

Quantification of Uncertainty in EPA's Estimates of Lung Function Impacts from Ozone Exposure



Prepared for:
American Petroleum Institute

March 17, 2015

Authors (*)

Anne E. Smith, Ph.D.

Garrett Glasgow, Ph.D.

(*) The opinions expressed herein do not necessarily represent the views of NERA Economic Consulting or any other NERA consultants.

NERA Economic Consulting
1255 23rd Street NW
Washington, DC 20037
Tel: +1 202 466 3510
Fax: +1 202 466 3605
www.nera.com

Report Qualifications/Assumptions and Limiting Conditions

Information furnished by others, upon which all or portions of this report are based, is believed to be reliable, but has not been independently verified, unless otherwise expressly indicated. Public information and industry and statistical data are from sources we deem to be reliable; however, we make no representation as to the accuracy or completeness of such information. The findings contained in this report may contain predictions based on current data and historical trends. Any such predictions are subject to inherent risks and uncertainties. NERA Economic Consulting accepts no responsibility for actual results or future events.

The opinions expressed in this report are valid only for the purpose stated herein and as of the date of this report. No obligation is assumed to revise this report to reflect changes, events or conditions, which occur subsequent to the date hereof.

All decisions in connection with the implementation or use of advice or recommendations contained in this report are the sole responsibility of the client. This report does not represent investment advice nor does it provide an opinion regarding the fairness of any transaction to any and all parties.

© NERA Economic Consulting

Contents

I.	OVERVIEW OF ANALYSIS GOALS AND KEY FINDINGS	1
II.	UNCERTAINTIES IN ESTIMATING LUNG FUNCTION IMPACTS, AND THE IMPORTANCE OF QUANTIFYING UNCERTAINTY	4
III.	OVERVIEW OF APPROACH FOR INCORPORATING STATISTICAL AND MODEL UNCERTAINTIES INTO LUNG FUNCTION RISK ESTIMATES.....	8
	A. Overview of How Statistical Uncertainty Is Estimated	8
	B. Overview of New Analyses in This Report	10
IV.	MSS MODEL SPECIFICATION UNCERTAINTY.....	12
	A. Extending the MSS Model to Ages 5-17	12
	B. The MSS Model with and without BMI	13
	C. Alternative MSS Model Specifications in APEX.....	15
V.	MSS MODEL STATISTICAL UNCERTAINTY.....	18
	A. Estimating Statistical Uncertainty in the APEX Simulations	18
	B. The Specific Procedure Used in this Analysis.....	18
VI.	SIMULATION NOISE WITHIN THE APEX RESULTS.....	20
VII.	RESULTS OF NERA’s UNCERTAINTY ANALYSES	22
	A. Density Plots	22
	B. Simulation Results	23
VIII.	IMPLICATIONS OF UNCERTAINTY ANALYSIS RESULTS.....	30
	APPENDIX.....	31

List of Tables

Table 1. Copy of Proposed Rule’s Table 2 Summarizing a Subset of the Lung Function Impact Estimates in the HREA	5
Table 2. MSS Model β Coefficients with and without BMI.....	14
Table 3. Estimated Simulation Noise Using 20 APEX Runs	20
Table 4. Estimated Probability that Effect Will Not Occur At All, by NAAQS Level and Age (rows) for Each Magnitude/Frequency Effect Category (columns)	28

List of Figures

Figure 1. MSS Age Term (Y) Across Age, Alternate Assumptions.....	13
Figure 2. Percent Experiencing Decrement at Least One Day at 75 ppb under Different Model Specifications, Atlanta 2010	16
Figure 3. Example of Histogram and Density Plot from One APEX Uncertainty Case (dFEV1>10% for at least 1 day, 5-18 years old, Atlanta 2010 at 75 ppb).....	22
Figure 4. Probability Densities for dFEV1> 10%, Atlanta 2010, MSS Model Including BMI ..	25
Figure 5. Probability Densities for dFEV1> 15% , Atlanta 2010, MSS Model Including BMI .	26
Figure 6. Probability Densities for dFEV1> 20%, Atlanta 2010, MSS Model Including BMI ..	27
Figure A - 1. Probability Densities for dFEV1> 10%, Atlanta 2010, MSS Model Excluding BMI	31
Figure A - 2. Probability Densities for dFEV1> 15%, Atlanta 2010, MSS Model Excluding BMI	32
Figure A - 3. Probability Densities for dFEV1> 20%, Atlanta 2010, MSS Model Excluding BMI	33
Figure A - 4. Probability Densities for dFEV1> 10%, Denver 2010, MSS Model Including BMI	34
Figure A - 5. Probability Densities for dFEV1> 15%, Denver 2010, MSS Model Including BMI	35
Figure A - 6. Probability Densities for dFEV1> 20%, Denver 2010, MSS Model Including BMI	36

I. OVERVIEW OF ANALYSIS GOALS AND KEY FINDINGS

The core of the Administrator's rationale for tightening the ozone NAAQS that is articulated in the Proposed Rule¹ is based on clinical studies of lung function responses to various levels of ozone. For example, the Proposed Rule states that "[t]he Administrator places the greatest weight on the results of controlled human exposure studies and on exposure and risk analyses based on information from these studies."² The specific measurement of lung function discussed is forced expiratory volume in one second (FEV₁), and FEV₁ changes are the lung function impact for which the Health Risk and Exposure Assessment (HREA) document provides quantitative risk estimates under alternative ozone NAAQS levels. A very large number of such estimates are provided in the HREA, including:

- For three age groups (5-18 years, 19-35 years, and 36-55 years),
- For three degrees of decrement in FEV₁ ($\geq 10\%$, $\geq 15\%$, and $\geq 20\%$),
- For three different numbers of days per season on which a given individual experiences such a decrement (≥ 1 , ≥ 2 , and ≥ 6),
- For five different years of ozone data (2006 through 2010), each year reflecting a different possible temporal variability in ozone levels, and
- For fifteen cities of the U.S., reflecting different population demographics and lifestyles as well as different temporal and spatial patterns of ozone concentrations.

Although each city and each year's variability in ozone levels results in different fractions of each age group projected to experience decrements in FEV₁, each individual risk estimate is made only as a point estimate; the HREA's methodology for lung function risk estimation did not include quantification of the range of uncertainty that surrounds those estimates. The large volume of many different point estimates suggests an accurate understanding of the frequency of lung function decrements under the current ozone NAAQS, and of the change in that frequency if the ozone NAAQS were to be tightened. However, this is misleading, since the scientific evidence that is the basis for these lung function risk estimates is subject to many sources of uncertainty beyond just the city to city and year to year differences that the HREA analyzes in detail.

The complexity of the lung function risk analysis methodology in the HREA makes characterization of statistical uncertainty associated with its point estimates a time consuming and challenging task. Although they are easier, sensitivity runs using alternative model

¹ 79 *Fed. Reg.* 75234, December 17, 2014.

² 79 *Fed. Reg.* 75234, December 17, 2014, at 75304.

specifications for the lung function exposure-response also take time and resources. However, the relative emphasis that the Proposed Rule assigns to estimates of frequency of lung function decrements heightens the importance and need for quantification of ranges of uncertainty on this particular portion of the HREA output. Representation of statistical errors and model uncertainties surrounding the lung function risk estimates would greatly assist the Administrator in being able to assess how much weight to assign to this particular element of her evaluation of the current and alternative ozone NAAQS, consistent with her recognition that her policy decision

...requires judgments based on an interpretation of the scientific evidence and exposure/risk information that neither overstates nor understates the strengths and limitations of that evidence and information, nor the appropriate inferences to be drawn therefrom.³

Even providing a range of uncertainty on lung function results for just one of the 15 cities in the HREA would be very informative in assessing appropriate inferences that can be drawn from the lung function risk analysis. This is because the most salient sources of uncertainty in those risk estimates – statistical and model uncertainty – apply similarly to the estimation method in all the cities.

This study was conducted specifically to provide such uncertainty ranges, using additional portions of the existing evidence base. It provides quantitative information about model uncertainty due to alternative specifications for the HREA’s lung function exposure-response relationship. It also provides a formal quantification of the statistical confidence intervals around the point estimates in the HREA. These additional analyses that we have prepared provide useful supplemental information about the lung function risk estimates that the Administrator proposes to rely on, in part, in setting the ozone NAAQS. Although we also have found time constraints to present a barrier to a comprehensive application of the uncertainty analysis to the entirety of Chapter 6 of the HREA, we have been able to provide a robust assessment of the patterns in the confidence ranges for all the lung decrement categories for two cities, Atlanta and Denver, for the year 2010 ozone pattern. The similarity of these patterns is such that these uncertainties can be expected to be repeated in other cities and for other years.

In brief, our analysis finds that there is a large degree of both statistical and model uncertainty in the lung function decrements reported in the HREA, and that uncertainty undercuts the confidence in the projected reductions in lung function impacts that are described in the Proposed Rule. For example:

- When these uncertainties are considered, there is as much noise in our understanding of the level of risk as there is in estimated benefits from tightening the level of the standard,

³ 79 Fed. Reg. 75234, December 17, 2014, at 75303-4.

even for the most sensitive category of children experiencing at least one decrement in FEV₁ that is more than 10% (dFEV₁≥10%).

- The “signal to noise” ratio is progressively weakened in every age group considering more decrements per season, and when considering larger FEV₁ decrements.
- The probability that there is no lung function impact even under the current 75 ppb standard increases for larger dFEV₁, and with age.
 - There is a non-trivial probability⁴ that no adults over the age of 35 years will experience a single dFEV₁ ≥15% under the current standard of 75 ppb.
 - Even for children, the population subgroup projected with the largest percent of individuals projected to experience decrements, there is non-trivial probability that none will experience 6 or more instances of dFEV₁ ≥15% under the current standard of 75 ppb.
 - Even for the smallest decrement, dFEV₁ ≥10%, there is a non-trivial probability that no adults will experience ≥ 6 such decrements under the current standard.

The results summarized above, and further detailed in the rest of this report, provide important information for the Administrator regarding imprecision in the HREA’s lung function risk estimates; they also provide an important context regarding the size of the projected improvements in lung function impacts due to tighter ozone standards relative to the underlying uncertainty in the magnitude of the impacts under the current standard of 75 ppb. For this reason, the contents of this report should be given close consideration in the Administrator’s evaluation of the need for tightening the current standard, as well as the amount of weight to give to the lung function risk estimates in deciding on the stringency of an alternative standard.

⁴ We define a “non-trivial probability” as at least 20% of the estimated density for the lung function decrement overlapping zero. This is explained in more detail below.

II. UNCERTAINTIES IN ESTIMATING LUNG FUNCTION IMPACTS, AND THE IMPORTANCE OF QUANTIFYING UNCERTAINTY

Uncertainty is an important aspect of any projection of the risk or change in risk due to altered concentrations of an ambient pollutant such as ozone, since these risk estimates are based on extrapolations of highly variable evidence from epidemiological studies. Even in the case of the lung function risk estimates, which are based on clinical studies, there is large interpersonal and intrapersonal variability in responsiveness to exposures of ozone. Although statistical models have been fit to these clinical data, there is substantial variability in the true values of the statistically-derived parameters – variability that comes from the differences in the responses of individuals in the controlled exposure studies, and the fact that these individuals are just a tiny sample of the general population physically fit enough to maintain the level and duration of exertion required by these clinical trials.

The statistical uncertainty associated with those model parameter values translates into uncertainty in the projected lung function impacts under different approximations of ambient ozone concentrations. On top of this “statistical uncertainty” associated with any single model specification, there is “model uncertainty.” Model uncertainty is the variability in projected risk estimates when a different model specification is used to fit the available clinical evidence. Seemingly small changes in how a model’s functional form is specified can result in very different estimates of population-wide risk, even when two statistical models provide an equivalently good fit to the original data.⁵ In the case of air pollution risk analyses, model uncertainty is typically much larger than statistical uncertainty.⁶

In an effort to bring some form of insight to what these FEV₁ risk estimates mean for setting a standard, the Proposed Rule focuses on summaries for a subset:

In evaluating these lung function risk estimates within the context of considering the current and alternative O₃ standards, the PA focuses on the percent of children estimated to experience one or more and two or more decrements > 10, 15, and 20%, noting that the percentage of asthmatic children estimated to

⁵ An example of vast differences in risk estimates from models that statistical tests find to offer roughly similar fits is found in another key part of the ozone HREA, in regards to estimates of the long-term respiratory mortality risk from ozone. For a full description of this model uncertainty, see: Smith, AE. Comments regarding effects thresholds for long-term mortality risk (for the May 28, 2014 CASAC teleconference on the ozone NAAQS), prepared on behalf of the Utility Air Regulatory Group and submitted to CASAC, May 27, 2014. Available: [http://yosemite.epa.gov/sab/sabproduct.nsf/69A6D3ADE58B82C485257CE5006B0829/\\$File/Smith+comments+on+Threshold+Sensitivities+for+May+28+2014+CASAC+meeting.pdf](http://yosemite.epa.gov/sab/sabproduct.nsf/69A6D3ADE58B82C485257CE5006B0829/$File/Smith+comments+on+Threshold+Sensitivities+for+May+28+2014+CASAC+meeting.pdf).

