18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Examining the impact of Ozone on lung health in a vulnerable population: a Bayesian modeling strategy.

Al Bartolucci[1], Carol Ballinger[2], Sejong Bae[3] , Karan P. Singh[3]
Michelle V. Fanucchi[2] and Edward Postlethwait[2]

[1] Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama 35294-0022, USA
[2] Department of Environmental health Sciences, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama 35294-0022, USA
[3] Department of Biostatistics, School of Public Health, University of North Texas Health Science Center, Fort Worth, Texas 76107-2699, USA

Abstract: The fundamental rationale for this research is that Air pollution continues to be a major health concern not only for its impacts on morbidity but also that acute mortality may be directly associated with exposures. Mounting evidence suggests that the postnatal lung is especially sensitive, receives a disproportionate dose, and that exposure may perturb normal lung development and lead to long-standing alterations in lung function, issues sufficiently noteworthy to have attracted the attention of researchers and health care workers alike. Notable is that these findings parallel observations by research teams that demonstrate remarkable exposure-induced impacts on normal growth and development. These, other observations, and a novel programmatic paradigm afford investigators and their associates on this project unique opportunities to develop new insights regarding susceptibility to environmental insult and how exposure alters the biology of the lung and to develop tools for evaluating its affect and predict outcomes.

Thus clearly, Ozone (O_3) remains an important public health problem especially with regards to sensitive populations such as children and infants with developing lungs. Numerous recent studies document that exposures induce both short and potentially long term impacts on the childhood lung. Responses to a given exposure demonstrate marked heterogeneity with respect to age, anatomic site, species, and exposure history. Infants may be particularly at risk due to a greater inhaled dose rate. Respiratory tract surfaces are covered by an aqueous layer (epithelial lining fluid; ELF) that inhaled gases first encounter and is a complex mixture containing significant concentrations of small molecular weight antioxidants (*e.g.*, ascorbic acid (AH_2), glutathione, and uric acid), lipids, and proteins. In this study we use an animal model controlling for age and anatomic site and examine the effect of ozone on the antioxidant values in the ELF after varying exposures. We compare the resultant antioxidant levels by two exposure levels to filtered air which is the control level. Because of the variation in these antioxidant levels across lung tissue sections as well as over time a Bayesian modeling approach with diffuse priors applies well to this data. Also the data is transformed to a log scale allowing us to assume underlying normality of the parameters of interest. Our last objective will be to determine if there is a difference in levels of these antioxidants examining the passage of exposed air through the right caudal lobe and the geographic compartments (sections) of this tissue.

The approach here is to apply a Markov Chain Monte Carlo strategy to coherently combine prior diffuse information on the normal response (level of antioxidant) with the distribution of the response probability over 3 levels of exposure (2 ozone exposure strategies plus filtered air) and derive posterior means of antioxidant level for exposure group and compare these exposure results further over regions of the lung to determine level of change in the antioxidant values over these regions. We will examine the credible regions for the parameters and the convergence properties as well. The posterior values of the parameters of the three exposures as well as the combined posterior parameters of the entire sample will be evaluated and examined. The overall goal is naturally to determine if there is significantly greater disparity in the exposure levels versus the filtered air in each of the antioxidants. The design of the study allows for section and exposure interaction results which will enhance the understanding of the mechanism of O_3 in the population of younger individuals.

***Key Words**: Ozone, lung, antioxidants, exposure, Bayesian*

Bartolucci *et al.*, l Examining the impact of Ozone on lung health in a vulnerable population: a Bayesian modeling strategy.

1. INTRODUCTION

Tropospheric O_3 has been identified as a significant risk factor in the development and exacerbation of lung disease.(Lippman, 1993, Kim,2004, Gauderman et.al. 2004) O_3 produces diverse pulmonary pathophysiology, including inflammation, compromises in immune function, and chronic reorganization of the normal lung architecture(Lippman, 1993**) .** Although the spectrum of responses displays marked heterogeneity relative to species, age, anatomic site, disease, and exposure-history, the explanation or these pronounced susceptibility variations has remained largely undefined. Characterizing the bases of divergent susceptibility in people is complicated by the inability to perform invasive mechanistic studies in target populations, such as during postnatal growth. Despite publication of ~ 2000 reports related to humans and O_3, how environmental exposures impact the development of lung disease, affect lung growth and development, and exacerbate the compromised lung remain poorly understood. Many of our insights regarding O_3 toxicology come from studies that are not directly applicable to most human exposures or biology.

