Appendix

This appendix provides a more detailed discussion of the traffic data, and of exercises we conduct in order to explore bias in our estimated effects that derive from measurement error in child blood lead levels (Table 5). We also provide a more detailed discussion of our analysis of selection on observables.

1. Construction of the Traffic Measure

Steps in the construction of the traffic measure:

1) For each household, we measure the number of meters of road within 25 and 50 meters by functional class (primary, secondary and tertiary roads) and urban/rural status. Primary roads refer to highways, secondary roads to state routes and tertiary roads to streets. In total, there are 6 road types: primary-urban, secondary-urban, tertiary-urban, primary – rural, secondary-rural, tertiary-rural. See below for more detailed description of the road types.
2) Using data from the Bureau of Transportation Statistics on traffic patterns by road type for 1980 we then generate a measure of traffic volume by multiplying each meter of road by functional class and urban/rural status with measures of vehicles miles travelled per lane-mile by functional class and urban/rural status in 1980.[1]
3) Finally, we sum up the traffic measures generated in step 2. Roads within 25 meters are given full weight, roads between 25 and 50 meters are given ¼ weight to reflect the highly localized nature of lead contamination.
4) For each child, we may have multiple addresses and therefore multiple measures of traffic exposure. Since we are estimating the impact of an average measure of lead on outcomes, we calculate an average measure of traffic exposure and use that measure in our analysis.

Description of road types: Primary roads are generally divided limited-access highways within the interstate highway system or under State management, and are distinguished by the presence of interchanges. These highways are accessible by ramps and may include some toll highways. Secondary roads are main arteries, usually in the U.S. Highway, State Highway, and/or County Highway system. These roads have one or more lanes of traffic in each direction, may or may not be divided, and usually have at-grade intersections with many other roads and driveways. They

[1]These data are found here: http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_01_36.html

usually have both a local name and a route number. Tertiary roads consist of paved non-arterial streets, roads, or byways that usually have a single lane of traffic in each direction.

2. Further Discussion of Table 5

Columns 1 to 5 of Table 5 focus on the subset of children who have both a capillary and a venous BLL measure. Since capillary measures are less accurate, we would expect the use of a capillary measure to result in an estimate that is smaller than one based on a venous measure. Indeed, estimates for girls increase by 77% and by 56% for boys. These increases are calculated by comparing columns 2 and 3: for girls, the estimated effect increases from 0.0038 to 0.00674 and for boys from 0.00565 to 0.0088. When we instrument for the capillary measures with the venous measures, the estimates increase by 179% for girls and 130% for boys (column 4 compared with column 2).

A drawback of this approach is that relatively few children in our sample (24,509) have both types of measures and they are not a random sample of all children. They have higher than average BLLs and a greater degree of family disadvantage. The OLS estimate of the impact of average lead on suspension for this sample generates smaller estimated effects for girls, but the same effects for boys as in the general sample (Table 5, column 1).

Another means of assessing how much measurement error in BLLs might bias OLS estimates is to exploit the presence of multiple BLL measures per child and, following the approach of Ashenfelter and Krueger (1994). That is, one can instrument for one measure of a child's BLL with the average of the child's other BLL measures. Since 72 percent of the full sample has multiple BLL measures, the estimates derived from this sample are more likely to generalize to the full population than those based on comparing capillary and venous measures. Indeed, when we estimate simple OLS models on this sample, the estimates are nearly identical to previous OLS results (Table 5 of column 6). Columns 9 and 10 show that these IV estimates are again much larger than the corresponding OLS estimates shown in columns 7 and 8. For example, the instrumental variables estimate in column 10 suggests that a one unit increase in the average BLL increases the probability of any suspension by 0.96 percentage points for girls, and by 1.7 percentage points for boys, compared with 0.49 and 0.96, respectively, in the OLS, which represents an increase of 100% for girls and 70% for boys. The much larger IV estimates are consistent with measurement error in child BLLs biasing down OLS estimates.