⁶ Smith AE and Gans W, “Enhancing the Characterization of Epistemic Uncertainties in PM_{2.5} Risk Analyses,” *Risk Analysis* 35(3) (forthcoming March 2015; available in early release on-line at DOI: 10.1111/risa.12236).

experience such decrements is virtually indistinguishable from the percentage estimated for all children.⁷

A table is then presented in the Proposed Rule (Table 2), which is replicated as Table 1 below. This table, consistent with the emphasis stated in the Proposed Rule, summarizes lung function decrement frequencies (after averaging over all the years 2006-2010), for children only, and for the frequencies of more than one, and more than two, decrements per season.

Table 1. Copy of Proposed Rule’s Table 2 Summarizing a Subset of the Lung Function Impact Estimates in the HREA

(Source: 79 Fed. Reg. 75234, December 17, 2014 at 75275)

TABLE 2—SUMMARY OF ESTIMATED O₃-INDUCED LUNG FUNCTION DECREMENTS FOR THE CURRENT AND POTENTIAL ALTERNATIVE O₃ STANDARDS IN URBAN CASE STUDY AREAS

Lung function decrement	Alternative standard level	Average % children ⁷⁸	Number of children (5 to 18 years) [number of asthmatic children] ⁷⁹	% Children worst year and area
One or more decrements per season				
≥10%	75	14–19	3,007,000 [312,000]	22
	70	11–17	2,527,000 [261,000]	20
	65	3–15	1,896,000 [191,000]	18
	60	5–11	1,404,000 [139,000] ⁸⁰	13
≥15%	75	3–5	766,000 [80,000]	7
	70	2–4	562,000 [58,000]	5
	65	0–3	356,000 [36,000]	4
	60	1–2	225,000 [22,000]	3
≥20%	75	1–2	285,000 [30,000]	2.8
	70	1–2	189,000 [20,000]	2.1
	65	0–1	106,000 [11,000]	1.4
	60	0–1	57,000 [6,000]	0.9
Two or more decrements per season				
≥10%	75	7.5–12	1,730,000 [179,000]	14
	70	5.5–11	1,414,000 [145,000]	13
	65	1.3–8.8	1,023,000 [102,000]	11
	60	2.1–6.4	741,000 [73,000]	7.3
≥15%	75	1.7–2.9	391,000 [40,000]	3.8
	70	0.9–2.4	276,000 [28,000]	3.1
	65	0.1–1.8	168,000 [17,000]	2.3
	60	0.2–1.0	101,000 [10,000]	1.4
≥20%	75	0.5–1.1	128,000 [13,000]	1.5
	70	0.3–0.8	81,000 [8,000]	1.1
	65	0–0.5	43,000 [4,000]	0.8
	60	0–0.2	21,000 [2,000]	0.4

The most important point we wish to make with respect to the results shown in the Proposed Rule table above is that there is no representation of uncertainty in the results shown in it. Some readers may confuse the ranges in the third column (“Average % children”) with uncertainty ranges, but these are in fact only ranges in the point estimates that were made for each of the 15

⁷ 79 Fed. Reg. 75234, December 17, 2014 at 75274. (Notably, the HREA has not estimated FEV₁ changes for asthmatic children, but the Proposed Rule is assuming that HREA results for children in general are indicative of impacts that asthmatic children will also experience. The veracity of this assumption is unknown, given that asthmatic children may have different average activity patterns than non-asthmatic children. For our paper, we will consider the uncertainty in these HREA FEV₁ risk estimates and note here that the uncertainties for asthmatic children are even more uncertain.)

urban study areas of the HREA. While HREA estimates do vary from city to city because the cities have different spatial surfaces for exposures and their populations may have different activity patterns, the uncertainty associated with the statistically-derived exposure-response function creates uncertainty around the point estimate in each individual city. That statistical variance is not represented at all for any of the city-specific estimates of the percent of children experience certain FEV₁ decrements.

It is widely understood among risk analysts that the purpose of a risk analysis is not just to characterize the quantitative magnitude of a risk, but also to inform decision makers about the confidence that alternative policy actions will mitigate an apparent risk.⁸ The proper role of the risk analysis professional is to guide and inform the regulator, and not to offer risk estimates that suggest a false degree of certainty.

It is also clear from the Proposed Rule that understanding such uncertainties plays a significant role in the policy decision on whether the current standard needs to be tightened and by how much. For example, although the HREA provides estimates of non-zero health risks even at an alternative standard as low as 60 ppb, the PR makes it clear substantial uncertainties exist in the supporting evidence, and that such uncertainties become greater as the alternative standard becomes more stringent. For example, in explaining why EPA is not proposing to set the standard as low as 60 ppb, the PR notes that these uncertainties are also a part of the decision:

...the Administrator notes that setting a standard below 0.065 ppm, down to 0.060 ppm, would inappropriately place very little weight on the uncertainties in the health effects evidence and exposure/risk information.⁹

Uncertainties come in many forms, and the statement above does not refer solely to uncertainties in the quantitative estimates of risk that the HREA provides. Nevertheless, given the substantial role that uncertainty plays in the Administrator's judgment about the level of the standard, it is important that uncertainty be appropriately characterized in the HREA.

In reviewing EPA's methods for conducting its risk analyses, a committee of the National Academy of Sciences identified representation of uncertainties as a substantial limitation:

There are several major barriers to broad acceptance of recent EPA health benefits analyses. One barrier is the large amount of uncertainty inherent in these analyses, and another is the manner in which the agency deals with this uncertainty.¹⁰

⁸ Stern PC, Fineberg HV (eds.). *Understanding Risk: Informing Decisions in a Democratic Society*, 1st ed. Washington, DC: National Academy Press, 1996.

⁹ 79 *Fed. Reg.* 75234, December 17, 2014, at 75236.

¹⁰ National Academies of Sciences. 2002. *Estimating the Public Health Benefits of Proposed Air Pollution Regulations*, p. 126 Available: http://www.nap.edu/openbook.php?record_id=10511.

In particular, the NAS committee identified emphasis on statistical uncertainty rather than model uncertainty in the health effect-pollutant function as a weakness in how EPA's risk analyses deal with uncertainty:

*Only one source of uncertainty (the random sampling error associated with the estimated concentration-response function) was incorporated into the analysis. EPA typically emphasizes only the mean value of the probability distribution. Because of the lack of consideration of other sources of uncertainty, the results of the primary analysis often appear more certain than they actually are.*¹¹

While the NAS committee spoke of the need for EPA to quantify model uncertainty in addition to statistical error in its risk analyses, the ozone HREA does not report even the statistical uncertainties in its lung function risk estimates. This omission is important because those estimates are based on a statistical model of very noisy (*i.e.*, highly variable) data from small samples of individuals who participated in a few clinical studies. Thus, the Administrator is supplied with an insufficient understanding of the uncertainties in the estimates of lung function impacts at the current standard, and how significant the changes in those impacts may be if the ozone NAAQS were to be tightened.

In light of the high policy-relevance of the lung function estimates, this analysis was conducted to develop the missing estimates of the uncertainty ranges associated with the HREA's lung function risk results.

¹¹ *Ibid.* pp. 10-11.

III. OVERVIEW OF APPROACH FOR INCORPORATING STATISTICAL AND MODEL UNCERTAINTIES INTO LUNG FUNCTION RISK ESTIMATES

Chapters 7 and 8 of the HREA contain estimates of morbidity and mortality risks that are based on population-wide epidemiological studies. All of those risk estimates are reported with statistical error bounds. The HREA also provides some evidence (albeit to an insufficient degree) of model uncertainty on those morbidity and mortality risk estimates. However, the estimates of risk of lung function decrements in Chapter 6 of the HREA are point estimates. The HREA provides no statistical uncertainties for those estimates, nor any indication of potential ranges of model uncertainty. Our analysis was conducted to provide the missing statistical uncertainty on the lung function results, and also to discuss other sources of uncertainty, which include model uncertainty and simulation uncertainty.

A. Overview of How Statistical Uncertainty Is Estimated

The error bounds in Chapters 7 and 8 of the HREA are based on the statistical variance of the concentration-response slope parameter estimated for the epidemiological model that is used in each respective risk calculation. These 95% confidence intervals are relatively straightforward to compute because all of those epidemiologically-based risk estimates are calculated with a formula that uses only a single statistically-estimated parameter, known as the slope parameter β . The formula is simply the following:

$$\% \text{ increase in population risk from concentration } C = e^{(\beta C)} - 1 \quad (1)$$

where the units for the concentration of the pollutant are the same as in the air quality concentration data that were used in the original epidemiological study. To calculate the upper and lower bounds of the statistical uncertainty for this risk estimate, one simply recomputes Eq. (1) using values of β that are increased and decreased, respectively, by 1.96 times the standard error of the statistical estimate of β .¹²

The simplicity of providing statistical confidence intervals for the mortality and morbidity risk estimates in the HREA may explain why they are readily supplied. However, the same principles are applicable to the risk estimates for lung function responses to ozone that are the subject of Chapter 6 of the HREA. The approach is somewhat more complex because the lung function risk estimates involve a *set* of statistically estimated parameters that are interdependent, but there is no reason the confidence intervals on the APEX-based results cannot be computed

¹² The standard error is the square root of the variance of an estimate. The standard error is multiplied by 1.96 because that defines the 95% confidence range for the true value around an estimate (under normal conditions and large enough sample sizes, which are assumed to apply to the epidemiological models used in the HREA).

from the same basic statistical outputs of the estimation of the model for lung function responses to ozone exposure levels.

A simple formula such as Eq. (1) is not possible to write down for how the lung function risk estimates are calculated. This is because this calculation involves use of a multivariate exposure-response function.¹³ Additionally, this function is applied in a more complicated manner that considers cumulative exposures over multiple hours, based on simulated patterns of individuals' activity levels and ozone exposure levels. The APEX model provides the simulations of many different types of individual activity and exposure patterns and applies the multivariate exposure-response function to each, to produce estimates of the potential changes in the lung function, FEV₁, across an entire population (categorized into subgroups such as children, young adults, and older adults). This calculation by APEX uses a multivariate lung function exposure-response formula that has been statistically derived from the available clinical studies on a small sample of healthy volunteers. For the current HREA, that exposure-response function is called the "MSS model," named after the authors of a paper that first reported the results of a statistical modeling of such a multivariate exposure-response function.¹⁴

Because the MSS model involves multiple parameters, rather than a single "slope parameter" such as the β in Eq.(1), derivation of the statistical error around the lung function risk estimates cannot be performed with just a "higher" and "lower" value of a parameter estimate, derived from the variance on that estimate. However, a direct extension of that technique is used in a multivariate case; it relies on both the variances of the individual parameter estimates, and the covariances among those parameter estimates. The covariances account for the interrelatedness of the parameter estimates in which, if one parameter is actually higher than its mean estimate, the other parameters would also take on different values.

EPA could have, but did not, apply this method to estimate the statistical errors on the lung function risk estimates in its HREA.¹⁵ To do so in this study, we obtained the

¹³ For pollutant risks that are based on clinically-observed responses, such as FEV₁ changes, the term for the formula used to estimate risk is "exposure-response" rather than "concentration-response" because the actual exposure of each individual is directly measured in the clinical setting. In contrast, non-clinical epidemiological associations of ozone and health outcomes typically use monitored ambient concentrations of the pollutant as a proxy for the exposures of individuals. The latter situation creates a problem with exposure misclassification error, which heightens concern that the statistical variance does not fully indicate the uncertainty in the estimates; this is signaled by the use of the different terminology for the response function. However, exposure-response and concentration-response functions are used in the same way in making estimates of risk under alternative pollutant standards.

¹⁴ MSS stands for McDonnell, Stewart, and Smith, authors of a 2010 paper that provided the first such function. The function actually in use in the HREA is from a 2012 update, and has more authors: McDonnell, WF; Stewart, PW; Smith, MV; Kim, CS; Schelegle, ES. (2012). Prediction of lung function response for populations exposed to a wide range of ozone conditions. *Inhal Toxicol* 24: 619-633. However, it continues to be referred to as the MSS model, in this case "MSS (2012)."

