Appendix E

Detailed Discussion of Massachusetts Exploratory Analyses

E. Relationship between Local Blood Lead Data and Explanatory Variables

E.1 Analyses of Local Blood Lead Data by Demographic Variables

As noted above, the identical set of 43 Census variables across nine general categories were explored to determine their association with the Massachusetts surveillance data.

E.1.1 Income Variables

Initial results from exploring and modeling the income-related variables are presented in Figures/Tables B1 to B9. Highlights from the analysis of each exploratory variable are listed below.

Median Family Income – Similar relationships between GM blood lead and Median Family Income are seen across the three time periods, with GMs decreasing as Median Family Income increases. Blood lead levels decline slightly between each consecutive time periods. The distribution of Median Family Income is stable across the three time periods. Median Family Income is significant in Models 1 to 5. The findings for Massachusetts were very similar to those seen in the national data.

Median Household Income – Similar relationships between GM blood lead and Median Household Income are seen across the three time periods, with GMs decreasing as Median Household Income increases. Blood lead levels decline steadily across the three periods. Median Household Income is stable across the three time periods. Median Household Income is significant in Models 1 to 5.

Median Per Capita Income – Consistent relationships between GM blood lead and Median Per Capita Income are seen across the four time periods, with GMs decreasing as Median Per Capita Income decreases. The distribution of Median Per Capita Income shows a very slight increase across the three periods, about 0.1 percent from the 2000-2002 period to the 2005-2006 period. Median Per Capita Income is significant in Models 1 to 5.

Percent Units with No Household Earnings – Consistent relationships between GM blood lead and Percent of Units with No Household Earnings are seen across the three time periods, with GMs increasing as percent of units with no household earnings increases. Mean and median Percent of Units with No Household Earnings do not change across the three time periods. Percent of Units with No Household Earnings is significant in Models 1 to 5.

Percent Units with No Household Wage – This variable performs similarly to the previous one. Consistent relationships between GM blood lead and Percent of Units with No Household Wages are seen across the three time periods, with GMs increasing as the percentage increases. Mean and median Percent of Units with No Household Wages is stable across the three time periods. Percent of Units with No Household Wages is significant in Models 1 to 5.

Percent Households on Public Assistance – Consistent relationships between GM blood lead and Percent of Households on Public Assistance are seen across the three time periods, with GMs increasing as Percent of Households on Public Assistance increases. Mean and median Percent of Households on Public Assistance are stable across the three periods. Percent of Households on Public Assistance is significant in Models 1 to 5.

Percent Households Below Poverty Line – Consistent relationships between GM blood lead and Percent of Households Below the Poverty Line are seen across the three time periods, with GMs increasing as Percent of Households Below the Poverty Line increases. Mean and median Percent of

Households Below the Poverty Line are stable across the three periods. Percent of Households Below the Poverty Line is significant in Models 1 to 5.

Percent Units with Family Income Below Poverty Line – Consistent relationships between GM blood lead and Percent of Units with Family Income Below the Poverty Line are seen across the three time periods, with GMs increasing gradually as percentage of units increases. Mean and median Percent of Units with Family Income Below the Poverty Line are stable across the three periods. Percent of Units with Family Income Below the Poverty Line is significant in Models 1 to 5.

Percent Units Spending Less than Five Years in Poverty – Consistent relationships between GM blood lead and Percent of Units Spending Less than Five Years in Poverty are seen across the three time periods, with GMs increasing as percent of units increases. Mean and median Percent of Units Spending Less than Five Years in Poverty do not change across the three periods, although the latter two periods have a significantly higher maximum value (17.3 vs. 12.7 percent). Percent of Units Spending Less than Five Years in Poverty is significant in Models 1 to 3 and 5, but failed to converge in Model 4.

Across these nine variables, very few differences were evident from the exploratory analyses. In all cases, each of the three time periods had similar distributions and similar relationships with predicted GM blood lead levels. Of the nine variables in the Income category, 8 were significant for Models 1 through 5, with only Model 6 (proportion of children with blood lead levels over 25 µg/dL) failing to converge. As seen in Table 4-2, the variable Median Household Income provided the best fit in Models 1 to 4, while Percent Units with No Household Wage provided the best fit for Model 5.

E.1.2 Race Variables

Results from the eight race variables explored are presented in Figures/Tables B10 to B17, with highlights listed below.

Percent American Indian and Alaskan Native Alone – As expected, most Census tracts in Massachusetts have very low percentages of American Indian and Alaskan Natives, with the 90th percentile being 0.5 percent even lower than the 1.2 percent nationally. The three time periods display similar relationships between blood lead levels and percent American Indian and Alaskan Native, as blood lead levels increase as Percent American Indian increases. The mean percent of American Indian and Alaskan Native is steady across the three time periods at 0.19 percent. Percent of American Indian and Alaskan Native is a significant predictor of the outcome measure in Models 1 to 5.

Percent Asian Alone – The relationship between GM blood lead and percent Asian appears to be changing slightly across the three time periods. While the relationship between GM blood levels and Percent Asian was nearly flat with just a slight negative slope, in each the two successive time periods the slopes are more negative. This indicates that in more recent years lower blood lead levels are associated with Census tracts with higher percentages of Asian people. The distribution of Percent Asian is stable across the three time periods. Percent Asian is significant in Models 1 to 3, but not in Model 5, while Model 4 failed to converge.