Importantly, the utility of this data is uncertain because our ability to predict human health effects is limited by the lack of understanding of how animal studies and differing exposure modalities may or may not translate to the human situation. Key research identified as critical to improving health risk assessments related to O_3 exposure **(**US Environmental protection Agency, 1998**)** includes the need to:

♦ Determine the biological roles of time course and temporal patterns of acute and chronic O_3 exposure.
♦ Determine the role of O_3 dose versus species sensitivity, mechanisms of action, and toxicity mediators.
♦ Expand research on extracellular lung lining fluids and augment respiratory tract anatomical data.
♦ Extend dosimetry models to include O_3 -substrate interactions and lung growth during chronic O_3 exposure.
♦ Formulate and parameterize local dose-response relationships as a basis for quantitative extrapolations.
♦ Improve understanding of the relationships between short-term O_3 effects and long-term health risk.

It has been recently emphasized that while children represent 30% of the population, only a very small percent of research funding is directed towards children's health. Within the past several years, important questions (Mustafa, 1990) have been raised with regard to:

• Are exposure studies appropriate for modeling children based on their biology, behavior, and activities?
• Do we have adequate knowledge of dose/response and mechanisms to assess risk in infants and children?
• Do exposures earlier in life have the same health consequences as exposures later in life?

Despite the extensive laboratory animal and human research on the health effects of O_3, the understanding of its potential long term human health consequences, especially in the developing lung, remains poorly defined. While many responses that occur due to O_3 exposure have been characterized (Thurston and Ito, 2001), the mechanisms that govern host-, age-, anatomic site-, and exposure history-dependent differences in susceptibility have not been adequately characterized. This report will focus on the development of models that explain and predict the exposure-induced effects on tissue, organization, and susceptibility.

2. METHODS

This is part a major interdisciplinary, multi-institution program project focused on delineating the impacts of postnatal O_3 exposure on lung biology, respiratory tract architecture, and susceptibility. It involves the collaboration of four major participating universities in the United States and a primary laboratory for imaging and 3 dimensional reconstructions. Fourteen rhesus macaques were exposed to 0.5 ppm of ozone for extended periods of time for either one cycle of 5 days x 8 hrs/days (5 animals), or 11 cycles of 5 days x 8 hrs/day with 9 days off intermittent with the cycles (4 animals). The first exposure of 11 cycle begins at 30 days of age, and so will be 6 months old at necropsy, or we hold the 1 cycle exposure until the last week of the respective age group we are studying. The control sample (5 animals) were animals with no exposure or to filtered air only. As one can see, the sample size is quite small as is the case in the ethical inclusions of animals for these studies. However, the anticipated differences in the exposed versus non exposed groups in the levels of antioxidants (e.g., ascorbic acid (AH_2), glutathione, and uric acid) measured in nmoles/mg protein was anticipated enough to allow for adequate power in the posterior comparative analysis of these levels across exposure groups. This was especially the case since the readings were replicated across 8 sites throughout the right caudal lobe of the lung. They were in anatomic order; the trachea,

generation 1 and 2, branch points, generation 3 and 4, mid small branch, distal axial, terminal respiratory bronchus, and parenchyma.

Statistically, we compare the resultant antioxidant levels by two exposure levels to filtered air which is the control level. Because of the variation in these antioxidant levels across lung tissue as well as over time a Bayesian modeling approach with diffuse priors applies well to this data. Also the data is transformed to a log scale allowing us to assume underlying normality of the parameters of interest. Test of normality was done by traditional methods such as Shapiro Wilk and it proved to fit this data well. The approach here is to apply a Markov Chain Monte Carlo (MCMC) strategy to coherently combine prior diffuse information on the normal response (level of antioxidant) with the distribution of the response probability over 3 levels of exposure (2 ozone plus filtered air) and derive a posterior mean of antioxidant level for exposure group and compare these exposure results further over regions or sections of the lung to determine level of change in the antioxidant values over these regions. We will examine the credible regions for the parameters and the convergence properties as well. The posterior values of the parameters of the three exposures as well as the combined posterior parameters of the entire sample will be evaluated and examined. The overall goal is naturally to determine if there is significantly greater disparity in the exposure levels versus the filtered air in each of the antioxidants.