¹⁵ The HREA mentions that uncertainty analysis would have been desirable, but implies EPA could not do it because EPA did not have all of the individual data on which the MSS model was estimated. (HREA, p. 6-40). However, it

variance/covariance matrix output from the original researchers who estimated the MSS model (it is a standard output of that estimation procedure, although only the standard errors were reported in their journal articles).¹⁶ We also obtained the APEX model and APEX input files necessary to replicate the lung function risk estimates in Chapter 6 of the HREA from EPA.¹⁷ After confirming that we could replicate the point estimates for lung function decrements reported in the HREA using the APEX files, we proceeded to use the MSS variance/covariance outputs to estimate statistical uncertainties associated with those estimates, as explained in more detail below.

We note that the HREA does provide a sensitivity analysis to the various parameters of the MSS model.¹⁸ However, this sensitivity analysis does not provide useful information about the range of uncertainty around the risk estimates in that same chapter. It varies the parameters individually, when in fact any change in one parameter will affect the others, so that the *apparent* sensitivity may not have any relationship to the actual sensitivity. (The application of covariances would have remedied this situation, but EPA did not obtain information on those covariances.) Additionally, this sensitivity analysis only changed parameters by 5% from their mean values, which in most cases is a very small change relative to the standard error. There is no reason to believe that a 5% change is likely to occur, or be large enough to reflect actual uncertainty. Sensitivity analyses to explore estimate uncertainty need to vary the inputs over a reasonably likely range, and not by some *ad hoc* percentage. Thus, although the HREA attempts to discuss sensitivity, the outcomes of the analysis reported are uninformative. Furthermore, one form of uncertainty that EPA could have done would have been to apply alternative forms of the MSS model, which are reported in the journal article. This sort of sensitivity analysis would have helped identify model uncertainty, but it was not done at all in the HREA.

B. Overview of New Analyses in This Report

Our simulations focus on the cities of Atlanta and Denver in 2010. Atlanta is used as an example by the EPA throughout the HREA, and is the focus of our examples in the text. However, the same results are provided for Denver in the Appendix. We examine three different ozone scenarios in our simulations – 75 ppb, 70 ppb, and 65 ppb.

The health outcome of interest that comes out of the APEX simulations is the percentage of individuals in various age groups that are predicted to suffer a lung function decrement of a given percentage amount. Lung function decrement is measured as the percentage decline in

is not necessary to have the raw individual-level MSS data to perform the uncertainty analysis described above; all that is needed is the standard output from the statistical estimation.

¹⁶ These were provided to NERA upon our request by Dr. Paul Stewart.

¹⁷ These were provided to NERA upon our request by John Langstaff, EPA Office of Air Quality Planning and Standards.

¹⁸ HREA, pp. 6-39 to 6-41.

forced expiratory volume in one second (FEV₁). Three different levels of lung function decrements were considered – a decrease in FEV₁ of at least 10%, at least 15%, and at least 20%. The three age groups considered were 5-18, 19-35, and 36-55 years old. (It was assumed that lung function decrements would be minimal for individuals above 55 years of age, so results for this group were not presented in the HREA.) Finally, we consider the frequency of FEV₁ decrements – at least 1, 2, or 6 days with a given decrement.

Thus, for each scenario one APEX run for this scenario will produce 27 results of interest – the percentage of individuals in each of the 3 age groups, that experienced each of the three levels of lung function decrement, for at least one, at least two, or at least 6 days. Below we discuss the uncertainty inherent in these estimates.

IV. MSS MODEL SPECIFICATION UNCERTAINTY

One source of uncertainty in modelling FEV₁ decrements through APEX is in the specification of the MSS model. There are two specification questions that we examine in detail here. First, the models presented in McDonnell *et al.* (2012) were estimated using subjects between the ages of 18 and 35. In order to extend these model results to younger individuals, the EPA assumed that individuals aged 5-17 were as responsive to ozone as the 18 year olds observed by McDonnell *et al.* Second, McDonnell *et al.* present two versions of the MSS model with a threshold, one that includes body mass index (BMI) as a variable and one that excludes it. The EPA elected to use the version of the MSS model that excluded BMI. As we will see, both of these specification decisions have a significant impact on the APEX simulation results.

This sort of model specification uncertainty is distinct from the statistical uncertainty we examine in the following section. Statistical uncertainty arises when we use a sample to estimate population parameters, and is reflected in the standard errors on our model coefficients. In contrast, the model specification uncertainty we examine here arises from uncertainty about which model specification (each with its own statistical uncertainty) is correct.

It is theoretically possible to calculate confidence intervals based on model specification uncertainty by making subjective decisions about the likelihood of various model specifications. However, we do not undertake that exercise here. Instead, we simply present evidence that slightly different MSS model specifications produce very different APEX simulation results.

A. Extending the MSS Model to Ages 5-17

The EPA justifies the decision to assume that all individuals from ages 5-18 are equally responsive to ozone as follows:

*Clinical studies data for children which could be used to fit the model for children are not available at this time. In the absence of data, we are extending the model to ages 5 to 18 by holding the age term constant at the age 18 level.*¹⁹

However, there *is* published data on the effect of ozone on children, and this information should be considered when extending the MSS model to ages 5-17 years. McDonnell *et al.* (1985) conducted a chamber study of 23 children between the ages of 8 and 11, and found they had a reduced responsiveness to ozone when compared to an identical chamber exposure protocol of 18-30 year old subjects.²⁰ Kinney *et al.* (1996) reanalyzed studies of children ages 7 to 17 in summer camps, and after making a learning curve adjustment for early pulmonary function tests

¹⁹ HREA, p. 6-12.

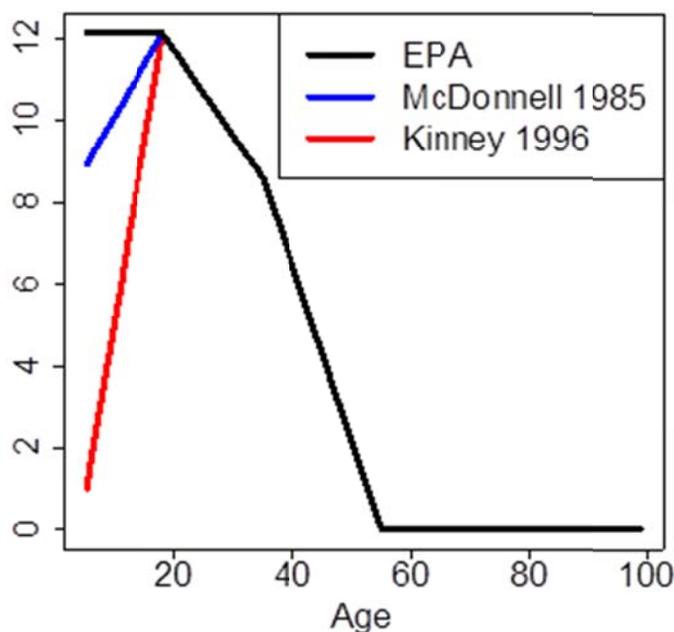
²⁰ McDonnell, WF; Chapman, RS; Leigh, MW; Strobe, GL; Collier, AM. (1985). Respiratory responses of vigorously exercising children to 0.12 ppm ozone exposure. *Am Rev Respir Dis* 132:875-879.

(which tended to inflate estimated FEV₁ decrements) also found a reduced responsiveness to ozone as compared to adults.²¹

One way to use this data in extending the MSS model to children is to take the ratio of the observed FEV₁ decrements for children and adults to generate a responsiveness factor, and then construct a piecewise linear age function in the same way as the EPA. A presumed ratio from McDonnell *et al.* was taken as 0.837 while for Kinney *et al.* it was 0.435. The MSS age term was then adjusted so that the piecewise linear function would run through these points at age 10

A comparison of the MSS age term (Y in the notation of McDonnell *et al.*) under these three different assumptions is presented in Figure 1.

Figure 1. MSS Age Term (Y) Across Age, Alternate Assumptions



B. The MSS Model with and without BMI

McDonnell *et al.* (2012) present two versions of the MSS model with a threshold, one that includes body mass index (BMI) as a variable, and one that does not. Both versions of the MSS model are presented in Table 2.

The APEX simulations presented in the HREA use the version of the MSS model that does not include BMI. This decision is discussed in the HREA:

²¹ Kinney, PL; Thurston, GD; Raizenne, M. (1996). The effects of ambient ozone on lung function in children: A reanalysis of six summer camp studies. *Environmental Health Perspectives* 104:170-174.

Some studies have found greater FEV₁ decrements to be associated with increasing BMI. BMI was included in some of the models of McDonnell et al. (2012); however, the BMI terms were found to be statistically insignificant, indicating that the effect of BMI on FEV₁ in the presence of O₃ is likely to be small, within the range of BMIs of the subjects studied.²²

While it is true that the coefficient on BMI is statistically insignificant in the MSS model, it should be noted that the coefficient on age is also statistically insignificant, regardless of whether BMI is included in the model. The HREA discusses this point on page 6-40. Despite this, the HREA attaches some importance to the coefficient on age, and develops a piecewise linear function of age to allow estimation of an age effect for age groups outside of those studied in McDonnell *et al.* No explanation is provided in the HREA for why one statistically insignificant coefficient is treated as important, while the insignificance of another is used as grounds to prefer a model that excludes that variable.

Table 2. MSS Model β Coefficients with and without BMI

	Without BMI	With BMI
β_1	10.916 (0.8446)	11.092 (0.866)
β_2 (Age)	-0.2104 (0.3100)	-0.2873 (0.2542)
β_3	0.01506 (0.00333)	0.01486 (0.00326)
β_4	13.497 (4.734)	13.442 (4.710)
β_5	0.003321 (0.000207)	0.003224 (0.000207)
β_6	0.8839 (0.0647)	0.8867 (0.0644)
β_8 (BMI)		0.5467 (0.3687)
β_9	59.284 (10.192)	59.963 (10.286)
Var(U)	0.9373 (0.0824)	0.9166 (0.0805)
Var(E)	17.0816 (1.1506)	17.076 (1.139)
AIC	49,594	49,583

Further, the version of the MSS model that includes BMI as a variable clearly fits the data better than the version that omits BMI. The Akaike Information Criterion (AIC) for the model that

²² HREA, p. 6-8.

includes BMI is 49,583, while the AIC for the model that excludes BMI is 49,594 – a lower AIC indicates a better fit. We also note that the latest publication from McDonnell *et al.* (2013) only considers the version of the model that includes BMI, suggesting that this is their preferred specification.

C. Alternative MSS Model Specifications in APEX

An obvious question is how the APEX simulation results change when these alternative model specification decisions are considered. The APEX simulation code can be modified to make different age adjustments for children, and to use the version of the MSS model that includes BMI to calculate FEV₁ decrements.²³

To test the effect of switching to the version of the MSS model that includes BMI, we ran APEX simulations with 200,000 simulated individuals for Atlanta in 2010 under the 75 ppb scenario. We examined the MSS model without BMI and with the age adjustment for children used by the EPA (model A), the MSS model with BMI and with the age adjustment for children used by the EPA (model B), the MSS model without BMI and with the age adjustment for children calculated using the results from McDonnell *et al.* 1985 (model C), and the MSS model without BMI and with the age adjustment for children calculated using the results from Kinney *et al.* 1996 (model D).

