Response to Internal Review of Target Lead Project (as of May 2, 2007)

Reviewer	Comment Number	Page Number
Michael Messner	1 – 25	1 – 5
John Schwemberger	26 – 38	6 – 11
Region 5	39 – 98	11 – 32
Will Smith	114 – 115	49 - 50
Ronnie Levin	116 – 120	50 - 51
Jade L. Freeman	121	52 – 53
Barry Nussbaum/ Cheng Ling	122 – 128	53 - 59

Please note that comments 99-113 received from Brody/Region 5 were completely duplicative of comments 39-98 received from Region 5 (and eliminated from this table).

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
1	Michael Messner (M. M)	1 a) Is the design adequately developed, well integrated, well reasoned, and appropriate to the aims of the project? Why or why not	Design	Yes	No response necessary.
2	M. M.	1 b) Are the methods adequately developed, well integrated, well reasoned, and appropriate to the aims of the project? Why or why not?	Methods	Yes	No response necessary.
3	M. M	1 c) Are the analyses adequately developed, well integrated, well reasoned, and appropriate to the aims of the project? Why or why not?	Analyses	Yes	No response necessary.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
4	M. M	1 d) Are potential problem areas acknowledged?	Problem areas	Yes	No response necessary.
5	M. M	1 e) Are appropriate resolutions suggested?	Resolutions	For the most part. Some additional work is needed to better understand what is shown in Chapter 5's plots of predicted vs. measured values.	We have added some additional information regarding the observed vs. predicted plots as a response to this comment.
6	M. M	2) Assess the appropriateness and relevance of the model selection process and the best fit model. Describe any issues with the models chosen, and if other alternative methodologies would have better met the objectives.	Model Selection Process	Choice of models seems well-reasoned and appropriate.	No response necessary.
7	M. M	3) Do the results give EPA and CDC the information they need to target resources to address problem areas? What information is still needed to select factors used for targeting? What do you suggest for further study to fill any remaining gaps in the information necessary to address problem areas?	Problem Areas and Suggestion for further study	It isn't clear how much of the predictive power of the models is due to county-specific parameters. The models may not be able to predict levels in areas which have no historical measurement data.	We have added some additional information regarding the log-likelihood from the following three types of models: (1) Full multivariate predictive model, including both fixed and random effects (2) Partial multivariate predictive model including fixed effects in Full Model but no random effects (3) Partial predictive model including random effects and regional parameters
8	M. M	4)What modifications would you suggest to the recommendations? What other direction of investigation would you suggest?	Modifications	Investigate systematic prediction error issues (as shown in plots of predicted vs. observed values).	Additional information is provided in the report based on this comment.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
9	M. M	5)Describe the extent to which this pilot project has developed a novel approach, methodology, tool for this environmental health issue.	Innovation	Models selected are appropriate to the problem at hand. Their choice seems natural (rather than novel).	No response necessary.
10	M. M	6) Are the models in the study usable for all geographic areas? How can they be extrapolated for use in areas with little or no data? Are there applications for this methodology to target other environmental health effects? If so, identify applications.	Applications	I think this is the big issue that hasn't been addressed in the report. I'm especially concerned about the large number of county-specific parameters and wonder whether they speak much more "loudly" than other parameters (the ones that get all the attention in the report) in predicting blood lead levels.	The additional information added, as described in the response to comment #7, should address some of these concerns
11	M. M	7)Based on your reading and analysis of the information provided, please identify your overall recommendation using one of the following statements;	Recommendations	acceptable with minor revision (as indicated); (Need to address area-specific parameters. See detailed comments/questions below.)	No response necessary.
12	M. M	Questions (1)		Bottom p6 The bullet mentions "geometric mean." Geometric mean of what, blood lead concentration?	Editorial comment – changes made to report
13	M. M	Questions (2)		Pg7, para 1 Is the "intercept term" beta0?	Editorial comment – changes made to report
14	M. M	Questions (3)		Pg7, matrices Shouldn't the subscripts be 00, 01, 10, and 11, so they relate to the subscripts for delta?	Editorial comment – changes made to report

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
15	M. M	Questions (4)		Pg7, GLMs Does the likelihood function therefore include many beta densities, with parameters E(Yij/nij)? We can expect the likelihoods to be quite different from those of the mixed model, which has normal error structure.	No – the model is fit within the context of Generalized Linear Models (McCullaugh & Nelder). Please see discussion in Apllied Longitudinal Analysis (Fitzmaurice, Laird and Ware) for a more detailed discussion of generalize linear mixed models.
16	M. M	Questions (5)		Pg8, imputation What does "imputed using the mean of the distribution" mean? Does this simply mean that the mean was used wherever a predictor variable was missing?	Yes – we used a very simple imputation method when necessary (which was a very rare occurrence, as shown in the nmiss column of the univariate summary statistics of Appendix A).
17	M. M	Questions (6)		Pg9 This is also a binomial model, right?	Yes – additional information is provided in the revised report to clarify the model.
18	M. M	Questions (7)		Pg29 Can information be added to help the reader interpret Table 4-1? Can likelihoods be compared across categories, so that, for example, one might conclude that the best single-variable to use in Model 1 is Housing Age / Year Built (has smallest number in the column)? Can likelihoods be compared across columns? Why not? Is there an adjustment one can make (such as adding penalty for effective number of parameters) so that the numbers are comparable? Do the much lower numbers under Model 1 indicate anything whatsoever about the superiority of that model over the others?	Additional information was integrated into the discussion of Table 4-1 (which was replaced by a series of tables). The log-likelihoods are comparable across rows – but not columns in the original Table 4-1, assuming that the df for the variables are the same. An adjustment can be made for differences in df. The likelihood statistics are not comparable across columns – because they relate to a different response vector.
19	M. M	Questions (8)		Pg33 Same questions as for Pg29.	Please see response to comment #18 above.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
20	M. M	Questions (9)		P37 "Tables 5-2 through 5-7 present the parameter estimates..." How were the estimates obtained? Maximum likelihood? Minimum variance?	The mixed models analysis of variance models for continuous data were fit using SAS Proc MIXED under a restricted maximum likelihood approach. The generalized linear mixed models logistic regression for binomial responses were fit using SAS Proc GLIMMIX, under a maximum likelihood approach with ridged Newton-Raphson optimization.
21	M. M	Questions (10)		P42 The error variance appears to be large by factor of 100. Has the decimal been misplaced?	No – this is an artifact of the weighting by the number of children screened.
22	M. M	Questions (11)		Fig 5-6, 5-8, 5-10 & 5-12 - Zeros are shown in these figures. Was there some nonzero reporting limit / limit of detection and, if so, were these properly treated as censored values? Does the likelihood function include these as probability masses (pr{censored})?	No – CDC did not provide information regarding limits of detection or reporting limits on the original data used for this project. Due to shared concerns about undocumented variability among programs that reported the data and laboratories that performed the tests, we chose to remove the National models for continuous data (GMs) and focus solely on the more robust logistic regression models.
23	M. M	Questions (12)		Pg62 Is there a censoring issue here? The plots look awful.	We address this concern in the revised report.
24	M. M	Questions (13)		Figures A. Why do Model 1 and Model 2 error variances not vary from table to table? Shouldn't p-value monotonically increase with -2 Log Likelihood (across tables for same model number)?	Model 1 is based on a Mixed Models ANOVA Approach where all county/quarter combinations are weighted equally in the analysis; whereas Model 2 utilizes a similar modeling approach in which the responses are weighted by the number of children tested in each county/quarter. The likelihood statistics (as well as the p-values associated with the explanatory variables) are affected by the difference in weighting.
25	M. M	Comments		Not nearly enough is said about the importance of county-specific parameters. I wouldn't be surprised if these are far-and-away the most important parameters. Please consider ways to explore this and how to communicate to readers. How do prediction intervals differ between counties with data (and therefore information on county median levels and trends) and counties without data? What other parameters would be important if no county-specific parameters are included? How well does a model with ONLY county-specific parameters perform? Consider including log-	We agree with this comment. Additional information was added to the report, as described in the response to comment #7, should address some of these concerns