Percent Black Alone – Percent Black Alone in each Census tract ranges from 0 to 91 percent. Consistent relationships between GM blood lead and Percent Black are seen across the three time periods, with GMs increasing as Percent Black increases. Mean and median Percent Black are stable across the three periods. Percent Black is a significant predictor of the outcome variable in Models 1 to 5.

Percent White Alone – The distribution of Percent White is also wide-ranging with Census tracts containing 0.1 to 100 Percent White. The relationship between GM blood lead and Percent White is consistent across the three time periods, with GMs declining gradually as Percent White increases. The distribution of Percent White is stable across the three time periods, as mean and median Percent White stay about the same. Percent White is a significant predictor of the outcome variable in Models 1 to 5.

Percent Native Hawaiian and Other Pacific Islander Alone – Very few Native Hawaiians live in Massachusetts as evidenced by a 90^{th} percentile of Percent Native Hawaiian of 0.0 percent. The distribution of Percent Native Hawaiian is stable across the three time periods with a mean of 0.02 percent for each period. The relationship between GM blood lead levels and Percent Native Hawaiian is consistently flat across the three time periods. Percent Native Hawaiian is not significant in Models 1 to 5.

Percent Other Race Alone – The relationship between Percent Other Race Alone and GM blood lead levels in Massachusetts is opposite of the national trend. GM blood lead levels increase steadily as percent Other Race increases across all three time periods; however, the increase appears slightly steeper in the 2000-2002 time period. Mean and median percent Other Race remain fairly stable across time periods, with an average mean percentage of 0.75 across the time periods. Percent Other Race is significant in Models 1 to 5.

Percent Multiple Races – Predicted GM blood lead levels consistently increase as Percent Multiple Races increases across the three time periods, although the 2000-2002 period again has a slightly more positive slope. Mean and median Percent Multiple Races are stable across the three periods. Percent Multiple Races is significant in Models 1 to 5.

Percent Hispanic – Predicted GM blood lead levels generally increase gradually as Percent Hispanic increases, however the 2000-2002 time period had a slightly less positive slope. Mean and median Percent Hispanic are stable across the three time periods. Percent Hispanic is significant in Models 1 to 5.

Similar to the national data, in the Massachusetts data the eight race variables were not as consistently predictive of blood lead levels as the income variables; however six of the eight variables were significant predictors in the five models that consistently converged – Models 1 to 5. The only two variables that yielded different results were Percent Asian and Percent Native Hawaiian. The race composition of the Census tracts included in each of the three time periods was consistent across the periods.

Two different variables provided best-fitting models as seen in Table 4-2. Percent Multiple Races provided the best fit for Models 1, 2, 3, and 4; and Percent Native Hawaiian provided the best fit for Model 5 (proportion of children with blood lead levels above 15 μg/dL), although it was not a significant predictor in that model. Percent Asian provided the second best fit in Model 5, but it also was not significant.

E.1.3 Housing Cost Variables

Two variables related to housing costs in Massachusetts, Median Rent and Housing Value, were analyzed to explore their relationships with children's blood lead levels. Figure/Table B19 and B20 contain the exploratory results from these two variables.

Median Rent – Predicted GM blood lead levels decline sharply as Median Rent increases across all three time periods, with similar slopes. The distribution of Median Rent is quite stable across the three time periods, as the means are within $1 of each other and medians are all $619. Median Rent is significant in Models 1 to 5.

Housing Value – Predicted GM blood lead levels consistently decline as Housing Value increases across all three time periods, although the slopes are not as steep as with Median Rent. Mean Housing Value in each time period is around $204,000 while median Housing Value is about $172,000 in each time period. Housing Value is significant in Models 1 to 5.

As in the national model, Median Rent provided a better fit than Housing Value in each model, as evidenced by the log-likelihood statistics in Table 4-2**.**

E.1.4 Occupancy Variables

Figure/Table B18 contains results of exploring Percent of Rental Units and Figure/Table B21 contains results of exploring Percent of Vacant Units.

Percent Rented Units – Consistent relationships between GM blood lead and Percent Rented Units are seen across the three time periods, with GMs increasing gradually as Percent Rented Units increases. The distribution of Percent Rented Units is stable across time periods, with means of about 39.6 percent and medians about 34.8 percent. Percent Rented Units is significant in Models 1 to 5.

Percent Vacant Units – The relationship between predicted GM blood lead and percent Vacant Units is similar across the three time periods, with blood lead levels increasing as percentage increases. The 2000-2002 time period has a more positive slope than the other two periods. The distribution of Percent Vacant Units is stable across the time periods, with means of about 5.35 percent and medians at 3.2 percent. Identical to the modeling results for Percent Rented Units, Percent Vacant Units is significant in Models 1 to 5.

The log-likelihood statistics in Table 4-2 indicate that Percent Rented Units provided a better model fit for Models 1 through 4. Model 5 was the exception in which Percent Vacant Units provided a better fit. This is counter to the national models, in which Percent Vacant Units achieved the better fit in all but one model.

E.1.5 Single Parent Status Variable

Exploratory analysis results are contained in Figure/Table B22.

Percent Single Parent Households – Predicted GM blood lead levels increase as Percent Single Parent Households increases across the three time periods. The distribution of Percent Single Parent Households is nearly identical across the three time periods. Percent Single Parent Households is significant in Models 1 to 5. For comparison, the fit of Models 1 to 3 evaluating Percent Single Parent Households is better than any models from the previous two variable categories.

E.1.6 Housing Age Variables

The results of the 12 variables explored are contained in Figure/Tables B23 to B34.