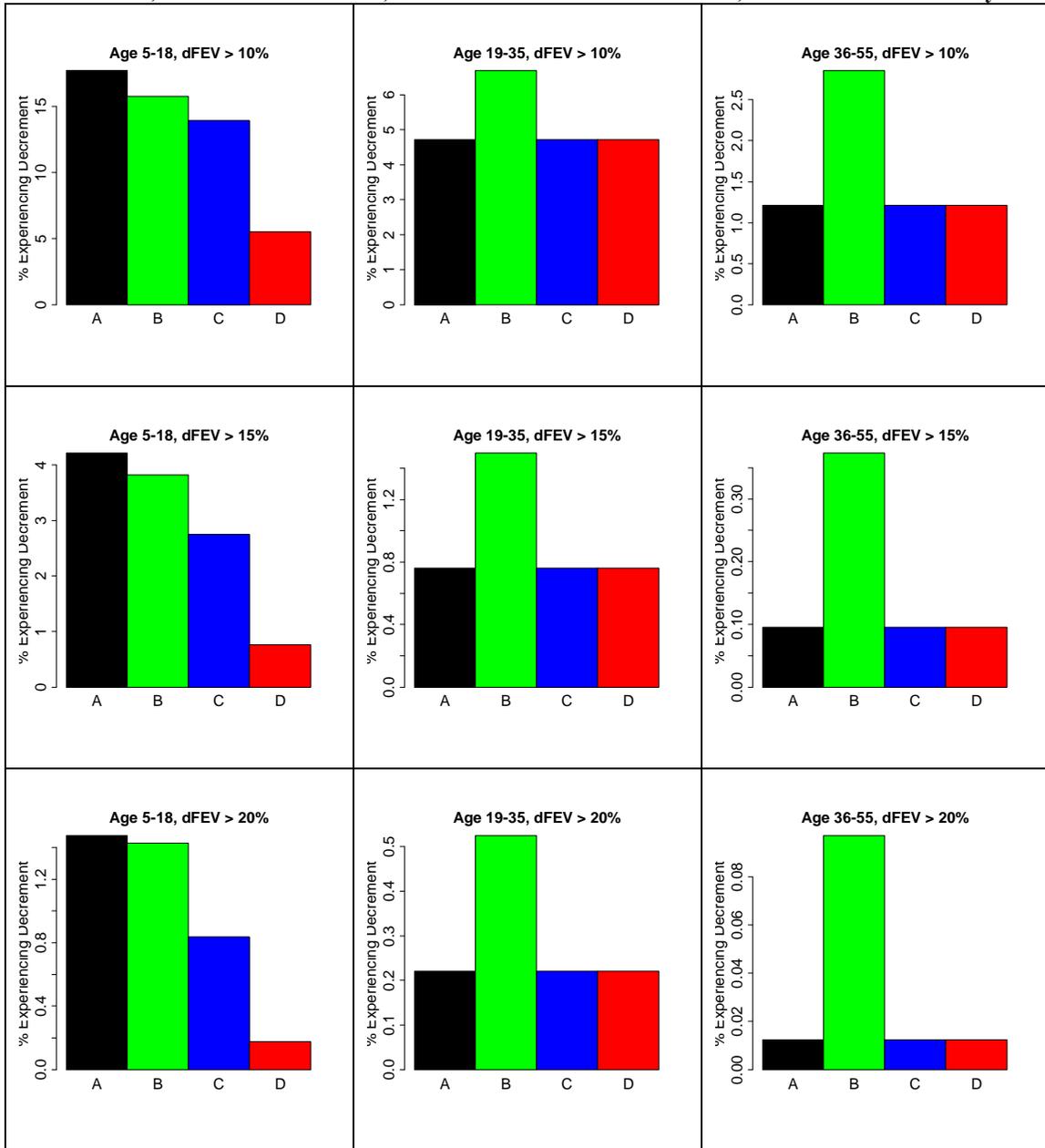
For each of these alternate model specifications, Figure 2 reports the percentage of individuals within the three age groups (5-18, 19-35, and 36-55 years) that were projected to experience FEV₁ decrements of greater than or equal to 10%, 15%, and 20% on at least one day during the 2010 O₃ season.

We first note that the results from model A (the version of the MSS model used by the EPA) match the corresponding numbers in the HREA Appendix for Chapter 6 (Table 6B-1). This demonstrates that we are able to successfully replicate the APEX simulations presented in the HREA.

²³ In fact, the BMI variable is included in the MSS formula used to calculate FEV₁ decrements, and is excluded from the model by simply setting the BMI coefficient to zero (and setting the other coefficients to the non-BMI MSS values reported in Table 6-1 of the HREA).

Figure 2. Percent Experiencing Decrement at Least One Day at 75 ppb under Different Model Specifications, Atlanta 2010

A = MSS 2012, B = MSS 2012 + BMI, C = MSS 2012 + McDonnell 1985, D = MSS 2012 + Kinney 1996



A comparison of the MSS models with and without BMI (models A and B) reveals that while age has a negative coefficient in both models, the coefficient on BMI is positive. Since BMI tends to increase with age, the result of switching to the MSS model that includes BMI in the APEX simulations is to reduce the differences across age groups. Specifically, the percentage of individuals aged 5-18 estimated to experience a decrement in FEV₁ drops, while the percent of older individuals projected to experience a decrement in FEV₁ rises. The 5-18 age group, which

is the group that has the largest projected risk of experiencing decrements, and which is the focus of the rationale for tightening the standard in the Proposed Rule, experiences the largest percentage point change as a result of using the version of the exposure-response model that includes BMI as an explanatory variable. Those percentage point changes represent a 15% to 21% reduction in the level of risk that the HREA has presented for the 5-18 age group. Although the use of the better-fitting exposure-response model increases the projected risk of FEV₁ decrements for the higher age groups, those increases are much smaller in absolute terms than the decreases in the children's group. Further, those older groups' total projected risk remains substantially smaller than for children.

The reduction in percentage of children projected to experience FEV₁ decrements from using the version of the MSS model that accounts for the effect of BMI is also substantial relative to the HREA's projected benefit from tightening the ozone NAAQS. For instance, the percentage of individuals ages 5-18 who are expected to experience a decrement in FEV₁ of at least 10% under the 75 ppb scenario drops by 1.92 percentage points, which is similar in magnitude to the HREA's projected improvement in risk by lowering the current 75 ppb standard to a 70 ppb standard (3.23 percentage points according to Table 6B-1).

As we would expect from an examination of Figure 1, the alternative assumptions about how to extend the MSS model to children based on the data from McDonnell *et al.* (model C) and Kinney *et al.* (model D) reduce the percentage of children expected to experience a decrement, and do not change the predictions for adults. The reductions in the percentages for children are substantial, and greater than the reductions seen when using the version of the MSS model that includes BMI. For instance, the percentage of individuals aged 5-18 who are expected to experience a decrement in FEV₁ of at least 10% under the 75 ppb scenario drops by 3.67 percentage points for model C, which is larger than the 3.23 percentage point decline predicted in the HREA from lowering the current 75 ppb standard to a 70 ppb standard. The drop for model D is 12.12 percentage points, which is larger than the decline predicted in the HREA from lowering the current 75 ppb standard to a 60 ppb standard.

A strong case could be made for any one of the three of the alternative model specifications considered here. The MSS model that includes BMI (model B) provides a better fit to the data than the version used by the EPA (model A), and the alternate assumptions when extending the MSS model to ages 5-17 (models C and D) are based on published observed data. This model specification uncertainty has a large influence on the results of the APEX simulations, with changes in model specification producing changes in APEX results that in some cases are larger than the differences between the different ozone standard scenarios considered by the EPA.

V. MSS MODEL STATISTICAL UNCERTAINTY

A separate source of uncertainty in the APEX simulations comes from the statistical uncertainty in the MSS model estimates. Statistical uncertainty in the MSS model arises from the fact that the model is estimated on a sample of individuals, and thus the model coefficients that relate other variables to lung function are not perfectly known. This statistical uncertainty is represented by the standard errors on the model coefficients. In the APEX model the EPA uses the MSS model coefficients to calculate lung function risk, but does not account for these standard errors. Here we describe our method for determining how the statistical uncertainty in the MSS model influences the predictions that come out of the APEX model.

A. Estimating Statistical Uncertainty in the APEX Simulations

Under maximum likelihood theory, the estimated coefficients and covariance matrix of a model define a multivariate normal distribution – each coefficient is a mean of a normal distribution, while the covariance matrix gives us the variance of and covariance between each of these normal distributions. The standard errors of each parameter estimate are just the square root of the diagonal of this covariance matrix.

We can represent the statistical uncertainty in our model coefficients by taking draws of coefficients from this multivariate normal distribution to produce different sets of possible model coefficients, which can then be used in further calculations. Economists know this method as the Krinsky-Robb method.²⁴

In our case, each set of coefficients that we draw from the multivariate normal distribution represents an alternative set of MSS model coefficients we might have seen if the model had been estimated on a different sample. We use each set of simulated coefficients as the input into an APEX simulation, and examine how the APEX simulation results differ across these alternative versions of the MSS model. The variation across the APEX simulation results reveals how statistical uncertainty in the MSS model causes uncertainty in the APEX predictions.

B. The Specific Procedure Used in this Analysis

In this analysis we consider both the versions of the MSS model that do and do not include BMI, and maintain the EPA assumption that all ages between 5 and 18 are equally responsive to ozone.²⁵

²⁴ Krinsky, I; Robb, AL. (1986), “On Approximating the Statistical Properties of Elasticities,” *The Review of Economics and Statistics*, 68:715-719.

²⁵ In the uncertainty calculations for the model that includes BMI we rely on the coefficients and covariance matrix resulting from the MSS model with BMI that was reported in McDonnell *et al.* (2013). The differences between these model results and the model results reported in McDonnell *et al.* (2012) are minimal.

For each of the simulations below, we took 100 Halton draws (quasi-random number sequences) from the multivariate normal distribution that is defined by the coefficients and covariance matrix from the MSS model, and created 100 alternate sets of the MSS model coefficients.²⁶ For each set of simulated coefficients we also recalculate the age term parameters to preserve the piecewise linear function of age constructed by the EPA in the original APEX calculations (see Table 6.2 in the HREA). These 100 versions of the MSS model coefficients were used for all simulations below in order to allow us to directly compare results across different ozone scenarios.

Our main set of simulations considered three different ozone scenarios for the city of Atlanta in 2010 – 75 ppb, 70 ppb, and 65 ppb. For each of these three scenarios we run 100 APEX simulations, each time replacing the original coefficients from the MSS model with one set of the simulated coefficients generated above. All inputs for these APEX simulations (air quality, meteorology, etc.) were held constant except for the MSS model coefficients.

Each of our APEX simulations had 10,000 simulated individuals. The random seed in APEX was held constant, so we are examining the same 10,000 simulated individuals for each set of MSS model coefficients. Thus, the only source of variation across our 100 APEX simulations is the variation in the MSS coefficients.

Note that in these APEX calculations we are using fewer simulated individuals than used by the EPA, which used 200,000 simulated individuals for each APEX calculation. The EPA used 200,000 simulated individuals for each APEX run in order to reduce the random variance in results (simulation noise) that could arise from generating different sets of simulated individuals. In our simulations we do not need such a large number of simulated individuals, because we hold the random seed constant and eliminate this source of random variation. Using 10,000 simulated individuals still gives us variation across age groups and other individual characteristics, while being more computationally tractable than using 200,000 simulated individuals.

Further, note that since our simulations use a different set of simulated individuals than were used in the HREA, the mean percentages for each age group and decrement will not necessarily exactly match those presented in Table 6B-1. We made some effort to calibrate our random seed to more closely match the percentages in that table, but some differences are inevitable; these differences are typically minor, and do not affect the overall uncertainty that is being found.

²⁶ We use 100 Halton draws to generate our draws of coefficients. Halton draws are quasi-random, and give better coverage of the space we are simulating than truly random draws. We generate one Halton sequence for each coefficient to be simulated, multiply it by the Cholesky decomposition of the covariance matrix to generate the correct variances and covariance, and add the correct estimate coefficient to give each sequence the correct mean. Research in other areas suggests that 100 Halton draws is about as accurate as 1000 random draws in estimating distributions. On this last point see Train, K. (2009), *Discrete Choice Methods with Simulation*, New York: Cambridge University Press.

VI. SIMULATION NOISE WITHIN THE APEX RESULTS

The discussion immediately above reveals a third source of uncertainty in the APEX simulations. The results of each APEX simulation rely in part on a set of random variables that are created by taking draws from probability distributions. These random variables include characteristics of the simulated individuals, such as age, location of residence, and individual responsiveness to ozone (U in the MSS model), as well as characteristics that vary across time, such as activity level. This means that the results of each APEX run will depend in part on the seed used to generate these random distributions, even if all of the input files for the APEX simulation are identical. This random variability is commonly known as “simulation noise,” or “jitter.”

The HREA examines the magnitude of this simulation noise,²⁷ and reports that with an APEX run using 200,000 simulated individuals “the range of results over the 40 APEX runs is less than one percent.” However, this is slightly misleading – the range of results reported is less than one *percentage point*, which represents considerably more variation in results.

In order to quantify this uncertainty more precisely, we examined 20 APEX runs for Atlanta 2010 under the 75 ppb scenario, using 200,000 simulated individuals. In each run we held everything constant except for the random seed, so that any variation across runs is solely due to simulation noise. We then constructed a 90% confidence interval for the simulation noise using the second-highest and second-lowest calculated percentages within each age group and FEV₁ decrement (dropping the largest and smallest values out of 20 gives us the middle 90% of simulated values). We also calculated the standard deviation for each category. The confidence intervals and standard deviation we calculated are presented in Table 3.

