Supplemental Appendix A. Analyses for the Full Cohort Sample

Methods

Several of the variables included in the analyses had missing values, especially those collected at the Age 18 laboratory visit (see Table S1). Despite the high proportion of missing values for some variables, concern over bias is somewhat mitigated because a large portion of missing outcome data are missing by design (i.e., only some remaining participants were invited to participate in the laboratory visit at Age 18, based in part on proximity). Regardless, as a form of sensitivity analysis,[1] we used missing data methods to conduct analyses in the Full Cohort Sample and compared the results to those for the Complete Case sample, which are reported in the Results section of the manuscript.

A large proportion of outcomes were missing, thereby reducing the efficiency advantages of multiple imputation (MI) and heightening concerns about correct specification of the imputation models. However, some of the predictors were also missing, preventing straightforward implementation of methods based on iterative proportional weighting (IPW). Thus, we combined MI with IPW, using multiple imputation to handle missing data in predictors and other variables predictive of outcome missingness, and then using IPW to handle missingness in the outcome variable (i.e., Late Adolescent Waist-to-Height).[2,3]

We used R version 3.2.0 to generate multiple imputations and evaluate which predictors should be used to calculate the IPW weights. We subsequently used R version 3.4.4 to implement weighted linear regression for the multiply-imputed datasets and to combine the results.

Multiple Imputation

We used the mice package for R (version 2.3.0) to implement multiple imputation by chained equations (also known as fully conditional specification)[4,5] under the standard assumption of missing at random (MAR). We followed standard recommendations for implementing multiple imputation by chained equations.[4,5]

Data were imputed separately for girls and for boys. The imputation dataset included all variables used in the analysis, including interaction terms (i.e., CSR x Childhood BMI). If a variable was derived from two or more other variables (i.e., Late Adolescent Waist-to-Height, Childhood Self-Regulation, Childhood SES, Adolescent Behavior Index), those variables (rather than the derived variable) were included in the imputation dataset whenever possible. However, for scales (e.g., EATQ Attention and CBQ Inhibitory Control), we included the total scores for each informant—rather than item-level data—in the imputation model because alphas were high and there were few missing items for informants who responded.[5] The imputation dataset also included auxiliary variables selected from among measures that, based on prior research, potentially predicted either the *values of* or *missingness in* the variables used in the analyses, especially the outcome variable.[6] Including auxiliary variables predictive of missingness in analysis variables, in particular the outcome, has utility for increasing the validity of the MAR assumption and for implementing IPW. The auxiliary variables included in the imputation model were: variables collected when the participant was *in utero* (e.g., number living in household, maternal age at birth of first child, maternal weight during pregnancy); maternal and paternal personality variables collected when the participant was *in utero* or an infant (i.e., Achievement and Nurturance from the Jackson Personality Inventory and Neuroticism and Extraversion from the Eysenck Personality Questionnaire); participant variables (internalizing scores, externalizing scores, school engagement, and stimulant use) collected at Age 9; maternal and paternal body mass index collected when the participant was Age 13; participant sedentary and physical activity variables (i.e., hours spent playing sports, doing active exercise, watching television or videos, using the computer, or playing video games) collected at Age 13; and participant alcohol use collected at Ages 15, 16, and 17.

Prior to performing multiple imputation, we used univariate Box-Cox transformations to transform continuous variables with markedly skewed distributions, although we expected predictive mean matching to be robust to less marked departures from normality.[4] After multiple imputation, we applied the inverse transformation to return variables used in analyses to their original scale before conducting analyses.

For imputation models, we used predictive mean matching for continuous variables and semicontinuous variables[7] and logistic regression for binary variables. As a form of sensitivity analysis, we instead used a Gaussian model for continuous data, which produced similar results. After generating preliminary multiple imputations,[4,5] we used Lasso, implemented via the R glmnet package (version 2.0-10), to decide which other variables to include in the imputation model for a given variable. More specifically, for each variable in the imputation dataset (e.g., "Variable *j*"), we used Lasso to determine which other variables were most predictive of either *missingness* in Variable *j* or the *values* of Variable *j*. We implemented Lasso with $\lambda = 30$ if Variable *j* was continuous, but with $\lambda = 5$ or $\lambda = 10$ if Variable *j* was binary; we chose a more conservative value of λ for binary variables to avoid problems with perfect prediction. For continuous variables, we conducted sensitivity analysis with respect to λ, which determines the penalty for including more variables, but varying λ had little effect on results. Variables selected by Lasso were included in imputation models used to generate the final multiple imputations (for use in IPW), with the caveat that outcome variables had to be included in the imputation models for all covariates, and vice versa.

