Science Advances
AAAS

Supplementary Materials for

Concentrated poverty, ambient air pollution, and child cognitive development

Geoffrey T. Wodtke *et al.*

Corresponding author: Geoffrey T. Wodtke, wodtke@uchicago.edu

Sci. Adv. **8**, eadd0285 (2022)
DOI: 10.1126/sciadv.add0285

This PDF file includes:

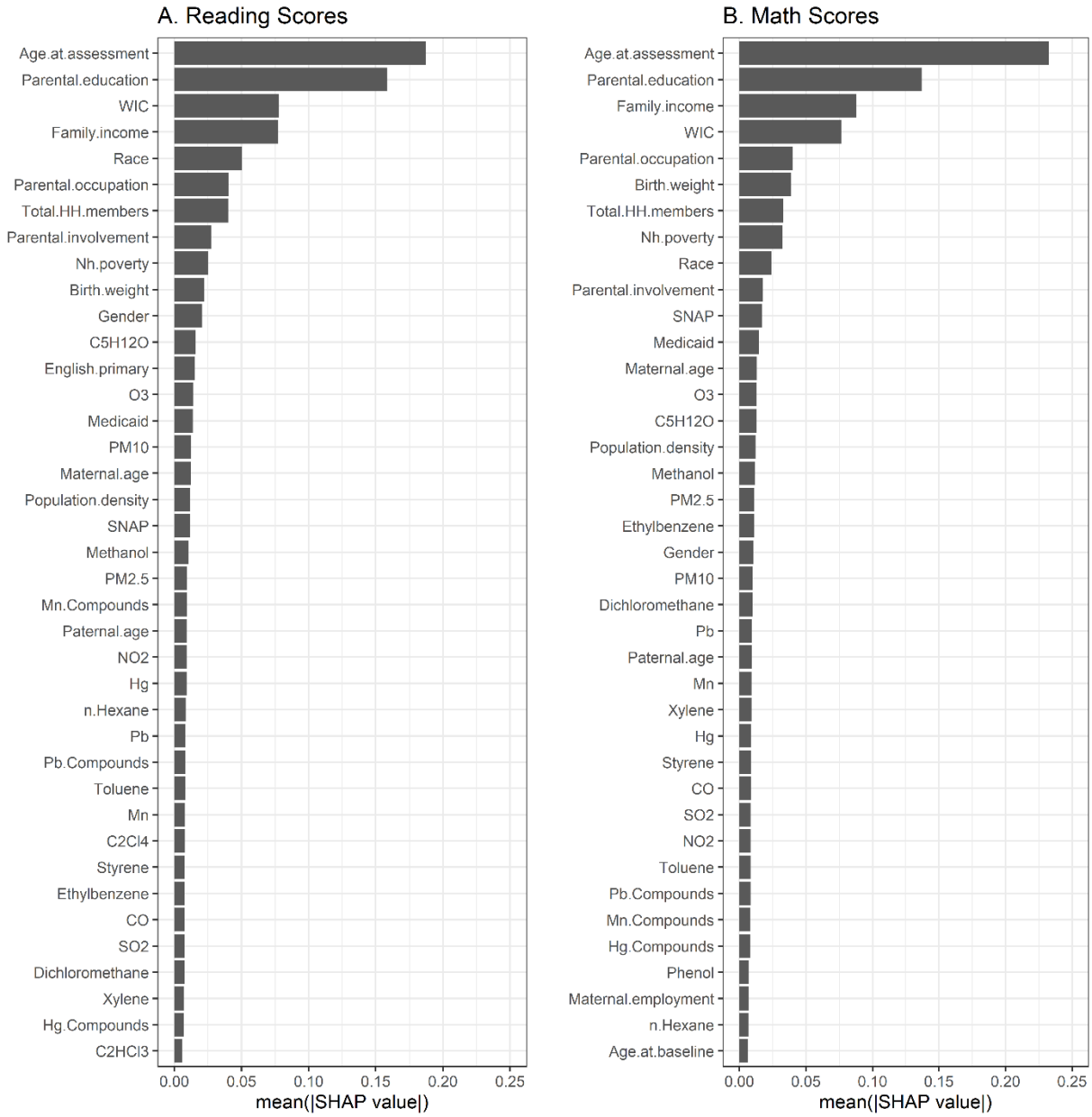

Fig. S1. Predictive Importance of All Covariates in Models for the Conditional Mean of Test Scores given the Confounders, Neighborhood Poverty, and Measured Air Toxics. This figure reports mean absolute SHAP values from random forests with neighborhood poverty, the full set of controls, and the full set of air toxics as predictors. Data sources: (*39-42*).

