Identification of chemical components of combustion emissions that affect pro-atherosclerotic vascular responses in mice

Steven K. Seilkop[1,3], **Matthew J. Campen**[2], **Amie K. Lund**[3], **Jacob D. McDonald**[3], and **Joe L. Mauderly**[3]

[1]SKS Consulting Services, Siler City, NC, USA

[2]University of New Mexico, College of Pharmacy, Albuquerque, NM, USA

[3]Lovelace Respiratory Research Institute, Albuquerque, NM, USA

Abstract

Combustion emissions cause pro-atherosclerotic responses in apolipoprotein E-deficient (ApoE$^{/-}$) mice, but the causal components of these complex mixtures are unresolved. In studies previously reported, ApoE$^{-/-}$ mice were exposed by inhalation 6 h/day for 50 consecutive days to multiple dilutions of diesel or gasoline exhaust, wood smoke, or simulated "downwind" coal emissions. In this study, the analysis of the combined four-study database using the Multiple Additive Regression Trees (MART) data mining approach to determine putative causal exposure components regardless of combustion source is reported. Over 700 physical–chemical components were grouped into 45 predictor variables. Response variables measured in aorta included endothelin-1, vascular endothelin growth factor, three matrix metalloproteinases (3, 7, 9), metalloproteinase inhibitor 2, heme-oxygenase-1, and thiobarbituric acid reactive substances. Two or three predictors typically explained most of the variation in response among the experimental groups. Overall, sulfur dioxide, ammonia, nitrogen oxides, and carbon monoxide were most highly predictive of responses, although their rankings differed among the responses. Consistent with the earlier finding that filtration of particles had little effect on responses, particulate components ranked third to seventh in predictive importance for the eight response variables. MART proved useful for identifying putative causal components, although the small number of pollution mixtures (4) can provide only suggestive evidence of causality. The potential independent causal contributions of these gases to the vascular responses, as well as possible interactions among them and other components of complex pollutant mixtures, warrant further evaluation.

Keywords

Atherosclerosis; combustion emissions; matrix metalloproteinases; multiple additive regression trees; air pollution; vascular responses

Introduction

This paper is the first of various planned reports communicating the final results of the National Environmental Respiratory Center (NERC) Program. NERC has been funded by government and industry since 1998 as a step toward identifying the physical–chemical

Address for Correspondence: Steven K. Seilkop, SKS Consulting Services, 3942 Rives Chapel Church Rd., Siler City, NC, 27344, USA. Tel: 919-742-5636. Fax: 919-742-5636. skseilkop@centurylink.net.

Declaration of interest

components driving various adverse health effects associated statistically with complex mixtures of air pollutants, regardless of the source of the components (www.nercenter.org). Intended as a departure from the historic single-pollutant, single-source orientation of air quality health research, NERC undertook development of a novel composition–concentration–response database by applying a consistent animal exposure and response measurement protocol to multiple laboratory-generated complex mixtures of air contaminants. Expert workshops guided selection of animal models, experimental design, and methods for generating four exposure mixtures representative of typical combustion-derived emissions: diesel (DE) and gasoline (GE) engine exhaust, wood smoke (WS), and a mixture simulating "downwind" coal combustion emissions (CE). Animal models were exposed to multiple concentrations of each mixture and to clean air as controls, and some models were exposed to the mixtures with particles removed by filtration. The four studies were conducted serially and study details and exposure–response results from each were published as they were completed. NERC is now conducting integrative analyses of the combined four-study database to identify the putative physical–chemical drivers of the different health outcomes that demonstrated significant response trends with exposure. The hypothesis underlying the program is that biological responses are caused by certain pollutants, or combinations of pollutants, regardless of the sources of the pollutants.

Results from the ApoE$^{-/-}$ mouse model of hypercholesterolemic central vascular responses to inhaled materials were selected as the trial data set for refining statistical analytical strategies for identifying exposure components most closely associated with the animal model's various indicators. This animal model demonstrated different patterns of significant responses to DE, GE, and WS, and little response to CE (Campen et al., 2010a, b; Lund et al., 2007, 2009). Moreover, exposures with and without particles demonstrated that nonparticulate components were largely responsible for responses of most indicators but could not reveal which gases or vapors were primarily responsible. After exploration of other analytical approaches, the Multiple Additive Regression Trees (MART) method (Hastie et al., 2001) was applied to the ApoE$^{-/-}$ response data, and the results point toward putative causal exposure components for evaluation by follow-on confirmatory studies.

Materials and methods

Animals

The animal protocol has been reported in detail (Campen et al., 2010a; Lund et al., 2007, 2009). Ten-week old male ApoE$^{-/-}$ mice (on C57BL/6J background, Taconic, Oxnard, CA, USA) were fed a high-fat diet (No. 88137, Harlan Teklad, Madison, WI, USA) beginning at the initiation of exposure. The mice were housed throughout the study in a facility fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International in whole-body inhalation chambers (H2000, Hazleton, Maywood, NJ, USA) at 30–60% relative humidity, 20°C–24°C, and a 12 h light cycle. Exposure chambers were maintained twice daily and washed weekly.

Exposures and characterization

The mice were exposed 6 h/day, 7 days/week for 50 consecutive days to one of three dilutions of one of the four atmospheres, to the highest dilution with particles removed by filtration or to clean air as study-specific negative controls. The generation, measurement, and compositions of the exposure atmospheres have been reported in detail. DE was generated by 2000 model Cummins 5.9L engines operated on the US Federal Test Procedure certification cycle and burning circa 2000 certification fuel (McDonald et al., 2004). Hardwood smoke was generated by burning split oak in a simple heating stove on a three-phase daily cycle simulating home heating (McDonald et al., 2006). GE was generated by

1996 model General Motors 4.3L engines equipped with catalysts, operated on the California Unified Driving Cycle and burning nonoxygenated fuel blended to 2001–2002 US average specifications (McDonald et al., 2008). A mixture simulating key components of downwind (atmospherically processed) emissions from coal-fired power plants was generated by burning low-sulfur sub-bituminous coal in an electric furnace and supplementing with sulfate and gases (McDonald et al., 2011a). Diluting air and the control atmosphere was charcoal and HEPA-filtered, conditioned ambient air.

Particle mass concentrations were measured gravimetrically and the particle size distribution was measured using a micro-orifice uniform deposit impactor (MOUDI, MSP, Minneapolis, MN, USA) or a fast mobility particle sizer (FMPS, TSI, St. Paul, MN, USA). Particle mass collected on quartz filters was analyzed for elemental and organic carbon by thermal optical reflectance and for inorganic ions by ion chromatography after aqueous extraction. Particle mass on Teflon filters was analyzed for metals by inductively coupled mass spectrometry. Semivolatile and particle-phase organics were analyzed by gas chromatography/mass spectrometry (GC/MS) of organic extracts of an XAD resin-coated denuder followed by a Teflon-coated glass filter. Inorganic gases were measured by chemiluminescence (NOx), photoacoustic infrared spectroscopy (CO), ion chromatography of adsorbants (NH_3, SO_2), liquid chromatography of adsorbants (volatile acids and carbonyls), and GC/MS of canister samples (volatile organic hydrocarbons).

Over 700 different physical–chemical components were measured. These were grouped into 45 predictor variables (Table 1).

Measurement of responses

Measurements of the vascular responses have been described in detail (Campen et al., 2010a, b; Lund et al., 2007, 2009). The mice were anesthetized with pentobarbital/phenytoin and euthanized by exsanguination. The proximal aortas were dissected, weighed, and frozen at −80°C until analysis. Total RNA was isolated (RNeasy fibrous tissue Mini Kit, Qiagen, Valencia, CA, USA) and real-time polymerase chain reaction (PCR) was performed (iCycler, Biorad, Hercules, CA; ABI 7500, Applied biosystems, Foster City, CA, USA) using the appropriate primers for each endpoint (Lund et al., 2007). A melt curve was added to each run to ensure product dimerization and absence of primer dimerization. Lipid peroxidation was assessed using a thiobarbituric acid reactive substances (TBARS) assay (OXItec, ZeptoMetrix, Buffalo, NY, USA). The eight response indicators listed in Table 2 all showed some evidence of response to one or more of the four exposure atmospheres and were thus used in this analysis.

Statistical approach

Assay results for control animals differed across the four emissions studies (CE, DE, GE, and WS); consequently, responses in each study were scaled (i.e., divided) by that study's control mean values. All endpoints exhibited skewed distributions, with treatment group variability increasing with increasing mean values. In analyses which depended on assumptions of homogeneity of variance and normality (e.g., standard regression-based lack of fit tests) weighted least squares reflecting the differing variances was utilized. Weights were calculated as the reciprocals of variance estimates obtained from fitted relationships between standard deviations and mean values (Neter et al., 1996).

The primary focus of the analysis was to identify likely causal linkages between exposures to the chemical components of emissions (irrespective of source) and observed pro-atherosclerotic central vascular responses. Preliminary evaluations were based on pairwise correlations between responses and individual chemical components. This was, however,

deemed to be an insufficient approach to understanding the underlying structural exposure–response associations due to the high degree of intercorrelation among chemical components. Other statistical methods which might be used to identify the strongest associations, such as stepwise regression, principal components, and partial least squares analysis, were also considered to be suboptimal, if not inappropriate, due to their use of inherently additive linear mathematical models to estimate underlying exposure–response functions that may be either highly nonlinear or contain synergistic (multiplicative) effects between chemical components. Although these additive linear model-based approaches might produce reasonably accurate predictions, this success would potentially be the result of using a linear combination of a given set of chemical components to model underlying nonlinear exposure–responses to a different combination of chemical components. This would result in misleading inferences concerning the relative importance of different chemical component concentrations in describing response variability. An example obtained from stepwise linear regression analysis of the present study's atherosclerosis response indicator/chemical component dataset is provided in Supplemental Material online.

A data mining technique known as Multiple Additive Regression Trees (MART) analysis, or more generally as "boosted regression trees" analysis (Friedman, 2001, 2002), was used to mitigate these problems. Discussions of the concepts and mathematical details underlying this approach are provided by Hastie et al. (2001) and Friedman and Muelman (2003). A good conceptual overview of boosted regression trees is given by Elith et al. (2008), and a recent application of the method in air pollution analysis was conducted by Carslaw and Taylor (2009). Another multiple regression tree data mining method known as Random Forests (Breiman, 2001) was also considered and its results were compared with those obtained from the MART analysis (see Supplemental Material online). The examination of the Random Forests method indicated that it was substantially more sensitive to outlier data than the MART approach; consequently the results of the MART analysis are provided here.

