

**Does the Type and Timing of Educational Attainment Influence Physical Health? A Novel Application of
Sequence Analysis**

Anusha M. Vable, Catherine dP Duarte, Alison K. Cohen, M. Maria Glymour, Robert K. Ream, and Irene H. Yen

Table of Contents: Web Tables and Figures

Web Table 1. Data Cleaning Rules 2

Web Table 1. How individuals excluded from analysis differed from those included 3

Web Figure 1. Modal plot of all 24 educational trajectories recommended using the Halpin approach to
calculate costs and the Duda-Hart cluster stopping rule 4

Web Table 3. Linear regression coefficients for 10 education trajectory clusters predicting physical health
(ref = HS only) 5

Web Figure 2. Linear regression results for 24 education trajectory clusters predicting physical health
(ref = HS only) 6

Web Figure 3. Linear regression results using dynamic hamming to calculate costs to create
educational trajectories (ref = HS only) 7

Web Table 4. Educational trajectories and physical health additionally adjusted for childhood health proxy
(ref = HS only) 8

Web Figure 4. Educational trajectories and physical health additionally adjusted for poverty and AFQT
(ref = HS only) 9

Web Figure 5. Educational trajectories and physical health by cSES, with cSES operationalized by father’s
education (ref = HS only) 10

Web Table 5. Educational trajectories and physical health after multiple imputation (ref = HS only) 11

Web Table 1. Data cleaning rules

Transforming the repeated interviews into analyzable sequences of educational status at each year of age 14-48 entailed a number of data cleaning decisions. To facilitate replication of the individual educational trajectories, these decisions are detailed here:

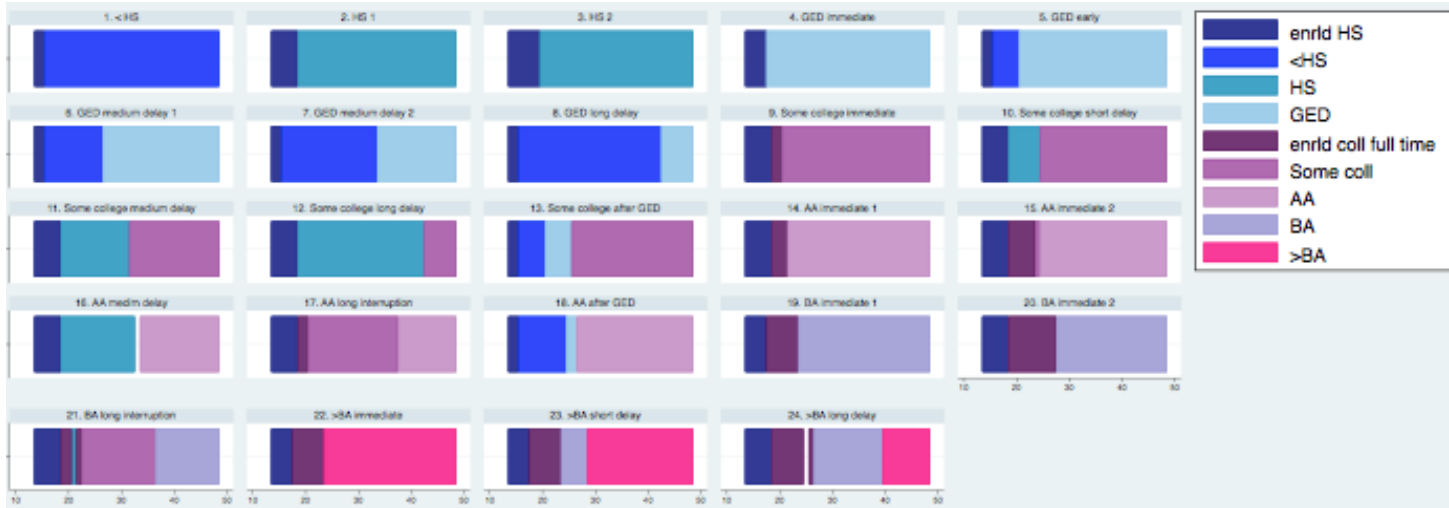
Rule 1: When two consecutive interviews occurred when the respondent was the same age (e.g. there were 2 interviews in a 12-month period, 0.6% of interviews), we assigned a one-year age increment compared to the prior interview age.
Rule 2: When plausible, assume education was continuous, such that respondents finished one year of schooling at each age
Rule 3: if respondents were older than 18 and had completed HS at study enrollment, assume graduated high school at age 18, and filled in data prospectively and retrospectively accordingly
Rule 4: if first observation in 1979 after age = 18 and highest grade completed in 1979 = 12, then assume graduated HS at age 18
Rule 5: if dropped out of HS before 1979: a. assume dropped out of HS at age 16, and highest grade completed = 10 based on CSL data for 1970s b. assume received GED at age 16, most states removed age requirements for GEDs in the 1970s
Rule 6: if < HS in 1979, then a. assume dropped out of HS at age 16 because most states required people to be in school until age 16 in the 70s, which corresponds to 10th grade, unless: b. they had less than 10 years of education in 1979; if so, then use that level of education
Rule 7: if respondent says they have a HS diploma, but their highest grade completed < 12, we assume they have a GED
Rule 8: if previous 3 years were < HS & highest grade completed increases by 1 compared to the previous year, and age > 20, assume working towards GED and not enrolled in high school or college
Rule 9: if respondent says they have a HS diploma, but they had < 12 years the previous 3 years, we assume they have a GED
Rule 10: The year you graduate, you're still considered "enrolled"; this will avoid people having an HS degree for 1 year before college
Rule 11: If highest grade completed is 10, 10, 11, 11 across 4 years (or similar) they may be repeating grades and therefore enrolled the whole time
Rule 12: if highest grade completed < 6th grade at age 16, assume not enrolled and <HS from ages 14 (14 is the youngest age they are in the dataset)