Finally, in column 11 we show estimates where one child in the household's average BLL is instrumented using a sibling's BLL (if there are multiple siblings, we use the average of the siblings' BLLs).[2] This estimation can only be done on the subset of children (34,252) who have a sibling with a BLL. This is a larger sample than that used for the sibling FE in Table 4 because it includes children whose preschool siblings have BLLs but do not yet have suspension data because they are too young. The IV estimates of 1.2 for girls and 2.98 for boys are somewhat larger than those discussed above.

Panel B of Table 5 performs the same estimations for detention/incarceration, with the exception of the sibling IV. The latter is infeasible because the sibling subsample was born in 1997 or later so that they are too young for the detention outcome to be meaningful. Again, these estimates suggest that OLS estimates are biased towards zero due to considerable measurement error in BLLs. For example, column 8 shows the OLS estimate of lead using the average of all other tests as the measure of lead. The estimated effect of a one-unit increase in BLL for males is

[2] To maximize sample size we do not restrict to contemporaneous measures of sibling BLLs, but rather use all the average of all sibling BLLs.

0.36 percentage points. Column 10 shows the IV estimate, where the instrument is a randomly drawn BLL. The estimated effect increases to 0.47 percentage points.

3. Selection on Observables.

In the text, we discuss a second way to test for the importance of omitted variable bias in the OLS estimates which exploits the decline in negative selection into lead over this period. We argue that if observables and unobservables are positively correlated, then a decline in negative selection on observables over time will imply a decline in selection on unobservables as well. And if negative selection on unobservables is driving OLS estimates of the relationship between lead and suspensions, then the strength of the estimated OLS relationship between lead and suspensions should decline over time.

This conjecture is explored in Appendix Table 1. We first present evidence of the decline in negative selection in elevated BLLs over time. The sample is limited to children born from 1993 to 1998 so that we can examine a single uniform outcome for all: The number of infractions in 9th grade. We regress lead levels on race, free lunch, gender, year of birth (entered linearly), and census block group fixed effects, as well as interactions between free lunch and year of birth and African-American race and year of birth. The interaction terms are intended to capture any changes over time in selection into elevated lead by race and income. The results shown in column 1 indicate that being on free lunch or African American is strongly predictive of elevated BLLs, but that this relationship has been declining over time, as evidence by the negative and significant interaction terms.

Column 2 of Appendix Table 1 presents estimates of a regression of the number of ninth grade infractions on BLLs, lead interacted with year of birth, and all of the demographic controls. While the coefficient on lead is positive and significant, the coefficient on the interaction term between BLL and year of birth is small and imprecisely estimated. This table suggests that although disparities in lead exposure are falling, the relationship between BLLs and suspensions is not changing over time. This finding is inconsistent with the estimated effects being driven by omitted variables bias.

4. The relationship between traffic volume and birth weight

There appears to be no relationship between birth weight and traffic volume or its interaction with birth cohort (Appendix Table 2, column 1). Moreover, there does not appear to be any meaningful or significant relationship between a child's future lead level and birth weight (columns 2 and 3): The OLS estimate of -.00135 would suggest that a 3.5 unit decline in early childhood lead (the decline witnessed over the 14 year period studied here) is associated with a 5 gram decline in birth weight (relative to a mean birth weight of 3,300 grams) and the IV estimate is similarly small in magnitude though positive and statistically insignificant. We get similar results when we examine prematurity (birth before 36 weeks). We interpret these estimates as suggestive evidence that the decline in lead levels over time in high traffic areas does not reflect improving conditions more generally as these might also be expected to affect birth weight.

5. The relationship between traffic and child characteristics

As a separate check of our identification assumptions, we regress an indicator for African American and Free lunch status on the measures of traffic and traffic*year of birth and examine the coefficients on the interaction term. These regressions are parallel to the first stage regressions shown in Table 6 and are shown in Appendix Table 3. This table shows that over time, high traffic areas became slightly less disadvantaged in terms of percent African-American and percent using free lunch. Since African-Americans and children on free lunch have higher BLLs, the question is whether these small changes in the composition of high traffic areas could explain our results?

