Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy

TITLE **Quantitative Risk Assessment of the Pulmonary Toxicity of Nano-particles by Machine-Learning-Enabled Meta-Analysis**

PRESENTED BY **Jeremy M. Gernand**

ACCEPTED BY THE DEPARTMENT OF

Engineering and Public Policy

ADVISOR, MAJOR PROFESSOR	DATE
DEPARTMENT HEAD	DATE

APPROVED BY THE COLLEGE COUNCIL

DEAN	DATE

Quantitative Risk Assessment of the Pulmonary Toxicity of Nanoparticles by Machine-Learning-Enabled Meta-Analysis

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy
in
Engineering and Public Policy

Jeremy M. Gernand

Department of Engineering and Public Policy
Carnegie Mellon University
Pittsburgh, Pennsylvania
August 2013

Dissertation Committee

Elizabeth A. Casman, PhD (chair)
Associate Research Professor
Department of Engineering and Public Policy
Carnegie Mellon University

Daniel B. Neill, PhD
Associate Professor of Information Systems
H. John Heinz III College
Carnegie Mellon University

Mitchell J. Small, PhD
H. John Heinz III Professor of Environmental Engineering
Department of Engineering and Public Policy and Civil and Environmental Engineering
Carnegie Mellon University

Jason M. Unrine, PhD
Assistant Professor
Department of Plant and Soil Sciences
University of Kentucky

Acknowledgements

I am very grateful for the efforts of my committee.

Liz, your accessibility, responsiveness, and approachability as a research advisor has been unmatched in my experience. Your limitless appetite for continuing to make each bit of research and each paper draft a little better is appreciated (if not always immediately in the moment). Our many discussions over the years have helped me put my own thoughts in better order, and our cordial disagreements, though not frequent, have always been enlightening.

Mitch, from our initial introduction when you played with my 18-month-old son and made clear your priority for family to our discussions focused on improving the quality of this research with better quantitative analyis, I have valued your perspective and priorities and often kept your example in mind as I made my way through the program.

Daniel, I enjoyed your focus in using machine learning and computer science to address real problems affecting real people that played a big part in the development and evolution of my methods, and I have been grateful for your perspective on my research, which helped me keep the appropriate frame of mind for the significance of my results.

Jason, thank you for offering the perspective of a toxicologist and helping me be more focused on the big questions and not being lost in the "forest" of my methodology, and thanks for the encouraging words at CEINT meetings over the past few years.

I also appreciate Amy Wang's perspective and significant efforts in improving chapter 3 and conversations about making this research in general more comprehensible and useful to toxicologists. I also would like to acknowledge the contributions of Amy Wang and Samantha Frady of the Environmental Protection Agency (EPA), who compiled the pulmonary toxicity data on titanium dioxide nanoparticles used in chapter 3.

I am grateful for the assistance of Sophie Grodsinsky, who collected and compiled the pulmonary toxicity data for the metal oxide nanoparticles used in chapter 4. I hope your summer research experience was as useful to you as it was to me.

Vicki, always cheerfully answering questions or lending aid whenever I asked, the tone you set for the administrative office and interactions with all of us graduate students is greatly appreciated. For all of the Engineering and Public Policy administrative staff, including several who have come and gone during my time here, I sincerely appreciate all your efforts in making this department a great place to work.

Alison, my wife, having previously walked this path just a few years ago, you have been understanding and patient. I am looking forward to the next chapter of life we will be writing together.

Ezra, you have grown up from a toddling baby into an active, inquisitive, and kind boy while your daddy has been engaged in this endeavor. I am glad that you have always been there to remind me of what is most important in my life.

Jude, born just 16 months ago, you have now grown into a very sweet, curious, and brave toddler during this time of study for your father. While you don't know it yet, you have lent me focus and drive in completing this work, and for that I am thankful.

Funding for this research was provided by the National Science Foundation (NSF) and the Environmental Protection Agency (EPA) under NSF Cooperative Agreement EF-0830093, Center for the Environmental Implications of NanoTechnology (CEINT), the Carnegie Institute of Technology (CIT) Dean's Fellowship, the Prem Narain Srivastava Legacy Fellowship, the Neil and Jo Bushnell Fellowship, the Bertucci Fellowship in Engineering, and the Steinbrenner Institute for Environmental Education and Research (SEER). Any opinions, findings, conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF, the EPA, or other funders. This work has not been subjected to EPA review and no official endorsement should be inferred.

Abstract

Accurately anticipating the toxic risks and specific factors contributing to the toxic risks of nanomaterials is a necessary step for the safe and effective proliferation, utilization, and regulation of these unique materials. This thesis addresses this problem through meta-analysis on existing nanomaterial pulmonary toxicity experiments as enabled by the use of machine learning algorithms including regression trees and random forests models at a time when the completeness of the data do not support traditional meta-analysis techniques like multiple linear regression. This thesis presents the results of analysis using these models to identify the most important nanomaterial characteristics contributing to toxicity as well as the magnitude of changes in toxicity expected from changes in those characteristics.

This thesis presents predictive models for the pulmonary toxicity of carbon nanotubes and titanium dioxide nanoparticles showing the degree to which changes in experimental design, nanomaterial dimensions, impurities, and aggregation might explain differences in observed toxicity.

Secondly, this thesis presents the predictions of random forest models revealing interactions between 2 or 3 nanomaterial characteristics and exposure attributes in a manner such that a material designer might minimize risk while continuing to meet functional objectives.