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				likelihoods for the Model 1 through 6 versions of this in the summary tables.	
26	John Schwemberger (J.S.)	1 a) Is the design adequately developed, well integrated, well reasoned, and appropriate to the aims of the project? Why or why not	Design	No. There is no section on design. Add a section to the report to explain where the surveillance data for the project were collected, from whom the data was collected, and how that data were collected.	Some of this information was already included in the report from a more global perspective. Much of the detailed information is not available (and would require substantive resources and contributions from CDC to fully document).
27	J. S.	1 b) Are the methods adequately developed, well integrated, well reasoned, and appropriate to the aims of the project? Why or why not?	Methods	In general, yes. However, there is at least an appearance of using data that is readily available for the models. Two areas for improvement: 1) add a list of variables that logically are candidates for a model, regardless of whether or not readily available, and 2) address whether the selection process could result in variables being deemed significant because a large number of variables were considered and tested.	All information sources that were uniformly available across the geographic areas represented in the project were integrated into the report. Other information sources (e.g. soil lead concentrations) would have been integrated if available across the Nation. The variable selection process (among those information sources available) is documented in the revised report.
28	J. S.	1 c) Are the analyses adequately developed, well integrated, well reasoned, and appropriate to the aims of the project? Why or why not?	Analyses	In general, yes, but the report should address how the surveillance data can be used to inform decisions for the best ways to meet the 2010 goal of eliminating childhood lead poisoning. The more difficult goal of trying to expand the surveillance data to a characterization lead levels in the general population would best be described as a more distance goal, and the report does address the steps that need considered and undertaken to achieve that goal. The report should include a quality assurance step in which estimates from the models are compared to a known independent value. For example, if the report could establish that the models predict that about 300,000 children aged 1-5 years old had elevated blood lead levels in the year 2000, that step would provide good evidence that the models are on target nationally. One area where I would like to see more	Additional work was conducted to ensure that the models would be able to inform these decisions. In particular – we altered the methodology for allowing the effects of specific predictor variables to vary over time by including linear and quadratic (over time) interaction terms, rather than the previously used categorical variable – which was not as informative for forecasting towards 2010. The report does include a comparison to NHANES data – however (as included in the discussion section) – the surveillance data is not Nationally Representative of all children (as NHANES). Rather – it is representative of children that were tested by lead poisoning prevention programs funded by CDC (and therefore may represent higher risk children).