The following illustrative calculation shows that the changes are too small to have had much of an impact on BLLs. African-Americans make up 9% of the sample. Table 1 shows that mean BLLs are 5.29 for African Americans, and 3.80 overall, which implies a BLL of 3.65 for non-African Americans. Appendix Table 3 shows that over a ten-year period, the fraction African American declined in high traffic areas by .0021. Using the figures above and changing only the share of African Americans would yield an overall BLL of 3.79 instead of 3.80. Similarly, if we use the Appendix 3 figure and say that the share of children in high traffic areas increased by .0157 over ten years, then this would yield an overall BLL of 3.78. Given the large changes in BLLs that we find, it seems that very little of it could be accounted for by changes in the composition of high traffic areas.

Appendix Table 1: Summary Statistics for Sample by Child Characteristic

	All	White	Black	Hispanic	Free Lunch	Paid Lunch	Born >=1997	Sibling Sample
Lead (Geometric Mean per Child)	3.80	3.40	5.29	4.50	4.50	3.00	3.18	3.18
Number of tests	2.60	2.40	3.10	3.30	2.70	2.30	2.80	2.80
Share capillary	0.32	0.37	0.27	0.18	0.28	0.37	0.28	0.28
Any school disciplinary infraction	0.24	0.20	0.41	0.35	0.35	0.12	0.18	0.17
Any suspension	0.20	0.15	0.36	0.29	0.29	0.09	0.15	0.14
Any detention/incarceration	0.01	0.01	0.03	0.02	0.02	0.00		
Any detention/incarceration - males only	0.02	0.01	0.06	0.03	0.03	0.00		
Number of years with school infraction data	5.30	5.30	5.20	5.40	5.40	5.20	6.02	6.20
Traffic Volume (% s.d. relative to mean)	0.00	-0.07	0.18	0.17	0.10	-0.12		-0.02
Black	0.09				0.16	0.02	0.09	0.09
White	0.69				0.49	0.94	0.70	0.72
Hispanic	0.18				0.31	0.02	0.18	0.16
Asian	0.03				0.04	0.02	0.03	0.03
Sometimes Free Lunch	0.29	0.26	0.40	0.37	0.55		0.30	0.28
Always Free Lunch	0.24	0.12	0.50	0.57	0.45		0.26	0.26
Mother <HS	0.15	0.05	0.15	0.23	0.17	0.01	0.15	0.18
Mother HS	0.29	0.18	0.23	0.20	0.24	0.13	0.29	0.31
Mother College+	0.25	0.21	0.07	0.04	0.06	0.28	0.25	0.36
Maternal age	28.40	29.40	26.30	26.20	26.40	30.90	28.40	28.60
Mother married at birth	0.67	0.76	0.40	0.49	0.49	0.90	0.67	0.66
Birth Weight	3.34	3.36	3.25	3.30	3.30	3.40	3.30	3.30
Birth order	1.9	1.8	2.1	2.1	2.0	1.8	1.9	2.0
Observations (1990-2004 birth cohorts)	124,579	90,813	11,527	22,239	66,710	57,869	80,078	27, 515
Observations for birth certificate data (birth cohorts 1997+)	80,078	53,833	7,805	15,586	44,732	35,346	80,078	28,040
Observations for detention/incarceration (birth cohorts 1991-1999)	70,681	54,722	6,602	12,700	37,040	39,606	-	-

Data on lead levels and suspensions are available for children born between 1990 and 2004. Data from birth certificates is available only for children born in RI in or after 1997. Data on detention/incarceration is only available for children born 1991-1999.

Appendix Table 2: Negative Seleciton into Elevated Lead and Omitted Variable Bias			
	(1)	(2)	(3)
Outcome:	Lead	9th Grade Infractions	
Geometric Mean of Lead		0.0254	0.0171
		[0.0059]	[0.0063]
Lead*year of birth		0.0001	0.0024
		[0.0015]	[0.0016]
African American	1.2690	0.3040	0.4140
	[0.1250]	[0.0537]	[0.0826]
Sometimes free/reduced lunch	0.9680	0.5000	0.6210
	[0.0696]	[0.0211]	[0.0463]
Always free/reduced lunch	1.5690	0.6480	0.7950
	[0.0805]	[0.0260]	[0.0542]
Black*year of birth	-0.1190		-0.0284
	[0.0254]		[0.0168]
Sometimes free lunch*year of birth	-0.0967		-0.0329
	[0.0166]		[0.0110]
Always free lunch*year of birth	-0.1410		-0.0389
	[0.0186]		[0.0126]
Observations	49,374	49,374	49,374
R-squared	0.233	0.067	0.067

Sample restricted to chidlren born 1993-1998. This was done so as to limit the sample to children who have a uniform outcome available: 9th grade infractions data. Also included are year of birth (linear) and neighborhood (census block group) FE.