The main features of the MART method that are important to understand in using and interpreting its results are: (1) The underlying structural relationships between outcome and predictor variables that it uncovers are independent of any *a priori* mathematical model (this differs from standard regression analysis, which is based on a model specifying a mathematical functional relationship between response and explanatory variables, dependent on unknown parameters); (2) it is resistant to the impacts of extreme observations (data outliers); (3) it does not depend on data distributional assumptions; (4) it ranks predictor variables by their relative importance in predicting the outcome variable (with a score of 100 for the highest ranking predictor and other variables scaled accordingly down to 0); and (5) the influence of a predictor variable is characterized by a "partial dependence" function, which measures the effect of a predictor after accounting for the average effects of all other predictors. Partial dependence functions are typically represented as two-dimensional graphs, representing the partial dependence of a response variable on a single predictor or three-dimensional graphs showing the effects of two predictors (and possible interactions). These functions show major features of the nature of the exposure–response function for a given predictor (approximate linearity versus substantial nonlinearity, threshold-like response, etc.).

MART analysis was applied using TreeNet (Salford Systems) software. A regression tree analysis of each of the response indicators in Table 2 was conducted, specifying the response indicators as "target" variables and all of the composition variables in Table 1 as "predictor" variables. The software's 10-fold cross-validation approach was used to test the multiple additive regression trees that were constructed. Variable importance indices were used to determine the most highly predictive chemical component variables, and two- and three-dimensional partial dependence plots were examined to assess relationships between

predictor and response indicators. As with any regression method, it is important to note that the results, in terms of predictor variable importance scores and the shapes of estimated exposure–response relationships (e.g., partial dependence functions) are invariant to linear scale changes in the predictor variables. For example, the results would be the same irrespective of whether chemical component measurements are expressed as mass concentrations (as used here) or molar concentrations.

Generally, it was clear that two or three predictors had substantially more importance than the remaining predictors in explaining the variability in responses. The amount of response variance explained by these predictors was measured individually and together using partial dependence functions as predictors of response. For each endpoint, the fits of the full MART model (all 45 predictors), models based on the individual predictor variables, and a model based on the three most important predictors were assessed by comparing their predictions to the experimental group means for the 18 treatment groups listed in Table 3. Fractions of response differences among the experimental groups explained by predictor variables (and combinations of variables) were estimated. The p values from F-tests for lack of fit of the models were used to assess the extent to which the deviations of model predictions from the observed cell means were attributable to underlying random between-animal response variation.

Results

Figure 1 summarizes the response indicator data from the emissions studies, reflecting clear differences between the effects of the different exposures. For several endpoints (HO-1, MMP3, MMP7, MMP9, TIMP2, TBARS, and VEGF), responses to GE and/or DE appear to be exposure-related with less compelling evidence, if any, of effects associated with either WS or CE. ET-1 exhibits significant exposure-related responses for GE, DE, and a marginally significant ($p = 0.1$) trend across CE exposures. While this is interesting and potentially useful information, it gives no direct insight into the chemical components of the exposures (predictors) that potentially caused the responses. We use ET-1 as an example to demonstrate the approach to finding and assessing the strength of associations between predictors and responses, regardless of exposure atmosphere.

Example analysis for ET-1

ET-1 data from animals exposed to the four atmospheres (CE, DE, GE, and WS) were matched to the corresponding concentrations of the chemical components in those atmospheres (Table 1). The MART analysis predictor scoring results (Table 4) indicate that SO_2 was clearly the most important variable in predicting ET-1 response, followed by NO_2, and then several variables (CO, NO, ammonia [NH_3], ammonium, etc.) with nearly equal importance scores that were substantially less than those for SO_2 or NO_2.

"Partial dependence" plots for the three most important predictor variables are shown in Figure 2. They depict the MART-estimated ET-1 exposure–response to these predictors after accounting for the average effects of all the other chemical predictors across their experimental exposure ranges. The points on the graphs correspond to MART-predicted ET-1 values in the 18 experimental groups in Table 3 (the four control group estimates are indistinguishable from each other, as the background exposure values were virtually equal – near zero across studies). The lines between the points reflect an unsmoothed version of an exposure–response relationship; thus, not every inflection of the resulting curve is necessarily a meaningful indication of a true underlying feature of the exposure–response function. Generally, it is not difficult to gain a sense of the shape of the function from these unsmoothed curves, but here a spline fit (dashed curve) was applied to smooth the SO_2 partial dependence plot as an additional aid for visualizing the shape of the function. This

degree of smoothing appears to be reasonably consistent with the degree of uncertainty in the mean values for the data shown in Figure 1.

As might be expected, the magnitudes of the exposure–response gradients reflected in the partial dependence plots relate to differences in the importance scores of the predictors, that is, increasing gradient with increasing importance score. The gradients can be used to estimate exposure-specific effects of individual predictors on a response indicator. For example, using the scale shown on the left vertical axis of Figure 2a, the MART-predicted increase in ET-1 attributable to exposure to 500 μg/m^3 of SO$_2$ is computed as the difference between the values at 500 μg/m^3 and 0 μg/m^3 (1.35 − 1.10 = 0.25), yielding an estimated 25% increase over the estimated baseline mean value for control animals at this SO$_2$ exposure. The partial dependence values at zero concentration differ across predictors, as well as from the expected value for control animals (1). The differences in the deviations from 1 for the predictors reflect differences in the average effects of the remaining variables. Each point on the graph reflects the estimated concentration-specific effect of the partial dependence predictor plus the average of the estimated effects of the remaining predictors calculated across the observed combinations of those predictors (in this case, 18 combinations corresponding to the experimental groups). Thus, the SO$_2$ partial dependence value at zero concentration (1.1) reflects the smaller average effects of NO$_2$ and CO (as well as the other predictors) across their exposure ranges, and the NO$_2$ partial dependence at zero concentration (1.2) reflects the larger effect of SO$_2$ (and smaller effects of other variables). CO's relatively large partial dependence value at zero concentration (1.25) reflects the larger average effects of both SO$_2$ and NO$_2$ (and smaller effects of other predictors) across their concentration ranges.

Partial dependence plots are typically depicted with the alternative scaling provided on the right vertical axis in Figure 2, namely as deviations from the predicted overall mean across all observations (for ET-1, this is 1.25). These "scaled partial dependence" values have two advantages over the unscaled values: (1) The magnitudes of exposure–response gradients are more easily compared across predictor variables, as they are centered on zero; and (2) the scaled values can be summed across predictors and then added to the overall mean to obtain estimates of the "main" effects of several variables (i.e., independent of the effects of interactions between predictors), simultaneously adjusted for the effects of the remaining predictors. Although the scaled values were used in computations to ascertain the explanatory power of predictors, the unscaled values are presented here in graphical characterizations of the exposure–response functions for individual predictors of the response indicators. This facilitates comparisons to the natural reference point for the response indicators (mean value of 1 for control animals in the different studies, due to the initial scaling of the data described above).

For ET-1, the SO$_2$ and NO$_2$ graphs increase monotonically (which makes intuitive sense). From its partial dependence plot, one could interpret the exposure–response for SO$_2$ to be moderately supralinear. Similarly, ET-1 appears to exhibit a saturation-based response to NO$_2$ exposure. For CO, the exposure–response function can be interpreted as being either flat (considering response indicator data uncertainty), or nonmonotonic (and counter-intuitive).

The absence of monotonicity in the CO partial dependence function and its lack of a substantial gradient (despite being deemed the third most important predictor) demonstrate that the MART results should be interpreted judiciously. It is important to understand that the method utilizes whatever exposure–response features it can detect in the data to improve the accuracy of predictions, irrespective of their plausibility (either in direction or smoothness). In particular, the covariation between outcome and predictor variables that

MART identifies is not necessarily monotonic, as is typically expected of toxicological exposure–response relationships. The method will often identify and accumulate minor contributions of many (monotonic or nonmonotonic) associations between predictor variables and predicted outcome response. The challenge in interpreting the results is to determine when the contributions of a predictor (or predictors) are sufficiently small to be considered inconsequential relative to the stronger associations between other predictors and the outcome variable. In the analysis that follows, this is addressed by considering the amount of variation explained by individual predictors and combinations of predictors.

Although the MART importance scores and partial dependence plots reveal potentially important features of exposure–response relationships for individual atmospheric components and combinations of these components, they are only semi-quantitative. This is particularly true for the importance scores, which quantify the contributions of the predictor variables on a relative scale. Furthermore, while the partial dependence functions give a sense of the shapes of the exposure–response functions, they do not provide a direct assessment of the degree to which the most important explanatory variables are capable of accurately describing the response functions. To make this assessment, the partial dependence functions were utilized as predictors of response and measured the extent to which the most important variables describe the observed systematic variation in experimental group mean values (shown in Figure 3a). First, however, the fit of the full MART model (based on all of the predictor variables) was assessed by examining the distributions of model error residuals (observed − predicted values) across the experimental groups (Figure 3b). If the model fits the data perfectly, the expected values of the residuals within experimental groups would be zero, with the observed averages varying around zero subject to the underlying degree of between-animal random variation in response. As shown in Figure 3b, in all of the experimental groups these residuals are distributed with average values near zero, and (with the exception of WS3, which shows an inconsistent negative response) the first and third quartiles all overlap zero. This indicates that the MART model provides a good fit, and a statistical F-test comparing all of these mean values simultaneously against zero confirms this with no evidence of lack of fit ($p = 0.83$; Table 5).

Although the MART model (using all 45 atmospheric components as predictors) provides accurate predictions, it is of most interest to determine the extent to which the most important predictors describe the differences among experimental group mean values shown in Figure 3a. This can be assessed by examining model residual errors from the partial dependence-based predictions of each of the top three MART predictors (SO_2, NO_2, and CO) shown in Figure 3c–3e respectively. A further assessment of the joint effects of the three predictors (under the assumption of no interaction between them) was obtained by summing their partial dependency predictions and comparing them to the observed responses (Figure 3f).

Figure 3c indicates that SO_2 alone is substantially less reliable than the full MART model in making predictions. Although it provides reasonably accurate predictions for CE and DE, it shows evidence of lack of fit for experimental groups GE3 and WS3. Figure 3d shows that NO_2 alone is less reliable than SO_2 as a predictor with further evidence of lack of fit for groups GE4 and DE3. As would be expected from the lower importance score for CO, relative to SO_2 and NO_2, it alone is the least reliable of the three predictors, with added evidence of lack of fit for group DE2 and no substantial difference from the response pattern for the data shown in Figure 3a. However, when applied together as predictors, the three atmospheric components provide a reasonably good fit to the data, as shown in Figure 3f. Although there is statistical evidence of lack of fit ($p = 0.02$), this is primarily due to the error in predicting the statistically significant negative group WS3 response; there is no evidence of lack of fit ($p = 0.47$) across the rest of the experimental groups. This brings into

question whether the negative response in group WS3 (which is inconsistent with strong evidence of positive response for other atmospheric exposures) is a chance finding, or whether it is indeed real and explainable due to exposure to emissions components other than SO_2, NO_2, or CO.

