Add Health
The National Longitudinal Study of Adolescent to Adult Health

Construction of Wave IV dbGaP GWAS Sample Weight

Report prepared by

Ping Chen

Chirayath M Suchindran

Joyce Tabor

Kathleen Mullan Harris

UNC CAROLINA POPULATION CENTER

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

Add Health is supported 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.

https://doi.org/10.17615/g35a-w564

Introduction

This document provides a brief overview of the steps in constructing GWAS sample weights. It also reports results of some statistical analysis using the constructed weights.

The total number of GWAS sample (*N=9,975*) who were actually assayed and number of those (*N=12,234*) who consented to be archived and assayed are different. When the proportions of subjects who have consented and who have actually been assayed are different for certain Add Health post-stratification domains, this might cause problems if we simply use grand sample weights for GWAS sample (Bethlehem 2002). We thus developed weights for this special sample.

Data Construction

First, we used post-stratification variables, including gender, grade (7-12), and race (black and non-black), to create 24 Add Health domains. Within each domain, we separately calculated the number of respondents who have been assayed (N_a) and number of respondents who have consented to be assayed (N_c).

Second, we calculated response rates within each domain by dividing N_a by N_c. **Table 1** displays the number of respondents who were assayed, number of respondents who have consented to be assayed, and response rate for each of the 24 post-stratification domains. It shows that response rates vary across different post-stratification domains.

Table 1. Response Rates of Wave IV Post-Stratification Domains for GWAS Sample

Black Respondents				
Gender	**Grade**	**# of Respondents Assayed**	**# of Respondents Consented to be Assayed**	**Response Rate**
Male	7	163	200	0.815
	8	165	190	0.868
	9	162	196	0.827
	10	179	217	0.825
	11	147	184	0.800
	12	147	170	0.865
	Total	**963**	**1,157**	---
Female	7	180	215	0.837
	8	190	234	0.812
	9	216	257	0.840
	10	244	285	0.856
	11	203	241	0.842
	12	171	203	0.842
	Total	**1,204**	**1,435**	---
Non-Black Respondents				
Gender	**Grade**	**# of Respondents Assayed**	**# of Respondents Consented to be Assayed**	**Response Rate**

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and 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.

Male	7	472	598	0.789
	8	510	599	0.851
	9	703	848	0.829
	10	744	922	0.807
	11	730	902	0.810
	12	616	738	0.835
	Total	**3,775**	**4,607**	**---**
Female	7	591	739	0.800
	8	545	680	0.801
	9	738	914	0.807
	10	744	943	0.789
	11	704	903	0.780
	12	711	856	0.831
	Total	**4,033**	**5,035**	**---**
Total Respondents				
Gender	**Grade**	**# of Respondents Assayed**	**# of Respondents Consented to be Assayed**	**Response Rate**
Male	Total	4,738	5,764	---
Female	Total	5,237	6,470	---
	Total	**9,975**	**12,234**	**---**

Third, we calculated the inverse of the response rates in each domain.

Fourth, we multiplied Wave IV *grand* sample weights by the inverse of the response rates, which produced GWAS sample weights for respondents who have been assayed and have grand sample weights (N=9,404).

Lastly, we also calculated weights for respondents (N=571) who were assayed but *missing* Wave IV grand sample weights. We assigned them with value 1 as the base weight. We then multiplied their base weight of 1 by the inverse of the response rate of their domain.

Data Summary Statistics

Table 2 provides summary statistics of the final constructed GWAS weights for a total of 9,975 respondents. We also calculated the range and mean of GWAS weights separately for two sub-groups. One is for GWAS respondents who are *not* missing Wave IV grand sample weight (N=9,404); and the other for those who are missing grand sample weights (N=571).

Table 2. Summary Statistics of Weight Variables

	Minimum	**Maximum**	**Mean**	**N**
Final Wave IV GWAS Sample Weight	**1.152**	**23185.16**	**1758.701**	**9,975**
GWAS Weight Only for Respondents Who Are *Not* Missing Wave IV Grand Sample Weight	24.154	23185.16	1865.143	9,404
GWAS Weight Only for Respondents Who are Missing Wave IV Grand Sample Weight	1.152	1.283	1,223	571

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and 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.

Note: *one case was dropped after the GWAS weights were constructed. The exclusion of the weight for this case won't affect analysis results. The final data file includes 9,974 respondents with weights.*

In addition, we conducted some descriptive statistical analysis of Wave IV outcome variables to compare results using three types of weights. **Table 3** shows the comparative results. We used two outcome variables. One is daily smoker which is a binary measure; and the other is BMI which is a continuous measure. We calculated the proportion for daily smoker and mean for BMI along with their standard errors.

Table 3. Proportions/Means and Their Standard Error of Two Wave IV Outcome Variables Using Three Types of Weights

	Weight 1		**Weight 2**		**Weight 3**	
	Final Wave IV GWAS Sample Weight		Wave IV GWAS Sample Weight When Respondents Who are Missing WIV Grand Sample Weights are excluded		Wave IV Grand Sample Weight with Subproportion Analysis	
	Proportion	**Standard Error**	**Proportion**	**Standard Error**	**Proportion**	**Standard Error**
WIV Daily Smoker	0.2651083	0.0105232	0.2651087	0.0105233	0.264839	0.0105311
N	9,679*		9,336		9,336	
	Mean	**Standard Error**	**Mean**	**Standard Error**	**Mean**	**Standard Error**
WIV BMI	29.16692	0.1480495	29.16692	0.1480517	29.18109	0.1484379
N	9,640*		9,294		9,294	

Note: *Total N is smaller than the final GWAS sample (9,975) because both outcome variables have missing data. All the analysis was adjusted for school clustering and stratification by region as well.*

The first analysis used the *final* GWAS sample weights for the full GWAS sample of 9,975 respondents. The second analysis excluded 571 respondents who were missing grand sample weights, and was limited to a sub-sample of 9,404 respondents whose GWAS weights were computed using the non-missing grand sample weights. The third one took the subpopulation analysis approach using the original *grand* sample weights. A binary subpopulation variable was constructed with 1 referring to those who were assayed and have grand sample weights and 0 otherwise (referring to those who are not in the genetic sample along and 571 respondents who are missing grand sampling weights).

The result using the *final GWAS sample weights* (*Weight 1*) is almost identical with the one that is limited to the subsample of respondents who have GWAS weights and are not missing grand sample weights (*Weight 2*). The result from grand sample weights (*Weight 3*) and subpopulation analysis is also similar to the results from the other two subsamples. The slight difference arises in the third or fifth decimal of the estimates for daily smoker; and appears in the second or fourth decimal of the estimates for BMI.

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and 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.

References

Bethlehem, J. G. (2002). “Weighting Nonresponse Adjustments Based on Auxiliary Information.” in Groves, R., Dillman, D., Eltinge, J., and Little, R. (eds.) Survey Nonresponse. New York, NY: Wiley.

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and 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.