We let X_{ij} represent the log level of an antioxidant and have a normal distribution,

$$x[ij] \sim \text{Norm}(\text{lambda}[i],\text{tau}), \tag{1}$$

where i=1,2,3 for the three exposure levels, 1=cycle 1, 2=cycle 11 and 3=filtered air and $j=1,\ldots,n_i$, for each observation of the antioxidant within exposure level, i. The prior structures on the parameters, lambda and tau are normal and gamma respectively, i.e.

$$\text{lambda}[i] \sim \text{Norm}(0,0.001) \tag{2}$$

$$\text{tau} \sim \text{Gamma}(0.001,0.001),$$

assuming both priors to be uninformative. We use the MCMC procedure in a simple one way analysis of variance to compare the three mean exposure levels within the three antioxidants across the entire lung. This procedure will yield posterior means and variances for each of the antioxidants in each of the exposure groups. Should we find overall posterior significance comparing the three exposure means via the posterior F statistic we then follow the same procedure for post hoc pair wise comparisons of the antioxidants. This initial step is fairly straight forward.

The next issue to be addressed and perhaps more informative is the change of the antioxidant value within exposure level as one maps the path of the antioxidant from the trachea which is the first section to the parenchyma which is the eighth and last section of the lung compartment This hierarchical model will take the form:

$$Y_{i\,jk} \sim \text{Normal}(\text{alpha}_i + \text{beta}_i(x_{j\,k}), \text{tau.c}) \tag{3}$$

$$\text{alpha}_i \sim \text{Normal}(\text{alpha.c}, \text{tau.a})$$

$$\text{beta}_i \sim \text{Normal}(\text{beta.c}, \text{tau.b})$$

$$\text{tau.c} \sim \text{Gamma}(0.001,0.001)$$

The i and the j are defined as above representing the exposure and the observation count within the exposure. The k , (k=1,….,8) represents the ordinal count from the first lung section to the eighth. Here the $Y_{i\,jk}$ is the antioxidant value for the jth value of ith exposure at position or lung section k. The equations apply separately for each of the three antioxidants. Also recall the normality assumption was validated using the log antioxidant value which we in fact do. Also note in the MCMC coding, the tau.c is given a gamma prior which translates into a non informative prior for the variance, sigma=1/tau.c. It is appropriate now to address the other parameter values. The alpha_i represents the contribution if the ith exposure, i=1,2,3. The beta_i represents the gradient of the ith exposure as the air

moves through the lung from the first to the last section. The interest here is to determine that as one maps the progress of the air through the lung is there a significant change in the antioxidant level. The alpha.c , tau.a, beta.c, and tau.b are the hyper parameters in the hierarchical model and are given the hyper priors,

$$\text{alpha.c} \sim \text{dnorm}(0.0,1.0\text{E-}6) \quad (4)$$

$$\text{tau.a} \sim \text{dgamma}(0.001,0.001)$$

$$\text{beta.c} \sim \text{dnorm}(0.0,1.0\text{E-}6)$$

$$\text{tau.b} \sim \text{dgamma}(0.001,0.001).$$

The strategy is thus to determine the role of the exposures on the antioxidants in the equations (1) and (2) above by examining the posterior mean differences of lambd[i], i=1,2,3 for each of the antioxidants. If the overall posterior F is significant for testing the composite hypotheses H_0: lambda[1]=lambda[2]=lambda[3] then the MCMC procedure will allow us to then determine the posterior credible regions for delta[i] which represent the difference of each lambda[i] from the common underlying posterior distribution of the lambda[i] 's under the assumption of H_0 . This will give us some idea of possible differences of the exposures on the antioxidants overall throughout the lung. Although this is certainly of interest, we will then apply the strategy of equations (3) and (4) to determine if there is a significant gradient, if any, and which exposures are most strongly associated with the gradient of change throughout the lung.