Rule 13: if they have 12 years of schooling and a GED, they have a GED when they have 12 years of schooling (also: they have a GED the year they say they have a GED)
Rule 14: assume enrolled in college full time unless otherwise stated
Rule 15: people are enrolled in high school, not college if they have <12 years of school or the year they got a GED
Rule 16: assume people have an AA when they say they have an AA, even if they have < 14 years of schooling
Rule 17: if they 14 years of schooling for > 3 years, and don't get any more, we assume they have an AA, and they got it the year first year they had 14 years of schooling

Web Table 2. How individuals excluded from analysis differed from those included

VARIABLES	Included (N = 7,501)		Excluded (N = 5,185)	
	mean	sd	mean	sd
Educational attainment (yrs)	12.88	2.739	13.34	2.544
Birth year	1,960	2.286	1,961	2.201
Southern birth	0.323	0.468	0.379	0.485
Mother's education (yrs)	10.24	3.99	10.13	4.1
Father's education (yrs)	9.551	5.169	9.277	5.353
Female	0.469	0.499	0.513	0.5
Immigrant	0.0687	0.253	0.0689	0.253
Rural residence at age 14	0.221	0.415	0.205	0.404
Non-Hispanic White	0.679	0.467	0.494	0.5
Non-Hispanic Black	0.152	0.359	0.305	0.46
Latino	0.104	0.306	0.155	0.362
Other Race	0.0476	0.213	0.0379	0.191
Mother work for pay at age 14	0.485	0.5	0.526	0.499
Father work for pay at age 14	0.735	0.441	0.746	0.435
AFQT percentile	43.4	27.28	39.16	28.07
Physical health component summary score	47.67	10.94	49.14	10.25
Mother's education missing	0.0615	0.24	0.0652	0.247
Father's education missing	0.131	0.337	0.15	0.357
Mother work for pay data missing	0.0268	0.162	0.0231	0.15
Father work for pay data missing	0.192	0.394	0.191	0.393
Mother's work unskilled	0.373	0.484	0.4	0.49
Mother's work skilled / unskilled missing	0.0287	0.167	0.0315	0.175
Father's work unskilled	0.492	0.5	0.504	0.5
Poverty in 1979 missing	0.202	0.402	0.192	0.394
Poverty in 1979	0.218	0.413	0.187	0.39
AFTQ percentile missing	0.0939	0.292	0.0428	0.202