Table 3. Estimated Simulation Noise Using 20 APEX Runs

Measure of Simulation Noise	FEV10 Ages 5 to 18	FEV15 Ages 5 to 18	FEV20 Ages 5 to 18	FEV10 Ages 19 to 35	FEV15 Ages 19 to 35	FEV20 Ages 19 to 35	FEV10 Ages 36 to 55	FEV15 Ages 36 to 55	FEV20 Ages 36 to 55
Std. Dev.	0.18	0.08	0.04	0.08	0.03	0.02	0.03	0.01	0.005
90% CI	17.56, 18.11	4.55, 4.78	1.68, 1.82	4.93, 5.20	0.83, 0.92	0.22, 0.26	1.23, 1.31	0.08, 0.12	0.01, 0.03

This simulation noise is not negligible. For the largest percentage in this scenario (ages 5-18, FEV₁ decrement at least 10%) the 90% confidence interval covers more than half a percentage point, and the standard deviation of 0.18 implies a 95% confidence interval of (17.74, 18.44). This 95% confidence interval covers a range of 0.7 percentage points, which corresponds well with the reported range of 0.88 percentage points for this same age group and FEV₁ decrement in

²⁷ HREA, p. 6-44 and *ff.*

the EPA's own tests of simulation noise. (The apparent asymmetry of the 90% confidence interval could be due to the relatively small number of APEX runs used in its calculation.)

Recall that our method for examining statistical uncertainty holds the APEX random seed constant, eliminating simulation noise. Since this simulation noise is based on a random seed that is uncorrelated with the MSS model coefficients, we can treat it as an independent source of uncertainty, and thus it is possible to simply add our estimates of simulation noise and statistical uncertainty together when calculating confidence intervals.

VII. RESULTS OF NERA’S UNCERTAINTY ANALYSES

A. Density Plots

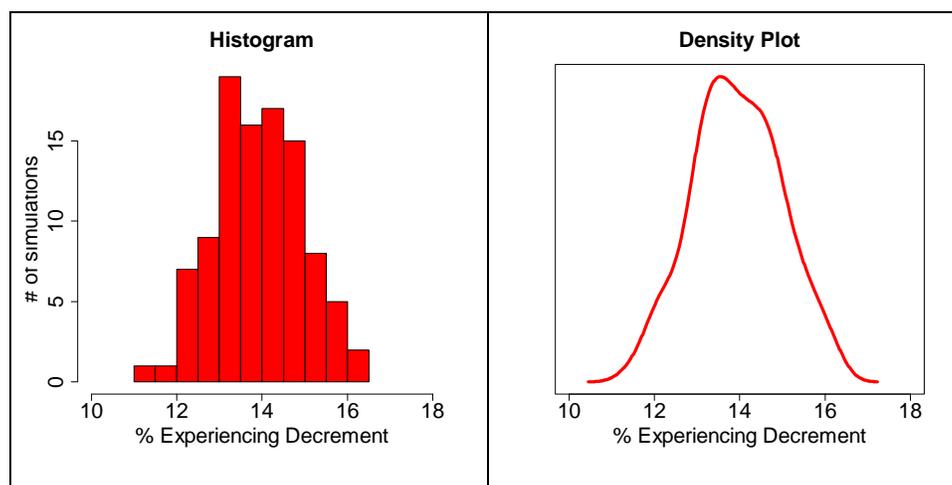
Here we present the results of our uncertainty analysis for the city of Atlanta in 2010, starting with the version of the MSS model that includes BMI and using EPA’s method to extend the model to younger ages (*i.e.*, children less than 18 years of age). Results for the version of the MSS model that does not include BMI are presented in the Appendix as Figures A-1 through A-3. These show the same overall range of uncertainty that we see for the with-BMI results.

For each age group, level of lung decrement, and frequency (≥ 1 , ≥ 2 , or ≥ 6 days), we present density plots that overlay the 75 ppb, 70 ppb, and 65 ppb ozone scenario results. These density plots, created from the statistical uncertainty simulations described above, are our estimate of the probability distribution for the percentage of individuals experiencing a given lung decrement.

As an example, consider the left hand panel in Figure 3, which presents a histogram that describes the results from 100 APEX runs that took statistical uncertainty into account. This histogram shows the variation across these APEX runs in the percentage of individuals in the 5-18 age group that were predicted to have experienced a decrease in FEV₁ of at least 10% for at least one day.

The right hand panel in Figure 3 presents a density plot for the same data. A density plot (or kernel density estimation) is a non-parametric way to estimate the probability distribution of a variable. An intuitive way to think of these density plots is as a smoothed representation of the underlying histograms – a comparison of the histogram and density plot will help to reinforce this understanding. It is easier to observe overlaps in distributions with density plots as compared to histograms, which is why we use them here.

Figure 3. Example of Histogram and Density Plot from One APEX Uncertainty Case (dFEV₁>10% for at least 1 day, 5-18 years old, Atlanta 2010 at 75 ppb)



B. Simulation Results

Our simulation results are grouped by age and level of lung function decrement. The percentage of individuals experiencing a given decrement for one or more days corresponds to the analysis presented in the appendix to Chapter 6 in the HREA, specifically Table 6B-1. The percentage of individuals experiencing a given decrement for 6 or more days is also presented.

Figure 4 presents the density plots for the percentage of individuals in each age group experiencing a lung function decrement of $\geq 10\%$ for at least one, at least 2, or at least 6 days. Age groups are arranged by column (ages 5-18 left, ages 19-35 middle, ages 36-55 right), and frequency of decrement by row (1 day top, 2 days middle, 6 days bottom). Three ozone scenarios are presented – 75 ppb (red), 70 ppb (green), and 65 ppb (blue).

Across all age groups and frequencies of decrements one sees significant overlap in the distributions for each ozone scenario, indicating there is uncertainty as to whether a change in ozone standards would lead to a significant reduction in the percentage experiencing this level of lung decrement.

For example, for the one-day decrement for ages 5-18, 43 out of 100 APEX simulations using the 75 ppb ozone scenario were less than or equal to the maximum value estimated by the 100 APEX simulations using the 70 ppb ozone scenario. Similarly, 23 out of 100 APEX simulations using the 70 ppb ozone scenario were less than the maximum value estimated by the 100 APEX simulations using the 65 ppb ozone scenario.

Despite this overlap, t-tests show that the difference in the means of these distributions are statistically significant. This is unsurprising, since the only changes in the inputs to produce these different distributions were in air quality.

With these estimates of statistical uncertainty we can compare our results to the point estimates of expected lung decrement presented in the HREA. For example, in Table 6B-1 the HREA predicts 18.09% of individuals aged 5 to 18 would experience a lung function decrement of at least 10% at least once during the ozone season at 75 ppb. The corresponding percentage at 70ppb is 14.86%, and at 65 ppb it is 12.10%.

T-tests reveal that the mean of our simulated distribution under the 75 ppb scenario (13.94%) is statistically significantly lower than the percentage predicted in the HREA for both the 75 ppb and 70 ppb scenarios. Put another way, once we account for model specification and statistical uncertainty in this way, the anticipated health benefits of moving to 70 ppb presented in the HREA for this age group may already have been realized at 75 ppb. Similarly, a t-test reveals that the mean of our simulated distribution under the 70 ppb scenario (11.59%) is statistically significantly lower than the percentage predicted in the HREA for the 65 ppb scenario.

An examination of the older age groups shows the opposite pattern, with the expected one-day lung decrement for each ozone scenario statistically significantly higher than the point estimate presented in the HREA.

Overall, we can interpret these results as evidence that this type of model specification uncertainty (whether or not to include BMI in the MSS model) leads to statistically significantly different predictions in the APEX model.

An examination of the simulated percentages of individuals experiencing a lung decrement for 2 or more or 6 or more days reveals a similar pattern, with obvious overlap between the distributions, but with statistically significant differences between the means. The HREA does not provide point estimates for the expected percentages experiencing a decrement for 2 or more or 6 or more days, so the t-tests we calculated above for the one-day decrements are not possible.

Figure 5 and Figure 6 present the simulation results for lung function decrements of $\geq 15\%$ (Figure 5) and $\geq 20\%$ (Figure 6). As with Figure 4, there is a notable amount of overlap between the distributions for each ozone scenario for every age group and frequency of decrement. This overlap becomes especially pronounced when considering decrements for two or more or six or more days, and in some cases the distributions become visually indistinguishable. T-tests reveal that the predicted decrements from our simulations for children (ages 5-18) are significantly lower than those predicted in the HREA, while they are significantly higher for adults.

Figure 4. Probability Densities for dFEV1 > 10%, Atlanta 2010, MSS Model Including BMI

75 ppb in red, 70 ppb in green, 65 ppb in blue

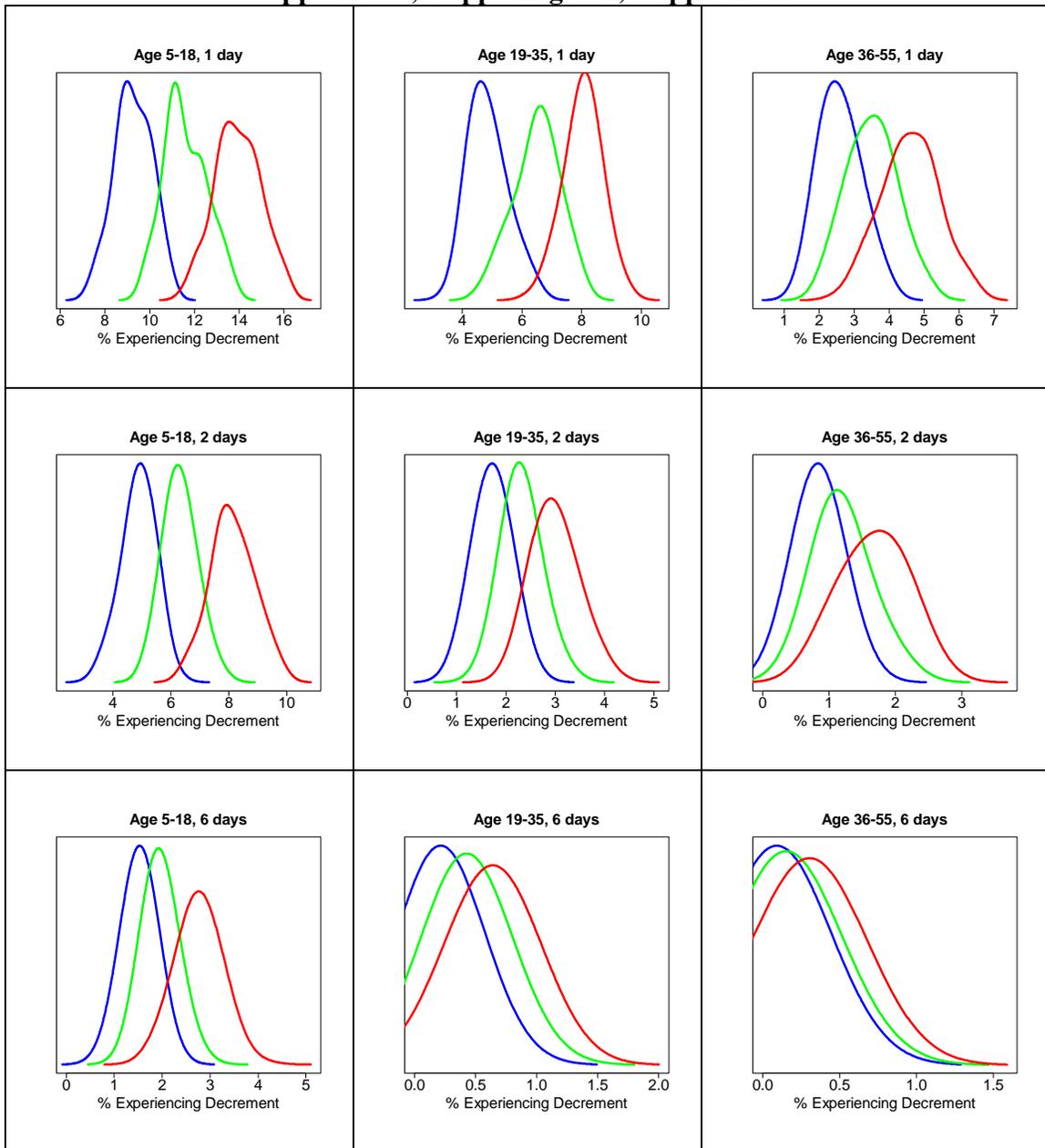


Figure 5. Probability Densities for dFEV1 > 15% , Atlanta 2010, MSS Model Including BMI