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				information is the performance of the models over time compared to NHANES statistics. Expanding Section 7.2 to cover other measures such as geometric mean blood lead level and 30percentage and number of EBLL children would be a good addition, and would provide more information on the differences between surveillance data and general population data from NHANES.	
29	J. S.	1 d) Are potential problem areas acknowledged?	Problems	Yes, these are covered in Chapter 7.	No response is necessary.
30	J. S.	1 e) Are appropriate resolutions suggested?	Resolutions	Yes, although mostly from the perspective of developing a model for the general population. Add more from the perspective of using the information from the surveillance data to make decisions to meet the 2010 goal of eliminating childhood lead poisoning.	No response is necessary.
31	J. S.	2) Assess the appropriateness and relevance of the model selection process and the best fit model. Describe any issues with the models chosen, and if other alternative methodologies would have better met the objectives.	Model Selection Process	The report could be improved by developing the list of variables the authors logically believe should be in the models, and then proceeding to what is available, and then how what is available can be used. As necessary, add a summary of major lead models in the report to provide background on what variables other researchers have used for models.	We have added some information regarding certain variables that would have been considered if available.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
32	J. S.	3) Do the results give EPA and CDC the information they need to target resources to address problem areas? What information is still needed to select factors used for targeting? What do you suggest for further study to fill any remaining gaps in the information necessary to address problem areas?	Problem Areas and Suggestion for further study	The use of percentages in the models as a response variable has both upsides and downsides. A percentage is a way of comparing that adjusts for differences in size. A percentage may mask other events that are taking place, such as 1) increases in the number of children being screened in a state or local program, 2) increases in population in an area over time, and 3) the overall number of children that live in a county. So when county lists are being presented, and possibly even for census tracks, lists by the estimated number of children in a category should be presented in addition to the percentage of children in a category to provide perspective. The number of children is a statistic that is germane for surveillance programs and corresponds to the goal of eliminating childhood lead poisoning by the year 2010. Geometric means and percentages are good measures of surveillance program efficacy, but both can be misleading if presented as representing county wide lead levels. It would be preferable to focus on numbers of children that have an EBLL or other higher level for this effort. In addition, the number or percentage of children screened is a potential model variable of interest, possibly on a lagged basis. Increases in screening rates should result fewer children with high blood lead levels in the following years. The age range for children in the national model is 6 months to 36 months. CDC's main statistics for monitoring progress have been based on children aged 1-5 years old in NHANES, which appears to translate to 12 to 71 months old. There should be more discussion of the reason for choosing 6 to 36 months for the age range in the national models. It would appear 12 to 71 months would be preferable. Finally, in large population counties, further breaking down the county into census tracks would be useful as permitted by the data, to more precisely identify the areas that need attention.	The logistic regression models are appropriately weighted based on the number of children screened in each county/quarter – however, they do not adjust for the number of children that are in the county (unless this variable is specifically included in the analysis). Unfortunately, due to time and resource constraints (as well as data availability) – we were not able to adjust for how population size changes over time based on census projections, etc. Focusing solely on the number of children found with EBL would serve a different purpose – and could lead to different results. Based on this comment, we included an additional covariate in the report that represents the proportion of children screened in each county/quarter combination. This variable penetrated several of the multivariate models. The age range of 6-36 months was chosen for this project, because this is the age-range of heightened vulnerability. Additionally, there are many screening programs throughout the country that mandate testing in this age range, while relaxing testing requirements for older children. We agree with the goal of enhancing the geographic specificity of the model – however, the blood-lead response information is only available at the county level from the CDC. Some state/local programs do have their surveillance data geocoded to the census tract level (e.g. Massachusetts) – and as shown in this work, models can be developed that make use of data on this scale.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
33	J. S.	4)What modifications would you suggest to the recommendations? What other direction of investigation would you suggest?	Modifications	The dominance of Maine and Iowa in some of the lists in Appendix F and some of the maps should be addressed, as well as the increases in EPA Region 1.	We have addressed this concern in discussions with CDC, and have made appropriate revisions to the report.
34	J. S.	5)Describe the extent to which this pilot project has developed a novel approach, methodology, tool for this environmental health issue.	Innovation	At some point, the current NHANES will no longer detect any children with elevated blood lead levels. The CDC blood lead surveillance data will probably replace NHANES as the major blood lead data set of interest. Developing an organized approach to using the surveillance data at this time and possibly using the surveillance data to meet the 2010 goal is innovative. At a more detailed level, inclusion of HUD and CDC funding as model variables and inclusion of a variable for the Massachusetts standard of care for housing are noteworthy. Assessing whether funding makes a difference is an important question, and having a potential variable which can be changed is valuable. For the Massachusetts models, developing a variable to represent environmental lead hazards in a home is an excellent addition.	No response is necessary.
35	J. S.	6a)Are the models in the study usable for all geographic areas?	Application	A purist would say no. The data on which the models were based were not collected in all geographic areas. The data that were collected were probably focused on finding children at risk from exposure to lead-based paint. The models are probably most usable in the Northeast and Midwest parts of the country, where lead-based paint is more likely to be found. Nevertheless, the models can be presented and caveated as estimates based on the data that were collected. The report needs to do a better job of characterizing the surveillance data that was used for this project.	Please see response to question number 26.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
36	J. S.	6b)How can they be extrapolated for use in areas with little or no data?	Application	This is always difficult if not impossible. If the models largely consisted of variables that have plausible cause and effect explanations for their inclusion, then extrapolation might be defensible. Even so, the coefficients in the models would be based on data from other parts of the country, and so the relative importance of variables in the model might be skewed to areas where data were collected. If the current models are shown to predict declines in blood lead levels in the areas where data was collected, that could be used as a modest rationale for extrapolating to other areas, although this approach would still have the caveat that the model results were based on the areas where the data were collected. For the current report, the predictions in areas where no data was collected could be characterized as the best estimate of blood leads that would be found by a surveillance program if surveillance programs were initiated in those areas. The estimates will come with caveats until more data are collected.	We have attempted to capture these points as appropriate in the discussion section. This is also related to comment #7.
37	J. S.	6c)Are there applications for this methodology to target other environmental health effects? If so, identify applications.	Application	I don't know of any area that has as much information as lead.	No response necessary.
38	J. S.	7)Based on your reading and analysis of the information provided, please identify your overall recommendation using one of the following statements;	Recommend ations	X acceptable with major revision (as outlined);	No response necessary.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
39	Region 5	Over-arching concerns or questions (1)		Overall, the study needs to have increased transparency, especially as to its underlying assumptions and potential utility. Some of the questions we're asking here may resolve if the report is made more transparent. The study seems to have multiple intended audiences (e.g. federal, state, local agencies) at multiple geographic levels, but potential uses need to be made much more apparent to state and local childhood lead poisoning prevention programs (CLPPPs) that are already extremely busy. CLPPPs need to see this as something they can use to advance their efforts. The Region 5 CLPPPs already have identified high risk zip codes and track testing and EBLL rates for those zip codes, usually within a GIS system. The executive summary and the text should be clarified as to variables included in the "final national model." The use of the term "final national model" as refers to 6 individual models (e.g. Table 5.1) is confusing. Also, the predicted variable should be included in Table 5.1 as text refers to Model 1 and 2 modeling GM, while Models 3 through 6 model proportion above some level of concern (e.g. 5, 10, 15, 25 ug/dL).	We believe that the report is extremely transparent in terms of describing the data sources, methods, results, and potential caveats. The study does have multiple audiences – and perhaps streamlined versions could be produced with more targeted audiences (perhaps in the form of a manuscript for publication in a peer-review journal). We have attempted to clarify the executive summary as appropriate.
40	Region 5	Over-arching concerns or questions (2)		The 3-D graphical presentation of results is both innovative and high-tech, but it can also be disorienting to the uninitiated, and does not seem intuitive. Please include more discussion of how this tool might be used as well as a better guide to understanding how to interpret the visualizations. Adding clearer reference points on the visualization may assist in this, but for the visualization tool to be effective, more people will have to understand how to use and interpret it. Please consider developing training materials for the use of the visualization tool.	Based on resource and time constraints, updates to the visualization tool were not yet pursued in the immediate follow-up work intended to revise the report. We will take these comments into consideration if/when such revisions are conducted as part of this project.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
41	Region 5	Over-arching concerns or questions (3)		We appreciate the fact that the national visualization tool restricts the user to the state level as these data are often more standardized at this level. It would be helpful if your movie could just show the data without the modeling results as one option. We are also concerned that the interpolation between county centroids may be misinterpreted as part of the modeled data.	Based on resource and time constraints, updates to the visualization tool were not yet pursued in the immediate follow-up work intended to revise the report. We will take these comments into consideration if/when such revisions are conducted as part of this project.
42	Region 5	Over-arching concerns or questions (4)		R5LP is not convinced that use of the Geometric Mean (GM) is appropriate for this type of model. Our rationale for this is that the GM may be influenced by analytical methods and data reporting policies (both by individual laboratories as well as by states). In addition, an inherent inaccuracy in laboratory testing of low blood lead levels exists, which has been cited by CDC as one of the reasons for not lowering the level of concern to less than 10ug/dL. If all the blood lead analyses had been conducted by a single laboratory, such as occurs in a national survey study, like NHANES, the use of GM might be warranted. One of the stated purposes of using the model(s) developed through this study is for resource allocation. When GM is used as the predicted variable, there is an implicit assumption that the higher the GM, the greater the need for lead poisoning prevention resources. Given some of the inherent concerns with use of the GM, this may not be true. Because of this, there needs to be a discussion about whether it might be more useful for the predicted independent variable to be the **number** of children with EBLL (different models could be run at different thresholds as was done for the proportion of children with EBLL) as opposed to the GM. Resources should first be directed at localities that have the greatest numbers of children with elevated blood lead levels. Additional questions we have concerning the GM, which may negatively impact their use in the type of models developed here are: a. How were values of “not	Due to shared concerns about undocumented variability among programs that reported the data and laboratories that performed the tests, we chose to remove the National models for continuous data (GMs) and focus solely on the more robust logistic regression models.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				detected" handled in the surveillance data; were they assigned a value, and if so, what value? If they were assigned a value, was this done by the reporting laboratory, the state or for the purpose of this model development? How these values were handled could have a profound impact on the GM observed. b. Was there any attempt to control for potential laboratory bias that might exist in blood lead testing, especially when there is only one major provider for a county or state? While this would not have much impact on EBLL rates, it could potentially result in bias reflected in the GM.	
43	Region 5	Over-arching concerns or questions (5)		How will model(s) be validated to assess prediction value for locations where there are few observed surveillance data? Was there any attempt to validate the models in locations where sufficient data of good quality exist? The report indicates "that models developed as part of this project may require validation" but we feel it would be inappropriate to use or release this report without validation.	Validation of the models is an important area of future work – but is not able to be conducted as part of the immediate follow-on for revising this report. The report still includes a discussion on this topic.
44	Region 5	Over-arching concerns or questions (6)		Due to the complexity of the modeling, and the complex issues surrounding lead testing, analysis and data reporting, R5LP feels strongly that the model(s) development would benefit from peer input, bringing together outside practitioners and academicians to better hone the study. Subsequently, the study should undergo external peer review before public release.	The research team intends on pursuing one or more publications in peer-reviewed journals following completion of this report – which should serve the suggested purpose of external peer review of this work.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
45	Region 5	Over-arching concerns or questions (7)		Some specialized characters are not consistently displayed in Adobe Acrobat ™ documents, especially in equations; blank boxes are shown instead. For example, see the discussion about excluding certain county-quarters on page 14. As a result, it is difficult to critically evaluate the methodology described without guessing what characters are missing.	We will attempt to resolve these problems with pdf conversion in the next iteration of this report.
46	Region 5	Over-arching concerns or questions (8)		Logistic regression is fully appropriate, given the general goal of the study to predict the risk of childhood lead poisoning, but with the caveat that the LOGIT function used yields *odds ratios*. Odds ratios can approximate relative risks, especially when the numbers are large, but are a measure of association and do not imply cause or effect. This proviso should be explicitly stated in the introduction to the models (p. 9). As a result, this modeling approach is *informative, but not predictive*, especially with respect to policy decisions.	The authors believe that the properties of logistic regression modeling is well understood by the intended audience(s) and have been careful to exclude any reference to relative risk in the discussion of the models. This will be continued in the report revision.
47	Region 5	Over-arching concerns or questions (9)		There is concern that this modeling approach might be at risk for the "ecological fallacy," in which a well-documented statistical association between the explanatory and the response variables is identified, but the underlying relationship doesn't exist. For example over the same period (1995-2005), air emissions of trichloroethylene dropped in a monotonic fashion with the number of children with elevated blood-lead levels, but the two phenomena are not related.	We appreciate the concerns regarding the ecological fallacy – which is a potential problem with respect to this type of modeling approach. We guarded against this potential problem by only investigating those available information sources that we thought could explain of children's blood-lead concentrations in a multivariate statistical model. The goal was for prediction – and not necessarily to interpret the relationships as causal.
48	Region 5	Over-arching concerns or questions (10)		The census data for one year (2000) represent the fixed effects in these models. They vary of course, from county to county, but are fixed in time. However, the blood-lead data (response variable) varies over time. County-specific trends, if any, are random effects variables on the independent side of the calculations. Might this be introducing a circular argument? Otherwise, the major covariance is geographic.	We agree that the model predictions could potentially improve if we had the ability to allow the census predictor variables to vary over time (rather than including them as a static time-invariant set of predictors). That being said, we believe that by allowing each county to have their own random intercept and slope over time, that we likely capture any systematic changes in the demographics that would influence blood-lead trends within these parameters.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
49	Region 5	Over-arching concerns or questions (11)		In the Major Findings you should use such language as "This pilot study has created screening level models intended to help predict the burden of childhood lead poisoning." rather than the more definitive "They can be used to predict the risk of childhood lead poisoning…" Try to use less definitive language throughout the document.	We believe that the term risk is the appropriate term. Burden can be interpreted in many different ways – including financial, loss of human potential, effects on school systems, etc. We believe that the terms risk of childhood lead poisoning or risk of elevated blood-lead concentrations is more consistent with the goals of the study.
50	Region 5	Less global comments/question s (1)		The project report should include an expanded discussion on the differences in reporting of lead data, with regard to: a, Potential variability and differences in methods that laboratories use to analyze samples and in policies affecting data reporting (including discussion of potential biases that could result from incorporating both venous and capillary samples and any differences in precision in assessing samples with low BLLs) b, Differences arising from reporting agency policies and practices, i.e. universal reporting, or reporting of BLLs above a locally defined threshold, and how this has changed over time.	Please see the response to question #26.
51	Region 5	Less global comments/question s (2)		Define the use of the word "screening" as blood lead testing if that's what in fact it means in the context of this study. For our partners / stakeholders from CLPPPs, screening is generally a series of questions to identify whether a child is at high risk and therefore in need of a blood lead test. The terms "universal" screening, testing and reporting should be defined.	We have tried to make editorial changes to the document consistent with this comment.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
52	Region 5	Less global comments/questions (3)		Consider incorporating a time-lag approach when the HUD funding variable is incorporated into model(s) (e.g. 2000 funding received by the grantee in 2001 and used in 2002-2003, resulting in some measurable changes in 2004).	In the report revision, we have incorporated several time-lagged funding variables as potential predictor variables based on this comment.
53	Region 5	Less global comments/questions (4)		What has been done to assess the potential data biases that might arise with respect to population density, sample size and aggregation by county? Data estimates may be biased because in non-dense rural areas, testing may only be done in children suspected of having EBLL. Has this practice changed over time, and if so, how, and how might it affect data inputs? Is it useful to aggregate by county in instances where there is great variability within the county? Might there be a situation where the county mean is fairly low, but hides important hot spots with high population density? Would it be possible to do some sensitivity analyses in counties where this might be suspected to be the case?	Given the time and resource constraints of this project, we were unable to specifically investigate the types of biases described in this comment. The data available for this study from CDC (and Massachusetts) do not lend themselves easily to investigate these potential biases. Some of these questions could be addressed through focused prevalence studies as described in the discussion section – and could be pursued by EPA or CDC in the future.
54	Region 5	Less global comments/question s (5)		How were any time of test and time of report discrepancies addressed in the model? Sample analysis and reporting times have improved greatly in the last 10 years and if this potential error wasn't addressed in the model, it could have an impact on the quarterly summaries and seasonality.	The summary statistics generated by the CDC and Massachusetts are based on the date that the sample was collected from the child – and not the date of chemical analysis. Therefore, we believe that this is a non-issue for this project.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
55	Region 5	Less global comments/questions (6)		How were potentially collinear variables addressed in the model, especially some of the income / poverty variables and family structure (e.g. single parent family) variables?	We performed some limited collinearity diagnostics prior to developing the multivariate models to understand which candidate predictors were likely to strongly compete for explanatory power. This included review of the correlation matrix of selected predictors, as well as a principal components analysis of a reduced X-matrix (many of the components of X were repeated over time). This helped guide our backwards elimination of variables from the multivariate statistical models.
56	Region 5	Less global comments/questions (7)		Some of the graphs in the appendices need to have units identified (e.g. pounds of lead emitted in A-55).	Due to limitations on project resources, we were unable to address this comment as part of the current work.
57	Region 5	Less global comments/questions (8)		Might there be a concern that the census data variable used is the percentage of children under 6 years of age, while the CDC data is limited to those between 0.5 and 3 years of age.	We are not concerned about this – as we expect that the fraction represented by the 6-36 month age group is relatively stable across counties and time. Also – the variable is being included as a potential predictor representing the size of the child population in the county.
58	Region 5	Less global comments/question s (9)		Please say more about the seasonality identified in blood-lead levels from quarter to quarter.	We have added some discussion of the seasonality trends, based on this and other reviewer comments