Median Year Built – Note that the minimum Year Built is 1939. According to the histogram in Figure B23, this category accounts for the average Median Year Built in over 30 percent of all Census tracts. Consistent relationships between GM blood lead and Median Year Built are seen across the three time periods with blood lead levels declining as age of housing increases, although it appears that the 2000-2002 period has a slightly steeper slope. The distribution of Median Year Built is identical across the three periods. Median Year Built is significant in all models.

Median Year Occupied Units were Built – The results for Median Year Occupied Units were Built is very similar to those observed for Median Year Built. Consistent relationships between predicted GM blood lead and Median Year Occupied Units were Built are seen across the three time periods with blood lead levels declining as age of housing increases, although it appears that the 2000-2002 period has a slightly steeper slope. The distribution of Median Year Occupied Units were Built across the three time periods are nearly identical. Median Year Occupied Units were Built is significant in Models 1 to 5.

Percent Units Built Before 1940 – Consistent relationships between GM blood lead and Percent Units Built Before 1940 are seen across the three time periods with blood lead levels increasing as the percentage increases, although it appears that the 2000-2002 period increases at a slightly faster rate. The distribution of Percent Units Built Before 1940 across the three time periods are nearly identical. Percent Units Built Before 1940 is significant in Models 1 to 5.

Percent Units Built Before 1950 – Similar results are seen with this variable as with Percent Units Built Before 1940. Consistent relationships between GM blood lead and Percent Units Built Before 1950 are seen across the three time periods with blood lead levels increasing as the percentage increases, although it appears that the 2000-2002 period increases at a slightly faster rate. The distributions of Percent Units Built Before 1950 across the three time periods are nearly identical. Percent Units Built Before 1950 is significant in Models 1 to 5.

Percent Units Built Before 1960 – Similar results are seen with this variable as with the previous two (percent of units built before 1940 and 1950). Percent Units Built Before 1960 also is significant in Models 1 to 5.

Percent Units Built Before 1970 – Again, these results are very similar to the other Percent Units Built Before variables. Each time period has a consistently positive relationship between blood lead levels and percent built before 1970, although the intercepts are slightly lower. The distributions of Percent Units Built Before 1970 are stable across the three time periods. Percent Units Built Before 1970 is significant in Models 1 to 5.

Percent Units Built Before 1980 – Results are nearly identical to those for Percent Units Built Before 1970. Percent Units Built Before 1970 is significant in Models 1 to 5.

Percent Occupied Units Built Before 1940 – The results of the exploratory analyses on this variable are very similar to those seen from the Percent Units Built Before 1940 variable, and most of the other Percent Units Built Before variables. Percent Occupied Units Built Before 1940 is significant in all models.

Percent Occupied Units Built Before 1950 – Similar results. Percent Occupied Units Built Before 1950 is significant in all models.

Percent Occupied Units Built Before 1960 – Similar results. Percent Occupied Units Built Before 1960 is significant in all models.

Percent Occupied Units Built Before 1970 – Similar results. Percent Occupied Units Built Before 1970 is significant in all models.

Percent Occupied Units Built Before 1980 – Similar results. Percent Occupied Units Built Before 1980 is significant in all models.

As with the national county-level data set, the exploratory results of all 10 of the percentage variables in the Census-tract level Massachusetts data appeared to be quite similar. Whether percent of all housing units or percent of occupied units was used did not seem to make a difference in the results. Similarly, the cutoff year used also did not seem to impact results. Three different percentage variables were among the best-fitting models. As seen in Table 4-2, Percent Units Built Before 1940 provided the best fit for Models 1, 2, and 3. Percent Occupied Units Built Before 1940 provided the best fit for Model 4, although Percent Units Built Before 1940 had a nearly identical log-likelihood statistic. Percent Occupied Units Built Before 1980 provided the best fit for Model 5, whose outcome measure is proportion of children with blood lead levels of 15 μg/dL or greater.

E.1.7 Children's Age Variables

Figures/Tables B35 and B36 contain the exploratory results for percentage and number of residents less than six years old, respectively.

Percent Less Than 6 Years of Age – During the 2000-2002 time period, it appears that slightly lower GM blood lead levels are associated with higher percentages of children less than six years old. For the other two time periods, the relationship between predicted GM blood lead level and Percent Less Than 6 is flat. The distributions of Percent Less Than 6 are stable across the three time periods. Percent Less Than 6 Years of Age is not significant in Models 1 to 3 and 5, while Model 4 failed to converge.

Number Less Than 6 Years of Age – Consistent relationships between GM blood lead and Number Less Than 6 Years of Age are seen across the three time periods with blood lead levels decreasing as the percentage increases, although it appears that the 2000-2002 period decreases at a slightly faster rate. Mean and median Number Less Than 6 Years of Age are fairly stable across the three time periods. Number Less Than 6 Years of Age is significant in Models 1 to 4, while Model 5 failed to converge.

As reported in Table 4-2, the Number Less Than 6 Years of Age variable provided the better fit for Models 1 to 4, while Percent Less Than 6 Years of Age provided the better fit for Model 5, although this is because the model with Number Less Than 6 Years of Age did not converge.

E.1.8 Education Level Variables

Four variables were constructed and analyzed related to education level. Results are detailed in Figures/Tables B37 to B40.