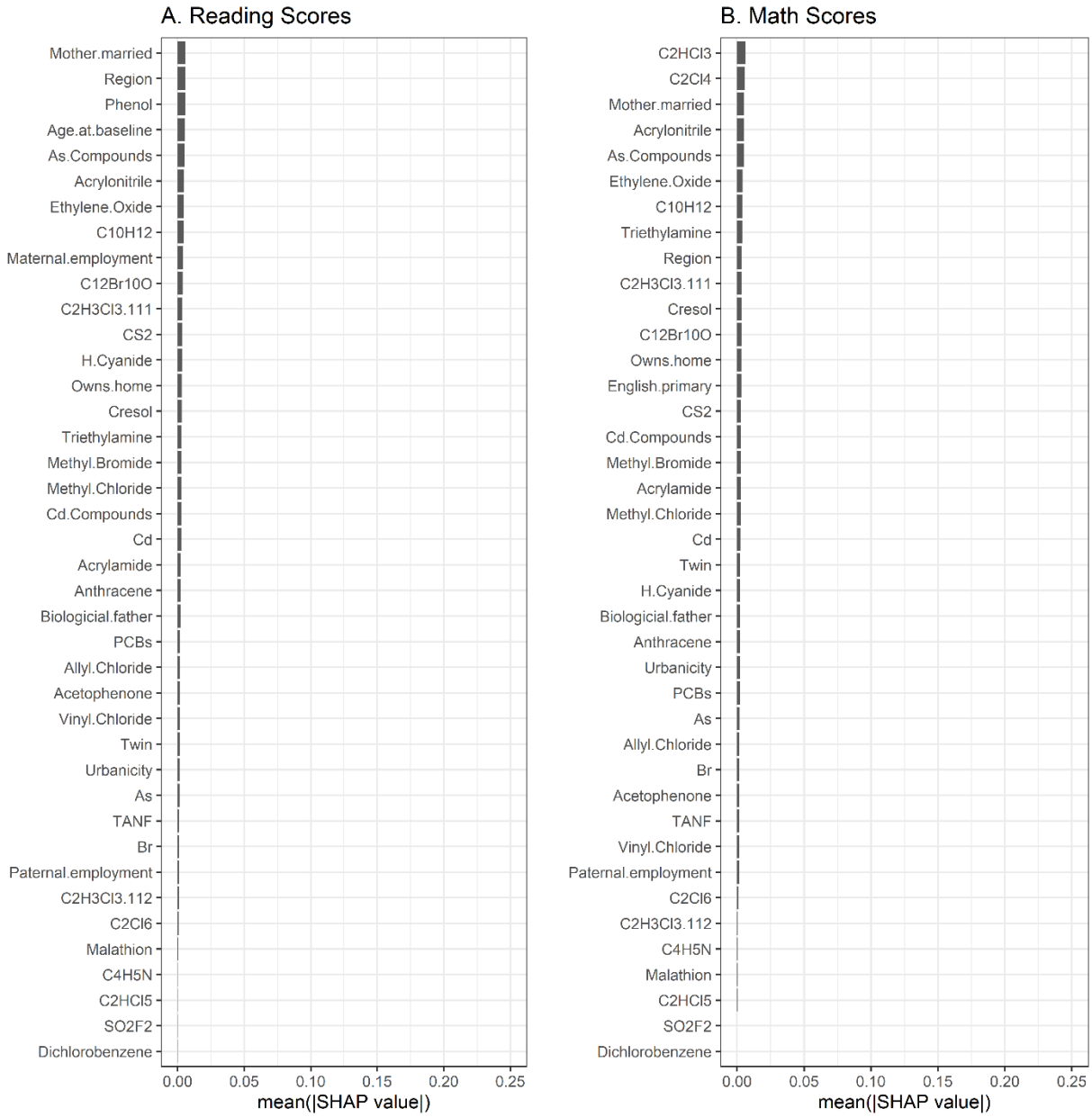

Fig. S1 continued.