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
59	Region 5	Less global comments/question s (10)		Was the state of Massachusetts used for local model development, based on availability of data only? They're the only state who can report their EBL rates in number per thousand rather than % since they've done such a good job decreasing EBL rates over the last 30+ years. What precautions have you taken or will you take to insure that the MA model approach isn't biased by its low EBLL rates (relative to the rest of the country) when this model is extended to other states?	The Commonwealth of Massachusetts was selected for several reasons, including (1) a willingness to participate, share data, and work collaboratively with our research team (2) a history of having excellent blood-lead surveillance data that has been used in prior research projects (3) availability of other unique information sources – most notably the housing inspection information. The modeling approach used for Massachusetts would perform similarly in States/Localities with higher observed blood-lead levels.
60	Region 5	Less global comments/questions (11)		How is population size taken into account? There may be a high population density in an urbanized county, but the potential number of affected children is still much lower than in larger urban area. Programs could target many of these areas with lower population and still not achieve the elimination goal because of sheer population differences. If the population size differences are accounted for by use of weighted variables, this should be further explained.	Population size is taken into account in two ways: (1) the logistic regression models and the weighted mixed model account for the number of children tested within each county/quarter response (2) there are covariates that were considered that are indicative of population size and/or density Accounting for differences in population size in using the modeling results is important to consider.
61	Region 5	Less global comments/questions (12)		How have models incorporated State-identified "high risk zip codes" or other State- or local-specific data such as rental housing units where multiple, repeated cases of childhood lead poisoning have been reported over time. These are the housing units that are causing the greatest harm, and posing the greatest risks. They are also targets where enforcement intervention can prospectively reap the greatest benefit. These housing data are similar to the "housing inspection" data described in Section 3.4.3 on page 25, but goes a step beyond determining compliance with the "standard of care." If the model could be expanded to add in local data, as available, the predictive value of the model might be improved. Alternatively, these elements might be able to be used in validation of models.	The National Model could not incorporate this type of information – as it is not universally reported. The complexities of the modeling approach require that the data be available and consistent across all areas – which can make addition of these types of predictors very challenging. The Massachusetts model attempts to adjust housing-based lead hazards via the inspection information (although it does not represent all housing units)

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
62	Region 5	Less global comments/questions (13)		How has local soil lead data been incorporated into models? While soil lead data is not always available, we've seen it be very important in communities with local legacy lead sources (e.g. smelters) in places like East St. Louis. It's often a much more important variable than air lead concentrations which tend to be pretty flat around the country.	Soil-lead information was not universally available in a consistent format – and was therefore not considered for this project.
63	Region 5	Less global comments/questions (14)		Was there any attempt made to assess state / local authority for addressing lead paint hazards in the models? The presence of strong state and / or local laws, with coordinated enforcement activity, might be related to trends in childhood lead poisoning and may mediate the affect of other variables used in the model such as amount of HUD funding. Perhaps this variable might be entered as a dichotomous variable.	This is an excellent comment, and one that we believe merits further investigation in future work.
64	Region 5	Less global comments/questions (15)		Was there any attempt to run the broad coverage models stratified by population size (not just density)? There may be some underlying issues when small, medium, large and very large entities are assessed within the same model. There may be some value in re-running the models stratified by population size.	There was no attempt to develop stratified analyses based on population size. In previous work conducted by the contractor (Battelle), there were attempts to develop separate models for large ciites, smaller cities, suburban areas, and rural areas for a series of models conducted at the census tract level ---- however, it is difficult to classify areas as large as counties in this manner. Instead, we tried to adjust for population size using a covariate in the models (when it added predictive power).
65	Region 5	Less global comments/question s (16)		How will model inputs and model be updated to reflect changing landscape in the elimination of lead poisoning? Given the number of variables, this is ambitious and will be resource-intensive, which limits its practical utility and likelihood of long-term success.	There are no known plans to make regular revisions to this work to reflect changes over time.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
66	Region 5	Less global comments/question s (17)		Were any regional or local models developed in locations where sufficient, good quality data exist? If not, this should be done. These regional or local models should take into consideration data from local assessments of screening rates and prevalence, e.g. Chicago's High Intensity Targeted Screening, which determined very good estimates of true prevalence of EBLL.	This was done in the Commonwealth of Massachusetts as described in the report. As resources permit, our research team would like to pursue additional high resolution models in other areas of the country with sufficient data.
67	Region 5	Less global comments/questions (18)		What precautions were taken to insure that the quarterly county point estimates are adequately representative of their respective data elements? Depending on the data distributions, use of single point estimates might introduce errors into the models. In the low-resolution models, the response variables for any given county (after exclusions have been made) will be a "profile" with a maximum of ten years (1995-2005) of seven quarterly data points, or 280 values. Is this profile sufficient, knowing the variability between counties, especially with respect to testing and reporting protocols over this period? We understand that this is data that we have to work with, but there may be a potential to mischaracterize certain counties in these models, especially if their aggregate testing "profiles" are sparse.	First, we ignored any county/quarter (or tract/quarter) combination that represented fewer than 5 children – due to data reporting requirements and confidentiality concerns at CDC and the Commonwealth of Massachusetts. Secondly, we worked collaboratively with our data sharing partners at CDC and MDPH to determine what data they were willing to report. These partners reported the data at the highest degree of spatial resolution that could be supported by their data sources while protecting confidentiality. The choice of quarterly summary statistics (rather than monthly) was based on CDC's recommendations and requirements.
68	Region 5	Less global comments/questions (19)		It seems pretty clear that the study leads to some unexpected results. The predicted high rates in Iowa and Maine, as well as the relatively low rates in areas of Arkansas, Missouri, and Oklahoma with historical contamination from lead mining, "chat" piles and smelting, should raise some concerns about the study. These data sets need to be investigated, and whatever lessons are learned should be applied to the rest of the data sets. There are also unexpected results for several dependent variables, which are identified below, e.g. comment 35. The unexpected results all need to be discussed, and the last paragraph on page 38 expanded.	The original data supporting the National (Low Resolution) Model have been jointly reviewed by both Battelle and CDC, which led to a decision to eliminate the continuous variable models that focused on geometric mean blood-lead levels. The decision was based on varying detection and reporting limits used by the laboratories and programs that provide information into the CDC National Surveillance Database over time. The more robust measures