Appendix Table 3: Exploring Bias from Measurement Error Exploiting Multiple BLL Measures Per Child or Family

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Dep. Var is any suspen	OLS	OLS	OLS	IV	IV	OLS	OLS	OLS	IV	IV	IV
Geometric Mean of Lead	0.0083					0.0069					0.0122
	[0.0013]					[0.0007]					[0.0028]
Lead*male	0.0021					0.0057					0.0191
	[0.0016]					[0.0009]					[0.0031]
Average of all capillary tests		0.0040		0.0107							
		[0.0008]		[0.0015]							
Capilary lead*male		0.0017		0.0022							
		[0.0011]		[0.0017]							
Average of all venous tests			0.0069		0.0122						
			[0.0012]		[0.0021]						
Venous lead*male			0.0018		0.0042						
			[0.0015]		[0.0023]						
First random draw							0.0038		0.0088		
							[0.0005]		[0.0009]		
First random*male							0.0034		0.0071		
							[0.0007]		[0.0010]		
Avg all other tests								0.0048		0.0103	
								[0.0006]		[0.0011]	
Avg all other tests*male								0.0052		0.0061	
								[0.0008		[0.0012]	
Observations	24,404	24,404	24,404	24,404	24,404	90,700	90,700	90,700	90,700	90,700	28,040
R-squared	0.204	0.203	0.204			0.196	0.195	0.196			
Sample			Children with capilary and venous BLLs					Chidlren with multiple BLLs			Siblings
Instrument	none	none	none	venous	capillary	none	none	none	other Lead	first lead	sibling lead
Panel B: Dep. Var. is any detention											
Geometric Mean of Lead	-0.0006					-0.0006					
	[0.0004]					[0.0002]					
Lead*male	0.0041					0.00433					
	[0.0008]					[0.0005]					
Average of all capillary tests		-0.0007		0.0000							
		[0.0002]		[0.0005]							
Capilary lead*male		0.0024		0.0044							
		[0.0004]		[0.0006]							
Average of all venous tests			-0.0004		-0.0013						
			[0.0004]		[0.0007]						
Venous lead*male			0.0036		0.0058						
			[0.0007]		[0.0008]						
First random draw							-0.0005		-0.0004		
							[0.0002]		[0.0003]		
First random*male							0.0028		0.0050		
							[0.0003]		[0.0004]		
Avg all other tests								-0.0005		-0.0006	
								[0.0002]		[0.0004]	
Avg all other tests*male								0.0036		0.0056	
								[0.0004]		[0.0004]	
Observations	16,015	16,015	16,015	16,015	16,015	50,684	50,684	50,684	50,684	50,684	
R-squared	0.079	0.078	0.079			0.048	0.046	0.047			
Sample			Children with capilary and venous BLLs					Chidlren with multiple BLLs			
Instrument	none	none	none	venous	capillary	none	none	none	other Lead	first lead	

All controls from table 3 column 2 are included in all regressions. Column 1-5 include children for whom we have both capillary and venous test results. We present OLS estimates for this sample in column 1 using the average over all BLL measures. We present OLS estimates based on either capillary or venous BLLs (columns 2-3, respectively) and then instrument for the capillary measures using the venous measures (column 4) or vice versa (column 5). Columns 6-8 include children with two or more lead tests. Column 6 shows the OLS estimates for this sample based on the average over all BLL measures. Columns 7-8 show OLS estimates of the impact of one randomly drawn lead level (column 7) and then the average of all remaining BLLs (column 8). Note that for the latter, there may sometimes be only one BLL. Columns 9-10 show IV estimates, instrumenting for one randonly drawn lead level with the average of the other BLLs and vice-versa. In column 11, panel A, the sample is limited to children with at least one sibling with a BLL. We present IV estimates using the average sibling BLL as an instrument for the child's own BLL. Robust standard errors are clustered at the Census tract level and appear in brackets below the coefficient estimate.