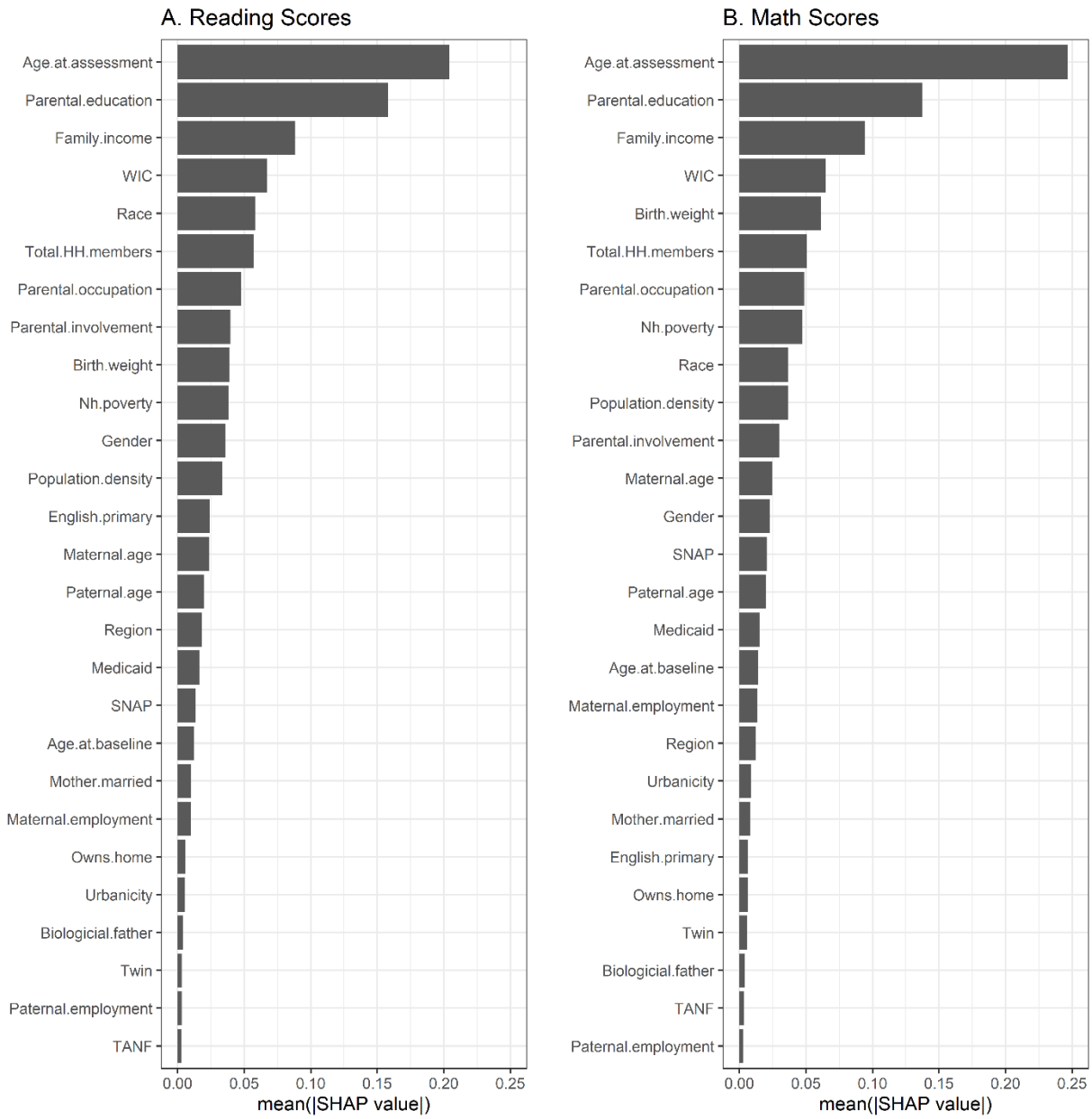

Fig. S2. Predictive Importance of All Covariates in Models for the Conditional Mean of Test Scores given the Confounders and Neighborhood Poverty. This figure reports mean absolute SHAP values computed from random forests with neighborhood poverty and the full set of controls, but not any air toxics, as predictors. Results are combined across 5 imputations. Data sources: (*39-42*).

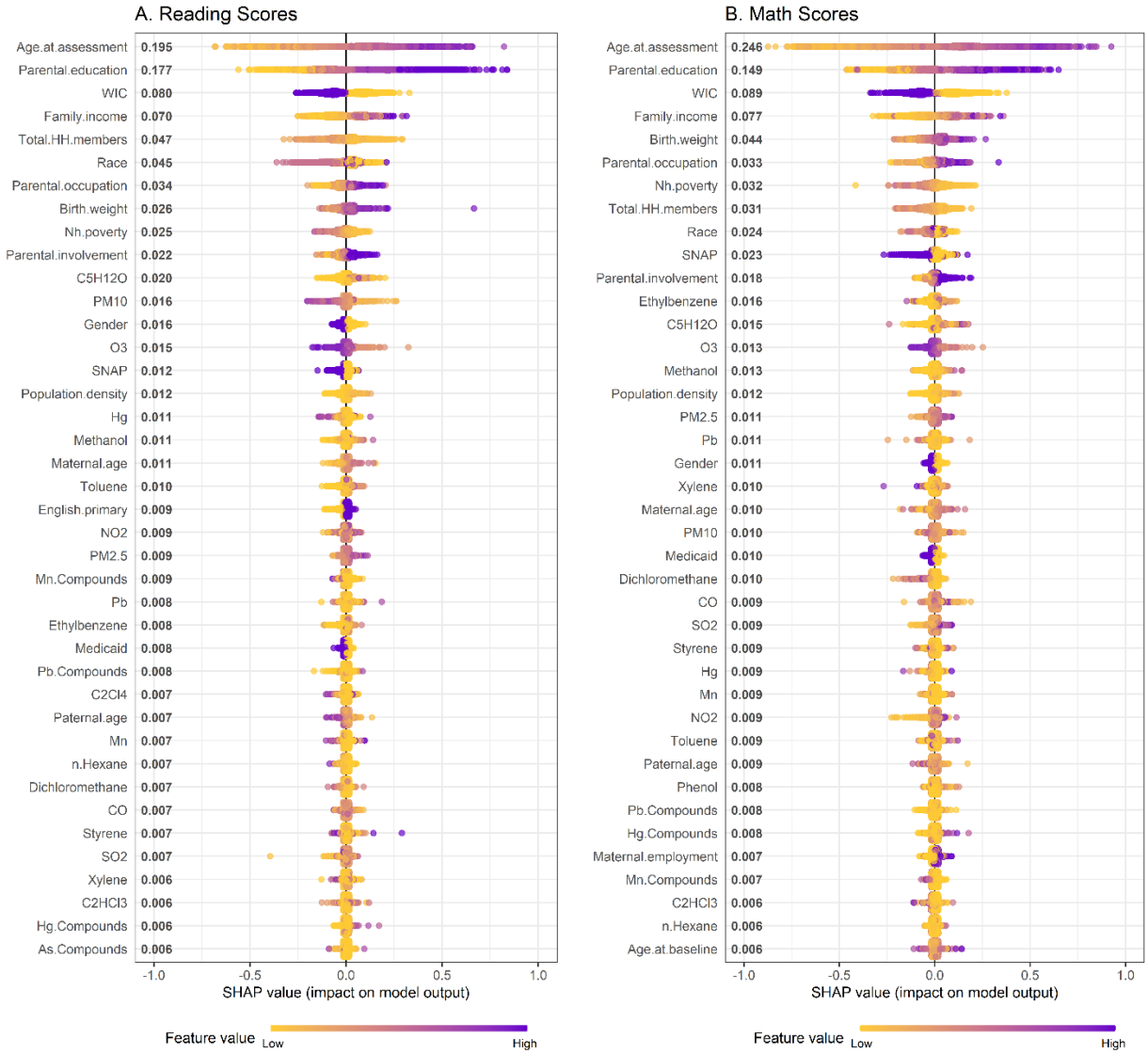

Fig. S3. SHAP Value Summary Plot for Models of the Conditional Mean of Test Scores given the Confounders, Neighborhood Poverty, and Measured Air Toxics. This figure reports individual SHAP values computed from random forests with neighborhood poverty, the full set of controls, and the full set of air toxics as predictors. Results are based on a single imputation. Data sources: (*39-42*).