Percent Residents with Less Than 9th Grade Education – Consistent relationships between predicted GM blood lead and Percent Residents with Less Than 9th Grade Education are seen across the three time periods with blood lead levels increasing as the percentage increases. The distribution of Percent Residents with Less Than 9th Grade Education is stable across the three periods. Percent Residents with Less Than 9th Grade Education is significant in all models.

Percent Residents Without a High School Degree – The exploratory results on this variable are similar to those for the less than 9th Grade variable. Consistent relationships between GM blood lead and Percent Residents Without a High School Degree are seen across the four time periods with blood lead levels increasing as the percentage increases. The distribution of Percent Residents Without a High School Degree is stable across the three periods. Percent Residents Without a High School Degree is significant in Models 1 to 5.

Percent Residents Without College Education – Consistent relationships between GM blood lead and Percent Residents Without College Education are seen across the three time periods with blood lead levels increasing as the percentage increases. Mean and median Percent Residents Without College Education change only slightly across the three periods. Percent Residents Without College Education is significant in Models 1 to 5.

Percent Residents Without College Degree – Similar results were obtained for this variable. The relationship between predicted blood lead levels and Percent Without College Degree is consistent across time, with higher lead levels associated with higher percentages. The distribution of Percent Residents Without a High School Degree is stable across the three periods. Percent Residents Without College Degree is significant in all models.

Very similar exploratory results were seen across these four variables focusing on parents' education level. In general, higher lead levels are associated with less education. Also, the percentage of residents included in the analyses without various education levels increases slightly across the four time periods. Table 4-2 reports that Percent Residents Without College Education provides the best model fits for Models 1 and 2, which model GM blood lead levels. Percent Residents without a High School Degree provided the best fit for Model 3. Percent Residents with Less Than 9th Grade Education provided the best fit for Models 4 and 5.

E.1.9 Population Variables

The detailed results of the exploratory analyses on the three variables related to population are included in Figures/Tables B41 to B43.

Total Housing Units – The relationship between predicted GM blood lead and Total Housing Units is mostly consistent across the three time periods with predicted blood lead levels decreasing slightly as number of housing units per Census tract increases. The relationship appears to be flattening slightly with each successive time period. The distribution of Total Housing Units within each Census tract is similar for each time period. Total Housing Units is significant in all models.

Total Population – The results for Total Population are very similar to Total Housing Units. The relationship between predicted GM blood lead and Total Population decreases slightly as number of housing units per Census tract increases, with the relationship flattening slightly with each successive time period. The distribution of Total Population within each Census tract is similar for each time period. Total Population is significant in Models 1 to 4, but Model 5 did not converge.

Housing Density – In each time period, predicted GM blood lead levels increase as housing density increases; however, it appears that the relationship flattens slightly with each successive time period. The distribution of Housing Density within each Census tract is similar for each time period. Housing Density is significant in all models.

Among the population variables, Total Population provided the best model fit for Models 1 to 4. For Model 5, that did not converge for Total Population, Total Housing Units provided the better fit.

E.2 Analyses of Local Blood Lead Data by Environmental Variables

Environmental data acquired for this project will include air and groundwater monitoring data aggregated at the county level for the low resolution model and at higher resolutions when possible for the MA analyses. In cases where the data are available for a limited number of air-monitoring stations or drinking water samples available for the region(s) being investigated, geo-spatial modeling techniques might be used as appropriate to develop predictions across the entire region. Existence of industrial sources of lead within each county as indicated by the toxics release inventory will also be included as an environmental data source. Currently, impacts of three air lead-related variables have been analyzed for their impacts on children's blood lead levels in Massachusetts. Additionally, twelve variables were generated from Toxics Release Inventory (TRI) data and analyzed in conjunction with Massachusetts childhood lead data.

E.2.1 Air Lead Variables

The variable Air Dispersion (ASPEN) Model captures the output from the ASPEN model for each Massachusetts Census tract investigated. Exploratory results are presented in Figure/Table B44. The Air Exposure (HAPEM5) Model variable captures the predicted exposure data from the HAPEM5 model. Results are presented in Figure/Table B45. The third air lead variable considered, Air Hazard Quotient (HQ), is derived from the 1999 National Scale Air Toxics Assessment data. This variable represents lifetime exposure for children at the centroids of each Census tract. Lifetime exposure is calculated based on considering annual exposures and yearly activity patterns. Results from the Air Hazard Quotient variable are in Figure/Table B46.

Air Dispersion (ASPEN) Model – Consistent relationships between predicted GM blood lead levels and Air Dispersion Model levels are seen across the three time periods, generally with higher air lead levels associated with slightly lower blood lead levels. The distribution of the Air Dispersion Model across the three time periods is identical. Air Dispersion Model was not significant in any models.

Air Exposure (HAPEM5) Model – The Air Exposure Model variable performs similarly to the Air Dispersion Model variable. Generally, higher exposure levels are associated with slightly lower predicted GM blood lead levels in each time period. The distribution of the Air Exposure Model across the three time periods is nearly identical. Air Exposure Model was not significant in any models.

Air Hazard Quotient – The Air Hazard Quotient also performs similarly to the other two air lead variables. Generally, higher Air Hazard Quotient is associated with slightly lower predicted GM blood lead levels in each time period, although the relationship is nearly flat. The distribution of the Air Exposure Model across the three time periods is nearly identical. Air Hazard Quotient was not significant in any models.

Table 4-2 reports that the log-likelihood ratios from the models containing the three air lead variables are very similar. Although the log-likelihoods are nearly identical across the three variables, Air Hazard Quotient had the lowest ratio in each model.