These results can be further quantified in terms of the amount of systematic variation that is explained by the different models. In Figure 3a, systematic (potentially explainable) variation is reflected by differences among experimental group means, while random (unexplainable) variation is represented by the distributions of response data around these group means. Table 5 (column 7) shows that differences in experimental group means for ET-1 represent 36% of the total (systematic + random) observed variation; the fractional representation of this percentage is typically denoted as R^2. The full MART model explains most (84%) of the systematic variation (column 5) and 31% of the total variation (column 7). SO_2, NO_2, and CO individually account for 55%, 17%, and 2% of the systematic variation, respectively, while together in the full MART model they are estimated to explain 66% of the systematic variation. That the latter estimate (66%) is less than the sum of the estimates for the three individual components (74%) is likely due to a modest degree of imperfection in adjusting the effects of each of the variables for the effects of the others. This is in contrast to standard regression, which is specifically designed to partition the explained variance into nonoverlapping sources.

The foregoing assessment of the amount of variability explained by the predictor variables is based on an assumption of additivity (i.e., lack of multiplicative interaction) among the effects the three predictors. Single variable partial dependence plots, such as those shown in Figure 2, are useful summaries of exposure– response when the effects of the predictor variables are independent (additive), that is, there are no synergistic (multiplicative) joint effects of the explanatory variables. With the sparse data available here, it is difficult to definitively assess possible interaction effects, but the three-dimensional graphs shown in Figure 4 give no obvious indication that the effects of NO_2 and CO were synergistic with those of SO_2. If this had been the case, we would have expected the partial dependence values for the solid bars to have been larger than those that are shown. These three-dimensional plots also demonstrate the substantial correlation between the exposure variables (0.63 for SO_2 and NO_2; 0.87 for SO_2 and CO), which contributes to the challenge of assessing the relative importance of individual exposure variables as well as interactions between them.

The extent to which two-way interactions between the predictor variables affect the assessment of explained variability was quantified by the calculated differences between two-variable partial dependencies and their components (e.g., partial dependence [SO_2,NO_2] – partial dependence [SO_2] – partial dependence [NO_2]). For the ET1 example, the estimated effects of two-way interactions between variables were negligible (as suggested by Figure 4a and b), with differences between the two-variable partial dependencies and their summed components ranging between –0.01 and 0.01 for pairwise combinations of the three predictors in all experimental groups. These differences are inconsequential relative to the magnitudes of the predicted values, indicating no substantial evidence of two-way interactions between predictor variables. This gives confidence in the assessment of explained variability that was dependent on this assumption.

The essence of this analysis for ET-1 can then be summarized as follows:

SO_2, NO_2, and CO were determined to be the three strongest predictors of ET-1 response. The estimated exposure–response functions for SO_2 and NO_2 are both nonlinear, with the relationship for SO_2 appearing to be supralinear, and the NO_2 exhibiting evidence of a more

distinct saturation-based response (linear to plateau). CO exhibited little, if any, evidence of a clear monotonic exposure–response relationship. SO_2 was the most highly predictive variable, describing more than half of the systematic response variation across experimental groups; together, SO_2, NO_2, and CO described 66% of the variation between experimental group means. A substantial amount of the unexplained variation appeared to relate to a possible spurious result in one of the experimental groups (WS3) that was inconsistent with the rest of the data.

Results for all response indicators

Variable importance scores from the MART analysis of atmospheric component predictors for the atherosclerosis markers are listed in Table 4. Across the response indicators, gases (NH_3, SO_2, NO_2, NO, and CO) predominate among the three most important predictors (16/24), followed by nonmethane volatile organics (4/24), volatile carbonyl (2/24), a single volatile acid (VOAALI), and a single particulate organic component (POHOP). Particulate components ranked third to seventh in relative predictive importance for the eight response indicators. Partial dependence plots for the three most important variables are shown in Figure 5, and the amount of response variance explained by these predictors is provided in Table 5. Partial dependence plots for the three most important predictors of each of the response indicators are shown in Figure 5. Individual interpretations of the results for each of the markers, similar in form to that provided for the ET-1 example follow:

VEGF—Table 4 indicates that NH_3 was the most important predictor of VEGF response, followed by NMVOOXY, which had a substantially lower importance score (41). Importance scores for remaining variables, starting with a value of 29 for CO, declined slowly. Partial dependence plots (Figure 5) for the three most important variables reflect the reduction in VEGF shown in Figure 1. The plots suggest a strong nonlinear exposure–response for NH_3 and much weaker effects for NMVOOXY and CO. Table 5 indicates an excellent fit for the full MART model ($p = 0.87$). Approximately 70% of systematic variation in VEGF was explained by the three most important variables, and most of this variation was explained by NH_3 exposure. The lack of fit for the three variables model likely relates to the exclusion of variables with scores similar to CO, which make individually small contributions to predictions.

MMP3—NH_3 was determined to be the most important predictive variable, followed by NMVOALKE and NMVOALKY, with importance scores of ~50 (Table 4). Importance scores for other variables were substantially lower. Although the full MART model described only about 70% of the observed variation in experimental mean values, there was no evidence of lack of fit due to the low signal-to-noise ratio ($R^2 = 0.18$ for experimental group means). Partial dependence plots indicated a nonlinear response for NH_3, a weaker linear response for NMVOOXY, and a weak threshold-like response for NMVOALKY. Approximately 60% of the variation between experimental group means was attributed to exposure to NH_3, NMVOALKE, and NMVOALKY, with most response variation associated with exposure to NH_3.

MMP7—Importance scores for NH_3 and VOALLI were both ~100, followed by a score of 45 for CARBDI; other predictors exhibited slowly declining scores from that point. There was evidence of a monotonically increasing response function for NH_3, but decreasing response functions for VOALLI and CARBDI. Both functions were highly nonlinear, exhibiting a hockey-stick form that bent sharply at values near zero. There was no substantial evidence of lack of fit for the full MART model ($p = 0.25$), although it explained only 33% of the systematic variability. This is likely to be related to the influence of outlying observations on the estimated mean values for groups CE2 and CE4 (and

corresponding inflation in the variation between mean values). Since the MART method downplays the importance of such data in predictions, it fails to produce predictions that come close to these mean values. The three predictors deemed to be the most important variables explained the same amount of variation as the full model. Other predictors exhibited partial dependence functions (data not shown), which became increasingly smaller in magnitude with limited interpretability. The counter-intuitive partial dependence functions for VOALLI and CARBDI act to increase predictions for experimental groups with low levels of these substances (GE, CE) and are believed to relate to the poor fit of the MART model.

MMP9—Ammonia and SO_2 were found to be the most predictive atmospheric components, with scores of 100 and 73, respectively. The next most important variable (CO) had a score of 30, with other predictors exhibiting scores that diminished slowly from that point. The partial dependence function for NH_3 was an S-shaped function and the SO_2 function was approximately linear with a more modest slope than NH_3 over the range of the observed data. The partial dependence function for CO had a negative, but inconsequential slope with respect to its impact on predictions. The full MART model had a good fit to the data, and the three variables model (NH_3, SO_2, CO) explained 80% of the systematic variation. There was evidence of lack of fit for the three variables model ($p = 0.04$), which related to its inability to accurately predict a reduction in MMP9 in group WS3 and an aberrantly higher mean for group CE2, relative to the monotonic trend shown across the remaining exposure groups.

TIMP2—The most predictive atmospheric components were NO_2 and NO (importance scores of 100 and 79, respectively). POHOP was the next most important predictor, with a score of 45, closely followed by SVOALK (importance score 42). Partial dependence functions for NO_2 and NO were both nonlinear, rising steeply to a plateau. POHOP showed small estimated increases in TIMP2 for the few experimental groups that had nonzero exposures (all DE groups and GE3 and GE4). The full MART model fit was excellent, and the three variables model described 65% of the systematic variation. There was marginal evidence of lack of fit for the three variables model ($p = 0.09$), which reflected an inability to predict reduced TIMP2 in group WS3.

HO-1—Ammonia was the strongest predictor of HO-1, followed by NMVOHAL and CARBALI, both with importance scores of ~40. There were several other variables that also had scores of nearly this magnitude. Ammonia had a strongly increasing nonlinear partial dependence function, while both NMVOHAL and CARBALI had comparatively modest decreasing functions. Although the full MART model explained only 63% of the differences in experimental mean values, much of this variation related to uncertain mean values in groups WS1 and WS3, consequently there was no evidence of lack of fit ($p = 0.93$). The three variables model, however, explained only 22% of the differences in group mean values, leading to some evidence of lack of fit ($p = 0.13$). The reduction in fit for the main effects model relative to the full model was attributable primarily to an inability to accurately predict increased HO-1 levels for groups WS1 and WS2. Both of these groups exhibited unusually high between-animal variability and inconsistency with absence of evidence of increased HO-1 at the highest exposure level (WS3). Overall, it appears that NH_3 exposure is the most reliable predictor of increased HO-1 response.

TBARS—The two most important predictors of TBARS response were SO_2 and NO_2 (scores of 100 and 90, respectively), followed by NO (75), NMVOALKE, CO (46), and POHOP (46). Partial dependence plots indicated a threshold-like response for SO_2 and nonlinear saturation functions for NO_2 and NO. The fit of the full MART model was excellent, and the three variables model (SO_2, NO_2, NO) explained 70% of the variation

between experimental group means. However, because of the relatively low degree of random variability ($R^2 = 0.57$), the three variables model was found to exhibit significant lack of fit ($p < 0.01$). This was due to an inability to accurately model the reduction in TBARS observed in group CE1 (which is likely a spurious finding), but more importantly, to underestimate the TBARS response in group DE3. It is interesting to note that the next three variables in the importance hierarchy were NMVOALKE, CO, and POHOP, which have their highest levels in group DE3 (see Table 1). When the predictive contributions of these variables were considered in an additional analysis, there was considerable improvement in the fit of the MART-based model (explaining 81% of the variation between experimental groups) but still statistically significant ($p < 0.01$) underestimation of the response in DE3. Thus, there is evidence that some of the remaining chemical components may be causal agents in inducing at least a portion of the observed response.

The foregoing summaries of the results for the response indicators have been predicated upon an assumption of additivity (as opposed to synergy) of effects associated with the predictors. Examinations of the two-way pairwise interactions between the three most important predictors for the response indicators gave no substantial indication of nonadditivity. However, with the high degree of intercorrelation among the predictors, and the relatively small number of experimental groups and their uneven and sparse coverage of the predictor space (as demonstrated in Figure 4), synergistic effects would be difficult, if not impossible to detect.