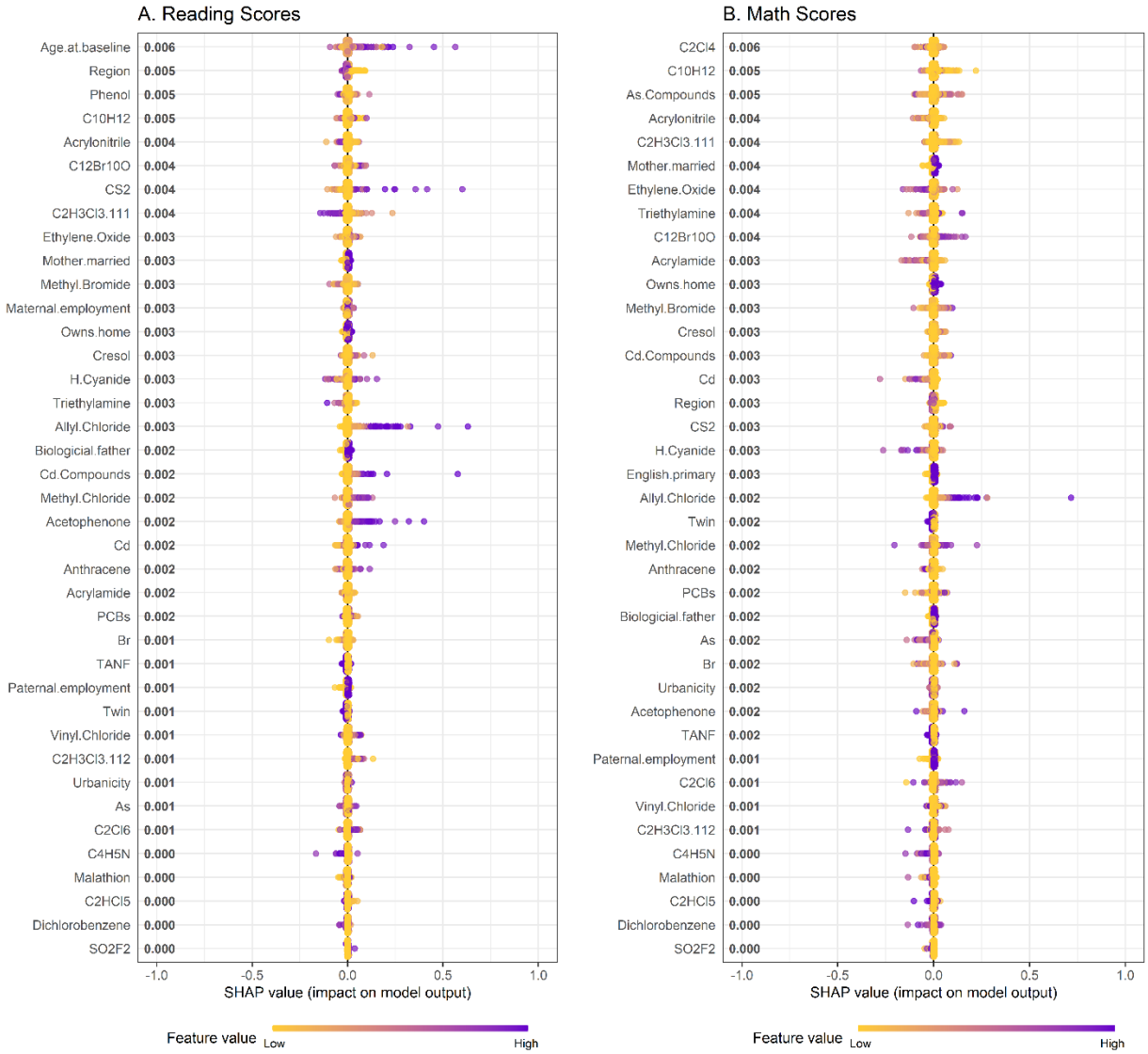

Fig. S3 continued.