3 RESULTS

The first antioxidant we consider using equations (1) and (2) is glutathione or GSH. Table 1 gives the results overall considering the global hypothesis tested by the F and the post hoc results comparing each of the exposure levels.

One can see from Table 1 that the overall posterior F statistics rather large as the posterior mean of 14.74 yields a plausible posterior probability of the null hypothesis of equal exposure values of less than 0.01. The table headings are the posterior Mean, Standard deviation and lower 2.5% and upper 97.5% value of the posterior credible interval. One can see for the post hoc analysis that there are differences for 11 cycle vs. 1 cycle and 11 cycle vs. filtered air, but not for 1 cycle vs. filtered air. The results for ascorbate (AH_2) are similar as seen in Table 2. The posterior F of 6.95 will result in a posterior probability of equal exposure means of about 0.02. One can see that the 95% posterior credible region does not contain one. Clearly as above for GSH it appears that the exposure effects differ for 11 cycles compared to 1 cycle and 11 cycles versus filtered air, but not for 1 cycle versus filtered air. Figure 1 gives the posterior density for the difference of the 11 cycle vs 1 cycle (labeled delta[2], with a 5000 iteration burn to compute)comparison and as one can see a zero value is certainly not plausible in this case. The same results were generated for uric acid or UA. However, the overall posterior F value was 2.452 with a posterior 95 credible region of (0.096, 7.449) in which case there was very weak evidence against the null hypothesis of similar posterior mean exposure values for UA. In summary considering the entire lung and the exposure results it appears that only GSH and AH_2 are significant.

Table 1. Results for glutathione, GSH.

GSH	Mean	SD	2.5%	97.5%
F	14.74	5.562	5.417	26.97
11 vs. 1	-0.193	0.056	-0.265	-0.123
11 vs. FA	-0.170	0.059	-0.251	-0.092
1 vs. FA	0.037	0.066	-0.055	0.126

Table 2. Results for AH2 ascorbate.

AH_2	Mean	SD	2.5%	97.5%
F	6.95	3.651	1.40	15.28
11 vs. 1	0.183	0.051	0.082	0.284
11 vs. FA	0.129	0.050	0.030	0.228
1 vs. FA	0.048	0.050	-0.049	0.146

The next task is to determine if there is plausible evidence in the data for a non-zero gradient of activity from the initial lung section to the last section over each exposure level. We start by considering the antioxidant GSH. As seen in Table 3 the significant results for the gradient are given for the 95% and 90% credible regions. One will note that the 1 cycle and filtered air are similar in values which are consistent with the results derived above for the

comparison of antioxidant levels for exposures in the entire lung. Table 3 gives the posterior density of the coefficient of the gradient value for exposure level of GSH through the lung sections.

One can see, for example, that the value of the posterior mean for cycle 11 is 0.101 and the value of zero or no slope is highly unlikely. All this does is demonstrate a possibly positive change in the gradient for GSH as one maps the progress of the air passage through the lung for this exposure. When one considers the 90% credible bound in Table 2 we note that for all three exposures the posterior interval for the mean gradient coefficient does not contain zero, although the gradient itself is not very steep. Also keep in mind that we are dealing with the log of the GSH values at each exposure level and this may contribute to the fact that the coefficient values are not very pronounced. It may be of interest to know the values of the parameters serving as the constant term, alpha, for each of the exposures. Using the MCMC technique the solution for these parameters are alpha[exposure 11]=1.366 for the 95% region with bounds (0.9364,1.74). For the 90% region the values with their respective interval bounds are:

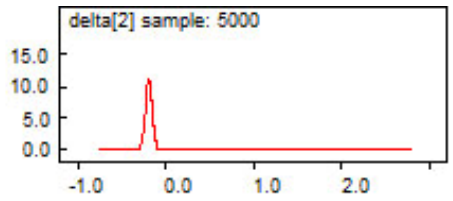

Figure 1. Posterior density of 11 vs. 1 cycle

alpha[exposure 1]=1.374 (1.037,1.681),
alpha [exposure 11]=1.592(1.255,1.951), and
alpha[exposure FA]=1.395(1.054,1.702).