We ran 80 cycles of multiple imputation because some variables were highly correlated, and we performed graphical checks to ensure that the distributions of the imputed variables had converged. Conservatively, we generated 60 imputed datasets, which exceeds the maximum percentage of missing data for any of the variables in the imputed dataset.[4] We compared the means and standard deviations of the imputed values to the analogous statistics for observed values, to check the appropriateness of imputations.[5]

Iterative Proportional Weighting

We followed standard recommendations for implementing IPW.[8] We identified variables potentially predictive of missingness in the outcome variable (Late Adolescent Waist-to-Height)[6] as well as variables strongly associated with values of the outcome variable. After transforming highly skewed continuous variables, we used Lasso (and logistic regression) to select the variables most predictive of missingness (or most predictive of the outcome variable), to avoid an overfitted model with small fitted probabilities. After fitting a logistic regression model for outcome missingness as a function of the variables selected via Lasso, we used Hinkley's method to check model fit, and added in interactions necessary to improve model fit. The best fitting model for outcome missingness in girls included: maternal Neuroticism, Achievement, and Nurturance during participant infancy; maternal education and paternal education at participant Age 3.5 and family income at participant Age 4.5; participant BMI z-scores, internalizing scores and squared internalizing scores for Age 9; maternal body mass index and mother-and child-reported TV, video, and computer use for the participant at Age 13; and participant alcohol use for Age 17. The best fitting model for outcome missingness in boys included: maternal Achievement and paternal Neuroticism during participant infancy; participant BMI z-scores, psychostimulant use, and teacher-reported Attention scores at Age 9; and mother- and child-reported video use, mother-reported sports and exercise, and mother-reported weekend exercise squared for the participant at Age 13. Of note, although these multivariable outcome missingness models did include certain predictors (e.g., Age 9 BMI z-scores) from the analyses of interest, univariate comparisons between participants with versus without Age 18 data revealed no significant differences in any predictors used in the analyses of interest, for either sex. Finally, we examined the predicted probabilities from the best-fitting model for each sex to check that the fitted probabilities were larger for those individuals with missing outcome compared to those individuals with observed outcomes, and to check whether any of the fitted probabilities were especially small, which would result in large weights in IPW.

Analyses

We used the IPW weights (i.e., the inverse of the fitted probabilities from the best fitting missingness models) to perform weighted linear regression. As recommended for IPW, standard errors were calculated using a sandwich estimator (specifically, the Huber-White sandwich estimator).

For all parameters of interest, Rubin's rules[9] were used to combine estimates and errors from the weighted linear regression models.

Results

The correlation between CSR and Childhood BMI is -0.15 (95% CI = [-0.29, -0.01]), p = 0.04) in girls and 0.04 (95% CI = [-0.10, 0.19]), p = 0.55) in boys, with a difference in correlations for girls vs. boys of -0.19 (95% CI = [-0.39, 0.01]), p = 0.06).

Results of regression analyses for the Full Cohort sample are presented in Table S2, which corresponds to analysis results for the Complete Case sample presented in Table 2.

The pattern of results for the Full Cohort sample replicates findings from the Complete Case analysis. However, the relationship between CSR and Late Adolescent Waist-to-Height is weaker for girls and stronger for boys in the Full Cohort sample, as compared to the Complete Case sample. The stronger relationship for boys in the Full Cohort sample is not surprising given that IPW resulted in the upweighting of data for several boys with lower CSR who experienced large gains in adiposity from childhood to late adolescence; notably, all of these boys were on psychostimulant medication at Age 10, but not Age 18. In contrast, the weaker relationship for girls in the Full Cohort sample was not surprising given that IPW resulted in the upweighting of data for several girls with high levels of alcohol use at Age 17; these girls had low CSR, but they did not experience large gains in adiposity from childhood and late adolescence, although their adiposity was high at both timepoints.

Supplemental References

1. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *Br Med J*. 2009;338:b2393.

2. Seaman SR, White I. Inverse probability weighting with missing predictors of treatment assignment or missingness. *Commun Stat - Theory Methods*. 2014;43(16):3499-3515.