Those excluded from analysis completed more years of schooling, a higher proportion were female, and minorities, and a lower proportion were in poverty in 1979. From the full sample of 12,686 NLSY respondents enrolled at baseline, 4,774 were loss to follow up, so we were not able to create educational trajectories for them. The remaining 411 respondents were missing data on the outcome or the covariates. These 411 individuals were included in the multiple imputation results presented in Web Table 4; results did not meaningfully vary after imputations.

Web Figure 1. Modal plot of all 24 educational trajectories recommended using the Halpin approach to calculate costs and the Duda-Hart cluster stopping rule



The x-axis of each plot is age. The modal plot shows the most common state at each educational trajectory. The educational trajectories are labeled as follows:

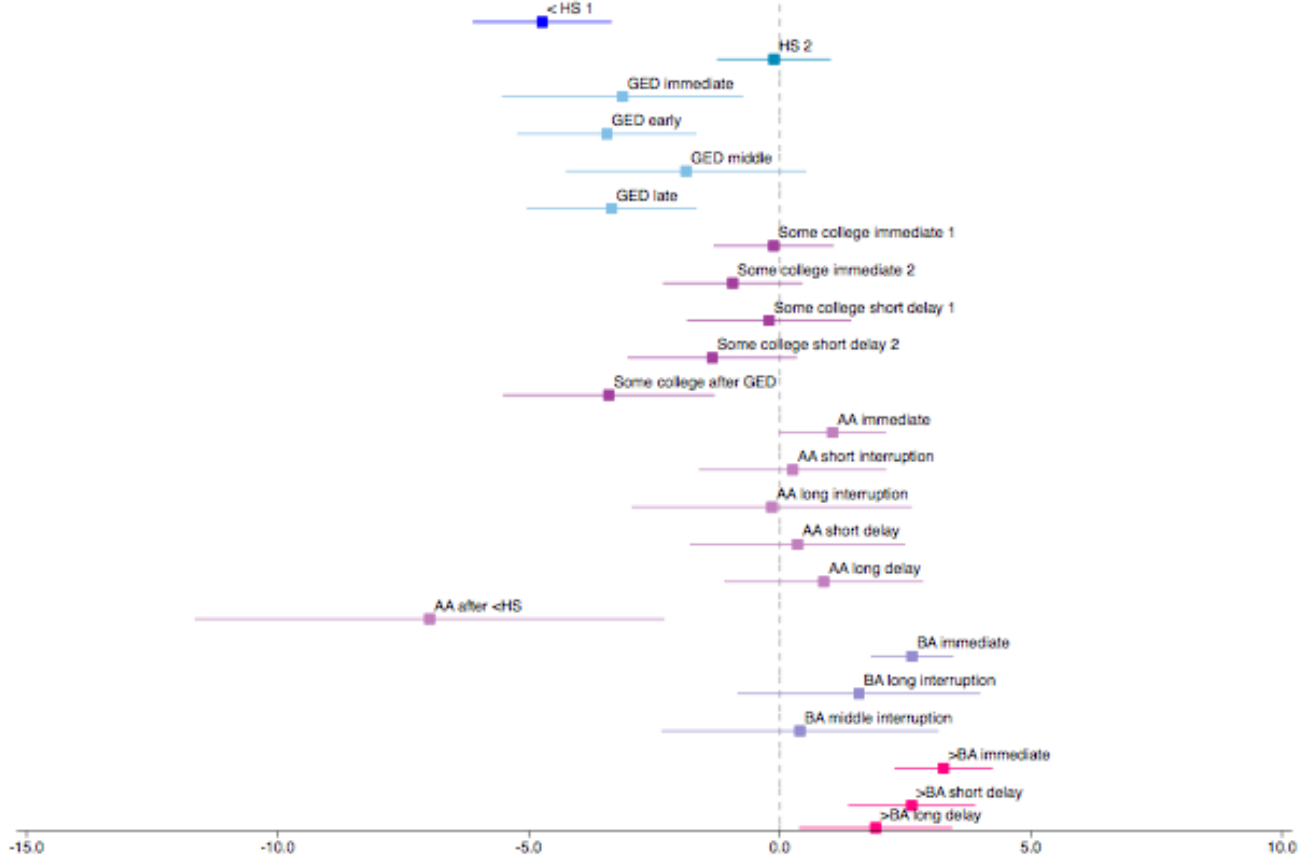
[1] < HS; [2] HS 1; [3] HS 2; [4] GED immediate; [5] GED early; [6] GED medium delay 1; [7] GED medium delay 2; [8] GED long delay; [9] Some college; immediate; [10] Some college short delay; [11] Some college medium delay; [12] Some college long delay; [13] Some college after GED; [14] AA immediate 1; [15] AA immediate 2; [16] AA medium delay; [17] AA long interruption; [18] AA after GED; [19] BA immediate 1; [20] BA immediate 2; [21] BA long interruption; [22] >BA immediate; [23] >BA short delay; [24] >BA long delay.