Table 3. Results for GSH gradient.

Cycle	Gradient	Interval	Lower	Upper
11	0.101	0.95	0.016	0.187
1	0.076	0.90	0.012	0.144
11	0.101	0.90	0.030	0.121
FA	0.074	0.90	0.011	0.141

A similar examination of the AH_2 antioxidant yield similar results as was seen with the 90% GSH gradient considerations. However, there was no significant gradient at the 95% level of consideration. At the 90% level the evidence of any gradient is rather weak with the 11 cycle exposure having a mean coefficient of 0.074 with credible interval (0.0025, 0.1458). Obviously not equal to 0. However, back transforming from the log to the base units, the upper bound of the gradient coefficient is 1.157. The exposure values of 1 cycle and FA show coefficient values and intervals as 0.069 (0.001, 0.1418) and 0.072 (0.0043, 0.1436), respectively. Solving for constant term in the model, we have:

alpha[exposure 1] = 1.064 (0.725,1.419),
alpha [exposure 11] = 1.209(0.870,1.583) and
alpha[exposure FA] = 1.115(0.777,1.448).

Examining the uric acid (UA) antioxidant yield similar results as was seen with the previous AH_2 gradient considerations. Likewise, there was no significant gradient at the 95% level of consideration. At the 90% level the evidence of any gradient is rather weak with the 11 cycle exposure having a mean coefficient of -0.0287 with credible interval (-0.07093, 0.01429). Obviously possibly equal to 0. However, back transforming from the log to the base units, the upper bound of the gradient coefficient is 1.0147. The exposure values of 1 cycle and FA give coefficient values and intervals as -0.03172 (-0.0737, 0.01097) ,possibly 0, and - 0.045 (-0.08679 -0.00188), respectively. Solving for the constant term in the model, we have:

alpha[exposure 1] = -0.7202 (-0.9233, -0.5236),
alpha [exposure 11] = -0.7115 (-0.9131,-0.5063) and
alpha[exposure FA] = -0.7158 (-0.9169,-0.5136).

Table 4. Probabilities of Gradient Effects.

Antioxixdant	Cycle	Interaction
AH_2	0.0413	0.5321
GSH	0.0230	0.7209
UA	0.1101	0.7337

Table 4 lists the summary probabilities for the Analysis of Covariance ANCOVA to determine if the gradients are parallel or differ over the three antioxidants. As confirmed by the initial ANOVA results above, there was some significant cycle effect over the sections except for UA. Not shown here, we considered the probability of a section effect in the ANCOVA. Clearly there was not, as all these probabilities (0.2330 for AH_2, 0.1145 for GSH , 0.9803 for UA) show that the passage of air through the lung sections was not significantly different in antioxidant levels across the sections which is consistent with the low gradient levels of the preceding analysis. The interaction column in Table 4 shows that the gradients, whether or not they were pronounced, were parallel and not different over the lung sections for the cycles.

This last exercise is certainly consistent with the preceding analysis in that having taken into account the section effect the gradients were somewhat similar as seen by the beta values represented by the gradient column in Table 3 for GSH and the accompanying discussion of AH_2 and UA which followed.

4. DISCUSSION AND CONCLUSIONS

One can see that the cycle of administration does, in fact, have an effect especially on the GSH and AH_2 antioxidants. For most of our analyses the UA was not affected. Thus when considering the entire lung we were able to differentiate in the cycle effect. When we considered the change in antioxidant value through the 8 lung sections, there was only a posterior significant gradient pattern for GSH at the 11 cycle delivery at the 95% probability level. However such was not the case for the other antioxidants. The gradient effect seemed somewhat evident only at the 90% posterior probability level for GSH and AH_2 at all cycle delivery levels – even filtered air. Also because of the posterior probability of the interaction effect in a covariance structure it is clear that the patterns of gradient were fairly parallel for all cycles and antioxidants. This is also seen by the similarity of the beta coefficients for the gradient in column 2 of Table 3 for GSH and the ensuing discussion of AH_2 and UA, although the UA clearly had only a weak cycle effect in Table 4 which is consistent with the previous analyses. One has to also be aware of the fact that in most studies of this kind our sample size is rather limited. As one gathers more data and pursues these types of analyses the subtle and not statistically apparent differences may become more statistically apparent.