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
69	Region 5	Less global comments/questions (20)		It was acknowledged that there might be some quality concerns with the data from Maine, and possibly other New England states as well. It might be helpful to remove Maine and other states with data concerns to see the sensitivity of the model to those data. As mentioned, a State based large resolution model for data rich States may be more appropriate (and helpful for comparison with the national model).	These QA/QC problems have been resolved jointly with CDC.
70	Region 5	Less global comments/question s (21)		Were other, more simplified models assessed? In terms of the ease of using these models in the future, it would be beneficial for the authors to recommend a much more simplified model, perhaps reducing the number of independent variables to 3 or 4, yet still resulting in a predictive model.	More simplified models were not assessed. Our goal was to provide a comprehensive tool that facilitated the best possible prediction across all areas of the country (at the county level). We believe that the report, combined with the visualization tool achieves this goal without necessitating more simplified (and less predictive) models.
71	Region 5	Less global comments/question s (22)		If possible to discuss meaningfully, it would be useful to include more discussion of the instances when air and water routes of exposure have a greater influence on blood lead levels, since this subject seems to be raised by Regional EPA folks around the country.	This is a population based study of the impact of predictor variables on the distribution of blood-lead concentrations across larger geographic areas. Other models that focus on the impact of exposure on individual child blood-lead concentrations might be more appropriate for the types of inferences suggested in this comment.
72	Region 5	Less global comments/question s (23)		How are negative intercepts being handled? What do they mean?	The negative intercepts in the logistic regression models suggest that the probability of exceeding the blood-lead threshold (e.g. 10 μg/dL) is less than 50% at the reference level for categorical variables (and at a level of zero for all continuous variables) included in the model.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
73	Region 5	Less global comments/question s (24)		There needs to be more discussion relating the findings of this study to previous studies of EBLLs. It would be helpful to go beyond just stating the findings; it would be useful to see explanations. As an example, why should percent black alone be significant in every model except Model 5? How does this compare to analysis of existing blood lead levels?	Due to the presence of many predictors variables and potential effects of some degree of colinearity, combined with the goal of developing a highly predictive series of models – we were less concerned about the interpretation of individual slopes in the final multivariate models and how it compares with other research. This is an avenue of discussion that could be added to the work if future resources permit.
74	Region 5	Less global comments/questions (25)		Section 2.3 and 2.4.1 have intertwined the models in their discussion. In some cases, they repeat the description. In some cases, they fall short of the description. Some points: 1. Section 2.3 is not listed in the table of contents. 2. In 2.3, the term geometric mean is referred to as "GM" before it is spelled out in the body of the document. 3. In 2.3, please further describe how Models 2-6 provide screening level weight. 4. In 2.3. please define t_{ij}, for Models 3 through 6. It is defined in 2.4.1. 5. In 2.3, n_{ij} is defined as the number of children screened in each county while it is defined in 2.4.1 as the number of children tested. Although these terms sound similar, they have important differences as described further below. Please choose the correct term. 6. It appears the models have a tendency to smooth the data. For instance, the seasonal differences of the actual data are greater than the modeled values in places like Cook County. Is this effect due to counties with similar predictor variables as Cook have lower levels of EBLL? In any event, the result should be explained in 7.3.4.	We have attempted to address these comments in the revision to the report. For comment #6 in particular – the model may appear to overly smooth the data on the original scale of measure – but please recall that the model is minimizing the deviations on the natural log scale for continuous measures and on the logit-scale for the binomial proportions. If you were to look at the differences on this scale of measure, they would not appear to be overly smoothing the observed data.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
75	Region 5	Less global comments/questions (26)		Page 13, Section 3.1. The discussion of the CDC data needs to be greatly expanded, including a discussion on the reporting requirements have changed over time. This discussion could also go, instead, in Section 4.1.	Please see response to comments #22 and #26.
76	Region 5	Less global comments/question s (27)		It might be helpful if you included a map of the counties where universal screening occurred, maybe with a gradient based on the percent actually tested, with another color indicating locations that were included because they met your decision criteria. That alone would be a helpful piece of information for groups that are trying to see where there are gaps in adequate surveillance.	We agree that this would be useful – however, we do not currently have access to this information.
77	Region 5	Less global comments/question s (28)		In Section 3.1, please thoroughly explain the algorithm by which you removed repeat tests from the quarterly data.	We have reviewed this section in the revised report.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
78	Region 5	Less global comments/question s (29)		It would also help to know if you were in fact able to effectively arrive at an annual prevalence. Region 5 found that Illinois included data from children who after screening had initially tested above elevated levels in prior years, while Wisconsin and Ohio included only new cases in the year. Thus the Illinois data provided the judgmentally sampled annual prevalence of EBLLs while Ohio and Wisconsin provided the judgmentally sampled incidence of EBLLs. Check with CDC to see if they were able to go back and get all the tests from States like Wisconsin and Ohio. If CDC did mix prevalence and incidence information in their data, states with incidence data will have counts that are less than or equal to their prevalence.	We reviewed this comment with CDC, and they believe that the National Surveillance Database includes all reported tests on children from these states (not just the incident cases) – and therefore can be used to represent prevalence.
79	Region 5	Less global comments/question s (30)		It would also be helpful if you could tell people to go to 7.3 for concerns with these data at the end of your paragraph in 3.0.	Not sure what part of 3.0 this comment addresses.
80	Region 5	Less global comments/question s (31)		Page 22. Section 3.3.2 : There is needs to be more of discussion of the limitations of using TRI data. TRI data can be reported using a detailed emission/release estimate, or it can be reported in a broad category. How did the authors address the reports that were categorized? Also, for purposes of modeling, and selecting leading variables for the model, why did they combine the releases from all media, instead of looking at which variables best predicted air, land, and water releases? What does the water by injection variable refer to? Why would we expect a relationship between this variable and EBLLs?	We believe our use of the TRI data is appropriately documented for the purposes of this report.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
81	Region 5	Less global comments/question s (32)		Page 36, Section 5.1 : How was the interaction of EPA Regions and the programs measured?	We did not understand what this comment refers to with respect to the term “interaction”
82	Region 5	Less global comments/question s (33)		Page 37 : the discussion of the models mentions that in different models, different variables were sometimes included or not included. It would be helpful to explain what was it about the model that would yield different variables being drivers.	This is clarified in the report revision.
83	Region 5	Less global comments/question s (34)		With respect to Appendix A Figures A.10 through A.17 (excluding A.13) Given that the data for the population as a whole tends to yield parallel linear relationships by year, there needs to be a discussion why this isn’t the case for this figure.	We discuss this in the main body of the report.
84	Region 5	Less global comments/question s (35)		Figure A.19 : Was the cost of living adjusted within the census data	No.
85	Region 5	Less global comments/questions (36)		Figure A.20: What is the significance of the linear relationships not maintaining a parallel relationship?	It means that the effect of the covariate on geometric mean blood-lead concentrations is varying over time.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
86	Region 5	Less global comments/questions (37)		Figure A.23: There needs to be a discussion on why a significant drop in blood lead levels isn't observed in housing built after 1978.	The figures in Appendix A were developed to allow us to assess the predictive ability of each candidate variable for possible use in the multivariate statistical models. We did not intend on expanding the discussion on each potential variable.
87	Region 5	Less global comments/question s (38)		Figure A.36: It is not clear what the hypotheses behind this correlation. Why should the blood lead levels increase as the number of residents < 6 years old rise? (and it doesn't for 2004-2006)	The figures in Appendix A were developed to allow us to assess the predictive ability of each candidate variable for possible use in the multivariate statistical models. We did not intend on expanding the discussion on each potential variable.
88	Region 5	Less global comments/questions (39)		Figure A.41: Why is there an inversion of the slope for 2004-2006?	The figures in Appendix A were developed to allow us to assess the predictive ability of each candidate variable for possible use in the multivariate statistical models. We did not intend on expanding the discussion on each potential variable.
89	Region 5	Less global comments/questions (40)		Figure A.42: Given that the data for the population as a whole tends to yield parallel linear relationships by year, there needs to be a discussion why this isn't the case for this figure.	The figures in Appendix A were developed to allow us to assess the predictive ability of each candidate variable for possible use in the multivariate statistical models. We did not intend on expanding the discussion on each potential variable.
90	Region 5	Less global comments/question s (41)		Figure A.47 and 48. In real life, the impact of HUD funding should lag behind the award year. This lag needs to be factored into the modeling	We investigated time-lagged funding variables in the report revision.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
91	Region 5	Less global comments/question s (42)		Figure A.49. What do negative numbers on the x axis signify?	The figures in Appendix A were developed to allow us to assess the predictive ability of each candidate variable for possible use in the multivariate statistical models. We did not intend on expanding the discussion on each potential variable.
92	Region 5	Less global comments/question s (43)		Figure A.53. One would expect parallel lines in this graph, because blood lead levels are generally declining over time. Why has the 2004-2006 slope leveled off?	The figures in Appendix A were developed to allow us to assess the predictive ability of each candidate variable for possible use in the multivariate statistical models. We did not intend on expanding the discussion on each potential variable.
93	Region 5	Less global comments/question s (44)		Figure A.55. There needs to be more discussion about this variable. Why should we expect a relationship between this variable and EBLLs?	The figures in Appendix A were developed to allow us to assess the predictive ability of each candidate variable for possible use in the multivariate statistical models. We did not intend on expanding the discussion on each potential variable.
94	Region 5	Less global comments/question s (45)		Figure A.65. The Y axis needs to be better defined and labeled.	The figures in Appendix A were developed to allow us to assess the predictive ability of each candidate variable for possible use in the multivariate statistical models. We did not intend on expanding the discussion on each potential variable.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
95	Region 5	Less global comments/question s (46)		Housing Area variables. D-5: The study states that the cutoff year did not impact results. Since this doesn't correlate with earlier studies, there needs to be more discussion. The discussion in D-8 also needs to address why there might be variability between models.	Appendix D provides an overview of the exploratory results found in this study, with focus maintained on how these variables might be considered for use in the multivariate models.
96	Region 5	Less global comments/question s (47)		Page D-9: There needs to be more discussion about the analysis of the impact of air concentrations. Why are there discrepancies within the models?	Appendix D provides an overview of the exploratory results found in this study, with focus maintained on how these variables might be considered for use in the multivariate models.
97	Region 5	Less global comments/questions (48)		Page D-9: TRI discussion. Aggregating air data by county may not make sense as it may minimize impacts of hot spots of lead contamination.	The TRI data was actually summed within counties based on reports from industrial sources. There was no other practical way to integrate these information sources with the county-level blood-lead data.
98	Region 5	Less global comments/question s (49)		Page D-12: As noted before, a time lag needs to be factored into the modeling for the impact of HUD funding.	We investigated time-lagged funding variables in the report revision.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
114	Will Smith	Peer Review for Target Lead Project (4/17/2007)		I have some concerns about the study 1. over 55% (some 83,484 of 150,768 observations) of the time series data for lead is missing data. I don't think you can fit the models the study proposes with such large missing data. 2. the model uses census 2000 data. There have been significant changes in racial and income composition of counties in the last 10 years. I believe these changes would significantly alter model predictions. I think it would be worthwhile to investigate model sensitivity to these changes. 3. the models have way too many parameters. I think they are over-specified. 4. if you want to predict missing values in the time series for a county then you would just use the observed time series data for that county. I don't see the need for such an elaborate model. 5. what are the prediction errors? how close is the model to the actual observed? I have a feeling the model is not even close. If it does not fit the observed at the county geography then I would not feel comfortable using it to impute missing data.	1. We respectfully disagree. There are many missing values within the dataset (assuming we want to predict for every single time-point and county combination) – however, there is plenty of information upon which to develop these models. 2. We agree that the model predictions could potentially improve if we had the ability to allow the census predictor variables to vary over time (rather than including them as a static time-invariant set of predictors). That being said, we believe that by allowing each county to have their own random intercept and slope over time, that we likely capture any systematic changes in the demographics that would influence blood-lead trends within these parameters. 3. While the addition of random effect intercept and slopes introduces a large number of additional parameters, our analysis of the log-likelihood of three nested versions of the model (see comment #7) suggest that the fixed and random effects of the model are both critical for prediction. 4. The combination of fixed and random effects in the complex model allows us to pursue multiple goals, including prediction in areas with scare data. 5. We have added some additional information regarding model performance based on this comment.
115	Will Smith	Peer Review for Target Lead Project (4/26/2007)		These are my final review comments. I think the report is acceptable with a minor revision. It really needs a section on Model Performance -- see my comment below. I don't know how good the models are because I have not yet received the model error runs, but I don't think it will change my recommendations. **Model performance -- How good is the Model?** The report needs to discuss the models' prediction accuracy in terms of model errors. I would use the 95% conf interval about the predicted value to gauge model error. In this section I would also discuss and give recommendations on when the model performs well and when it does not. The writers of the report need to plot the model against the observed values to get some idea on model	We have added some additional information regarding model performance based on this comment. The methodology suggested here is one that requires additional careful thought – as the accuracy based on the prediction intervals from the model need further scrutiny. For example – in some cases – the observed data might have included 0 EBLS out of 10 observed (in which case, the prediction interval would have no chance at capturing the observed value). We agree that the model predictions could potentially improve if we had the ability to allow the census predictor variables to vary over time (rather than including them as a static time-