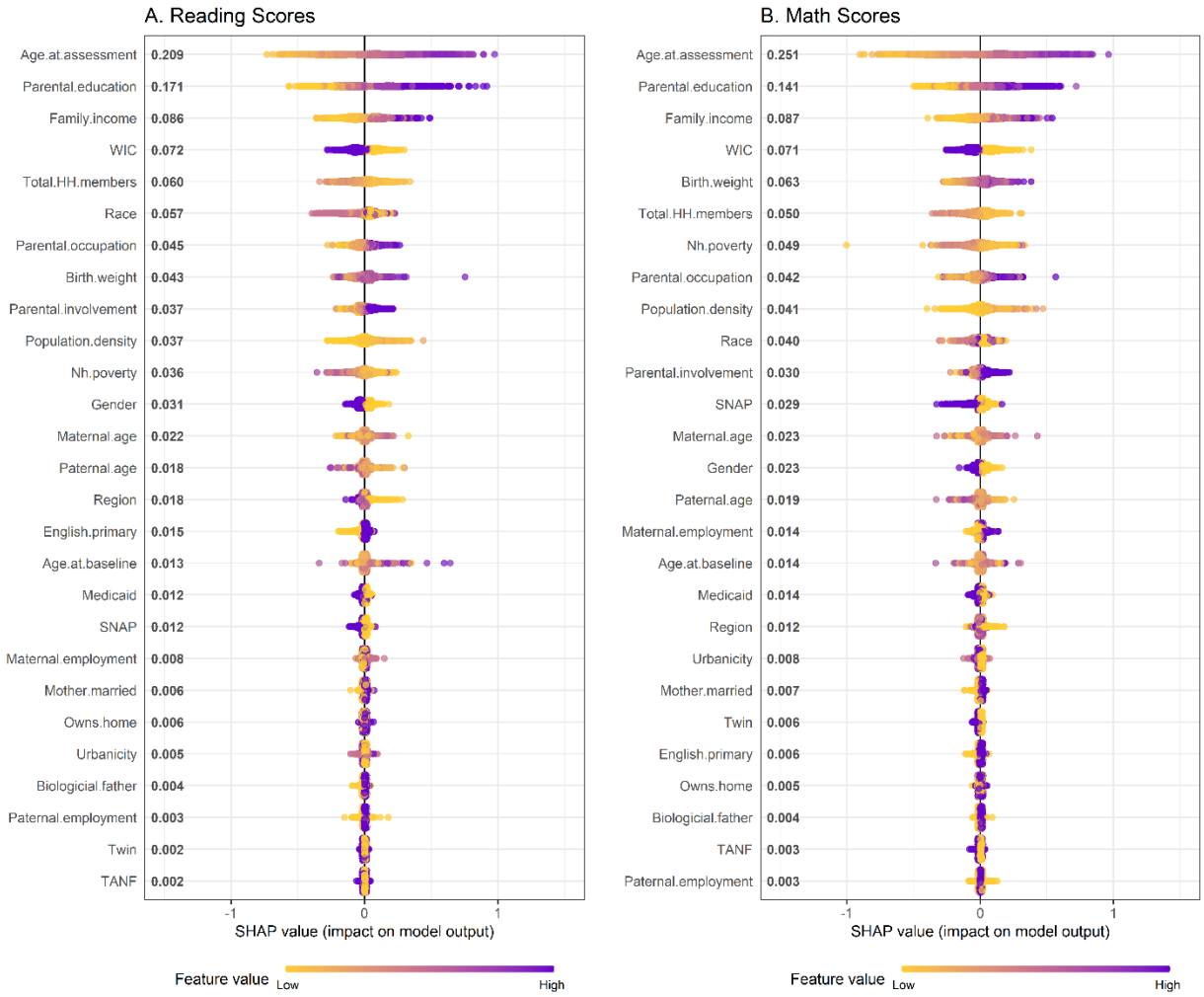

Fig. S4. SHAP Value Summary Plot for Models of the Conditional Mean of Test Scores given the Confounders and Neighborhood Poverty. This figure reports individual SHAP values computed from random forests with neighborhood poverty and the full set of controls, but not any air toxics, as predictors. Results are based on a single imputation. Data sources: (*39-42*).

Table S1. Descriptive Statistics for Measured Air Toxics among the Total ECLS-B Sample and Separately among Quartiles of the Neighborhood Poverty Distribution

Air toxic (units)	Abbreviated label	Total sample		Neighborhood poverty quartile (mean)			
		Mean	SD	Q1	Q2	Q3	Q4
Criteria air pollutants, CACES							
Carbon monoxide (ppm)	CO	0.469	0.180	0.422	0.448	0.464	0.544
Nitrogen dioxide (ppb)	NO2	14.149	7.356	13.899	13.378	12.836	16.486
Ozone (ppb)	O3	49.497	6.283	50.452	49.554	49.218	48.765
PM10 ($\mu g/m^3$)	PM10	23.157	7.306	21.286	22.641	23.073	25.627
PM2.5 ($\mu g/m^3$)	PM2.5	12.713	3.581	12.350	12.335	12.515	13.653
Sulfer dioxide (ppb)	SO2	3.001	1.537	3.045	2.931	2.839	3.191
Industrial neurotoxics, RSEI-GM							
Acetophenone (ng/m^3)	Acetophenone	0.056	0.276	0.061	0.035	0.037	0.092
Acrylamide (ng/m^3)	Acrylamide	0.003	0.012	0.003	0.003	0.002	0.002
Acrylonitrile (ng/m^3)	Acrylonitrile	0.091	0.364	0.054	0.112	0.095	0.104
Allyl chloride (ng/m^3)	Allyl.Chloride	0.003	0.028	0.000	0.000	0.008	0.003
Anthracene (ng/m^3)	Anthracene	0.023	0.121	0.031	0.019	0.012	0.029
Arsenic (ng/m^3)	As	0.000	0.000	0.000	0.000	0.000	0.000
Arsenic compounds (ng/m^3)	As.Compounds	0.012	0.042	0.018	0.014	0.007	0.010
Bromine (ng/m^3)	Br	0.002	0.015	0.002	0.003	0.002	0.002
Bromomethane (ng/m^3)	Methyl.Bromide	0.361	1.190	0.367	0.333	0.258	0.487
Cadmium (ng/m^3)	Cd	0.000	0.001	0.000	0.000	0.000	0.000
Cadmium compounds (ng/m^3)	Cd.Compounds	0.006	0.026	0.005	0.006	0.004	0.010
Carbon disulfide (ng/m^3)	CS2	0.168	0.661	0.118	0.110	0.117	0.325
Chloromethane (ng/m^3)	Methyl.Chloride	0.628	2.328	0.658	0.536	0.517	0.799
Cresol (ng/m^3)	Cresol	0.057	0.229	0.042	0.056	0.065	0.065
Decabromodiphenyl ether (ng/m^3)	C12Br10O	0.007	0.026	0.011	0.006	0.004	0.005
Dichlorobenzene (ng/m^3)	Dichlorobenzene	0.000	0.000	0.000	0.000	0.000	0.000
Dichloromethane (ng/m^3)	Dichloromethane	13.814	32.823	14.578	13.419	12.830	14.432
Dicyclopentadiene (ng/m^3)	C10H12	0.035	0.173	0.026	0.026	0.026	0.064
Ethylbenzene (ng/m^3)	Ethylbenzene	5.372	9.768	5.144	5.339	4.368	6.639
Ethylene oxide (ng/m^3)	Ethylene.Oxide	0.301	0.846	0.346	0.298	0.219	0.341
Hexachloroethane (ng/m^3)	C2Cl6	0.000	0.003	0.000	0.000	0.000	0.000
Hydrogen cyanide (ng/m^3)	H.Cyanide	0.131	0.586	0.139	0.116	0.113	0.157
Lead (ng/m^3)	Pb	0.234	0.578	0.216	0.247	0.173	0.302
Lead compounds (ng/m^3)	Pb.Compounds	0.481	1.163	0.365	0.377	0.417	0.764
Malathion (ng/m^3)	Malathion	0.000	0.000	0.000	0.000	0.000	0.000
Manganese (ng/m^3)	Mn	0.489	1.642	0.364	0.397	0.524	0.671
Manganese compounds (ng/m^3)	Mn.Compounds	0.523	1.595	0.434	0.455	0.475	0.727
Mercury (ng/m^3)	Hg	0.006	0.018	0.004	0.004	0.006	0.011
Mercury compounds (ng/m^3)	Hg.Compounds	0.011	0.023	0.011	0.010	0.008	0.015
Methacrylonitrile (ng/m^3)	C4H5N	0.000	0.000	0.000	0.000	0.000	0.000
Methanol (ng/m^3)	Methanol	56.794	113.993	32.108	50.343	61.982	82.762
Methyl tert-butyl ether (ng/m^3)	C5H12O	7.245	17.029	5.444	5.908	5.615	12.014

Table S1 continued.