Appendix Table 4: First Stage Regression Estimates		
	(1)	(2)
	Geometric Mean of Lead	Lead*male
Ln(traffic)*time trend	-0.0140	0.0001
	[0.0020]	[0.0015]
Ln(traffic)	0.2800	-0.0005
	[0.0274]	[0.0208]
Predicted lead*male		1.1440
		[0.0129]
African American	1.9860	-0.1070
	[0.3140]	[0.2320]
White	0.0101	0.0572
	[0.0472]	[0.0351]
Hispanic	-0.6120	0.0853
	[0.0519]	[0.0383]
Sometimes free/reduced lunch	2.5220	-0.1750
	[0.136]	[0.0992]
Always free/reduced lunch	2.8840	-0.2420
	[0.136]	[0.0990]
Male	0.1960	-0.5630
	[0.0146]	[0.0466]
Observations	124,371	124,371
R-squared	0.277	0.570

All regressions include the full set of controls included in the third column of Table 3.
Robust standard errors in parentheses. See equation (2) in text.

Appendix Table 5: Traffic and Birth Weight					
	(1)	(2)	(3)	(4)	(5)
	Reduced Form	OLS	IV	OLS	IV
Geometric Mean of Lead		-0.0031	0.0741	-0.0026	0.0784
		[0.0009]	[0.0663]	[0.0012]	[0.0684]
Lead*male				-0.0009	-0.0082
				[0.0015]	[0.0066]
Ln(traffic)*time trend	-0.0011				
	[0.0009]				
Ln(traffic)	0.0090		-0.0132		-0.0131
	[0.0126]		[0.0076]		[0.0076]
African American	0.4320	0.4530	0.1920	0.4540	0.2020
	[0.2680]	[0.2680]	[0.3500]	[0.2680]	[0.3470]
White	0.1290	0.1300	0.1220	0.1300	0.1220
	[0.0096]	[0.0096]	[0.0133]	[0.0096]	[0.0133]
Hispanic	0.1310	0.1300	0.1670	0.1300	0.1670
	[0.0102]	[0.0102]	[0.0336]	[0.0102]	[0.0336]
Sometimes free/reduced lunch	-0.1890	-0.1940	-0.2470	-0.1940	-0.2470
	[0.1540]	[0.1540]	[0.1720]	[0.1540]	[0.1720]
Always free/reduced lunch	-0.1890	-0.1930	-0.2650	-0.1940	-0.2660
	[0.1530]	[0.1530]	[0.1780]	[0.1530]	[0.1780]
Male	0.0979	0.0981	0.0864	0.1010	0.1130
	[0.0038]	[0.0038]	[0.0111]	[0.0063]	[0.0178]
Mother <HS	-0.0351	-0.0339	-0.0572	-0.0339	-0.0573
	[0.0073]	[0.0073]	[0.0211]	[0.0073]	[0.0211]
Mother HS grad	0.0170	0.0170	0.0176	0.0170	0.0176
	[0.0053]	[0.0053]	[0.0056]	[0.0053]	[0.0056]
Mother College+	0.0509	0.0510	0.0546	0.0510	0.0546
	[0.0057]	[0.0056]	[0.0068]	[0.0056]	[0.0068]
Married at birth	0.0610	0.0608	0.0759	0.0608	0.0758
	[0.0057]	[0.0057]	[0.0145]	[0.0057]	[0.0145]
Maternal age	-0.0044	-0.0044	-0.0023	-0.0044	-0.0022
	[0.0005]	[0.0005]	[0.0020]	[0.0005]	[0.0020]
Birth order	0.0509	0.0517	0.0358	0.0517	0.0359
	[0.0026]	[0.0026]	[0.0137]	[0.0026]	[0.0136]
Observations	79,951	80,079	79,951	80,079	79,951
R-squared	0.042	0.043		0.043	