Our motivation for using the Bayesian approach was that we were able to impose rather vague and somewhat dispersed prior variation on the parameters of interest as this was the case with our data. That is to say that the variation in our sample of antioxidants was rather large. One is tempted to classically look for outliers and then eliminate them. However, without good experimental rationale for doing so, this really makes no sense. Thus all the data contributes to our approach. Computationally the use of MCMC allows us to numerically circumvent high dimensional integrations and approximate the solutions with fairly good accuracy. Several authors have pointed out that a major limitation towards more widespread implementation of Bayesian approaches is that obtaining the posterior distribution often requires the integration of high-dimensional functions. This can be computationally very challenging, but several approaches have been proposed by several investigators (Smith 1991, Evans and Swartz 1995, Tanner 1996) and more recently in text form (Gelman, Carlin, Stern and Rubin, 2000). We focused in our work on Markov Chain Monte Carlo (MCMC) methods, which attempt to simulate direct draws from some complex distribution of interest. MCMC approaches are called such because one uses the previous sample values to randomly generate the next sample value, generating a Markov chain (as the transition probabilities between sample values are only a function of the most recent sample value). Today's modern computing technology allows us to take this approach. Like most algorithms in which iterative procedures are required one must also take care to be sure that initial values are reasonable and that the underlying model is realistic for the application proposed. Some limitations of the MCMC are fairly well described (Anrieu,de Freitus, Doucet, and Jordan, 2003). However, for most applications the convergence pattern can be examined and accuracy to final posterior parameter estimates be well established. For our application in this work, stable estimates were derived easily between 5000 and 10,000 iteration burns.

ACKNOWLEDGEMENTS

This research was supported in part by NIEHS grant P01 ES 011617.

REFERENCES

Andrieu, C., De Freitus, A., Doucet, A., Jordan, M.I.. (2003), An introduction to MCMC for machine learning. *Machine Learning,* 50, 5-43.

Bartolucci, A.A, Bae, S.J. and Singh, K.P. (2008), Establishing a Bayesian predictive survival model adjusting for random effects. *Mathematics and Computers in Simulation*, 78, 328-334.

Evans, M., and T. Swartz. (1995), Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science,* 10, 254–272.

Gauderman, W.J., Avol, E., Filliland, F., Vora, H., Thomas, D., Berhane, K., McConnel, R.,Kuenzli, N., Lurmann, R.,Rappaport, E., Margolis, H., Bates, D., and Peters, J. (2004), The effect of air pollution on lung development from 10 to 18 years of age. from 10 to 18 years of age. *New Eng. J. Med*., 351,1057-1067.

Gelman, A., Carlin,J.B., Stern, H.S., Rubin,D.B. (2000), Bayesian Data Analysis: second edition. 285-310.

Mustafa, M.G. Biochemical basis of ozone toxicity. Free Rad Biol. Med. (1990).9,245-265.

Kim, J.J. (2004), Ambient air pollution: health hazards to children. *Pediatrics,* 114,1699-1707.

Lippman. M. (1993),Health effects of tropospheric ozone: Review of recent research findings and their implications to the ambient air quality standard. *J. Exposure Anal. Environ. Epi*., 3, 103-129.

Smith, A. F. M. (1991), Bayesian computational methods. *Phil. Trans. R. Soc. Lond.*

Tanner,M.A. (1996), Tools for statistical inference, 3rd ed. Springer-Verlag, New York.

Thurston, G.D. and Ito, K. (2001), Epidemiological studies of acute ozone exposures and mortality. J. Expo. Anal. Environ.Epidemiol.,11,286-294.

U.S. Environmental Protection Agency, Ozone research needs to improve health and ecological risk assessments. (1998) EPA/600/R-98/031.

.