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				performance. I would guess that the model performance depends on which model, the blood lead characteristic, the sample size in the unit (county/time), geo location, etc. I'm sure there are cases when the model just cannot give an acceptable prediction -- these need to be determined and pointed out as rules to use for prediction. **Census Demographic Changes** The model uses census 2000 data. There have been significant changes in racial and income composition of counties in the last 10 years. I believe these changes may significantly alter model predictions. I think it would be worthwhile to investigate model sensitivity to these changes.	invariant set of predictors). That being said, we believe that by allowing each county to have their own random intercept and slope over time, that we likely capture any systematic changes in the demographics that would influence blood-lead trends within these parameters. This would be an area to investigate more fully in future work.
116	Ronnie Levin Region 1	Peer Review for Target Lead Project (3/23/2007)	US air lead emission data	The most recent 2 years of US air lead emission data is showing a slight upward trend after years of significant decreases. This is worthy of further investigation.	We agree that this would be worthy of further investigation – especially if we can determine that these increases cause (perhaps subtle) detectable increases in children's blood lead concentrations.
117	Ronnie Levin Region 1	Peer Review for Target Lead Project (3/23/2007)	EBLLs in Manchester NH	Refugee children from Africa living in NH exhibited very EBLLs. Refugee children in NH and other parts of the country may have nutritional deficiencies that cause their bodies to more rapidly absorb lead.	We would welcome your thoughts on how to integrate information on nutrition