Web Table 3. Educational trajectories and physical health for 10 education trajectory clusters predicting physical health (ref = HS only)

Educational trajectories	Point estimate
< HS	-4.69
	(-6.10, -3.28)
GED	-3.07
	(-4.07, -2.07)
Some college	-0.58
	(-1.54, 0.37)
Some college after GED	-3.37
	(-5.50, -1.24)
AA	0.69
	(-0.17, 1.55)
AA after < HS	-6.94
	(-11.56, -2.31)
BA immediate	2.64
	(1.90, 3.38)
BA interrupted	0.97
	(-0.72, 2.67)
Grad school	2.68
	(1.84, 3.51)

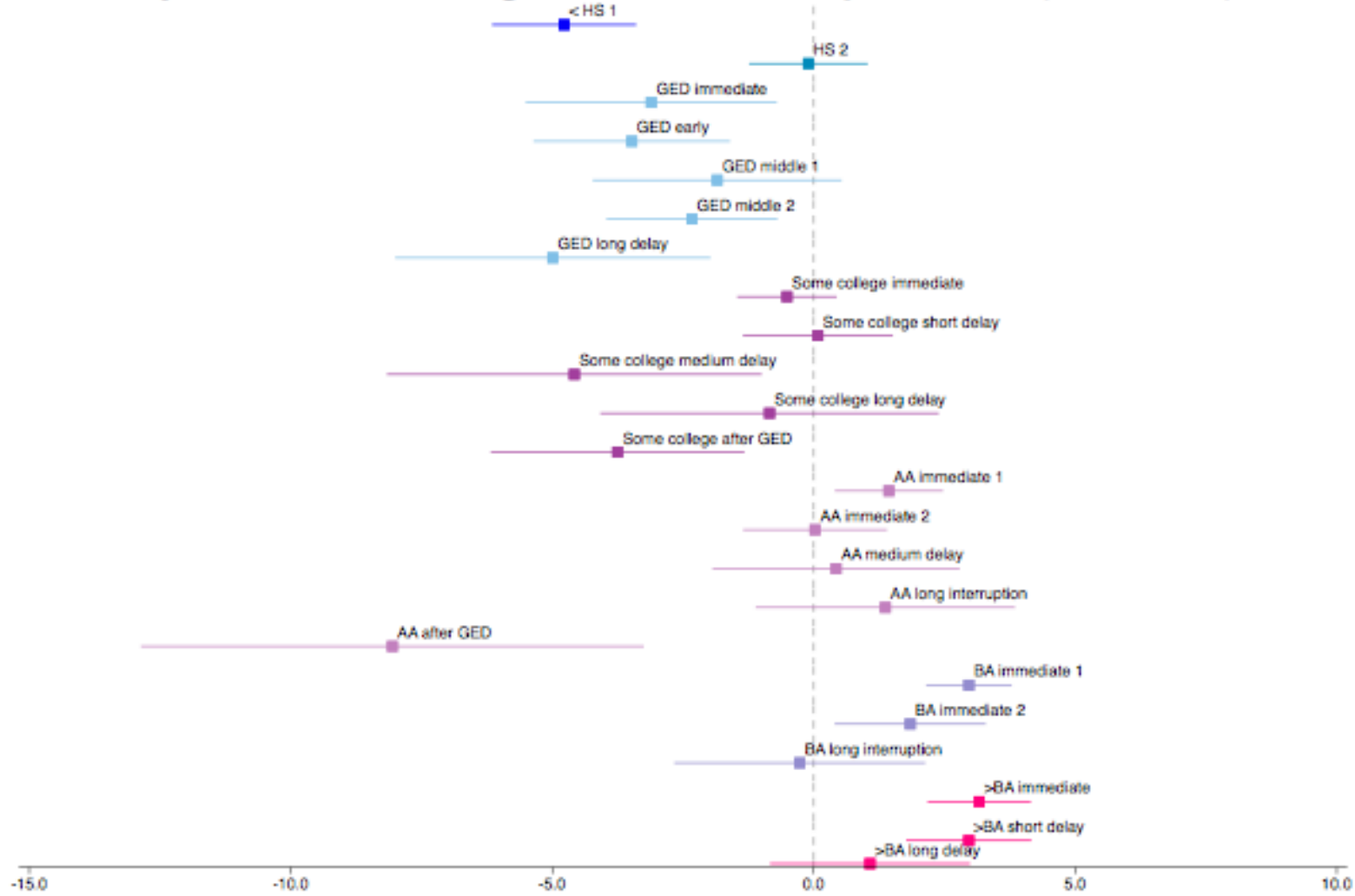
Web Figure 2. Educational trajectories and physical health for 24 education trajectory clusters predicting physical health (ref = HS only)

Physical health by all educational trajectories (ref = HS)



Web Figure 3. Educational trajectories and physical health using dynamic hamming to calculate costs to create educational trajectories (ref = HS only)

Dynamic Hamming educational trajectories (ref = HS)