Discussion

The data mining technique (MART) used here to identify the strongest predictors of response appears to be useful, but optimal application of this methodology requires: (1) differential responses across different combinations of chemical components and (2) enough different combinations to confidently discriminate between patterns of exposure–response among the components. Clear differences among the patterns of response to the four atmospheres facilitated the identification of components that contributed most strongly to the responses. However, with only 18 experimental points in a 45-dimensional predictor space, the study was at the low end of the discriminatory spectrum for confidently predicting likely causal relationships.

Further experiments are required to validate the causality of the components that were most predictive of the observed responses. First, it would be useful to determine, at least at a qualitative level, whether or not the observed responses to a given component can be replicated independent of the mixtures. It is possible that components that were not ranked highly in importance were necessary to cause or enhance the effects of those that had more systematic associations with the responses. Replication of the same quantitative exposure–response for a given predictor administered alone would verify the independence of its causal effect. This approach has already been taken to a very limited extent, as described below, with mixed results. Some responses were duplicated quantitatively, but only at higher concentrations of the single gases. However, exposures of the ApoE$^{-/-}$ model to single gases have not yet encompassed a sufficient matrix of gas species and concentrations to resolve their potential independent effects.

Second, it would be useful to determine whether or not a mixture of only the top few predictors might reproduce the responses observed in this study. Based on the relative importance scores in Table 4, one might select a mixture of SO_2, NO, NO_2, NH_3, and CO at concentrations indicated by partial dependence plots (Figure 5) to have exerted substantial impact on responses. In conjunction with the single gas exposures, this step would reveal whether or not a combination of major predictors was necessary to elicit the responses and/

or how the response to a simple mixture differed quantitatively from responses to single gases or the more complex mixtures. This step would be more practical than moving directly into a full factorial study of the top predictors.

The vascular responses examined in this study include markers for pathways leading to atherosclerosis, including inflammation (MMPs), angiogenesis (VEGF), oxidative stress (HO-1, TBARS), and endothelial dysfunction (ET-1). These markers coincide with atherosclerotic lesion development in ApoE$^{-/-}$ mice following exposures to ambient PM (Sun et al, 2005), ultrafine PM (Araujo et al., 2008), and DE (Bai et al., 2011; Campen 2010b).

We do not have full pathology data for all exposure atmospheres. In a recent study of diesel emissions (Campen et al., 2010b), the authors principally saw increases in plaque inflammation and composition, not plaque size, but TBARS was the most robust endpoint in terms of signal-to-noise ratio. More recently, in a large study sponsored by HEI, it was again found that TBARS exhibited the highest degree of systematic exposure-related variation, while only small effects on plaque size and inflammation were observed. The authors are not aware of any study in the literature that has used plaque pathology to discriminate between more than two exposure scenarios and doses. Araujo et al. (2008) found a small but significant difference in plaque size between fine and ultrafine atmospheres in one comparable study, but it was clear that the difference was too small in relative to the underlying variability to productively characterize concentration–response trends. Quan et al. (2010) used the ApoE model to explore differential effects between combinations of concentrated PM and diesel emissions on plaque pathology but encountered a similar problem – too little toxicological effect and too much individual variability. In the present study, the pathology indices served to reinforce the trends observed in the biomarkers, but because of the small effects and high interindividual variability in these indices, a greater emphasis has been placed on the more dynamic exposure-related changes in qPCR and lipid peroxidation metrics.

Moreover, we have observed cross-species coherence for MMP9 and ET-1 with acute human responses to DE (Lund et al., 2009), as have others (Calderón-Garcidueñas et al., 2007; Peretz et al., 2008). Thus, while more work is necessary to understand the impact of individual components and combinations on plaque modification and vascular function, the responses observed in the present study suggest that such outcomes would result from longer exposure.

Several of the top chemical predictors of cardiovascular effects have been tested in the ApoE$^{-/-}$ model, albeit after 1 week instead of 50-day exposures (Campen et al., 2010a). The CO and NO were found to exhibit some effects, such as inducing increases in ET-1 and MMP9, but the degree of induction was less than that seen for whole gasoline emissions. Thus, it is not surprising that these two gases are represented in the top five predictors for most of the endpoints analyzed. In that 1 week study, neither NOx nor CO independently induced measurable lipid peroxidation (TBARS), yet they are among the top five predictors in the present MART analysis. While it may be that the 1 week exposure was insufficient to elicit an effect from the gases independently, the 1 week model was found to be responsive in recent studies of combined gasoline and diesel emissions (Lund et al., 2011; McDonald et al., 2011b). In addition to NOx and CO, MART analysis pointed toward SO$_2$, PM-associated hopanes, and nonmethane volatile alkenes as potential drivers of lipid peroxidation. These predictions are consistent with the concept that particle–gas interactions can drive systemic vascular lipid peroxidation.

While exposure to vehicular emissions have been reported to be associated with increased vascular expression of TBARS, ET-1, and various MMPs, the roles of individual components (e.g., NH_3, SO_2, NO_2, and NO) identified in this study in mediating the expression of these vascular factors, and their subsequent role in vascular disease, have not yet been fully elucidated. For example, NH_3, which predicts expression of MMP3, -7, and -9, and HO-1 and VEGF, is present in the body during normal homeostatic conditions and plays a key role in DNA and protein synthesis. Inhalation of NH_3 may impact the respiratory tract in a manner that leads to secondary systemic vascular effects, but this effect has not been studied. Based on preceding knowledge, no long-term effects of such low levels of inhaled NH_3 (maximum of ~2.6 ppm) was expected in the cardiovascular system. However, the potential vascular effects of NH_3 in combination with other gases and/or PM, as in the present exposures, warrant further investigation.

Three key gases, NO, NO_2, and NH_3, may theoretically impact nitric oxide homeostasis. It is plausible that the observed effects of subchronic NH_3 exposure on expression of vascular MMPs, HO-1, and VEGF in the animal model may have resulted from its ability to act as a nitrogen donor, through involvement in nitric oxide (NO) production. In conditions of increased vascular oxidative stress, as is the case with the hypercholesterolemic model, NO is subsequently converted to peroxynitrite (ONOO·), a highly reactive protein-damaging oxidant (Gryglewski et al., 1986). Decreased bioavailability of NO is known to impair vasoreactivity and mediate endothelial dysfunction, which is associated with altered ET-1, HO-1, and MMP production (Amiri et al., 2004; Bonetti et al., 2003; Lund et al., 2009; Rajagopalan et al., 1996). The same rationale may apply to the role of NO_2 and NO in TIMP-2 expression, as TIMP-2 is the primary tissue-level inhibitor of MMP9, as well as other MMPs (Brew et al., 2000), and is thus likely upregulated to attenuate effects of increased MMP expression in the vasculature. Given that NO_2 reacts in the pulmonary surfactant layer (Postlethwait et al., 1991), it is likely that the role of NO_2 is indirect via biologically active intermediates that activate airway neural (Hazari et al., 2011) or immunomodulatory receptors (Kampfrath et al., 2011; Lund et al., 2011).

The role of SO_2 in mediating the increased vascular TBARS may be due, at least in part, to the increase in reactive oxygen species (ROS) that are reported to result from inhaled SO_2. Increased ROS have been shown to interact with the polyunsaturated fatty acids in biomembranes, thereby altering the structure of the lipid bilayer membrane, which can result in increased lipid peroxidation (Yargicoglu et al., 2007). Importantly, SO_2 is emerging as an important gaseous signaling molecule in the cardiovascular system (Wang et al., 2011), although its uptake and disposition from inhalation exposures complicate direct consideration of a parallel effect. Multiple epidemiological studies have reported that SO_2 is associated with increased risk of developing cardiovascular disease (Chen et al., 2011) and increased oxidative stress (TBARS), ET-1, and MMP9 expression in humans.

Conclusions

In summary, there is evidence that the most highly ranked chemical components in the MART analysis could have plausibly mediated pathways involved in cardiovascular disease. Further investigation is needed to elucidate the pathophysiologic significance of the individual components and to determine the extent to which their effects are dependent on co-exposure to other components (such as PM) that occur in environmental exposures. It is important to recognize that the suite of combustion-derived mixtures used in the initial NERC exposure matrix did not include many primary and secondary pollutants common in the environment (e.g., ozone). The extent to which the key predictors identified in this study would also be among the most important in the presence of additional pollutants is unknown. Of course, it is also important to recognize that the present findings resulted from

only one of the many disease models included in the NERC program and other air pollution research. Nevertheless, the present results demonstrate the utility of the general research strategy for apportioning causality among the components of complex exposures. Follow-on experiments using much simpler combinations of the pollutants most highly ranked by MART could test the practical utility of the MART results and determine the extent to which the effects of the mixtures can be reproduced by only a few key components. A better understanding of the extent to which effects of highly complex exposures can be attributed to a few key pollutants would bolster the foundation for multipollutant air quality management.

Acknowledgments

References

Amiri F, Virdis A, Neves MF, Iglarz M, Seidah NG, Touyz RM, Reudelhuber TL, Schiffrin EL. Endothelium-restricted overexpression of human endothelin-1 causes vascular remodeling and endothelial dysfunction. Circulation. 2004; 110:2233–2240. [PubMed: 15466627]

Araujo JA, Barajas B, Kleinman M, Wang X, Bennett BJ, Gong KW, Navab M, Harkema J, Sioutas C, Lusis AJ, Nel AE. Ambient particulate pollutants in the ultrafine range promote early atherosclerosis and systemic oxidative stress. Circ Res. 2008; 102:589–596. [PubMed: 18202315]

Bai N, Kido T, Suzuki H, Yang G, Kavanagh TJ, Kaufman JD, Rosenfeld ME, van Breemen C, Eeden SF. Changes in atherosclerotic plaques induced by inhalation of diesel exhaust. Atherosclerosis. 2011; 216:299–306. [PubMed: 21435644]

Bonetti PO, Lerman LO, Lerman A. Endothelial dysfunction: a marker of atherosclerotic risk. Arterioscler Thromb Vasc Biol. 2003; 23:168–175. [PubMed: 12588755]

Breiman L. Random forests. Mach Learn. 2001; 45:5–32.