E.2.2 Toxics Release Inventory Variables

EPA's Toxic Release Inventory (TRI) catalogs various sources of lead, based on information provided by industrial facilities. This data source was used to generate Census tract-level estimates within Massachusetts of the total amount of lead and/or lead-containing compounds that are released by industrial facilities into the environment via air, surface water, or underwater injection. Three types of TRI variables were utilized – total compounds, lead only, and total lead. Within each types, four pollution variables were explored – total lead in the air, lead in fugitive air, lead from smokestacks, and lead in surface water. Thus, 12 total TRI data variables were evaluated. The results from investigating the 12 TRI data variables are presented in Figures/Tables B63 to B74.

TRI Compounds (Total Air) – Generally, lower predicted blood lead levels are associated with higher levels of lead compounds in total air. The 2000-2002 and 2003-2004 time periods have slight negative slopes, while the 2005-2006 relationship is nearly flat. The distributions of lead compounds in total air are very similar across the three time periods, with the 90^{th} percentiles of the distributions equal to 0.00. TRI Compounds (Total Air) is significant in Model 4 and borderline significant in Model 3 (p-value=.064), but is not significant in the other three models.

TRI Compounds (Fugitive Air) – The relationship between TRI Compounds in fugitive air and predicted GM blood lead levels is more consistent across the three time periods than the prior variable. Again, there is a negative relationship as lower blood lead levels are associated with higher levels of fugitive air lead. The distributions of the three time periods are nearly identical, as again the 90^{th} percentiles all equal 0.00. TRI Compounds (Fugitive Air) is significant in Models 3 and 4, but not significant in the other models, although p-values from those three models range from 0.069 to 0.11.

TRI Compounds (Air Lead from Stacks) – Predicted blood lead levels decline slightly as TRI Compounds (Air Lead from Stacks) increases. The 2000-2002 and 2003-2004 periods have slight negative slopes between the two variables while the 2005-2006 period appear to be flat. The distributions of the three time periods are nearly identical, as again the 90^{th} percentiles all equal 0.00. TRI Compounds (Air Lead from Stacks) is only significant for Model 4, but not in the other four models.

TRI Compounds (Water Surface) – The relationship between predicted GM blood lead levels and TRI Compounds (Water Surface) appears to be flat in all three time periods. The distribution of TRI Compounds (Water Surface) is identical across the three time periods, as again the 90^{th} percentiles all equal 0.00. TRI Compounds (Water Surface) is not significant in any of the five models.

TRI Lead Only (Total Air) – The relationship between predicted GM blood lead levels and TRI Lead Only (Total Air) is positive as blood levels increase with air lead levels. The distribution of TRI Lead Only (Total Air) is identical across the three time periods, as again the 90^{th} percentiles all equal 0.00. TRI Lead Only (Total Air) is not significant in any of the five models.

TRI Lead Only (Fugitive Air) – The relationship between predicted GM blood lead levels and TRI Lead Only (Fugitive Air) is positive as blood levels increase with air lead levels in all three time periods. The distribution of TRI Lead Only (Fugitive Air) is nearly identical across the three time periods, as again the 90^{th} percentiles all equal 0.00. TRI Lead Only (Fugitive Air) is only significant at the .05 level for Model 4, while the p-values from Models 1 to 3 range from .056 to .081.

TRI Lead Only (Air Lead from Stacks) – The relationship between predicted GM blood lead levels and TRI Lead Only (Air Lead from Stacks) is negative over the 2000-2004 data as blood levels decrease as air lead levels increase, while the relationship is flat in the 2005-2006 data. The distribution of TRI

Lead Only (Air Lead from Stacks) is identical across the three time periods, as again the 90th percentiles all equal 0.00. TRI Lead Only (Air Lead from Stacks) is not significant in any of the five models.

TRI Lead Only (Water Surface) – The relationship between predicted GM blood lead levels and TRI Lead Only (Water Surface) is positive across the three time periods with the 2000-2002 period have a slightly more positive slope than the other two period. The distribution of TRI Lead Only (Water Surface) is identical across the three time periods, as again the 90th percentiles all equal 0.00. TRI Lead Only (Water Surface) is significant for Models 1 to 4, but not significant for Model 5.

TRI Total Lead (Total Air) – Generally, lower predicted blood lead levels are associated with higher levels of Total Lead in total air. The 2000-2002 and 2003-2004 time periods have slight negative slopes, while the 2005-2006 relationship is nearly flat. The distributions of TRI Total Lead (Total Air) are very similar across the three time periods, with the 90th percentiles of the distributions equal to 0.00. TRI Total Lead (Total Air) is not significant in any of the five models.

TRI Total Lead (Fugitive Air) – The relationship between predicted GM blood lead levels and TRI Total Lead (Fugitive Air) is positive as blood levels increase with air lead levels in all three time periods. The distribution of TRI Total Lead (Fugitive Air) is nearly identical across the three time periods, as again the 90th percentiles all equal 0.00. TRI Total Lead (Fugitive Air) is not significant in any of the five models.

TRI Total Lead (Air Lead from Stacks) – The relationship between predicted GM blood lead levels and TRI Total Lead (Air Lead from Stacks) is slightly negative over the 2000-2004 data as blood levels decrease as air lead levels increase, while the relationship appears to be flat in the 2005-2006 data. The distribution of TRI Total Lead (Air Lead from Stacks) is identical across the three time periods, as again the 90th percentiles all equal 0.00. TRI Total Lead (Air Lead from Stacks) is significant in Model 4, but is not significant in any of the other four models.