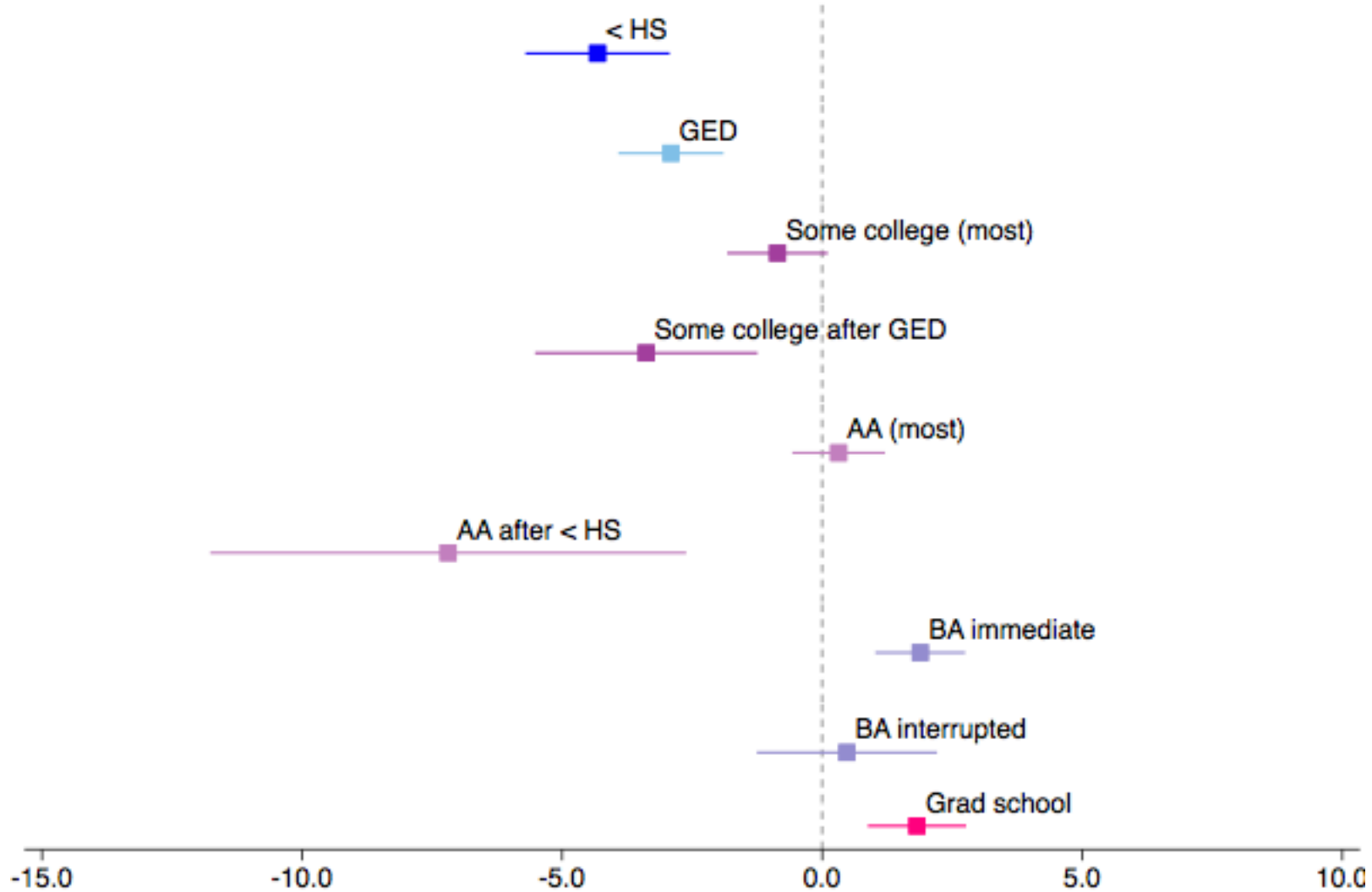
Main results were similar using the dynamic hamming approach to calculate costs

Web Table 3. Educational trajectories and physical health additionally adjusted for childhood health proxy (ref = HS only)

< HS	-4.70
	(-6.11, -3.28)
GED	-3.10
	(-4.10, -2.11)
Some college	-0.61
	(-1.56, 0.35)
Some college after GED	-3.31
	(-5.42, -1.20)
AA	0.65
	(-0.23, 1.52)
AA after < HS	-6.65
	(-11.37, -1.92)
BA immediate	2.61
	(1.87, 3.35)
BA interrupted	1.04
	(-0.67, 2.74)
Grad school	2.64
	(1.81, 3.47)