3. Seaman SR, White IR, Copas AJ, Li L. Combining multiple imputation and inverse-probability weighting. *Biometrics*. 2012;68(1):129-137.

4. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-399.

5. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? *Int J Methods Psychiatr Res*. 2011;20(1):40-49.

6. Stuart EA, Azur M, Frangakis C, Leaf P. Multiple imputation with large data sets: A case study of the Children’s Mental Health Initiative. *Am J Epidemiol*. 2009;169(9):1133-1139.

7. Vink G, Frank LE, Pannekoek J, van Buuren S. Predictive mean matching imputation of semicontinuous variables: PMM imputation of semicontinuous variables. *Stat Neerlandica*. 2014;68(1):61-90.

8. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22(3):278-295.

9. Rubin DB. *Multiple Imputation for Nonresponse in Surveys (Wiley Series in Probability and Statistics)*. Hoboken, NJ: John Wiley & Sons; 1987.

Table S1. Percentage Missing for Each Variable in the Full Cohort Sample (*n* = 435)

	Girls (*n* = 221)	**Boys (*n* = 214)**
Late Adolescent Waist-to-Height	43.4%	50.0%
Childhood Self-Regulation	2.7%	1.4%
Race/Ethnicity	0.0%	0.0%
Childhood SES	8.6%	7.0%
Childhood Verbal Ability	18.1%	17.8%
Childhood BMIz	10.0%	9.3%

Abbreviations: BMI = Body Mass Index; SES = Socioeconomic Status

Table S2. Regression Results for Late Adolescent Waist-to-Height[a] in the Full Cohort Sample

Predictors[b]	Girls[c]			Boys[d]			Difference for Girls vs. Boys[e]		
	Est.	**95% CI**	***p*-value**	**Est.**	**95% CI**	***p*-value**	**Est.**	**95% CI**	***p*-value**
	Model 1: Adjusted for Race/Ethnicity								
Intercept	-2.207	[-2.772, -1.641]	<0.001	-1.911	[-2.896, -0.926]	<0.001	-0.295	[-1.430, 0.839]	0.62
Childhood Self-Regulation (CSR)	-0.189	[-0.358, -0.021]	0.03	-0.124	[-0.330, 0.083]	0.25	-0.066	[-0.334, 0.202]	0.63
Race/Ethnicity	-0.240	[-0.828, 0.348]	0.43	-0.505	[-1.495, 0.485]	0.32	0.265	[-0.884, 1.415]	0.66
Childhood SES	--	-- --	--	--	-- --	--	--	-- --	--
Childhood Verbal Ability	--	-- --	--	--	-- --	--	--	-- --	--
Childhood BMIz	--	-- --	--	--	-- --	--	--	-- --	--
ΔR^2 for CSR	--	-- --	6.1%[f]	--	-- --	2.6%[g]	--	-- --	0.2%[h]
	Model 2: Adjusted for Race/Ethnicity, Childhood SES, and Childhood Verbal Ability								
Intercept	-2.935	[-4.260, -1.610]	<0.001	-1.765	[-3.630, 0.100]	0.07	-1.170	[-3.440, 1.100]	0.32
Childhood Self-Regulation (CSR)	-0.158	[-0.317, 0.000]	0.06	-0.111	[-0.321, 0.099]	0.31	-0.047	[-0.313, 0.219]	0.73
Race/Ethnicity	-0.172	[-0.783, 0.439]	0.59	-0.482	[-1.483, 0.518]	0.35	0.310	[-0.860, 1.480]	0.61
Childhood SES	-0.195	[-0.389, 0.000]	0.06	-0.048	[-0.300, 0.205]	0.72	-0.147	[-0.469, 0.174]	0.37
Childhood Verbal Ability	0.006	[-0.005, 0.016]	0.28	-0.002	[-0.019, 0.015]	0.86	0.007	[-0.012, 0.027]	0.47
Childhood BMIz	--	-- --	--	--	-- --	--	--	-- --	--
ΔR^2 for CSR	--	-- --	3.8%[f]	--	-- --	1.8%[g]	--	-- --	0.2%[h]
	Model 3: Adjusted for Race/Ethnicity, Childhood SES, Childhood Verbal Ability, and Childhood BMIz								
Intercept	-2.489	[-3.499, -1.479]	<0.001	-2.267	[-3.409, -1.126]	<0.001	-0.222	[-1.716, 1.273]	0.78
Childhood Self-Regulation (CSR)	-0.096	[-0.214, 0.021]	0.11	-0.173	[-0.317, -0.030]	0.02	0.077	[-0.111, 0.265]	0.43
Race/Ethnicity	0.010	[-0.568, 0.588]	0.98	-0.204	[-0.589, 0.181]	0.31	0.214	[-0.479, 0.907]	0.55
Childhood SES	-0.155	[-0.286, -0.024]	0.03	0.034	[-0.145, 0.213]	0.72	-0.189	[-0.413, 0.035]	0.10
Childhood Verbal Ability	-0.001	[-0.009, 0.006]	0.76	-0.002	[-0.013, 0.009]	0.73	0.001	[-0.012, 0.014]	0.91
Childhood BMIz	0.454	[0.322, 0.586]	<0.001	0.548	[0.431, 0.665]	<0.001	-0.094	[-0.271, 0.082]	0.30
ΔR^2 for CSR	--	-- --	1.4%[f]	--	-- --	4.2%[g]	--	-- --	0.3%[h]
	Model 4: Adjusted for Race/Ethnicity, Childhood SES, Childhood Verbal Ability, Childhood BMIz, and CSR x Childhood BMIz								