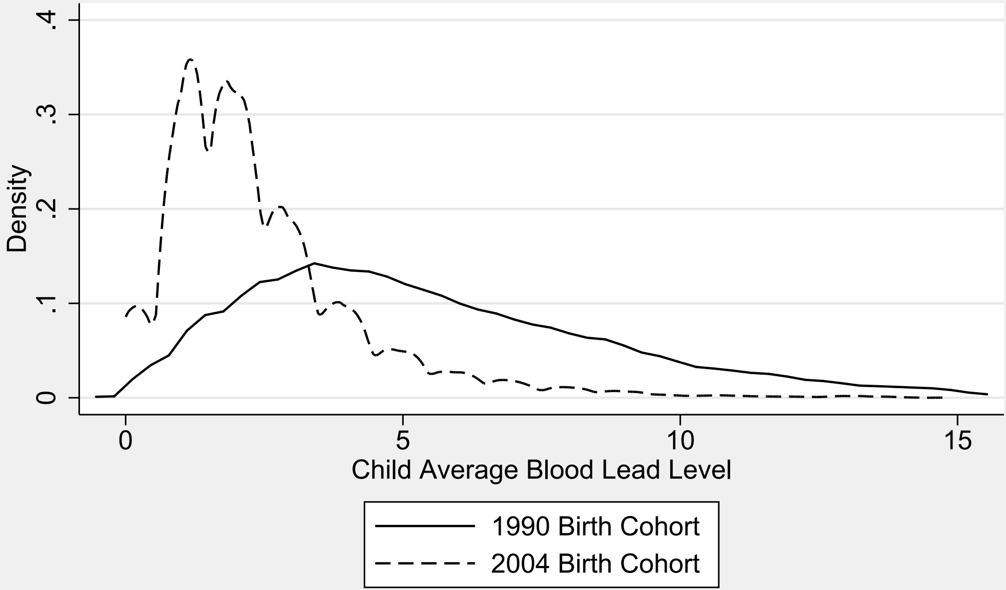
Sample includes births 1997+ that occurred in RI. Birth weight is measured in kilograms. Regressions include all other variables listed in Table 3 column 3. Robust standard errors clustered at the level of the census tract are shown in parentheses.

Appendix Table 6: Socio-Economic Status and Exposure to Traffic Over Time

Dependent Variable:	(1) African American	(2) Never free/reduced lunch	(3) Always free/reduced lunch	(4) Sometimes free/reduced lunch
Ln(traffic)*time trend	-0.00021	0.00157	-0.00026	-0.00131
	[0.00004]	[0.00008]	[0.00023]	[0.00024]
Observations	124,371	124,371	124,371	124,371
Average dependent variable	0.09	0.47	0.24	0.29
R-squared	0.980	0.980	0.358	0.429
Controls:				
Race specific quadratic time trends	Y	Y	Y	Y
Free lunch specific quadratic time trends	Y	Y	Y	Y
Same specification as the first stage	Y	Y	Y	Y

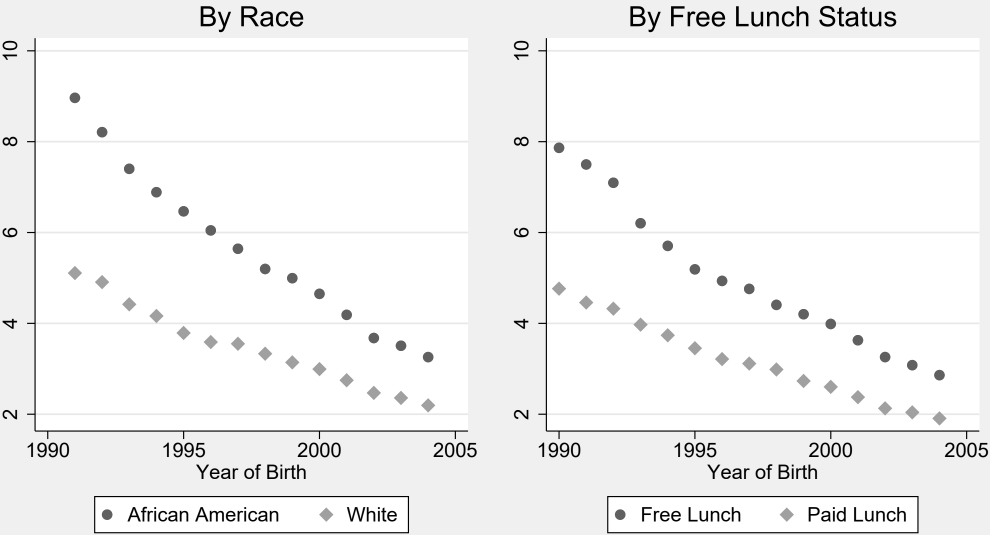
All regressions include gender, year of birth FE, neighborhood FE and controls for years of infractions data. Regressions with African American (free lunch) as the dependent variable include controls for free lunch status (race).