Air toxic (units)	Abbreviated label	Total sample		Neighborhood poverty quartile (mean)			
		Mean	SD	Q1	Q2	Q3	Q4
Pentachloroethane (ng/m^3)	C2HCl5	0.000	0.000	0.000	0.000	0.000	0.000
Phenol (ng/m^3)	Phenol	1.572	4.323	1.224	1.201	1.574	2.289
Polychlorinated biphenyls (ng/m^3)	PCBs	0.000	0.001	0.000	0.000	0.000	0.000
Styrene (ng/m^3)	Styrene	32.899	74.039	23.409	29.466	40.690	38.041
Sulfuryl fluoride (ng/m^3)	SO2F2	0.001	0.025	0.001	0.002	0.000	0.002
Tetrachloroethylene (ng/m^3)	C2Cl4	5.324	16.798	3.634	4.328	4.558	8.780
Toluene (ng/m^3)	Toluene	51.790	87.360	40.394	48.660	49.260	68.849
Trichloroethane 1,1,1 (ng/m^3)	C2H3Cl3.111	0.151	0.719	0.054	0.079	0.113	0.359
Trichloroethane 1,1,2 (ng/m^3)	C2H3Cl3.112	0.000	0.003	0.000	0.000	0.001	0.001
Trichloroethylene (ng/m^3)	C2HCl3	12.455	34.078	15.781	11.387	8.284	14.367
Triethylamine (ng/m^3)	Triethylamine	0.122	0.464	0.142	0.146	0.107	0.092
Vinyl chloride (ng/m^3)	Vinyl.Chloride	0.108	0.871	0.055	0.066	0.119	0.192
Xylene (ng/m^3)	Xylene	42.442	70.673	37.208	43.308	34.348	54.904
n-Hexane (ng/m^3)	n.Hexane	24.310	62.406	20.592	29.800	18.994	27.852

Notes: results are weighted to adjust for the ECLS-B complex sample design and combined across 5 multiply imputed datasets; toxic measures are annual average estimates of the amount of the pollutant in a sample member's zip code tabulation area; ppm = parts per million; ppb = parts per billion; µg/m3 = micrograms per cubic meter; ng/m3 = nanograms per cubic meter. Data sources: (*39-42*)

Table S2. Descriptive Statistics for Outcomes, Exposure, and Covariates

Variable	Freq. to nearest 100	Percent	Mean	SD
Outcomes				
Reading scores (unstandardized)			-0.51	0.74
Math scores (unstandardized)			-0.49	0.78
Exposure				
Neighborhood poverty rate			0.13	0.09
Covariates				
Gender, %				
Male	5500	51.11		
Female	5200	48.89		
Race				
White (non-hispanic)	5700	53.36		
Black (non-hispanic)	1500	13.72		
Hispanic	2700	25.50		
Asian	300	2.80		
Other	500	4.62		
Twin				
Yes	10400	97.06		
No	300	2.94		
Homeownership				
Owns home	5100	47.82		
Does not own home	5600	52.18		
Biological father in household				
Yes	8400	78.73		
No	2300	21.27		
Maternal marital status				
Married	7400	69.30		
Not married	3300	30.70		
Parental education				
<HS degree	1500	13.83		
HS degree or equivalent	2700	25.04		
Technical degree OR some college	3200	29.53		
Bachelor's degree	2100	19.75		
Graduate degree	1300	11.85		

Table S2 continued.

Variable	Freq. to nearest 100	Percent	Mean	SD
Maternal employment				
Working >=35 hours per week	3400	32.22		
Working <35 hours per week	2100	19.76		
Not in labor force	5100	48.02		
Paternal employmnet				
Working	9900	92.36		
Not working	800	7.64		
Primary home language				
English	8700	81.25		
Not English	2000	18.75		
Parent reads to child				
Not at all	1300	12.12		
Once or twice per week	3500	33.16		
Three to six times per week	2400	22.21		
Every day	3500	32.51		
WIC				
Recipient	5600	52.29		
Not a recipient	5100	47.71		
SNAP				
Recipient	2200	20.40		
Not a recipient	8500	79.60		
Medicaid				
Recipient	4100	37.96		
Not a recipient	6600	62.04		
TANF				
Recipient	900	8.23		
Not a recipient	9800	91.77		
Urbanicity				
Urban	7900	73.70		
Suburban	1300	11.84		
Rural	1500	14.46		

Table S2 continued.

Variable	Freq. to nearest 100	Percent	Mean	SD
Census division				
Northeast	1800	16.74		
Midwest	2400	22.33		
South	4000	37.21		
West	2500	23.72		
Birthweight (g)			3311.8	598.8
Child's age at baseline (months)			10.5	1.9
Child's age at assessment (months)			52.4	4.1
Mother's age at baseline (years)			28.3	6.3
Father's age at baseline (years)			31.1	7.1
Family income (usd)			50509.5	46795.5
Parental occupational status			43.4	10.9
Household size (count)			4.3	1.5
ZCTA population density (count/km^3)			3177.2	5163.8

Notes: results are weighted to adjust for the ECLS-B complex sample design and combined across 5 multiply imputed datasets; cell counts are rounded to the nearest 100 in accordance with disclosure risk requirements from the U.S. Institute of Education Sciences; ZCTA = zip code tabulation area. Data sources: (*39-42*).

