final weight (gswgt1) used in both the denominator and numerator of equation 3, these factors cancel out of the equation. Hence, the only information needed to compute the level-1 weight component is $M_{j(gsr)}$ estimated from the in-school level-1 weight component and the final weight for a respondent available from each wave of data. Formulas for each in-home weight component can be obtained by substituting the appropriate population-average sampling weight into the above formula.

The value of $M_{j(gsr)}$ cannot be estimated for the four schools that did not allow the in-school data collection. For these schools the in-home weight component for student i attending school j will be computed as:

$$wt\_comp_{i|j} = \frac{E_j}{\sum_i gswgt1_{i,j}} * gswgt1_{i,j} \qquad (4)$$

where E_j is the number of students enrolled in any grades at the school that were within the 7th to 12th grade range of eligible grades.

After these computations were done there were a few very small weights. We trimmed these small weights (imposing a minimum value of 1 and redistributing the deficit over all of the weights so that the sum of the weights remained a constant).

Wave II In-Home Level-1 Weight Component: w2_wc

The level-1 weight component for adolescents participating in the Wave II Survey was computed by making a non-response adjustment to the level-1 weight component for eligible respondents from Wave I. This adjustment was done for each sex by race (black/non-black) cell within each school. Let Q_j be an indicator variable with a value of 1 if the eligible adolescent from Wave I participated at Wave II and a value of 0 otherwise. The level-1 weight component ($w2\_wc_{i|j}$) for adolescent i who belongs to the sex-race cell s,r sampled from school j becomes:

$$w2\_wc_{i|j(s,r)} = w1\_wc_{i|j(s,r)} * \frac{\sum_{i=1}^{n_{j(sr)}} w1\_wc_{i|j(s,r)}}{\sum_{i=1}^{n_{j(sr)}} w1\_wc_{i|j} Q_{j(s,r)}} \qquad (5)$$

Thus, the level-1 weight component at Wave II is the corresponding Wave I weight component inflated by the inverse of the weighted response rate for a particular sex-race cell at each school.

There were two schools that did not have eligible adolescents participating in the Wave II interview from each of the four sex-race cells used in equation 5.   For adolescents sampled from these schools the non-response adjustment was done separately for each sex sampled from the school:

$$w2\_wc_{i|j(s)} = w1\_wc_{i|j(s)} * \frac{\sum_{i=1}^{n_{j(s)}} w1\_wc_{i|j(s)}}{\sum_{i=1}^{nj(s)} w1\_wc_{i|j} Q_{j(s)}} \tag{6}$$

We trimmed these small weights (imposing a minimum value of 1 and redistributing the deficit over all of the weights so that the sum of the weights remained a constant).

Wave III In-Home Level-1 Weight Component: w3_wc and w3_2_wc

The level-1 weight components for Wave III were calculated from the Wave I weight components in a manner identical to the Wave II level-1 weight component.

APPENDIX A. IMPUTING SEX AND GRADE

In-School Imputation Rules

Imputing Grade

Less than 1 percent of the participants who had participated in the In-School survey had missing information on grade at time of interview. Table 1 summarizes the rules to impute grade for these participants.

Table 1. Imputation of Grade at time of In-School Survey

Rule	N	Percent
0 = No imputation, Grade is from In-school Questionnaire	82,374	99.08
1 = Assign Grade according to value of sampling weight for grade-sex-race cell for sampled school [1]	621	0.75
2 = No imputation, Grade is from Wave I Questionnaire	13	0.02
3 = Assign modal grade for age of case. If this grade exceeds the maximum grade in the school, then assign the maximum grade in school.	100	0.12
4 = Assign modal grade for years school was attended by case.	2	0.00
5 = Random assignment to grades sampled from School	25	0.03
Total	83,135	100.00

[1] Because of the manner in which the In-School weights were computed, all students with in the same grade-sex-race cell for a school had the same value of the sampling weight.

Imputing Sex

Less than 1 percent of the participants were missing information on their biological sex. Table 2 summarizes the rules to impute biological sex for these participants.

Table 2. Imputation of Sex for In-School Survey

Rule	*N*	*Percent*
0 = No imputation, Sex is from In-School Survey	82,541	99.29
1 = Assign Sex according to value of In-School sampling weight for grade-sex-race cell for sampled school [1]	445	0.54
2 = No imputation, Sex is from Wave I Survey[2]	17	0.02
3 = Use friends with whom Respondent talked about problems[3]	131	0.16
4 = Participant attends All Boy School	1	0.00
Total	83,135	100.00

[1] Because of the manner in which the In-School weights were computed, all students with in the same grade-sex-race cell for a school had the same value of the sampling weight.
[2] Sex assignment is based on the reported biological sex in the Wave I In-Home Survey
[3] Sex assignment is based on number and sex of friends with whom adolescent reports discussing problems.
  a) If adolescent reported not discussing problems with any friends or only male friends then sex of adolescent was assumed to be male.
  b) If adolescent reported discussing problems with only female friends or friends of both sexes, then sex of adolescent was assumed to be female.

Wave I Imputation Rules

Every attempt was made to keep the grade assignment for a participant consistent across all surveys. Grade assignment used in creating the Wave I weight components was chosen from the In-School survey for 76% of the Wave I respondents. For respondents not interviewed during the In-School Survey, grade was selected from the Wave 1 Survey when available and within the range of grades selected from the school.

Table 3. Rules used to impute grade at Wave I.

Rule	N	Percent
1=Use grade from In-School Survey	14,395	76.07
2=Use grade from computing Wave 1 sampling weights, this grade is within grade range for sampled school	3838	20.28
3= Use grade from computing Wave 1 sampling weights, grade range for sampled school is unknown	528	2.79
4=Set to minimum grade from sampled school because grade used for computing Wave 1 sampling weights is less than minimum grade sampled from school	70	0.37
5= Set to maximum grade from sampled school because grade used for computing Wave 1 sampling weights is less than maximum grade sampled from school	92	0.49
6=Randomly assign to grade sampled from School.	1	0.01
Total	18,924	100.00