75 ppb in red, 70 ppb in green, 65 ppb in blue

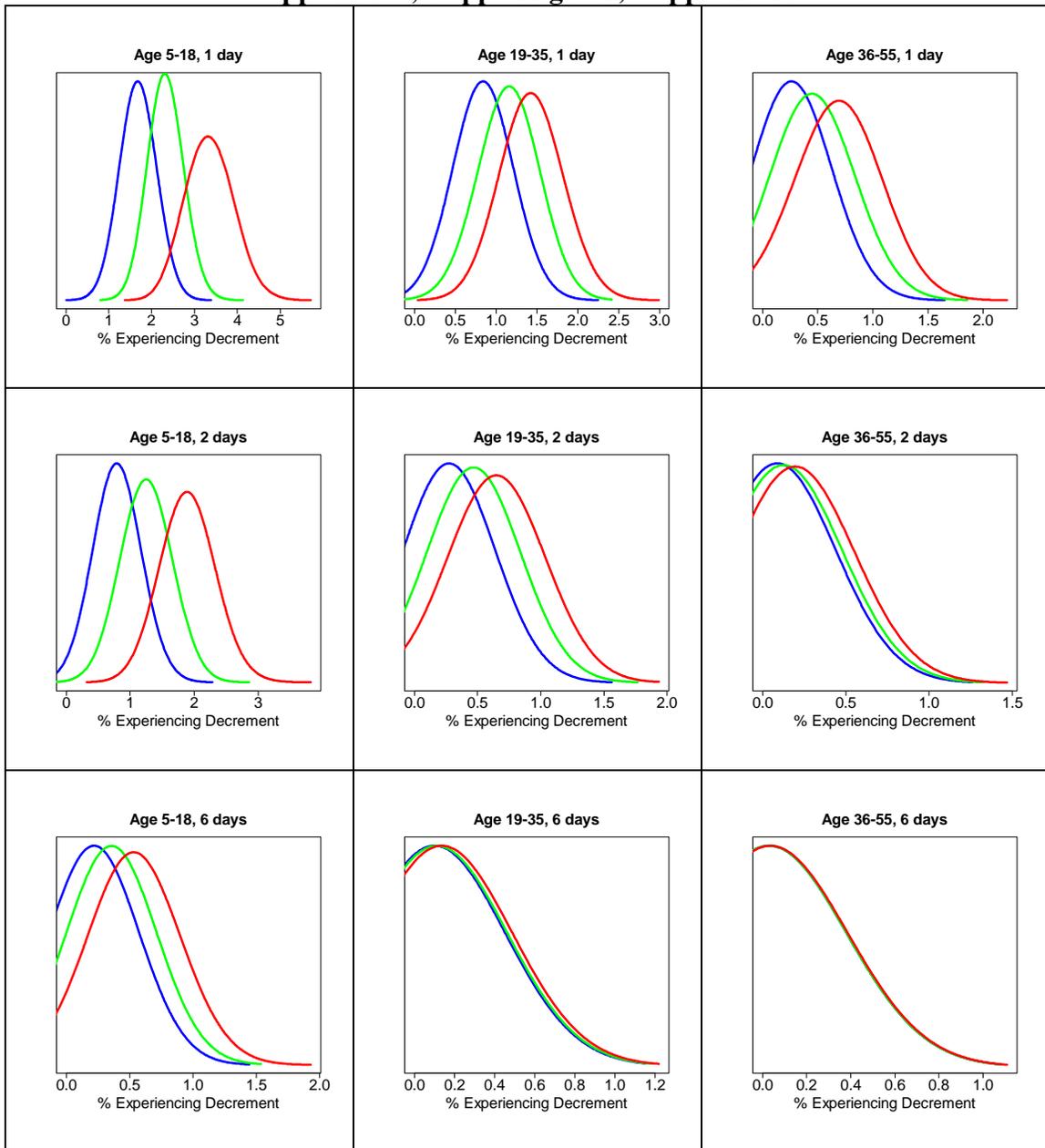
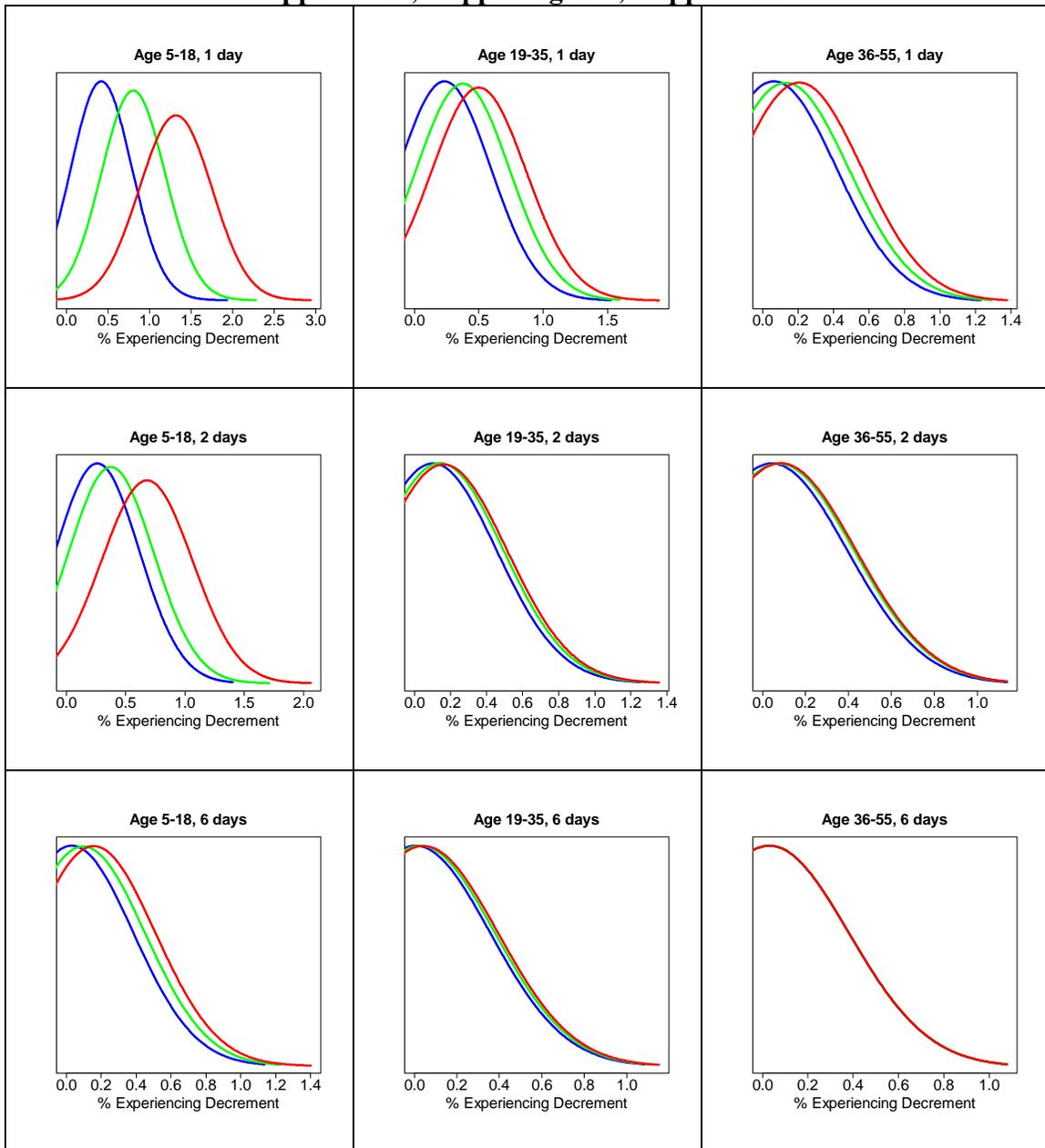


Figure 6. Probability Densities for dFEV1 > 20%, Atlanta 2010, MSS Model Including BMI

75 ppb in red, 70 ppb in green, 65 ppb in blue



Across all three figures the density plots show some overlap with zero for many age groups and ozone concentrations, particularly for the multi-day decrements. This suggests that the percentage of individuals in these categories experiencing a lung function decrement may be zero. That is, there is some non-zero probability that there are no effects on individuals at all in those effects categories. We cannot directly test for the statistical significance of any of these estimated decrements, since it is not possible to estimate a percentage of individuals experiencing a lung function decrement that is less than zero. However, we can use the percentage of our density plots that overlap zero as a measure of the likelihood that a meaningful percentage of individuals in a category will experience a given lung function decrement.

There is no standard way to calculate the area of a nonparametric density plot that falls above or below a certain number, but we can approximate it. We calculated the overlap with zero by assuming that each density plot was a normal distribution – from there we can use the mean and standard deviation of the density plot to calculate the percentage of the distribution that falls below zero. In some cases the long tails of the normal distribution may magnify these numbers, so we conservatively regard an overlap with zero as “non-trivial” if at least 20% of our normal approximation to the density plot falls below zero. Table 4 presents these calculations, and reveals that once statistical uncertainty is taken into account, there is a non-trivial chance that no health benefits will be realized by tightening ozone standards for many of the effects measures (*i.e.*, those shaded in grey)

Table 4. Estimated Probability that Effect Will Not Occur At All, by NAAQS Level and Age (rows) for Each Magnitude/Frequency Effect Category (columns)

(Source: Summary of Density Plots in Figure 4 through Figure 6)

		≥ 1 day per season			≥ 2 days per season			≥ 6 days per season		
		dFEV1 ≥10%	dFEV1 ≥15%	dFEV1 ≥20%	dFEV1 ≥10%	dFEV1 ≥15%	dFEV1 ≥20%	dFEV1 ≥10%	dFEV1 ≥15%	dFEV1 ≥20%
At 75 ppb NAAQS	5-18 years	0%	0%	8%	0%	2%	19%	1%	20%	37%
	19-35 years	0%	4%	24%	0%	21%	37%	21%	41%	47%
	36-55 years	0%	21%	39%	7%	36%	46%	33%	47%	48%
At 70 ppb NAAQS	5-18 years	0%	1%	15%	0%	7%	25%	1%	30%	43%
	19-35 years	0%	9%	30%	1%	29%	40%	27%	42%	48%
	36-55 years	1%	27%	42%	12%	39%	46%	36%	48%	48%
At 65 ppb NAAQS	5-18 years	0%	4%	22%	0%	15%	33%	5%	36%	47%
	19-35 years	0%	15%	33%	3%	33%	42%	34%	46%	49%
	36-55 years	2%	32%	44%	17%	43%	46%	42%	48%	48%

Note that the probability densities presented in this section are based solely on statistical uncertainty, and the differences between these results and the numbers presented in the HREA are based on both statistical uncertainty and model specification uncertainty (specifically switching to the version of the MSS model that includes BMI). Simulation uncertainty has not been included in these results, and would only increase that variance of these distributions and make it more difficult to distinguish between different ozone concentration scenarios.

In order to be sure that these results were not model specific, we also estimated statistical uncertainty for Atlanta 2010 using the version of the MSS model that excluded BMI. These results are very similar to the results presented here (although shifted up for children and down for adults). To confirm that Atlanta was not an unusual case, we also estimated statistical uncertainty for Denver 2010 using the version of the MSS model that includes BMI, and again, the results were very similar. All of these results are presented in the Appendix.

VIII. IMPLICATIONS OF UNCERTAINTY ANALYSIS RESULTS

The estimates of lung function benefits presented in the HREA do not include any measures of uncertainty. As we have seen in the analysis above, this uncertainty is considerable. Statistical uncertainty in the estimation of the MSS model and model specification uncertainty in selecting which version of the MSS model to use both have large influences on the HREA results for lung function decrement incidences.

In particular, model specification uncertainty has a dramatic impact on the potential lung function benefits for children, with the estimates of lung function decrement from the version of the MSS model that includes BMI at 75 ppb statistically significantly smaller than the estimates produced using the version of the MSS model that excludes BMI at 70 ppb

The opposite pattern is observed for older age groups, with the BMI version of the MSS model producing estimates of lung function decrement that are significantly higher than those produced by the version of the MSS model that excludes BMI at an ozone level that is 5 ppb higher.

Put another way, model specification uncertainty has more influence on the estimates of lung function decrement than a 5 ppb change in the ozone NAAQS. As both versions of the MSS model are regarded as credible by researchers, this represents a dramatic source of uncertainty that is completely ignored in the HREA. Alternate assumptions about how to extend the MSS model to children that are based on observed data also lead to dramatically lower estimates of lung function decrement for children.