Brew K, Dinakarpandian D, Nagase H. Tissue inhibitors of metalloproteinases: evolution, structure and function. Biochim Biophys Acta. 2000; 1477:267–283. [PubMed: 10708863]

Calderón-Garcidueñas L, Vincent R, Mora-Tiscareño A, Franco-Lira M, Henríquez-Roldán C, Barragán-Mejía G, Garrido-García L, Camacho-Reyes L, Valencia-Salazar G, Paredes R, Romero L, Osnaya H, Villarreal-Calderón R, Torres-Jardón R, Hazucha MJ, Reed W. Elevated plasma endothelin-1 and pulmonary arterial pressure in children exposed to air pollution. Environ Health Perspect. 2007; 115:1248–1253. [PubMed: 17687455]

Campen MJ, Lund AK, Doyle-Eisele ML, McDonald JD, Knuckles TL, Rohr AC, Knipping EM, Mauderly JL. A comparison of vascular effects from complex and individual air pollutants indicates a role for monoxide gases and volatile hydrocarbons. Environ Health Perspect. 2010a; 118:921–927. [PubMed: 20197249]

Campen MJ, Lund AK, Knuckles TL, Conklin DJ, Bishop B, Young D, Seilkop S, Seagrave J, Reed MD, McDonald JD. Inhaled diesel emissions alter atherosclerotic plaque composition in ApoE(−/−) mice. Toxicol Appl Pharmacol. 2010b; 242:310–317. [PubMed: 19891982]

Chen SS, Tang CS, Jin HF, Du JB. Sulfur dioxide acts as a novel endogenous gaseous signaling molecule in the cardiovascular system. Chin Med J. 2011; 124:1901–1905. [PubMed: 21740851]

Carslaw DC, Taylor PJ. Analysis of air pollution data at a mixed source location using boosted regression trees. Atmospheric Environment. 2009; 43:3563–3570.

Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008; 77:802–813. [PubMed: 18397250]

Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of Statistics. 2001; 29:1189–1232.

Friedman JH. Stochastic gradient boosting. Computational Statistics & Data Analysis. 2002; 38:367–378.

Friedman JH, Meulman JJ. Multiple additive regression trees with application in epidemiology. Stat Med. 2003; 22:1365–1381. [PubMed: 12704603]

Gryglewski RJ, Palmer RM, Moncada S. Superoxide anion is involved in the breakdown of endothelium-derived vascular relaxing factor. Nature. 1986; 320:454–456. [PubMed: 3007998]

Hastie, T.; Tibrishani, R.; Friedman, J. The Elements of Statistical Learning; Data Mining, Inference, and Prediction. New York: Springer Science+Business Media; 2001.

Hazari MS, Haykal-Coates N, Winsett DW, Krantz QT, King C, Costa DL, Farraj AK. TRPA1 and sympathetic activation contribute to increased risk of triggered cardiac arrhythmias in hypertensive rats exposed to diesel exhaust. Environ Health Perspect. 2011; 119:951–957. [PubMed: 21377951]

Kampfrath T, Maiseyeu A, Ying Z, Shah Z, Deiuliis JA, Xu X, Kherada N, Brook RD, Reddy KM, Padture NP, Parthasarathy S, Chen LC, Moffatt-Bruce S, Sun Q, Morawietz H, Rajagopalan S. Chronic fine particulate matter exposure induces systemic vascular dysfunction via NADPH oxidase and TLR4 pathways. Circ Res. 2011; 108:716–726. [PubMed: 21273555]

Lund AK, Knuckles TL, Obot Akata C, Shohet R, McDonald JD, Gigliotti A, Seagrave JC, Campen MJ. Gasoline exhaust emissions induce vascular remodeling pathways involved in atherosclerosis. Toxicol Sci. 2007; 95:485–494. [PubMed: 17065432]

Lund AK, Lucero J, Lucas S, Madden MC, McDonald JD, Seagrave JC, Knuckles TL, Campen MJ. Vehicular Emissions Induce Vascular MMP-9 Expression and Activity via Endothelin-1 Mediated Pathways. Arterio Thrombosis Vasc Biol. 2009; 29:511–517.

Lund AK, Lucero J, Harman M, Madden MC, McDonald JD, Seagrave JC, Campen MJ. The oxidized low-density lipoprotein receptor mediates vascular effects of inhaled vehicle emissions. Am J Respir Crit Care Med. 2011; 184:82–91. [PubMed: 21493736]

McDonald JD, Barr EB, White RK, Chow JC, Schauer JJ, Zielinska B, Grosjean E. Generation and characterization of four dilutions of diesel engine exhaust for a subchronic inhalation study. Environ Sci Technol. 2004; 38:2513–2522. [PubMed: 15180045]

McDonald JD, White RK, Barr EB, Zielinska B, Chow JC, Grosjean E. Generation and characterization of hardwood smoke inhalation exposure atmospheres. Aerosol Sci Technol. 2006; 40:573–584.

McDonald JD, Barr EB, White RK, Kracko D, Chow JC, Zielinska B, Grosjean E. Generation and characterization of gasoline engine exhaust inhalation exposure atmospheres. Inhal Toxicol. 2008; 20:1157–1168. [PubMed: 18951232]

McDonald JD, White RK, Holmes T, Mauderly JL, Laumb J, Zielinska B, Chow JC, Grosjean E. Simulated downwind coal emissions for laboratory inhalation exposure atmospheres. Inhal Toxicol. 2011a (In Press).

McDonald, JD.; Lund, AK.; Campen, MJ.; Vedal, S. National Particulate Component Toxicity Initiative; Vascular Effects of Fresh Emissions and Secondary Particulates. Poster presented at Health Effects Institute Annual Conference; March 2011; Boston MA. 2011b. Accessible at http://www.healtheffects.org/archive/Annconf2011_Poster_pdfs/McDonald%20NP

Neter, J.; Kutner, MH.; Nachtsheim, CJ.; Wasserman, W. Applied Linear Statistical Models. 4th Edition. Chicago: Irwin; 1996.

Postlethwait EM, Langford SD, Bidani A. Transfer of NO2 through pulmonary epithelial lining fluid. Toxicol Appl Pharmacol. 1991; 109:464–471. [PubMed: 1853345]

Peretz A, Sullivan JH, Leotta DF, Trenga CA, Sands FN, Allen J, Carlsten C, Wilkinson CW, Gill EA, Kaufman JD. Diesel exhaust inhalation elicits acute vasoconstriction in vivo. Environ Health Perspect. 2008; 116:937–942. [PubMed: 18629317]

Quan C, Sun Q, Lippmann M, Chen LC. Comparative effects of inhaled diesel exhaust and ambient fine particles on inflammation, atherosclerosis, and vascular dysfunction. Inhal Toxicol. 2010; 22:738–753. [PubMed: 20462391]

Rajagopalan S, Meng XP, Ramasamy S, Harrison DG, Galis ZS. Reactive oxygen species produced by macrophage-derived foam cells regulate the activity of vascular matrix metalloproteinases *in vitro*. Implications for atherosclerotic plaque stability. J Clin Invest. 1996; 98:2572–2579. [PubMed: 8958220]

R-Development Core Team. R Foundation for Statistical Computing. Vienna, Austria: 2008. A language and environment for statistical computing. ISBN 3-900051-07-0, URL http://www.R-project.org

Sun Q, Wang A, Jin X, Natanzon A, Duquaine D, Brook RD, Aguinaldo JG, Fayad ZA, Fuster V, Lippmann M, Chen LC, Rajagopalan S. Long-term air pollution exposure and acceleration of atherosclerosis and vascular inflammation in an animal model. JAMA. 2005; 294:3003–3010. [PubMed: 16414948]

Wang XB, Jin HF, Tang CS, Du JB. The biological effect of endogenous sulfur dioxide in the cardiovascular system. Eur J Pharmacol. 2011; 670:1–6. [PubMed: 21925165]