Intercept	-2.433	[-3.365, -1.502]	<0.001	-2.117	[-3.340, -0.894]	<0.01	-0.316	[-1.828, 1.195]	0.69
Childhood Self-Regulation	-0.067	[-0.172, 0.039]	0.22	-0.142	[-0.289, 0.004]	0.06	0.076	[-0.107, 0.258]	0.42
Race/Ethnicity	0.070	[-0.410, 0.549]	0.78	-0.237	[-0.705, 0.231]	0.33	0.307	[-0.362, 0.975]	0.37
Childhood SES	-0.161	[-0.289, -0.032]	0.02	0.042	[-0.142, 0.226]	0.66	-0.203	[-0.430, 0.025]	0.09
Childhood Verbal Ability	-0.002	[-0.010, 0.005]	0.54	-0.003	[-0.015, 0.008]	0.61	0.001	[-0.013, 0.014]	0.92
Childhood BMIz	0.432	[0.305, 0.559]	<0.001	0.534	[0.417, 0.650]	<0.001	-0.102	[-0.274, 0.071]	0.25
CSR x Childhood BMIz	-0.122	[-0.264, 0.021]	0.10	-0.061	[-0.171, 0.050]	0.29	-0.061	[-0.241, 0.119]	0.51

Abbreviations: BMIz = BMI z-Scores; CI = Confidence Interval; CSR = Childhood Self-Regulation; Est. = Estimate; SES = Socioeconomic Status

[a] Late Adolescent Waist-to-Height was transformed using the Box-Cox transformation with $\lambda = -2.5$, i.e., (Waist-to-Height$^{(-2.5)}$ - 1)/(-2.5).

[b] All models include Sex as a predictor, as well as interactions between Sex and all other predictors.

[c] Coefficients for girls (e.g., the coefficient for CSR in girls) correspond to the main effect of the predictor (e.g., CSR) when the Sex variable is coded as 0 for Girls and 1 for Boys.

[d] Coefficients for boys (e.g., the coefficient for CSR in boys) correspond to the main effect of the predictor (e.g., CSR) when the Sex variable is coded as 1 for Girls and 0 for Boys.

[e] Coefficients for differences in girls vs. boys (e.g., the coefficient for the difference in CSR for girls minus boys) correspond to the interaction between the predictor (e.g., CSR) and Sex when the Sex variable is coded as 1 for Girls and 0 for Boys.

[f] ΔR^2 for CSR = R^2 for *girls only* from model *including CSR* (with all coefficients sex-specific) –
R^2 for *girls only* from model *excluding CSR* (with all coefficients sex-specific)

[g] ΔR^2 for CSR = R^2 for *boys only* from model *including CSR* (with all coefficients sex-specific) –
R^2 for *boys only* from model *excluding CSR* (with all coefficients sex-specific)

[h] ΔR^2 for CSR = R^2 for *girls and boys* from model *including CSR* (with all coefficients sex-specific) –
R^2 for *girls and boys* from model *including CSR* (with all coefficients except CSR sex-specific)