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
118	Ronnie Levin Region 1	Peer Review for Target Lead Project (3/23/2007)	AVGAS	Small planes use leaded gas. Over 5,000 small airports in the US put residents in proximity of the airports at risk of EBLLs. Also, being investigated by OAR.	We have discussed the possibility of integrating proximity to airports in as a possible explanatory variable – but were unable to integrate this new information as part of the current project. This is something that the project team would like to pursue if the scope of the project is able to be expanded in the future.
119	Ronnie Levin Region 1	Peer Review for Target Lead Project (3/23/2007)	Seasonality	Strong seasonal patterns persists, with the downward trend, in all areas of the country. Why?	Investigation of the causal factors for the strong seasonal trends (peaks in the summer, etc.) is of high interest to the project team, and represents an area of research to pursue if the scope of the project is able to be expanded in the future.
120	Ronnie Levin Region 1	Peer Review for Target Lead Project (3/23/2007)	Central Tendency	A majority of Levin's comments encourage us to look at the tail of the distribution of EBLLs, and not the GM. Avoid Central Tendency Fallacy.	The geometric mean blood-lead concentration was removed as a response variable from the National (Low Resolution) Model as a result of concerns expressed by the CDC as well as this reviewer.
121	Jade L. Freeman, PhD Memo 4/10/07	Comment from SDWIS/OGWDW	Exclude the drinking water data from Models due to potential biased inference from the analyses Or thoroughly document use	The lead data in SDWIS/FED has limitations. In lead sampling from the drinking water, the number of samples means the number of sampling locations, i.e., the one sample per site, which varies depending on the size of the water system and the results of earlier testing. This means that the 90^{th} percentile of the samples is not the 90^{th} percentile of the repeated measurements from a sampling location, but it is the 90^{th} percentile of the samples from various locations. Samples are collected at locations that are at highest risk of lead contaminations generally because they are served by copper pipes with lead solder installed after 1982, lead pipes, and lead service lines. Additionally, the samples are collected	Because the SDWIS data was such a strong predictor of blood-lead concentrations in the National (Low Resolution) Models, the project team was reluctant to exclude this information from the Model. EPA/OEI worked with the EPA technical staff responsible for the SDWIS data to make sure that the data source was properly explained and caveated in the report.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				during the warmer months for the highest level of lead. Therefore, the 90th percentiles represent the samples from the upper tail of the distribution of the lead level in a county. It may be appropriate to estimate the population 90th percentile from a linear combination the 90th percentiles from random samples from the population. Because the 90th percentiles of the lead samples represent the high risk areas and time frame, the 90th percentiles from these samples from the upper tail of the population will overestimate the 90th percentile of the population distribution. If the analyses show the significant relationship between the 90th percentiles and the blood-lead level based on these data, it can potentially link higher lead level in the drinking water to the blood-lead level in humans. Further, the states are required to report the 90th percentiles only from the medium and large systems to SDWIS/FED since 2002. This means that there is a lack of information on the 90th percentiles of lead levels in a county with small communities such as rural, homeowners associations, or mobile homes areas. These communities may have different characteristics in their drinking water supply and distribution systems from other communities. Prior to 2002, only the 90th percentiles exceeding the action levels were to be reported to SDWIS/FED. It is questionable to analyze the relationship between the health risk and lead in the drinking water when a certain population, i.e., small community residents, is specifically excluded from the data. Therefore, recommend excluding the lead data from SDWIS/FED for this study. If data used for these analyses, document these limitations of the data in a thorough manner in the report to assist readers to make necessary caveat to the interpretation of the results.	

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
122	Barry Nussbaum / Cheng Ling	Introduction		Battelle's analysis, "Targeting Elevated Blood Lead Levels in Children" is an excellent analysis of where to look for children with elevated blood lead levels by geography and by indicator variables. It fills a need for having ancillary information to use resources wisely in the targeting process. I have three concerns to express in general, and then below I have added (in no particular order) specific comments.	No response necessary
123	Barry Nussbaum / Cheng Ling	General Comments		Battelle is quite aware that the data sources hardly comprise a random sample. In fact, they discuss this in section 7.2 where they compare their data with the more random approach of NHANES. This may or may not be a concern. It is true that the results cannot be projected up to give snapshots or estimates of percentage of population with elevated blood lead. But, that was not the purpose of the study. The purpose was to find ways to target the areas of high lead levels. As such, these non-random data are quite useful. However, since protocols differ in different areas on the non-random selection, it is difficult to accurately assess the relative importance of the indicator variables. Since most of the indicators were no surprise, this is probably not a big concern. But, this does affect the influence that missing or incorrect data can have, as discussed in the next paragraph. Noting in particular the states of Maine and Iowa, it is apparent that some missing or possibly incorrect data have an inordinately large effect on the resultant analysis. It is hard to figure out what makes Maine different from its neighboring states, and yet for 2005, it is unclear if it has a low, high, or no influence on neighboring states. The problem is Maine falls in the highest level of geometric blood lead seemingly dragging in New Hampshire, while close-by Vermont and Massachusetts have the lowest. This situation in Iowa seems to be predicited high lead levels which stop sharply at the border. All of this calls for some further investigation as to what is happening in these two states. A similar concern, albeit on the no data side, affects the New York area. There are no data for New York State, and the model tends to estimate quite low levels. In summary, the concern here, for	We worked with the CDC to attempt to resolve many of the data quality issues, including the possible reporting issues in Maine and the missing information for New York State. We believe that the revised report represents improved (yet imperfect) data – and that additional scrutiny of the county/quarter summary data by CDC Program Managers and their state/local grantees would likely result in subsequent improvements for the accuracy of the data. Unfortunately, this additional review was beyond the scope of this project (we tried to make use of existing data sources), and would have been difficult to achieve within the time and resource constraints of the current effort. The comment regarding predictor variables and their direct and indirect effects on lead exposure is well noted. However, we were not trying to develop a model that explains the causal pathway for childhood lead exposure. Rather, we were trying to develop a model that best predicted aggregate summary information over time. We attempted to avoid strong ill effects of collinearity in building the model – but know that many of the variables compete for explanatory power in the final models. This is why we have not made a significant investment in trying to explain the parameter estimates in the final models.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				high, erroneous, or no data is that it has a powerful effect on the model's predictions. The model considers many variables and categorizes them into categories such as income, housing, race, etc. many of which are concomitant variables, meaning they are, of course, related to each other**.** In this regard, I think some of the variables which appear significant are significant solely because they are related to another variable. In looking for variables that can serve as indicators for elevated blood lead this can be a problem since you might possible track the indirect variable, when the more direct correlation is with another indicator variable. As an easy example, if age of housing is important, and many immigrants live in older houses, it is likely that focusing on immigrants would be a very indirect, and perhaps costly, way to get to housing.	
124	Barry Nussbaum / Cheng Ling	Specific Comments		If instead of dividing the data by the four time groups, suppose the regressions actually used time as a variable. In such a case, time would definitely be significant, but would the demographic variables still be significant. I'm not sure how this works since the I think the demographic variables are fixed in time (1990 census data). Considering the race variable, the models seem to differ on their influence. For models 1 and 2, percentage multiple race appears most significant, for models 3 and 4 it is the percentage Asian, yet for model 5 it is either percentage Black(if you look at page 39) or percentage White (if you look at page 30). How can all this be explained? In fact, in model 4, why is percentage Asian population significant? This doesn't seem to make much sense given the relatively small percentage of the population that is Asian. Also, I'm not sure which direction lead will go based on the "independent" variable, Asian. Another concern regarding this is that percentage Asian seems to be the most important "race" variable in Model 4 according to page 39; yet on page A-11, for model 4, percent Asian is not significant. This is also confirmed in the discussion on page D-3. Also on the variable race, Cheng looked at several of the minority groups and tried to plot their distribution of lead, and found nothing very different about any of them. That is, percentage Asian or Hispanic by county showed the median	Based on this comment, we revised our strategy for dealing with predictor variables whose effect on blood-lead concentration varies over time. Rather than using a 4-level categorical variable (which is suboptimal for a number of reasons, including use of the model for predicting into the future), we now look at time-varying effects by using an interaction with time (both linear and quadratic).