Yargicoglu P, Sahin E, Gümüslü S, Agar A. The effect of sulfur dioxide inhalation on active avoidance learning, antioxidant status and lipid peroxidation during aging. Neurotoxicol Teratol. 2007; 29:211–218. [PubMed: 17197156]
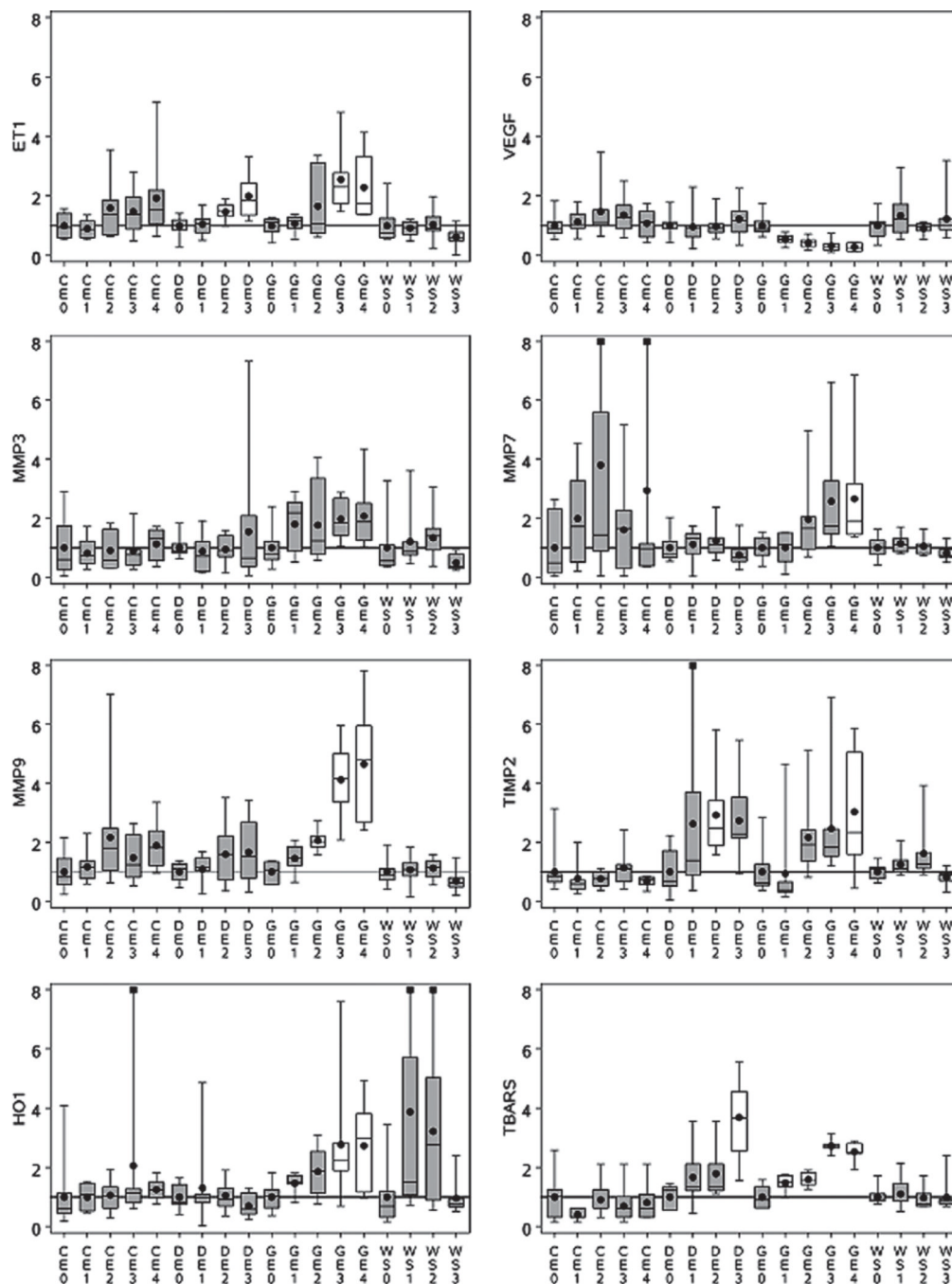
Figure 1.
Observed response patterns for response indicators. Data are expressed as fractions of individual study control mean values. Horizontal bars in boxes represent 25th and 75th percentiles and medians; dots depict mean values. Narrow vertical lines depict data ranges (maximum and minimum). Squares at maximum indicate a value above 8. Horizontal axis provides experimental group (CE, GE, WS, DE) and exposure designations (0 = control, 1 = low concentration, 2 = mid concentration, 3 = high concentration, 4 = high concentration with particle filtration). Boxes without shading indicate a statistically significant ($p < 0.05$) difference from the study control mean by Dunnett's multiple comparison test.
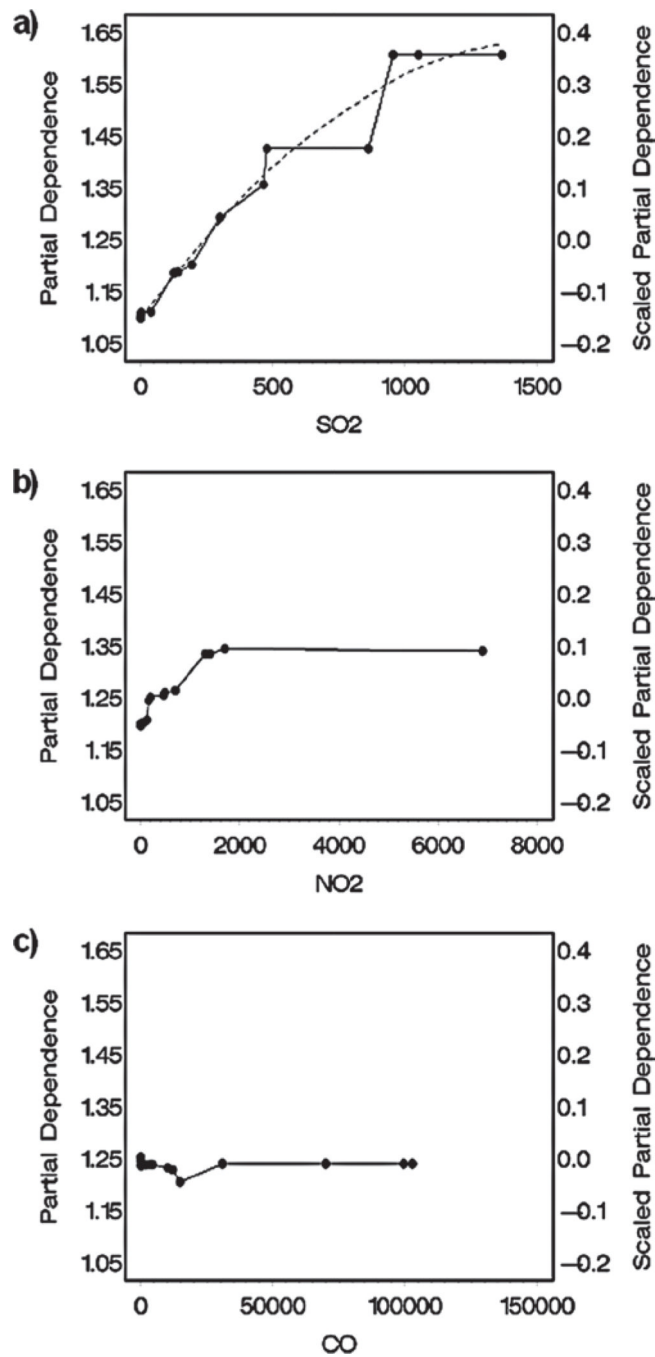
Figure 2.
Estimated partial dependence of ET-1 on (a) SO_2 exposure, (b) NO_2 exposure, and (c) CO exposure. Partial dependence for each of these atmospheric constituents is adjusted for the effects of all the remaining atmospheric components in the MART model (see text). Dashed line in Figure 2a represents a smoothed spline fit of the partial dependence function.
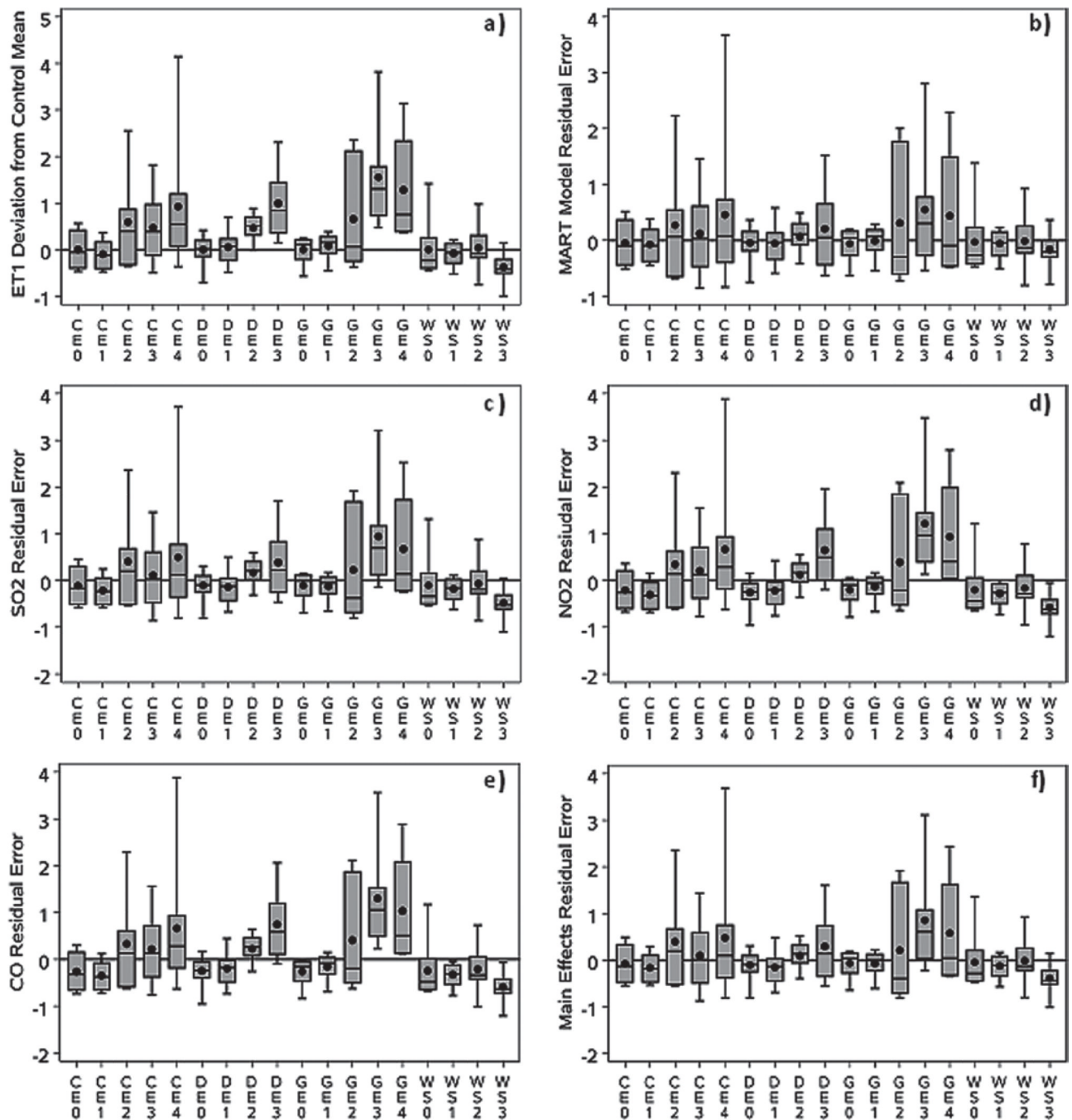
Figure 3.
(a) Observed response pattern for ET-1 expressed as deviation of fraction of control mean from 1 ([experimental group mean/control mean] – 1); (b) Residual error of full MART model based on all atmospheric components; (c) Residual error of MART predictions based on effect of SO_2 alone; (d) Residual error of MART predictions based on effect of NO_2 alone; (e) Residual error of MART predictions based on effect of CO alone; and (f) Residual error of MART predictions based on joint effects of SO_2, NO_2, and CO.
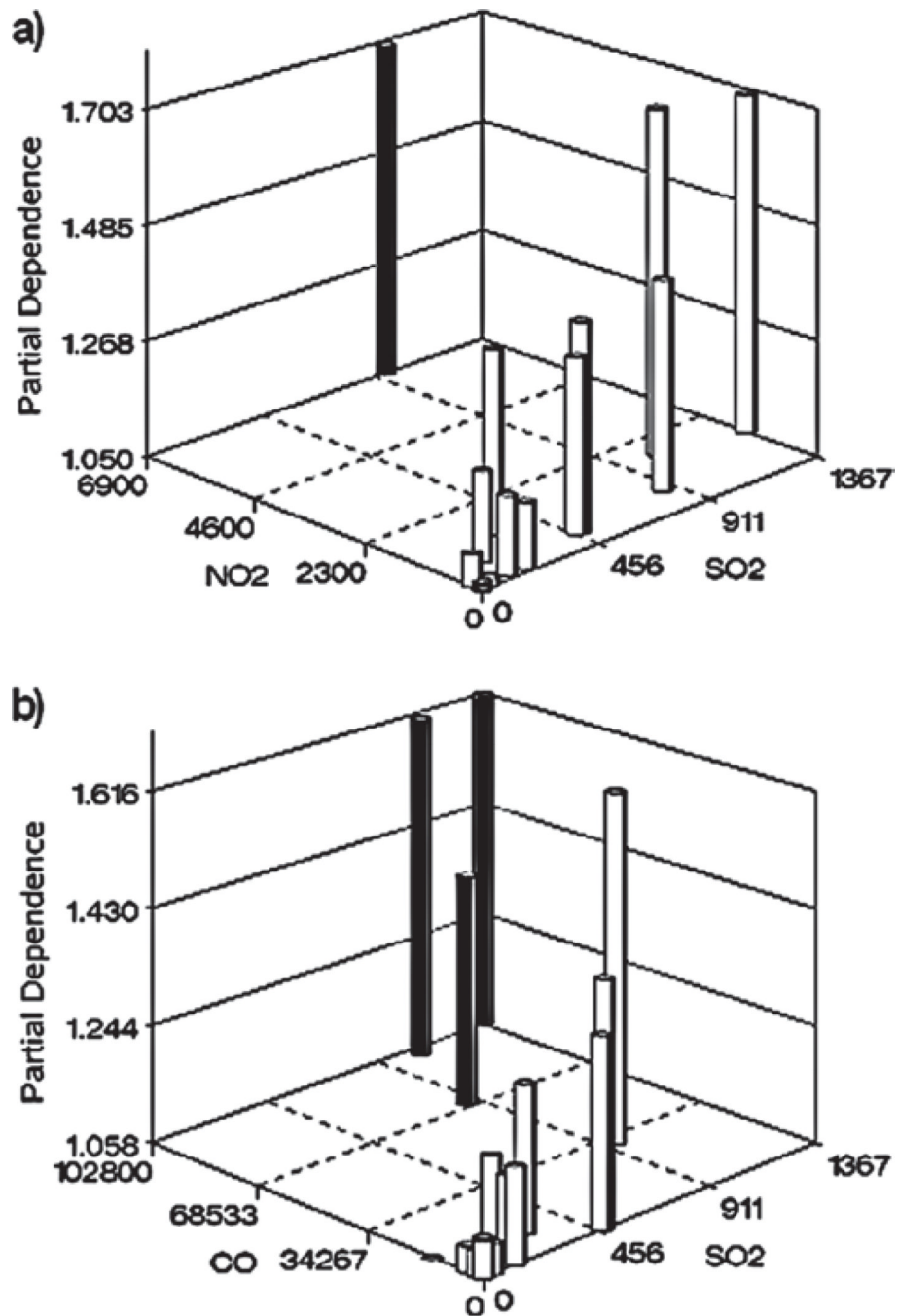
Figure 4.
Estimated partial dependence of ET-1 on: (a) $SO_2 \times NO_2$ exposure and (b) $SO_2 \times CO$ exposure. Solid bars indicate points in the predictor space which give little, if any, indication of synergistic response.
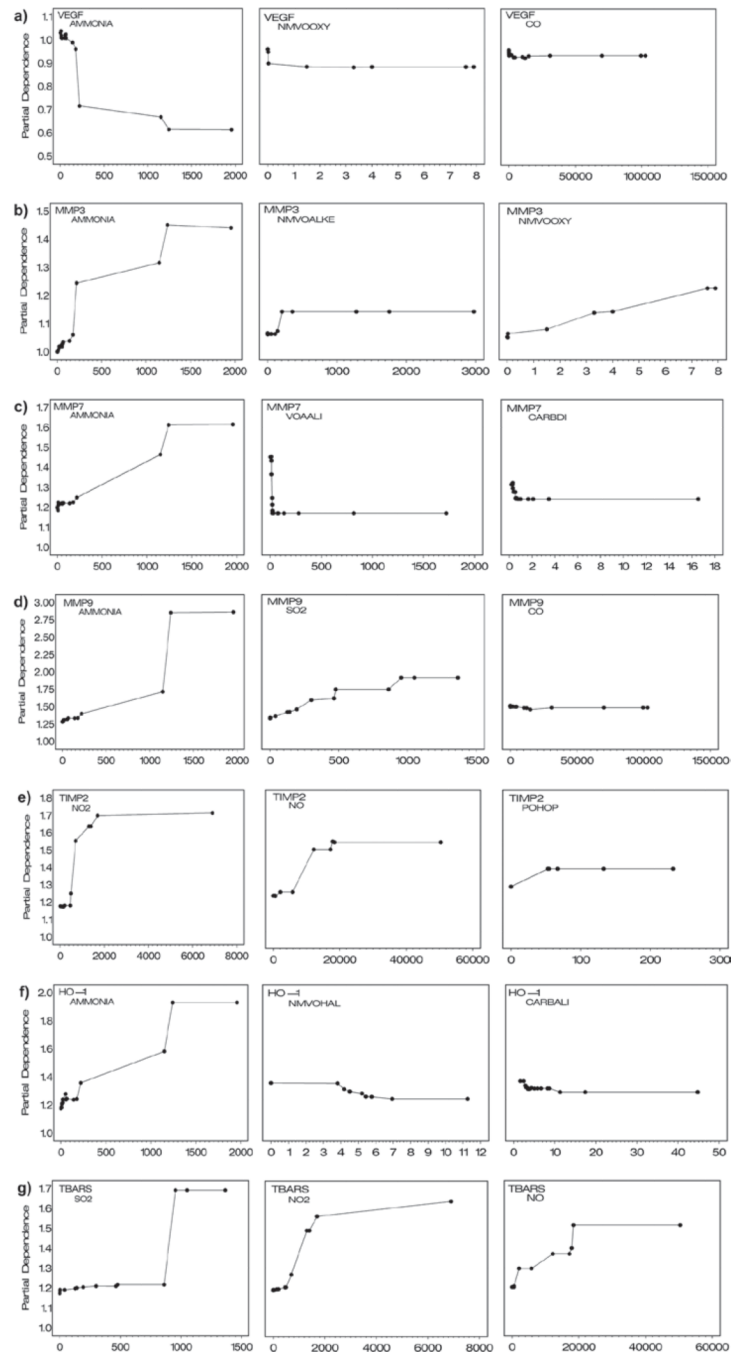
Figure 5.
Estimated partial dependence of response indicators on the three most important predictors in the MART model for (a) VEGF, (b) MMP3, (c) MMP7, (d) MMP9, (e) TIMP2, (f) HO1, and (g) TBARS. Partial dependence for each of the three predictors is adjusted for the effects of all the remaining predictors in the MART model.