Table S3. Total, Direct, and Indirect Effects of Neighborhood Context based on a Multidimensional Index of Socioeconomic Disadvantage

Estimands	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval
Reading Test Scores		
Average Total Effect	-0.098	[-0.187, -0.046]
Natural Direct Effect	-0.078	[-0.135, -0.037]
Natural Indirect Effect	-0.019	[-0.067, 0.000]
Math Test Scores		
Average Total Effect	-0.099	[-0.176, -0.037]
Natural Direct Effect	-0.074	[-0.124, -0.033]
Natural Indirect Effect	-0.025	[-0.064, 0.008]

Notes: estimates are reported in standard deviation units and are computed using regression-imputation and random forests; they are weighted and combined across 5 imputations; confidence intervals are based on the [2.5, 97.5] percentiles of a pooled sampling distribution simulated via the repeated half-sample bootstrap with 1000 replications per imputation; the index of neighborhood disadvantage is the first principle component from a PCA of the poverty rate, the unemployment rate, the proportion of adult residents with less than a high school education, the proportion of households that are female-headed, and the proportion of the population that identifies as black in a ZCTA; effect estimates contrast living a neighborhood at the 90^{th} percentile, rather than the 20^{th} percentile, of the neighborhood disadvantage index. Data sources: (*39-42*).

Table S4. Estimates for Total, Direct, and Indirect Effects of Alternative Contrasts in Neighborhood Poverty

Estimands	Reading Test Scores		Math Test Scores	
	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval
10% versus 5% Poverty				
Average Total Effect	-0.015	[-0.046, 0.010]	-0.037	[-0.081, 0.000]
Natural Direct Effect	-0.007	[-0.021, 0.006]	-0.021	[-0.052, 0.002]
Natural Indirect Effect	-0.008	[-0.031, 0.009]	-0.016	[-0.038, 0.002]
15% versus 5% Poverty				
Average Total Effect	-0.036	[-0.082, -0.006]	-0.066	[-0.124, -0.017]
Natural Direct Effect	-0.026	[-0.050, -0.008]	-0.043	[-0.090, -0.013]
Natural Indirect Effect	-0.010	[-0.037, 0.006]	-0.022	[-0.044, 0.001]
20% versus 5% Poverty				
Average Total Effect	-0.066	[-0.125, -0.029]	-0.102	[-0.163, -0.042]
Natural Direct Effect	-0.042	[-0.078, -0.020]	-0.067	[-0.118, -0.026]
Natural Indirect Effect	-0.024	[-0.054, -0.003]	-0.035	[-0.056, -0.010]
30% versus 5% Poverty				
Average Total Effect	-0.085	[-0.161, -0.041]	-0.119	[-0.190, -0.041]
Natural Direct Effect	-0.060	[-0.105, -0.033]	-0.084	[-0.142, -0.032]
Natural Indirect Effect	-0.025	[-0.066, -0.002]	-0.036	[-0.066, -0.002]

Notes: estimates are reported in standard deviation units and are computed using regression-imputation and random forests; they are weighted and combined across 5 imputations; confidence intervals are based on the [2.5, 97.5] percentiles of a pooled sampling distribution simulated via the repeated half-sample bootstrap with 1000 replications per imputation. Data sources: (*39-42*).

Table S5. Estimates for Total, Direct, and Indirect Effects of Neighborhood Poverty by Race

Estimands	Reading Test Scores		Math Test Scores	
	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval
Whites (n ≈ 4400)				
Average Total Effect	-0.060	[-0.134, -0.018]	-0.124	[-0.186, -0.061]
Natural Direct Effect	-0.056	[-0.098, -0.028]	-0.088	[-0.147, -0.042]
Natural Indirect Effect	-0.004	[-0.050, 0.017]	-0.036	[-0.064, -0.002]
Blacks (n ≈ 1700)				
Average Total Effect	-0.075	[-0.153, -0.016]	-0.075	[-0.154, 0.000]
Natural Direct Effect	-0.042	[-0.086, -0.009]	-0.042	[-0.105, 0.005]
Natural Indirect Effect	-0.032	[-0.078, 0.002]	-0.033	[-0.067, 0.003]
Hispanics (n ≈ 2200)				
Average Total Effect	-0.159	[-0.280, -0.070]	-0.123	[-0.206, -0.036]
Natural Direct Effect	-0.060	[-0.125, -0.017]	-0.065	[-0.132, -0.012]
Natural Indirect Effect	-0.099	[-0.171, -0.041]	-0.058	[-0.089, -0.019]
Another Race (n ≈ 2400)				
Average Total Effect	-0.070	[-0.274, -0.055]	-0.117	[-0.203, -0.038]
Natural Direct Effect	-0.074	[-0.124, -0.019]	-0.095	[-0.132, -0.014]
Natural Indirect Effect	0.004	[-0.169, 0.005]	-0.021	[-0.088, 0.001]

Notes: estimates are reported in standard deviation units and are computed using regression-imputation and random forests; they are weighted and combined across 5 imputations; confidence intervals are based on the [2.5, 97.5] percentiles of a pooled sampling distribution simulated via the repeated half-sample bootstrap with 1000 replications per imputation. Data sources: (*39-42*).

Table S6. Estimates for Total, Direct, and Indirect Effects of Neighborhood Poverty by Family Income

Estimands	Reading Test Scores		Math Test Scores	
	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval
> Two-thirds of Median Income (n ≈ 7800)				
Average Total Effect	-0.087	[-0.153, -0.048]	-0.128	[-0.189, -0.067]
Natural Direct Effect	-0.063	[-0.104, -0.035]	-0.087	[-0.143, -0.042]
Natural Indirect Effect	-0.023	[-0.061, -0.006]	-0.041	[-0.065, -0.013]
<= Two-thirds of Median Income (n ≈ 2900)				
Average Total Effect	-0.092	[-0.187, -0.026]	-0.083	[-0.179, 0.017]
Natural Direct Effect	-0.037	[-0.091, -0.005]	-0.048	[-0.125, 0.013]
Natural Indirect Effect	-0.055	[-0.104, -0.010]	-0.035	[-0.076, 0.010]

Notes: estimates are reported in standard deviation units and are computed using regression-imputation and random forests; they are weighted and combined across 5 imputations; confidence intervals are based on the [2.5, 97.5] percentiles of a pooled sampling distribution simulated via the repeated half-sample bootstrap with 1000 replications per imputation. Data sources: (*39-42*).