TRI Total Lead (Water Surface) – The relationship between predicted GM blood lead levels and TRI Total Lead (Water Surface) is slightly positive across the three time periods as predicted blood lead level increases very slightly as lead in water increases. The distribution of TRI Total Lead (Water Surface) is identical across the three time periods, as again the 90th percentiles all equal 0.00. TRI Total Lead (Water Surface) is not significant in any of the five models.

Across the 12 TRI variables, the vast majority of measurements for Census tracts in Massachusetts are 0.00, over 90 percent for each variable. Of the 60 variable/model combinations, the TRI variables were only significant predictors of blood lead levels for 10 of those models. Although the log-likelihood ratios from each model are similar across the 12 variables for each model, TRI Lead Only (Water Surface) provided the best model fit for all five models.

E.3 Analyses of Local Blood Lead Data by Programmatic Variables

Most of the explanatory variables being explored in this project are considered risk factors for childhood lead poisoning. It is anticipated that the level and characteristics of programmatic support from either federal, state, or local sponsors will contribute towards meaningful reductions in the prevalence of childhood lead poisoning. In the high resolution models, the various characteristics of the programs (information from housing inspections and case management services) are being explored within the statistical models.

E.3.1 Programmatic Funding Variables

For the high-resolution model in the Commonwealth of Massachusetts, information on within-state funding levels was obtained and analyzed. Within-state funding data were available down to the township level. That funding was then allocated to Census tracts in two ways. First, in the Not Standardized version, total dollars were allocated to a Census tract proportionally based on population and analyzed. These models are identified as $ per Census Tract in each variable name. Second, in the Standardized version, the dollars allocated to each tract were divided by the number of children less than six years old and the funds per child were analyzed. These models are identified as $ per Child in each variable name.

The HUD funding and CDC funding information was also explored with the Massachusetts data. Four variables were generated from these data and analyzed – current and cumulative funding allocated to each Census tract in Massachusetts to combat childhood lead poisoning, both Standardized by number of children per tract and Not Standardized. The state, HUD, and CDC funding data also were combined to create Total Funding variables, including both current and cumulative levels and both standardized and not standardized versions. The detailed exploratory analysis results for the 16 state, HUD, and total funding variables are presented in Figures/Tables B47 to B62. Results for each variable are discussed below.

Current HUD Funding ($ per Child) – The relationship between predicted GM blood lead levels and Current HUD Funding is not consistent across the three time periods. In the earlier time periods (2000-2002 and 2003-2004), the relationship appears to be nearly flat, with a slight negative slope in the first time period and a slight positive slope in the second time period. Meanwhile, the most current time period has a more negative slope, with lower lead levels associated with higher funding amounts. The distributions of Current HUD Funding also change across the time periods. The 2000-2002 period has higher mean and median Current HUD Funding levels than the other two periods. Current HUD Funding (Standardized) is found to be a significant predictor of GM blood lead levels only in Model 3, while it is not significant in the other four models.

Cumulative HUD Funding ($ per Child) – The relationship between Cumulative HUD Funding levels and predicted GM blood lead levels is flat in each of the three time periods. The distribution of Cumulative HUD Funding levels increases over the three time periods as expected given the nature of this variable. Cumulative HUD Funding (Standardized) is significant in Models 1 to 4, it was not significant in Model 5.

Current State Funding ($ per Child) – The relationship between Current State Funding levels and predicted GM blood lead levels is consistent and positive across the three time periods, with increases in state funding associated with slightly higher blood lead levels. The distribution of Current State Funding is identical across the three time periods, as all Census tracts are included in each time period. Current State Funding (Standardized) is a significant predictor of children’s blood lead levels in all models.

Cumulative State Funding ($ per Child) – The relationship between Cumulative State Funding levels and predicted GM blood lead levels is consistent across the three time periods, with increases in state funding associated with slightly higher blood lead levels, although there appears to be a slight flattening in that relationship with each successive time period. The distribution of Cumulative State Funding increases over the three time periods as expected given the nature of this variable. Cumulative State Funding (Standardized) is a significant predictor of children’s blood lead levels in all models.

Current CDC Funding ($ per Child) – The distribution of Current CDC Funding differs between the 2000-2002 time period and the other two periods. While the range of Current CDC Funding ($ per Child) across Census tracts in Massachusetts is about $4.50 per child in the 2000-2002 period, the range is only $0.05 and $0.08 in the other two periods. From Figure B51, it appears the relationship between Current CDC Funding levels and predicted GM blood lead levels is slightly positive for the 2000-2002 period; however, the relationship is unable to be determined for the other two periods. Current CDC Funding ($ per Child) is only significant in Model 5, but is not significant in the other four models.

Cumulative CDC Funding ($ per Child) – The relationship between Cumulative CDC Funding ($ per Child) and predicted GM blood lead levels is slightly positive in the 2000-2002 time period, but appears to be nearly flat in the subsequent two periods. The distribution of Cumulative CDC Funding ($ per Child) increases over the three time periods as expected. Cumulative CDC Funding ($ per Child) is significant in Models 3 to 5, but not in Models 1 and 2.

Current Total Funding ($ per Child) – The relationship between Current Total Funding levels and predicted GM blood lead levels is generally flat across the three time periods, although there is a slight positive slope in the middle time period. Mean and median levels of Current Total Funding ($ per Child) across Massachusetts Census tracts decrease from the 2000-2002 period to the later two periods. Current Total Funding ($ per Child) is significant in Models 1, 2, and 5, but Models 3 and 4 failed to converge.