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				blood lead of about 2.5 ug/dl whether they come from a high density or low density minority. This leads me to think that there is some confounding co-variate involved, and that it is not Asian, per se, that is significant	
125	Barry Nussbaum / Cheng Ling	Specific Comments (continued)		On figures B-84 and B-85, it appears contradictory that blood lead goes up as a function of housing units passing standard of care, and also goes up as a function of housing units failing standard of care. This doesn't make common sense, but even mathematically, why are both passing and failing used as "independent" variables. Aren't they exactly related? Cheng actually did some research into this, and developed the following hypothetical, but very plausible inspection result scenario **MA State Housing Units Inspection Scores and Children BLL Measurements Relationship (for illustration only)** Census Tract ID A B C D No. Housing (n) 10,000 5,000 2,500 1,000 No. Passing (p) 200 150 150 90 Passing Rate (P) P = p/n * 0.02 0.03 0.06 0.09 No. Failing (f) 20 30 50 90 Failing Rate (F) F = f/n ** 0.002 0.006 0.02 0.09 Total No. Inspected (N) N = p + f 220 180 200 180 % of Passing = [p / (p + f)] *** 90.9% 83.3% 75.0% 50.0% GM BLL 2 4 6 8 (p/f) Ratio 10 5 3 1	There is a third category that is missing from the analysis of this reviewer – namely the category of housing units that were never inspected by the Massachusetts Department of Public Health. Thus, we can only include up to two of these variables (fraction passing, fraction failing, and fraction never inspected) in the model.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				* The values of P increased from A through D same as that of GM BLL. However, it is misleading to look at P alone without considering F jointly. ** The values of F increased from A through D same as that of GM BLL. However, it is misleading to look at F alone without considering P jointly. *** This is the true measure of the quality of housing units inspected. The % of Passing Housing Units are decreased from A through D while GM BLL are increasing. Cheng's example includes four "tracks" with different number of housing units in each track, and some tracks showing better passing rates than others. I think that tables on pages B-76 to B-86 show values geometric blood lead as functions of P and F. If we look at Cheng's scenario, the geometric mean blood lead level increases as F increase as you would expect, but it also increases as P increases. This is the result found in pages B-76 through B-86 as well. However, with Cheng's model in mind it can well be an artifact of the percentage of housing inspected, and having nothing to do with the actual passing or failing rates. I think the problem is as stated on p. 26 of the report, the summed results are then divided by the number of housing units reported within each census tract from the 2000 census. This leads to the anomaly, I think you would get the correct rate of passing by dividing by the number of houses in the sample, not the total number of houses in the census tract. To further compound this problem, is there a practical distinction to the two statements in the penultimate paragraph on page vi of the executive summary? One says risk of lead poisoning was reduced as housing in compliance increases; the other says the risk of poisoning increased as proportion of house out of compliance increased; and are these sentences even consistent with the observations?	

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
126	Barry Nussbaum / Cheng Ling	Specific Comments (continued)		Is there a problem with confounding variables? For example, on A-1, blood lead level vs. income. While it looks like the relationship goes as one would expect it to , suppose a state like Nebraska with low income (but also with a better standard of living per income) also has lower lead. Then the regression would go the other way. Is it proper to cover different spatial areas in this type of graph? In many of the regressions, most of the data are clustered around zero. For instance, in A-42 and A-43, looking at lead vs. population and lead vs. housing density, the bulk of the data in the histogram are very low. Yet the regressions are plotted for a range that seems to go well past where the data are. Are these slopes appropriate to test for significance? In table 4-2 of the predictor variables, there seems to be very little difference in some of the values used to select the most important variables. This may be OK since they are using log-likelihoods; but it should be explained that an innocuous looking difference between 139627 and 139433 between the explanatory power of median family income compared with median household income is really a big deal. I am unsure about the inclusion of all the points in the linear regressions. Specifically, at x=0, ie, zero ambient lead, do you get a lot of points with coordinate (0, blood lead) or is it just the total geometric mean for all the x=0 points. If it is the latter, then only one point would represent lots of observations. If it is the former, with a whole lot of points at zero, I'm not sure how appropriate the regressions are.	The predictor variables (such as income) were not adjusted for the effects of confounding variables per se – however, in the revised report, we adjust each of the single covariate models for separate intercepts, slopes and seasonality terms within each of the 10 EPA regions (which might account for some of the cost-of-living type information that is a potential confounding variable as suggested in this comment). The problem of some of the continuous predictors clustering near a value of zero was also deemed important. We therefore investigated several binary indicator variables based on these continuous predictors (e.g. indicator that the TRI air emissions were above the 95th percentile). The authors of the report believe that inclusion of the entire distribution of each predictor is reasonable in most instances – as these zero values get integrated into the intercept estimate.
127	Barry Nussbaum / Cheng Ling	Specific Comments (continued)		I am curious what the R^2 is for many of the many of the blood lead vs. individual variables, especially when the demographic variables are bunched near zero. If these are low, is it still meaningful to have a linear regression and how can the results point to a significant variable in the six models. In counties in which had the highest ambient lead, the blood lead level seems widely scattered. So does this do anything for the analysis. For models 3-6, what does it mean that Y_{ij} is the number of kids over x ug/ml in county i at time j. Are we not looking at proportions or rates?	Model Diagnostics is a challenging issue for complex longitudinal models such as those utilized in this work – and could be the subject of future work. For the logistic regression models, the data represent the observed binomial proportions – however, the linear model operates on the logit expected value of these binomial proportions. Please see Generalized Linear Models by McCullaugh and Nelder for a more detailed discussion. The issues with Maine have been largely

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				Regarding Region I, pages G-1 and G-2 look inconsistent. In particular, why is the geometric mean blood lead so high in Maine, but very few subjects with blood lead greater than or equal to 10? . . I know the drinking water relationship is controversial, but I am not sure what to make of the graphs on p. A-49. For some years there is a positive correlation, for another set of years, it is negative, and for a third set it is essentially flat. So how can this be significant, and in which direction?	resolved with input from CDC. The previous problem with the geometric mean was based on Maine adopting new minimal reporting limits of 5 ug/dL (which artificially raised the observed GM without increasing the proportion observed above 5 or 10 ug/dL). This artifact of reporting contributed to our decision to eliminate the GM as a potential response variable in the series of National (Low Resolution) Models. We have now allowed for the use of improved time-varying (quadratic) effects of the SDWIS predictor variables in the National Model. The authors do not have specific insight that explains the manner in which the effects of these variables vary over time in the way – however, it is important to note that the SDWIS data was integrated into the model using a 2-level variable that captures the measured concentration when available, and provides an intercept to explain how the response variable (blood-lead) changes in county/quarter combinations that did not contain a SDWIS monitored facility.
128	Barry Nussbaum / Cheng Ling	Specific Comments (continued)		On. P. 10, the cluster of data around zero is discussed. To circumvent this difficulty, a series of indicator values were used. How does this work? On p. 23, where no drinking water data existed, how does the indicator variable work, and how does it allow you to get an intercept and slope? In chapter 4, the relationship between mean blood lead and the individual variables are plotted. Why is there such a difference concerning which variable was for further inclusion between the national and local models? I understand the model validation issues of section 7.4, but I am concerned about another issue. As I understand it, the 60 or so environmental, demographic, income independent variables were all considered individually for inclusion in the models. I am more familiar with stepwise regression where variables go in and out and only the most important ones remain. I am uncertain if individual variables might be used here because they are the "leader" of their category, yet they may be highly collinear with other variables, and may only be significant because of the colinearity. This point is really reflective of the commentary in Appendix D where variables are discussed in their groups. For instance, income variables always seem to have a marked effect, with	The indicator variables are constructed based on the observed distribution of the original predictor variable. Additional information is contained in the detailed appendices. For the drinking water data – an indicator variable was constructed so that we can attain a slope value for the influence of the observed water lead concentration (where measured) on the blood-lead response variable, and an intercept shift (in blood-lead response) for the county/quarter combinations where no water monitoring data existed. Our model fitting approach approximated a backwards elimination routine – in which the full model included the best predictors from each category. The model was then reduced by eliminating terms from the full model that were not contributing significant predictive power (using the Type III test statistics). We have not conducted any specialized analyses (such as discriminant function analysis) to differentiate what makes one county different from another. It should be noted that the longitudinal nature of the data might make this type of investigation somewhat challenging.

Comment Number	Peer Review	Charge to the Panel	Topic	Comment	Response
				"percent no household wage" generally being the best for most models. Similarly, the housing variable group had variables all measuring similar things. The education variables were found to have an effect, but I would guess they are positively correlated with income. In section 7.1, the second paragraph discusses a type of discriminant analysis concerning the rate of decline of elevated blood levels over time. Other than the mapping tool, have we done such an investigation of what makes one county different from another?	