Table 1

Atmospheric component categories and measured concentrations in high exposure groups[a].

Atmospheric component category	Variable	Coal (CE)	Diesel (DE)	Gasoline (GE)	Wood smoke (WS)
Particulate matter					
Particle mass	PM	1015	996	61	1041
Ammonium	AMMONIUM	94	31	8	2
Elements	ELEMENTS	8470	1271	1544	2200
Nitrate	NITRATE	0	32	0.5	1
Sulfate	SULFATE	785	55	14	0.7
Elemental carbon	EC	0.9	557	31	39
Organic carbon	OC	19	193	13	825
Organic acids	POACID	0	16928	816	5439
Organic alkanes	POALK	0	8684	170	485
Organic hopane	POHOP	0	233	67	0
Organic pahs	POPAH	0	2021	775	465
Organic phenols	POPHEN	0	0	0	3069
Organic steranes	POSTER	0	133	0	0
Organic sugars	POSUB	0	0	0	93288
Gases					
Ammonia	NH_3	4	9	1958	55
Carbon monoxide	CO	40	30900	102800	14888
Nitrogen monoxide	NO	624	50400	18400	0
Nitrogen dioxide	NO_2	488	6900	1400	0
Sulfur dioxide	SO_2	465	955	1367	0
Non-methane volatile organics					
Halogenated	NMVOHAL	0	5	0	11
Alcohols	NMVOALC	0	6	8	24
Alkanes	NMVOALKA	46	141	8855	185
Alkenes	NMVOALKE	3	359	1757	111
Alkynes	NMVOALKY	0	82	578	0
Aromatics	NMVOARO	2	103	4572	96

Atmospheric component category	Variable	Coal (CE)	Diesel (DE)	Gasoline (GE)	Wood smoke (WS)
Furans	NMVOFUR	0	0	0	369
Oxygenated	NMVOOXY	0	0	8	0
Volatile acids					
Aliphatic acid	VOAALI	6	279	20	1724
Aromatic acid	VOAARO	0.1	7	3	3
Volatile carbonyl					
Aliphatics carbonyl	CARBALI	3	45	8	11
Alkanals	CARBALKA	12	259	68	394
Alkenals	CARBALKE	0.4	55	14	21
Aromatic aldehydes	CARBARO	0.8	21	28	11
Dicarbonyls	CARBDI	0.2	2	0.5	17
Furans	CARBFUR	0	0.3	0.4	146
Hydroxycarbonyls	CARBHYD	0.1	2	2	112
Ketones	CARBKET	10	93	40	194
Other carbonyls	CARBOTH	0.3	3	1	50
Vapor phase semivolatile organic					
Acids	SVOACID	0	3670	500	311
Alkanes	SVOALK	0	63940	1209	200
Hopanes	SVOHOP	0	19	0	0
Pahs	SVOPAH	0	16057	18846	11060
Phenols	SVOPHEN	0	0	0	21482
Steranes	SVOSTER	0	30	0	0
Sugars	SVOSUG	0	0	0	1714

[a]Concentrations are in $\mu g/m^3$ for all substances, except for elements, which are expressed in ng/m^3.

Table 2

Response indicators of atherosclerosis.

Response indicator	Variable name	Description/relevance
Vascular endothelin-1	ET-1	Key mediator of vascular tone (vasoconstrictor); displays mitogenic properties in cardiovascular system. Found upregulated in atherosclerosis.
Vascular endothelin growth factor	VEGF	Stimulates vasculogenesis and angiogenesis (including after injury); chemotatic for leukocytes. Contributes to atherosclerotic plaque progression and destabilization.
Matrix metalloproteinase-3	MMP3	Collagenase involved in connective tissue remodeling and activation of other MMPs, such as MMP-7 and -9. Genetic variants of MMP3 are associated with plaque rupture and myocardial infarction.
Matrix metalloproteinase-7	MMP7	Degrades proteoglycans, fibronectin, elastin, and casein and is involved in tissue remodeling and wound repair. Found upregulated in atherosclerosis.
Matrix metalloproteinase-9	MMP9	Gelatinase associated with tissue remodeling and mobilization of hematopoietic progenitor cells. Found upregulated in atherosclerosis and activity is associated with plaque destabilization/rupture.
Tissue inhibitor of metalloproteinase-2	TIMP2	Tissue inhibitor of MMPs; found upregulated in atherosclerotic vessels.
Heme-oxygenase-1	HO-1	Enzyme that catalyzes degradation of heme; inducible in response to oxidative stress and cytokine expression.
Thiobarbituric acid reactive substances	TBARS	Formed as a byproduct of lipid peroxidation; indicator of oxidative stress. Found increased in atherosclerotic vessels.

Table 3

Experimental groups.

	Experimental group designations (No. of animals) Relative dilution				
Atmosphere	**Control (0)** 0	**Low (1)** 1/10	**Mid (2)** 1/3	**High (3)** 1	**High-Filt (4)**[A] 1
Coal emissions (CE)	CE0 (10)	CE1 (10)	CE2 (10)	CE3 (10)	CE4 (10)
Diesel exhaust (DE)	DE0 (10)	DE1 (10)	DE2 (10)	DE3 (10)	—[a]
Gasoline exhaust (GE)	GE0 (8)	GE1 (8)	GE2 (8)	GE3 (8)	GE4 (8)
Wood smoke (WS)	WS0 (10)	WS1 (10)	WS2 (10)	WS3 (10)	—[b]

[a]Same gas concentrations as high group, but with particles filtered out.