Cumulative Total Funding ($ per Child) – The relationship between Cumulative Total Funding ($ per Child) levels and predicted GM blood lead levels is consistent across the three time periods, with increases in state funding associated with slightly higher blood lead levels, although there appears to be a slight flattening in that relationship in the later two time periods. The distribution of Cumulative Total Funding ($ per Child) increases over the three time periods as expected given the nature of this variable. Cumulative Total Funding ($ per Child) is significant in Models 1, 2, and 3, but Models 4 and 5 failed to converge.

Current HUD Funding ($ per Census Tract) – Lower predicted GM blood lead levels are associated with higher levels of Current HUD Funding ($ per Census Tract) in the 2000-2002 and 2005-2006 time periods, while the middle period displays a mostly flat relationship. The distributions of Current HUD Funding ($ per Census Tract) also change across the time periods. The 2000-2002 time period has higher mean and median Current HUD Funding levels than the other two periods. Current HUD Funding ($ per Census Tract) is significant in Model 5, but not significant in the other four models.

Cumulative HUD Funding ($ per Census Tract) – Generally, slightly lower blood lead levels are associated with higher levels of Cumulative HUD Funding ($ per Census Tract), however, the relationship appear to be flattening with each successive time period. As with all the cumulative variables, the distribution of Cumulative HUD Funding ($ per Census Tract) levels increases over the three time periods. Cumulative HUD Funding ($ per Census Tract) is significant in all models, except Model 4 that failed to converge.

Current State Funding ($ per Census Tract) – The relationship between Current State Funding ($ per Census Tract) levels and predicted GM blood lead levels is mainly consistent across the three time periods, with increases in state funding associated with slightly higher blood lead levels, although the slope of this relationship increases with each time period (because blood lead levels are decreasing over time while the Current State Funding distribution remains constant. The distribution of Current State Funding ($ per Census Tract) is identical across the three time periods, as all Census tracts are included in each time period. Current State Funding ($ per Census Tract) is significant in all models.

Cumulative State Funding ($ per Census Tract) – The relationship between Cumulative State Funding ($ per Census Tract) levels and predicted GM blood lead levels is consistent across the three time periods, with increases in state funding associated with slightly higher blood lead levels, although the slope appears to be more positive with each successive time period. The distribution of Cumulative State Funding ($ per Census Tract) increases over the three time periods as expected given the nature of this variable. Cumulative State Funding ($ per Census Tract) is significant in all models.

Current CDC Funding ($ per Census Tract) – The relationship between Current CDC Funding ($ per Census Tract) levels and predicted GM blood lead levels is mainly consistent across the three time periods, with increases in state funding associated with slightly higher blood lead levels, although the slope of this relationship increases with each time period. The distribution of Current CDC Funding ($ per Census Tract) is nearly identical across the three time periods. Current CDC Funding ($ per Census Tract) is significant in Models 1 and 2, not significant in Models 3 and 5, while Model 4 failed to converge.

Cumulative CDC Funding ($ per Census Tract) – The relationship between Cumulative CDC Funding ($ per Census Tract) levels and predicted GM blood lead levels is consistent across the three time periods, with increases in state funding associated with slightly higher blood lead levels. The distribution of Cumulative CDC Funding ($ per Census Tract) increases over the three time periods as expected given the nature of this variable. Cumulative CDC Funding ($ per Census Tract) is significant in all Models 3 to 5, but is not significant in Models 1 and 2.

Current Total Funding ($ per Census Tract) – The relationship between Current Total Funding ($ per Census Tract) levels and predicted GM blood lead levels is inconsistent across the three time periods. During the 2000-2002 time period, predicted GM blood lead levels decline slightly as Current Total Funding ($ per Census Tract) increases, while the other two time periods have slight positive slopes. The distribution of Current Total Funding ($ per Census Tract) across Massachusetts Census tracts generally decrease from the 2000-2002 period to the later two periods, although the median decreases from $2,491 to $2,183 between the first time periods and returns to $2,389 in the most current period. Current Total Funding ($ per Census Tract) is significant only in Model 3, and is not significant in the other four models.

Cumulative Total Funding ($ per Census Tract) – The relationship between Cumulative Total Funding ($ per Census Tract) levels and predicted GM blood lead levels is slightly different across the three time periods, with a flat relationship in 2000-2002 but becoming slightly more positive in the successive time periods. The distribution of Cumulative Total Funding ($ per Census Tract) increases over the three time periods as expected given the nature of this variable. Cumulative Total Funding ($ per Census Tract) is significant in all models.

Analyses of these programmatic funding variables provided surprising results for Massachusetts, although potential explanations can be suggested. For the State funding variables, higher amounts of funding do not appear to be associated with lower blood lead levels. A few of the HUD funding variables did indicate some association between increased funds and lower blood lead levels, but for some time periods there does not appear to be a relationship between them. The total funding analyses show mainly flat to positive relationships, perhaps caused by the state funding data. When increased funding is associated with higher lead levels, a potential explanation is that funds are being successfully targeted at areas that have worse lead poisoning problems.

Table 4-2 reports that Current State Funding ($ per Child) provided the lowest log-likelihood statistics for Models 1 to 3. Note that as levels of this variable increase, predicted GM blood lead levels increase consistently across each time period (see Figure B49). Cumulative CDC Funding ($ per Census tract) provided the best model fits for Models 4 and 5.