Table S7. Estimates for Total, Direct, and Indirect Effects of Neighborhood Poverty by Region

Estimands	Reading Test Scores		Math Test Scores	
	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval
Northeast (n ≈ 1600)				
Average Total Effect	-0.080	[-0.148, -0.037]	-0.121	[-0.193, -0.053]
Natural Direct Effect	-0.070	[-0.110, -0.035]	-0.086	[-0.148, -0.038]
Natural Indirect Effect	-0.010	[-0.052, 0.009]	-0.035	[-0.067, -0.004]
Midwest (n ≈ 2500)				
Average Total Effect	-0.058	[-0.119, -0.025]	-0.109	[-0.167, -0.045]
Natural Direct Effect	-0.047	[-0.084, -0.023]	-0.074	[-0.128, -0.030]
Natural Indirect Effect	-0.011	[-0.049, 0.009]	-0.035	[-0.065, -0.002]
South (n ≈ 3800)				
Average Total Effect	-0.096	[-0.169, -0.047]	-0.113	[-0.175, -0.046]
Natural Direct Effect	-0.055	[-0.096, -0.028]	-0.071	[-0.129, -0.026]
Natural Indirect Effect	-0.040	[-0.083, -0.012]	-0.042	[-0.063, -0.012]
West (n ≈ 2700)				
Average Total Effect	-0.110	[-0.174, -0.048]	-0.126	[-0.174, -0.048]
Natural Direct Effect	-0.057	[-0.097, -0.027]	-0.082	[-0.127, -0.027]
Natural Indirect Effect	-0.053	[-0.086, -0.013]	-0.045	[-0.064, -0.012]

Notes: estimates are reported in standard deviation units and are computed using regression-imputation and random forests; they are weighted and combined across 5 imputations; confidence intervals are based on the [2.5, 97.5] percentiles of a pooled sampling distribution simulated via the repeated half-sample bootstrap with 1000 replications per imputation. Data sources: (*39-42*).

Table S8. Estimates for Total, Direct, and Indirect Effects of Neighborhood Poverty by Homeownership Status

Estimands	Reading Test Scores		Math Test Scores	
	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval
Homeowners ($n \approx 4900$)				
Average Total Effect	-0.083	[-0.155, -0.041]	-0.134	[-0.203, -0.068]
Natural Direct Effect	-0.060	[-0.104, -0.031]	-0.088	[-0.150, -0.044]
Natural Indirect Effect	-0.023	[-0.063, -0.001]	-0.047	[-0.076, -0.010]
Renters ($n \approx 5800$)				
Average Total Effect	-0.093	[-0.168, -0.045]	-0.100	[-0.175, -0.021]
Natural Direct Effect	-0.053	[-0.098, -0.023]	-0.066	[-0.129, -0.013]
Natural Indirect Effect	-0.040	[-0.081, -0.014]	-0.034	[-0.060, -0.003]

Notes: estimates are reported in standard deviation units and are computed using regression-imputation and random forests; they are weighted and combined across 5 imputations; confidence intervals are based on the [2.5, 97.5] percentiles of a pooled sampling distribution simulated via the repeated half-sample bootstrap with 1000 replications per imputation. Data sources: (*39-42*).

Table S9. Total, Direct, and Indirect Effects of Neighborhood Poverty Estimated from Alternative Random Forest (RF) algorithms and Super Learners

Estimands	RFs with Default Hyperparameters		RFs with Tuned Sampling Fraction		Super Leaners	
	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval
Reading Test Scores						
Average Total Effect	-0.095	[-0.158, -0.055]	-0.114	[-0.172, -0.070]	-0.108	[-0.161, -0.057]
Natural Direct Effect	-0.062	[-0.101, -0.033]	-0.088	[-0.134, -0.047]	-0.076	[-0.116, -0.042]
Natural Indirect Effect	-0.033	[-0.067, -0.015]	-0.026	[-0.059, -0.012]	-0.032	[-0.058, -0.005]
Math Test Scores						
Average Total Effect	-0.120	[-0.178, -0.056]	-0.115	[-0.175, -0.056]	-0.116	[-0.178, -0.056]
Natural Direct Effect	-0.081	[-0.129, -0.036]	-0.085	[-0.126, -0.041]	-0.085	[-0.126, -0.038]
Natural Indirect Effect	-0.039	[-0.061, -0.014]	-0.029	[-0.058, -0.009]	-0.032	[-0.062, -0.012]

Notes: estimates are reported in standard deviation units and are computed using regression-imputation and random forests; they are weighted and combined across 5 imputations; confidence intervals are based on the [2.5, 97.5] percentiles of a pooled sampling distribution simulated via the repeated half-sample bootstrap with 200 replications per imputation; weights for the super learners are computed using 4-fold cross-validation and non-negative least squares. Data sources: (*39-42*).

Table S10. Unweighted Estimates for the Total, Direct, and Indirect Effects of Neighborhood Poverty

Estimands	Point Estimate	[2.5, 97.5] Percentile Bootstrap Interval
Reading Test Scores		
Average Total Effect	-0.108	[-0.159, -0.067]
Natural Direct Effect	-0.074	[-0.107, -0.040]
Natural Indirect Effect	-0.035	[-0.063, -0.020]
Math Test Scores		
Average Total Effect	-0.103	[-0.151, -0.054]
Natural Direct Effect	-0.071	[-0.104, -0.033]
Natural Indirect Effect	-0.033	[-0.059, -0.016]

Notes: estimates are reported in standard deviation units and are computed using regression-imputation and random forests; they are unweighted and combined across 5 imputations; confidence intervals are based on the [2.5, 97.5] percentiles of a pooled sampling distribution simulated via the repeated half-sample bootstrap with 1000 replications per imputation. Data sources: (*39-42*).