[b]Response data reported in other publications, but a full atmospheric characterization of the components in Table 1 was not done for this group, precluding its use in this analysis.

Table 4

MART variable importance scores for predictors of atherosclerosis response indicators.

ET-1		VEGF		MMP3		MMP7		MMP9		TIMP2		HO-1		TBARS	
SO₂	100	NH₃	100	NH₃	100	NH₃	100	NH₃	100	NO₂	100	NH₃	100	SO₂	100
NO₂	49	NMVOOXY	41	NMVOALKE	41	VOAALI	50	SO₂	99	NO	73	NMVOHAL	79	NO₂	90
CO	38	CO	29	NMVOOXY	29	CARBDI	47	CO	45	POHOP	30	CARBALI	45	NO	75
NO	37	AMMONIUM	28	SO₂	28	AMMONIUM	33	CARBFUR	44	SVOALK	26	CARBFUR	42	NMVOALKE	56
NH₃	36	CARBALI	24	CO	24	CARBOTH	31	NO₂	41	SVOACID	26	NMVOFUR	36	CO	48
AMMONIUM	30	NMVOALKE	23	ELEMENTS	23	CARBALI	30	AMMONIUM	38	NH₃	25	NMVOALKA	32	POHOP	46
ELEMENTS	29	PM	23	NMVOHAL	23	EC	30	CARBDI	35	CO	23	EC	32	NH₃	38
EC	22	SO₂	22	NMVOALKA	23	SULFATE	29	NMVOALKE	35	EC	23	CO	29	NMVOALKY	36
CARBDI	20	ELEMENTS	20	SVOACID	23	ELEMENTS	28	VOAALI	35	ELEMENTS	23	OC	24	NMVOOXY	34
OC	20	NMVOHAL	20	CARBALI	23	OC	27	OC	30	CARBOTH	18	CARBKET	24	CARBALKA	34
NMVOHAL	18	OC	18	AMMONIUM	21	SO₂	27	EC	29	NMVOALKE	17	SVOACID	23	SVOALK	33
VOAALI	17	VOAALI	17	OC	21	CO	27	NMVOHAL	29	NMVOALKY	16	CARBHYD	22	CARBARO	30
PM	17	SVOACID	17	EC	20	NITRATE	25	NO	26	OC	16	CARBDI	21	EC	29
NMVOALKY	16	NMVOALKA	16	PM	18	CARBARO	20	PM	23	SULFATE	16	VOAARO	20	SVOACID	29
CARBALKE	15	NITRATE	15	NO₂	18	NMVOHAL	19	ELEMENTS	22	CARBHYD	15	ELEMENTS	20	NMVOALKA	28
NMVOARO	14	NMVOARO	14	NO	17	NO₂	19	CARBOTH	22	NMVOHAL	15	AMMONIUM	19	AMMONIUM	28
SVOACID	14	SVOALK	14	NITRATE	15	NMVOALKA	19	CARBALI	19	NMVOOXY	13	NMVOOXY	18	ELEMENTS	27
CARBALI	13	NO	13	SULFATE	15	NMVOARO	18	NITRATE	19	CARBALKA	12	PM	18	SVOHOP	25
SULFATE	13	CARBDI	13	VOAARO	14	VOAARO	16	SVOACID	18	AMMONIUM	11	NMVOALKE	17	NMVOARO	23
NMVOALKE	12	SVOSUG	12	CARBALKA	14	SVOACID	15	NMVOOXY	17	VOAARO	10	SULFATE	17	CARBOTH	23
POHOP	11	CARBALKA	11	NMVOARO	13	PM	15	VOAARO	17	CARBKET	10	SO₂	16	CARBALI	21
CARBOTH	11	NO₂	11	CARBOTH	13	CARBFUR	14	NMVOALKA	16	POACID	8	NMVOALKY	15	CARBFUR	20
NMVOALKA	11	CARBOTH	11	SVOALK	12	NMVOOXY	13	CARBALKE	15	NMVOARO	8	NITRATE	15	VOAALI	18
VOAARO	10	NMVOALKY	10	CARBDI	12	NO	12	NMVOARO	14	NMVOALKA	7	VOAALI	15	OC	16
SVOALK	10	VOAARO	10	CARBFUR	12	CARBHYD	10	CARBALKA	13	SO₂	7	NO₂	14	NMVOHAL	16
SVOSUG	9	EC	9	SVOSUG	12	CARBKET	10	SVOALK	13	CARBALI	7	CARBOTH	14	SULFATE	16
NMVOOXY	8	CARBARO	8	VOAALI	12	NMVOALKE	10	CARBKET	12	NITRATE	7	POACID	13	PM	15
CARBALKA	8	SVOHOP	8	NMVOALKY	11	CARBALKA	9	CARBHYD	7	CARBFUR	7	CARBALKA	12	VOAARO	15

ET-1		VEGF		MMP3		MMP7		MMP9		TIMP2		HO-1		TBARS	
CARBFUR	11	CARBALKE	7	NMVOFUR	8	NMVOFUR	8	POHOP	5	PM	7	SVOSUG	11	SVOPAH	13
NITRATE	11	CARBFUR	7	POACID	8	SVOALK	7	SVOHOP	5	CARBALKE	7	NO	9	CARBALKE	12
CARBARO	10	CARBHYD	6	CARBARO	7	NMVOALKY	6	SULFATE	2	POALK	6	NMVOARO	9	NITRATE	12
CARBKET	10	SULFATE	6	CARBALKE	7	POALK	6	SVOSUG	0	VOAALI	6	SVOALK	9	CARBKET	11
POACID	8	CARBKET	5	CARBHYD	4	POHOP	2	NMVOALKY	0	CARBARO	5	CARBARO	8	CARBDI	8
SVOPAH	6	POALK	5	POPHEN	4	POPAH	0	CARBARO	0	SVOSUG	4	POALK	8	POALK	7
SVOHOP	5	NMVOFUR	5	POSTER	3	POPHEN	0	POACID	0	CARBDI	4	SVOHOP	6	CARBHYD	7
CARBHYD	5	POACID	3	POSUG	2	POSTER	0	NMVOFUR	0	SVOHOP	4	CARBALKE	5	SVOSUG	2
NMVOFUR	2	POSUG	3	POALK	0	POSUG	0	SVOPAH	0	NMVOFUR	2	VOAALK	0	POACID	0
VOAALK	0	POSTER	0	CARBKET	0	CARBALKE	0	POPHEN	0	POPHEN	0	POHOP	0	POPAH	0
POSTER	0	POPHEN	0	POHOP	0	POACID	0	VOAALK	0	POSUG	0	POSTER	0	POSTER	0
POPHEN	0	POHOP	0	SVOHOP	0	SVOHOP	0	POSUG	0	POSTER	0	SVOPAH	0	NMVOFUR	0
SVOPHEN	0	SVOPAH	0	SVOPAH	0	SVOPAH	0	POALK	0	SVOPAH	0	SVOPHEN	0	SVOPHEN	0
SVOSTER	0	SVOPHEN	0.	SVOPHEN	0	SVOPHEN	0	SVOPHEN	0	SVOPHEN	0	SVOSTER	0	SVOSTER	0
POALK	0	SVOSTER	0	SVOSTER	0	SVOSTER	0	SVOSTER	0	SVOSTER	0	POPHEN	0	VOAALK	0
POPAH	0	VOAALK	0	POPAH	0	SVOSUG	0	POPAH	0	POPAH	0	POPAH	0	POSUG	0
POSUG	0	POPAH	0	VOAALK	0	VOAALK	0	POSTER	0	VOAALK	0	POSUG	0	POPHEN	0

Table 5

Performance of MART-selected predictors in explaining response variation.

Response indicator	Model	Model-explained variation	Unexplained (random) variation	Model explained variation/group mean explained variation	p value for lack of fit to experimental group means	Model explained variation/total variation (R^2)
ET-1	Experimental group means	39.0	66.9	1.00		0.36
	MART – all variables	32.9		0.84	0.83	0.31
	Main effects – 3 variables	25.7		0.66	0.02	0.24
	SO_2	21.4		0.55		0.20
	NO_2	6.5		0.17		0.06
	CO	0.7		0.02		<0.01
VEGF	Experimental group means	17.4	35.4	1.00		0.33
	MART – all variables	16.0		0.92	0.87	0.30
	Main effects – 3 variables	12.1		0.69	<0.01	0.23
	NH_3	10.5		0.60		0.20
	NMVOOXY	2.6		0.15		0.05
	CO	0.4		0.02		<0.01
MMP3	Experimental group means	28.2	124.4	1.00		0.18
	MART – all variables	19.2		0.68	0.76	0.13
	Main effects – 3 variables	17.6		0.62	0.15	0.12
	NH_3	12.7		0.45		0.08
	NMVOALKE	3.9		0.14		0.03
	NMVOOXY	4.4		0.15		0.03
MMP7	Experimental group means	112.2	445.9	1.00		0.20
	MART – all variables	36.6		0.33	0.25	0.07
	Main effects – 3 variables	37.7		0.33	0.33	0.07
	NH_3	12.7		0.11		0.02
	VOALLI	22.9		0.20		0.04
	CARBDI	5.3		0.05		<0.01
MMP9	Experimental group means	145.3	110.3	1.00		0.57
	MART – all variables	135.7		0.93	0.99	0.53

Response indicator	Model	Model-explained variation	Unexplained (random) variation	Model explained variation/group mean explained variation	*p* value for lack of fit to experimental group means	Model explained variation/total variation (R^2)
	Main effects – 3 variables	116.2		0.80	0.04	0.45
	NH_3	92.3		0.64		0.36
	SO_2	42.0		0.29		0.16
	CO	< 0.01		<0.01		<0.01
TIMP2	Experimental group means	110.0	205.6	1.00		0.35
	MART – all variables	92.7		0.84	0.99	0.29
	Main effects – 3 variables	71.7		0.65	0.09	0.23
	NO_2	40.5		0.37		0.13
	NO	26.2		0.23		0.08
	POHOP	10.6		0.10		0.03
HO-1	Experimental group means	123.3	400.5	1.00		0.24
	MART – all variables	77.7		0.63	0.93	0.15
	Main effects – 3 variables	26.7		0.22	0.13	0.05
	NH_3	22.1		0.18		0.04
	NMVOHAL	3.2		0.03		<0.01
	CARBDI	2.6		0.02		<0.01
TBARS	Experimental group means	103.8	55.3	1.00		0.65
	MART – all variables	93.3		0.90	0.99	0.57
	Main effects – 3 variables	72.8		0.70	< 0.01	0.46
	SO_2	35.0		0.34		0.21
	NO_2	30.9		0.30		0.19
	NO	23.1		0.22		0.15