E.3.2 Housing Inspection Variables

As discussed in Section 3.4.3, Massachusetts supplied a data set of housing inspections conducted within the state. The detailed exploratory analysis results for the 12 housing inspection variables are presented in Figures/Tables B75 to B86. Results for each variable are discussed below.

P1: Proportion of Housing Units Passing Massachusetts Standard of Care, Naïve Method 1 – The relationship between P1 variable levels and predicted GM blood lead levels is consistently positive across the three time periods, as higher blood lead levels are associated with a higher percentage of units passing the standard of care. The distribution of the P1 variable is similar across the three time periods, although the mean and median are one percentage point less in the 2000-2002 time period. P1 is significant in all models.

F1: Proportion of Housing Units Failing Massachusetts Standard of Care, Naïve Method 1 – Across all three time periods, predicted GM blood lead levels increase as levels of the F1 variable increase. The distribution of the F1 variable is similar across the three time periods, although the maximum proportion of failing units decreases with each successive time period. F1 is significant in all models.

N1: Proportion of Housing Units Assessed, Naïve Method 1 – The relationship between N1 variable levels and predicted GM blood lead levels is consistently positive across the three time periods, as higher blood lead levels are associated with a higher percentage of units assessed. The distribution of the N1 variable is nearly identical across the three time periods. N1 is significant in all models.

P2: Proportion of Housing Units Passing Massachusetts Standard of Care, Naïve Method 2 – The relationship between P2 variable levels and predicted GM blood lead levels is positive across the three time periods, as higher blood lead levels are consistently associated with a higher percentage of units passing the standard of care. The distribution of the P2 variable is similar across the three time periods. P2 is significant in all models.

F2: Proportion of Housing Units Failing Massachusetts Standard of Care, Naïve Method 2 – Across all three time periods, predicted GM blood lead levels increase as levels of the F2 variable increase, with the 2000-2002 appearing to have a slightly more positive slope than the other two periods. The distribution of the F2 variable is similar across the three time periods, although the maximum proportion of failing units decreases with each successive time period. F2 is significant in all models.

N2: Proportion of Housing Units Assessed, Naïve Method 2 – The relationship between N2 variable levels and predicted GM blood lead levels is consistently positive across the three time periods, as higher blood lead levels are associated with a higher percentage of units assessed. The slopes appear to be a bit less positive with each successive time period. Mean and median levels of the N2 variable increase across the three time periods as more units are assessed each year. N2 is significant in all models.

P3: Proportion of Housing Units Passing Massachusetts Standard of Care, Naïve Method 3 – The relationship between P3 variable levels and predicted GM blood lead levels is positive across the three time periods, as higher blood lead levels are consistently associated with a higher percentage of units passing the standard of care, although the slopes appear to be declining over time. The distribution of the P3 variable is similar across the three time periods, although the mean and median increase by a percentage point in the 2005-2006 period. P3 is significant in all models.

F3: Proportion of Housing Units Failing Massachusetts Standard of Care, Naïve Method 3 – Across all three time periods, predicted GM blood lead levels increase as levels of the F3 variable increase, with slopes appearing to be consistent across time. The distribution of the F3 variable is similar across the three time periods, although the maximum proportion of failing units decreases with each successive time period. As with all the previous housing inspection variables, F3 is significant in all models.

N3: Proportion of Housing Units Assessed, Naïve Method 3 – The relationship between N3 variable levels and predicted GM blood lead levels is positive across the three time periods, as higher blood lead levels are consistently associated with a higher percentage of units passing the standard of care, although the slopes appear to be declining over time. The distribution of the N3 variable increases across the three time periods as more units are assessed each year. N3 is significant in all models.

P4: Proportion of Housing Units Passing Massachusetts Standard of Care, MDPH Method – The relationship between P4 variable levels and predicted GM blood lead levels is positive across the three time periods, as higher blood lead levels are consistently associated with a higher percentage of units passing the standard of care, although the slopes appear to be declining over time. Mean and median levels of the P4 variable increase across the three time periods as higher percentages of units pass the MA standard of care. P4 is significant in all models.

F4: Proportion of Housing Units Failing Massachusetts Standard of Care, MDPH Method – Across all three time periods, predicted GM blood lead levels increase as levels of the F4 variable (% failing) increase, with slopes appearing to be consistent across time. Mean, median, and maximum failing rates decrease with each successive time period. F3 is also significant in all models.

N4: Proportion of Housing Units Assessed, MDPH Method – The relationship between N4 variable levels and predicted GM blood lead levels is positive across the three time periods, as higher blood lead levels are consistently associated with a higher percentage of units passing the standard of care, although the slopes appear to be declining over time. The distribution of the N4 variable increases slightly across the three time periods as the mean, median, and maximum go up. N4 is significant in all models, except Model 4 that failed to converge.

In summary, positive relationships are evident between all the housing inspection variables and predicted GM blood lead levels – passing rates, failure rates, and assessment rates. All variables were significant for predicting blood lead levels across all models, except the N4 variable for Model 4. As reported in Table 4-2, F4 provides the best fit for Models 1 and 2, although the F1 and F3 variables were very close behind. This reversed for Model 3 with F1 and F3 providing the best fit and F4 close behind. The P2 and P3 variables provided the best fit for Models 4 and 5. Note that P2 and P3 yielded identical fit statistics for all models, as did F1 and F3.