The statistical uncertainty that arises from estimating the MSS model on small samples of individuals also has an obvious influence on the estimates of lung function decrement, with the probability distributions for percentage of the population with a given lung function decrement clearly overlapping across ozone scenarios. Further, for older age groups and higher levels of lung function decrement, a large number of our simulations predict no negative health effects at all.

Any decision to tighten ozone standards must be informed by an analysis of the uncertainty in the benefits such a tightening would bring. Unfortunately, this uncertainty was ignored in the HREA. Our analysis reveals that uncertainty plays a major role in our understanding of the potential benefits from tightening ozone standards, and in fact the benefits may be non-existent for some age groups. These represent meaningful limitations in the scientific evidence and information that affect the strength of inferences that can be drawn regarding the lung function decrement risk estimates that are under consideration in the Proposed Rule.

APPENDIX

Figure A - 1. Probability Densities for dFEV1 > 10%, Atlanta 2010, MSS Model Excluding BMI

75 ppb in red, 70 ppb in green, 65 ppb in blue

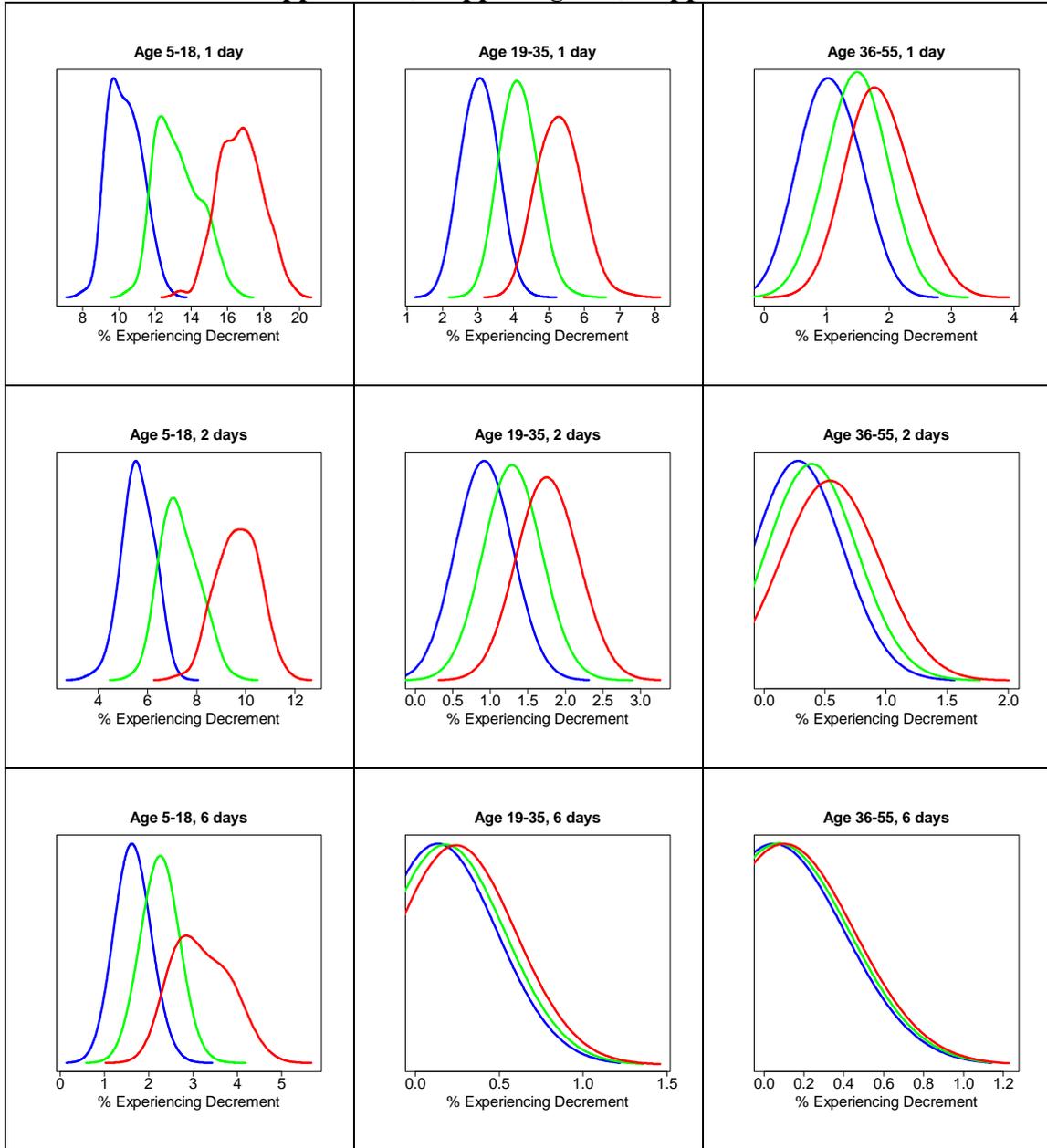


Figure A - 2. Probability Densities for dFEV1 > 15%, Atlanta 2010, MSS Model Excluding BMI

75 ppb in red, 70 ppb in green, 65 ppb in blue

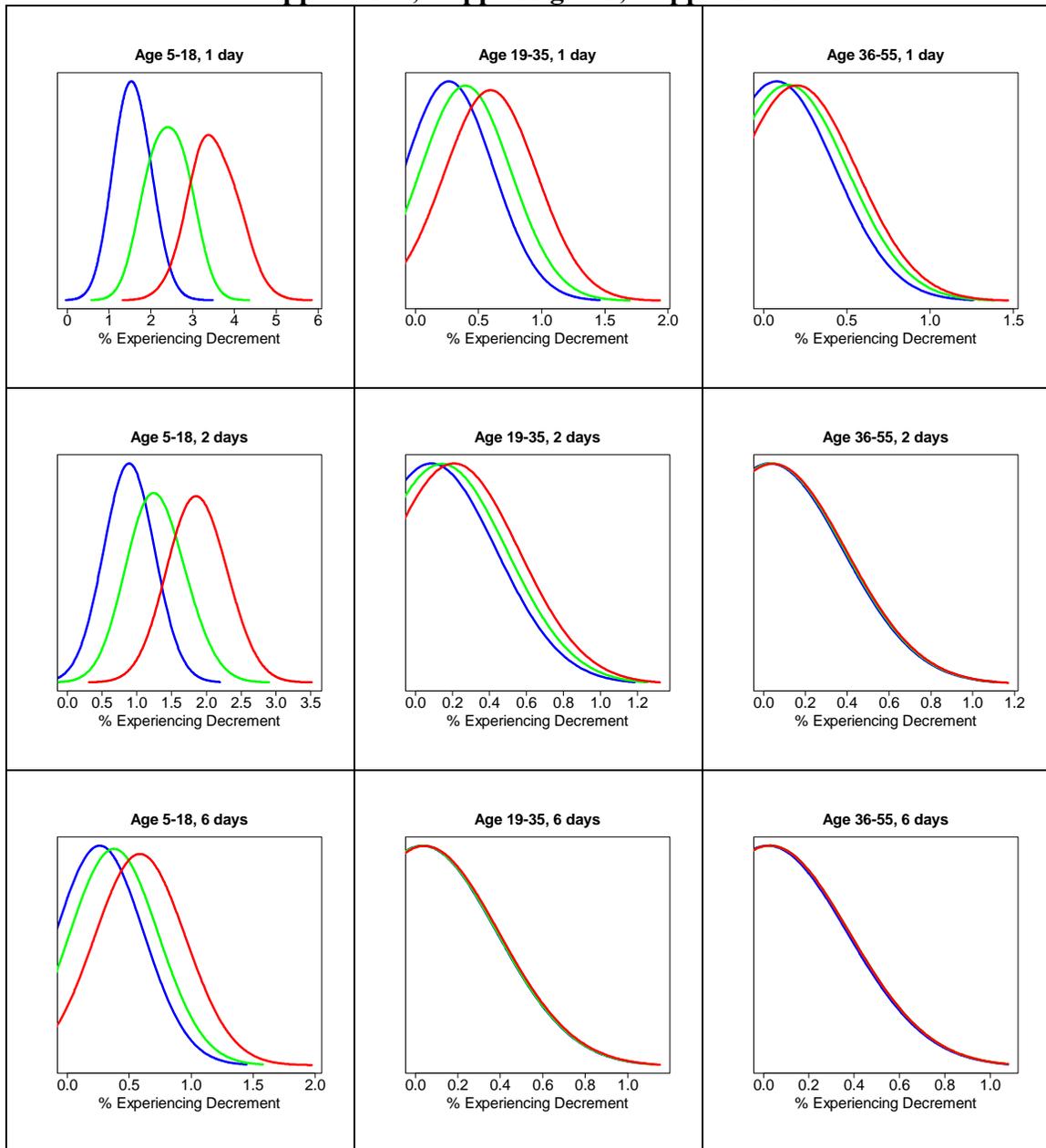


Figure A - 3. Probability Densities for dFEV1 > 20%, Atlanta 2010, MSS Model Excluding BMI

75 ppb in red, 70 ppb in green, 65 ppb in blue

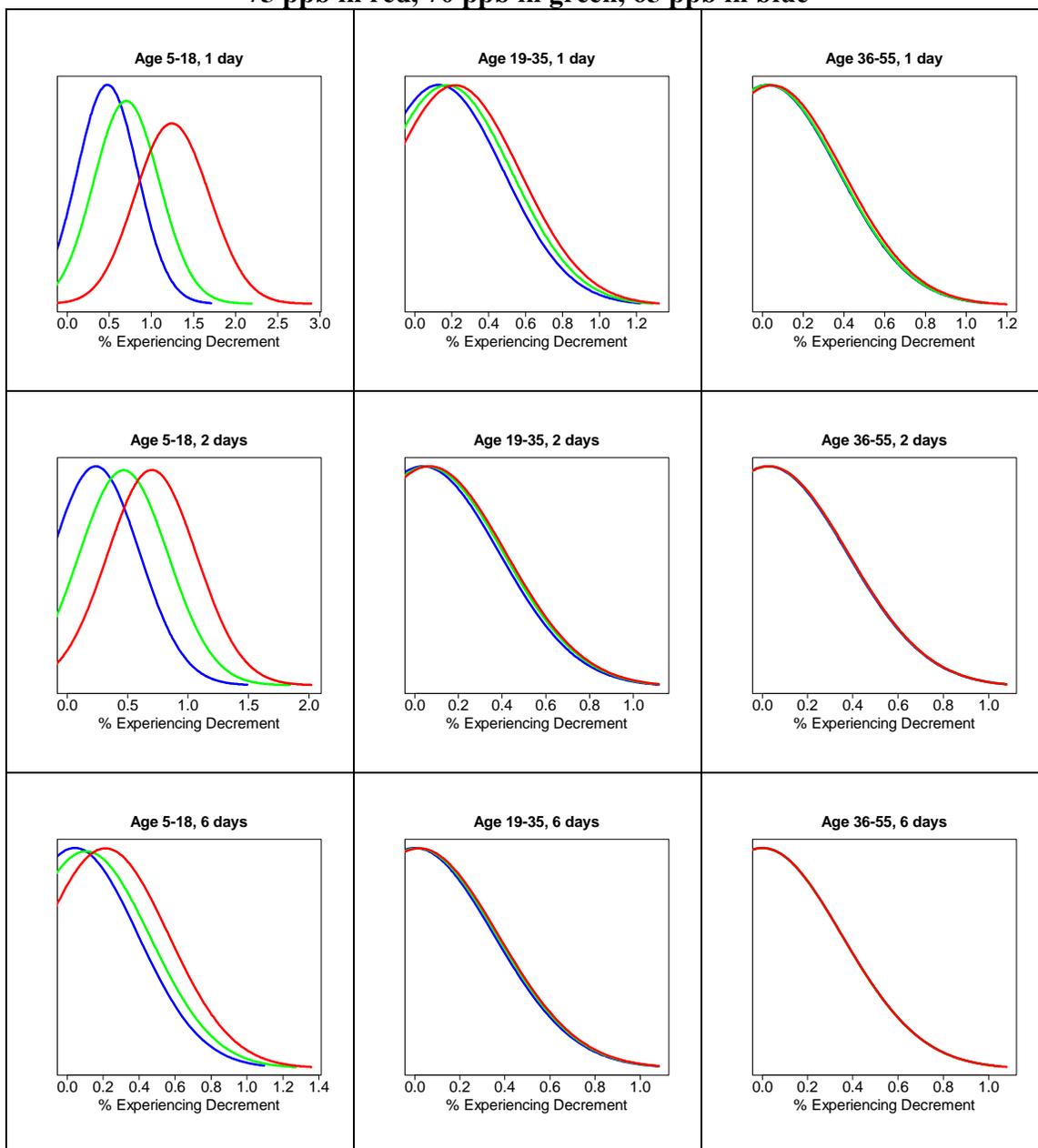


Figure A - 4. Probability Densities for dFEV1 > 10%, Denver 2010, MSS Model Including BMI