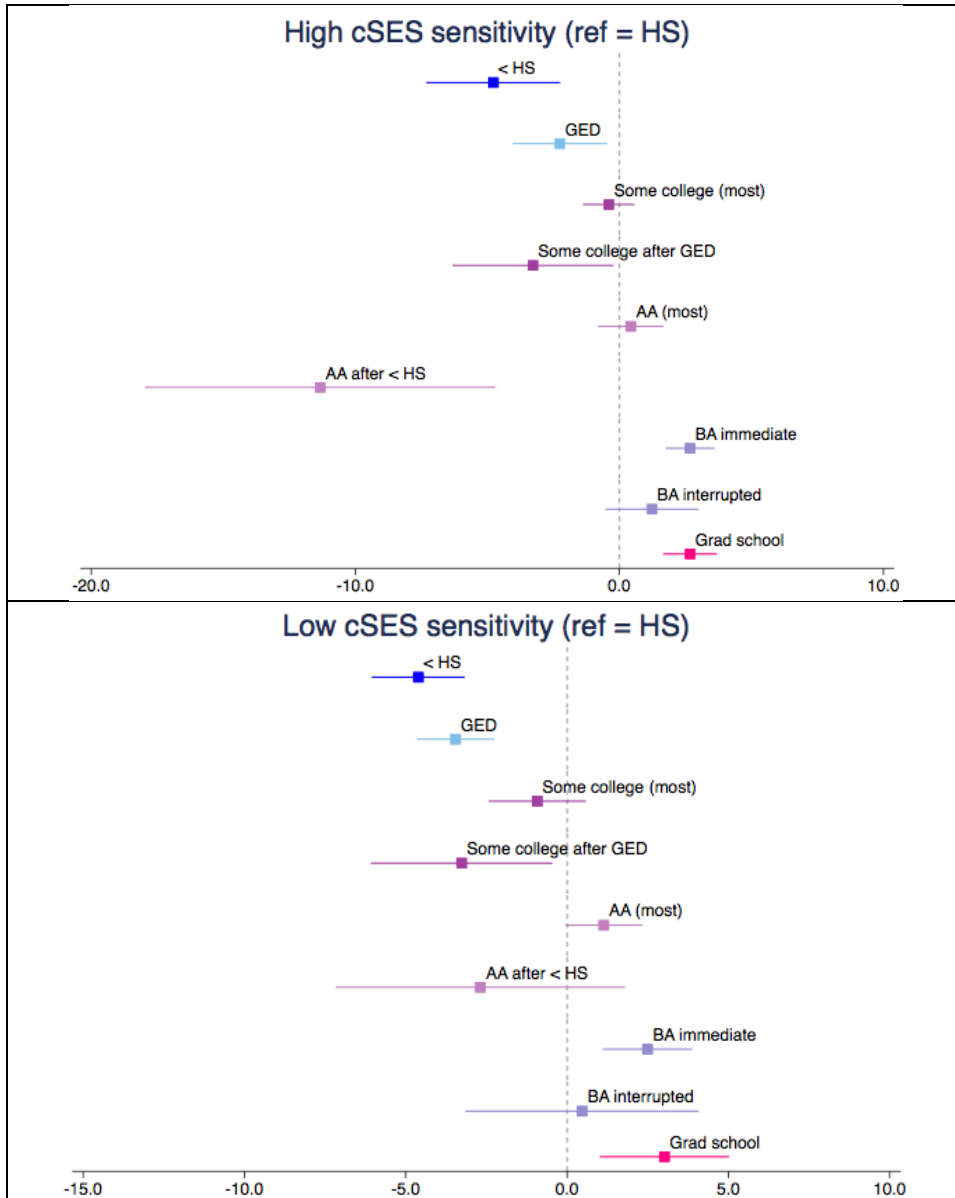
Web Figure 4. Educational trajectories and physical health additionally adjusted for poverty and AFQT (ref = HS only)

Additional adjustment for poverty and AFQT (ref = HS)



Results were similar after additional adjustment for poverty and AFQT.

Web Figure 5. Educational trajectories and physical health by cSES, with cSES operationalized by father's education (ref = HS only)



Results were similar when cSES was operationalized as mother's education (main paper Figure 5) and father's education

Web Table 5. Educational trajectories and physical health after multiple imputation (ref = HS only)

< HS	-4.55
	(-5.93, -3.18)
GED	-2.91
	(-3.90, -1.92)
Some college	-0.53
	(-1.47, 0.41)
Some college after GED	-3.30
	(-5.37, -1.23)
AA	0.60
	(-0.28, 1.47)
AA after < HS	-5.84
	(-10.02, -1.66)
BA immediate	2.57
	(1.81, 3.34)
BA interrupted	1.00
	(-0.69, 2.70)
Grad school	2.61
	(1.73, 3.48)