Appendix Figure 1: Child Lead Levels by Birth Cohort

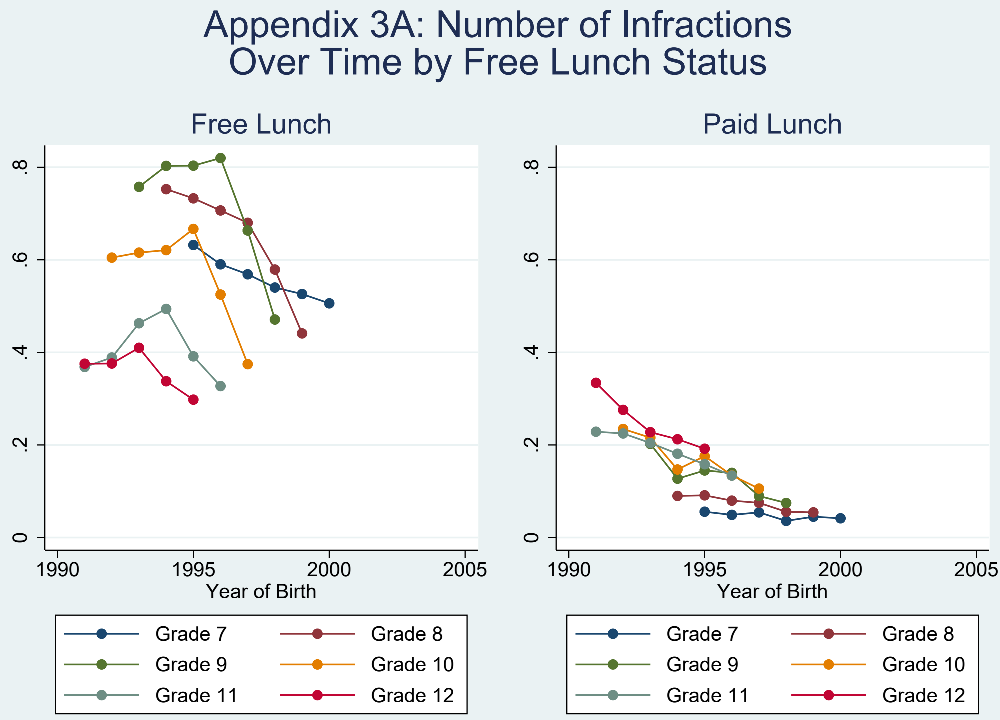


The distribution of the geometric mean of each child's BLL over the first 72 months of life is presented for children born in 1990 and 2004 corresponding to the oldest and youngest children in our sample.

Appendix Figure 2: Trends in Lead Levels by Race & Free Lunch Status



The geometric mean over all BLLs is computed for each child.
The average over all geometric means is then calculated
by child birth cohort and either race or free lunch status.


Appendix 3A: Number of Infractions
Over Time by Free Lunch Status
Free Lunch
Paid Lunch
0
.2
.4
.6
.8
1990
1995
2000
2005
Year of Birth
Grade 7
Grade 8
Grade 9
Grade 10
Grade 11
Grade 12

Appendix Figure 3B: Number of Infractions Over Time by Race