75 ppb in red, 70 ppb in green, 65 ppb in blue

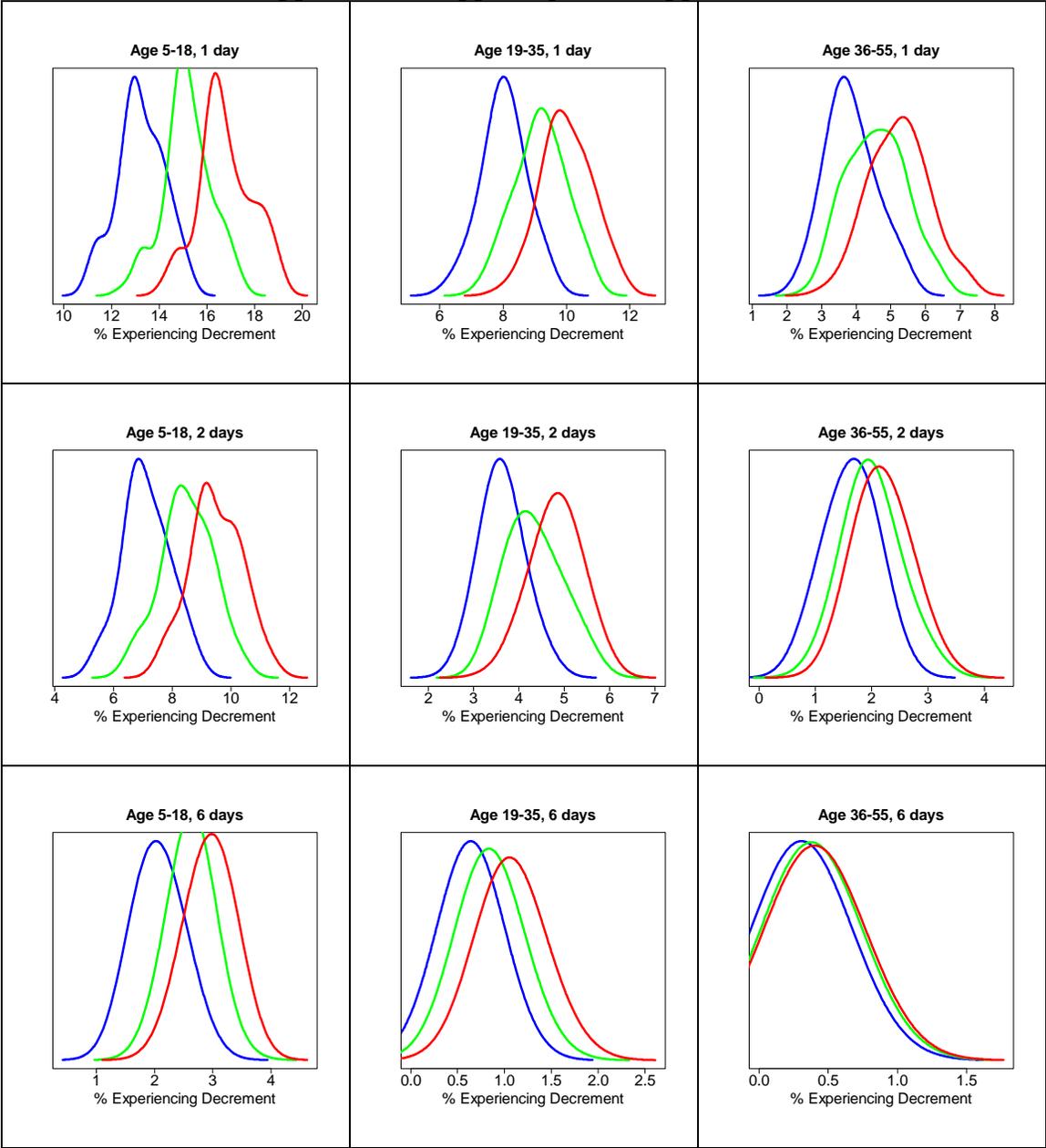


Figure A - 5. Probability Densities for dFEV1 > 15%, Denver 2010, MSS Model Including BMI

75 ppb in red, 70 ppb in green, 65 ppb in blue

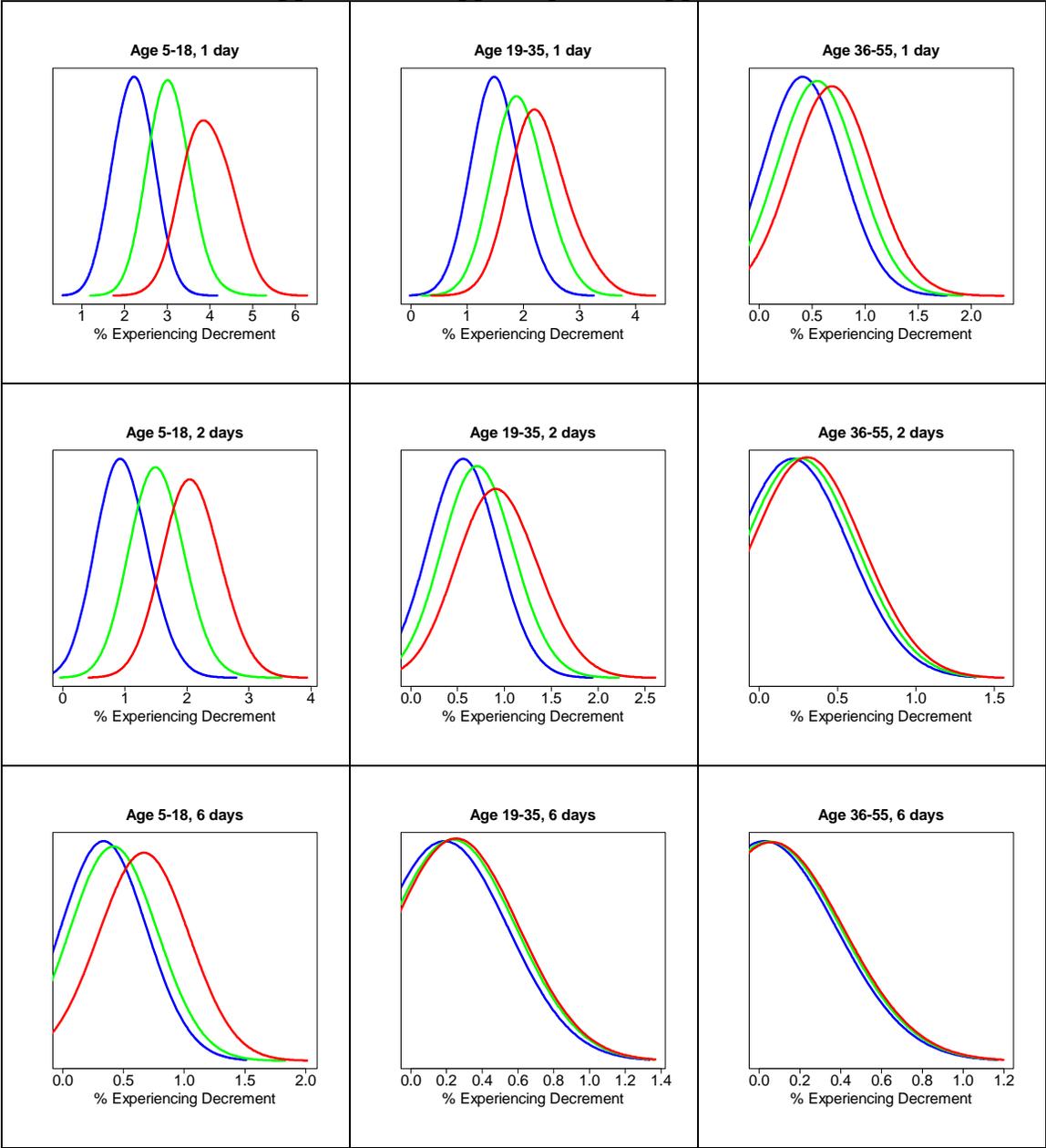
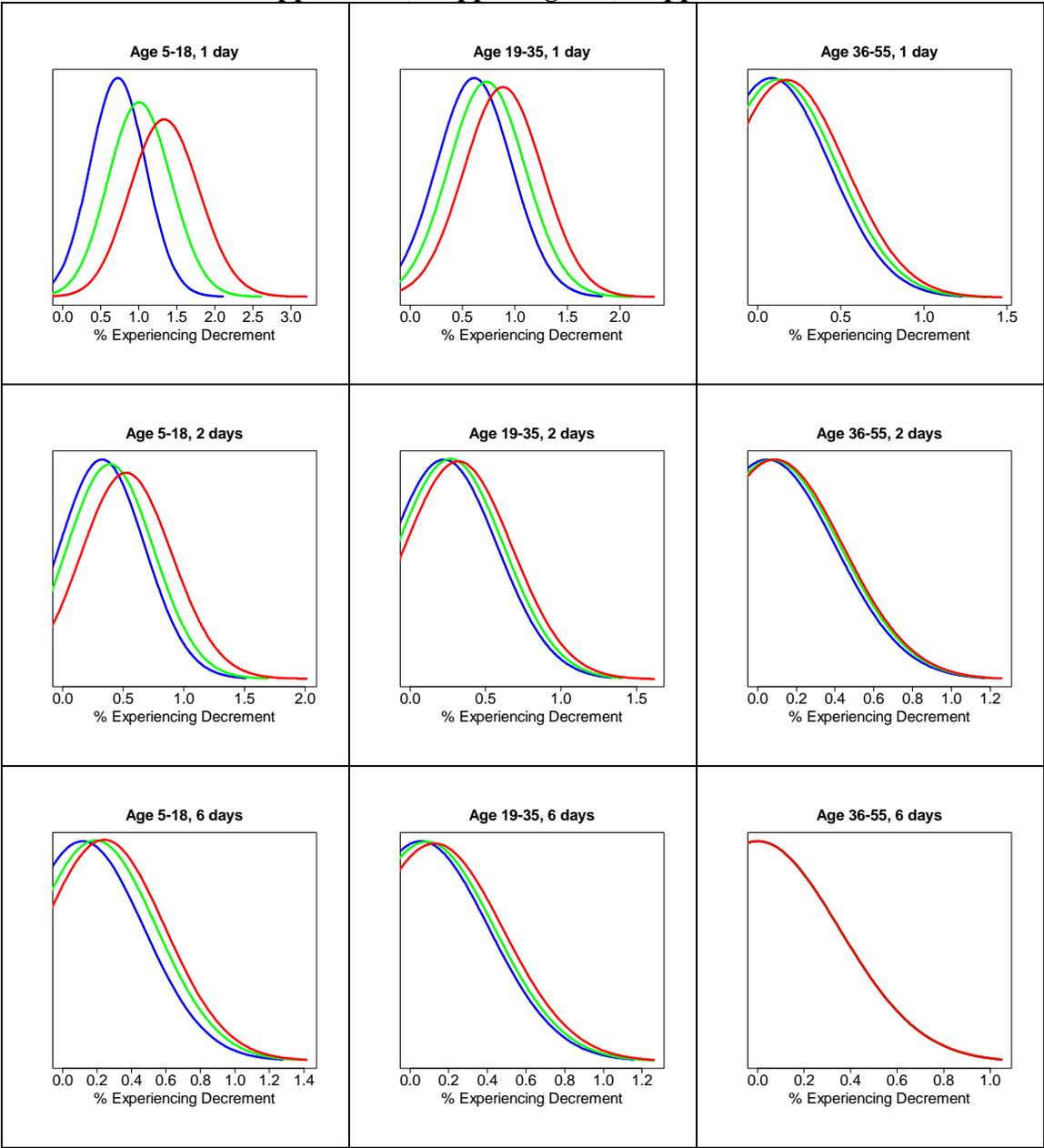


Figure A - 6. Probability Densities for dFEV1 > 20%, Denver 2010, MSS Model Including BMI

75 ppb in red, 70 ppb in green, 65 ppb in blue





NERA
ECONOMIC CONSULTING

NERA Economic Consulting
1255 23rd Street NW
Washington, DC 20037
Tel: +1 202 466 3510
Fax: +1 202 466 3605
www.nera.com