Table of Contents

List of Tables

List of Figures

1 Introduction

Nanotechnology is a relatively new innovation and advances our capability to control the properties of materials at a more fundamental level. Materials scientists have found nanomaterials to display many unique and potentially advantageous properties not shared by solutions or bulk solids of the same materials. The manufacture of nanomaterials is generally anticipated to accelerate both in terms of varieties and quantities [1], [2]. Since, these materials vary from one to another not only in molecular composition, but also in terms of size, shape, and surface coatings among other attributes, the sheer variety of particle configurations presents a problem for the study of nanotoxicology: to what extent can the variants be considered a as members of similarly behaving groups, or must each type be evaluated individually.

Understanding the toxic potential of these nanoparticles is important for accurately assessing the risks and protecting the health and safety of workers and consumers. One of the earliest and most significant exposures to nanomaterials by people will be through inhalation by workers in factories and consumers using aerosol products. However, our understanding of these risks based on published experiments on pulmonary exposures in rodents is impeded by missing nanomaterial characterization measurements and by the relatively large number of variables as compared to the number of experiments.

A modeling framework that can anticipate the toxic potential of a specific nanomaterial's design would be of great benefit to manufacturers and regulators. Nanomaterial designers and manufacturers could use such risk models to reduce the

likelihood of future problems arising, while continuing to employ these materials to their greatest benefit.

The pursuit of Quantitative Structure Activity Relationship (QSAR) toxicity models for nanomaterials has so far been limited to *in vitro* exposures to metal oxide nanoparticles of a very confined size range [3], [4]. What are needed are predictive toxicity models that account for differences in nanomaterial size and shape as well as differences in chemistry. Of greater benefit would be toxicity models based on mammalian *in vivo* exposures, since the correspondence between *in vitro* and *in vivo* exposure results remains unclear [5].

Although many studies exploring pulmonary exposures to nanoparticles exist, the number of differences between the nanomaterials used rivals the number of separate experiments performed. This combined with the fact that due to evolving laboratory measurement capabilities over time and differing investigator priorities, some characteristics of the tested nanomaterials are unknown.

Linear models and other algebraic models have difficulty discerning trends from this kind of information. Casting aside data records with missing input attributes, these models can discard the majority of available information in the quest to resolve statistically significant coefficients. Random Forests (RF) and Regression Trees (RT), on the other hand, are machine learning models based on hierarchal series of decision rules and learn from and produce predictions even when specific input attributes are missing [6], [7]. Relying on these methods for a meta-analysis of existing knowledge on the pulmonary risks of nanomaterials can yield risk models that begin to ascribe differences in observed toxicity to differences in specific nanomaterial attributes.

This dissertation is composed of 3 papers that utilize machine learning models based on meta-analysis of rodent pulmonary toxicity experiments to illuminate the contribution of nanomaterial attributes to risk. The first in Chapter 2, explores the degree to which physical dimensions, aggregation, and metallic impurities affect the toxicity of carbon nanotubes. The second in Chapter 3, describes the contribution of particle size and aggregation to the toxicity of titanium dioxide nanoparticles. The last in Chapter 4, explores interactions between changes in 2 or 3 nanomaterial physicochemical attributes and pulmonary toxicity for a set of nanoparticles including carbon nanotubes and metal oxides. Chapter 5 discusses recommendations for the direction of future research in this area as well as implications for policy makers.

2 A Meta-Analysis of Carbon Nanotube Pulmonary Toxicity Studies – How Physical Dimensions and Impurities Affect the Toxicity of Carbon Nanotubes

2.1 Abstract

This paper presents a regression-tree-based meta-analysis of rodent pulmonary toxicity studies of uncoated, non-functionalized carbon nanotube (CNT) exposure. The resulting analysis provides quantitative estimates of the contribution of CNT attributes (impurities, physical dimensions, aggregation) to pulmonary toxicity indicators in bronchoalveolar lavage fluid: neutrophil and macrophage count, and lactate dehydrogenase and total protein concentrations. The method employs classification and regression tree (CART) models, techniques which are relatively insensitive to data defects that impair other types of regression analysis: high-dimensionality, non-linearity, correlated variables, and significant quantities of missing values. Three types of analysis are presented, the regression tree, the random forest, and a random forest based dose-response model. The regression tree shows the best single model supported by all the data and typically contains a small number of variables. The random forest shows how much variance reduction is associated with every variable in the data set. The dose-response model is used to isolate the effects of CNT attributes from the CNT dose,

showing the shift in the dose-response caused by the attribute across the measured range of CNT doses. It was found that the CNT attributes that contribute the most to pulmonary toxicity were metallic impurities (cobalt significantly increased observed toxicity, while other impurities had mixed effects), CNT length (negatively correlated with most toxicity indicators), CNT diameter (significantly positively associated with toxicity), and aggregate size (negatively correlated with cell damage indicators and positively correlated with immune response indicators). Increasing CNT specific surface area decreased toxicity indicators.

2.2 Introduction

At one time the fibrous structure of carbon nanotubes (CNTs) led some to suspect that inhaled CNTs might cause pathologies similar to those associated with inhaled asbestos[8]. While CNT pulmonary exposure has not been observed to cause mesotheliomas, it does lead to the formation of granulomas[9], lung inflammation, and fibrotic responses[10]. Mesothelioma has been induced in p53+/- mice by injecting CNTs into their peritoneal cavities[11] and in Fischer 344 rats via intra-scrotal injection[12], but not by pulmonary routes of exposure[13]. Studies of long term (> 90 day) chronic exposures to CNTs have yet to be published, however. Recently researchers have observed CNTs acting as a promoter (but not an initiator) of cancer in mice [14]. Risks of cancer will likely dominate future regulatory decisions on the safety of CNTs.

The expected proliferation of products containing engineered carbon nanotubes in many configurations will mean that efforts to test toxicity of these variants through conventional animal experimentation will be burdensome. If the attributes of exposure and nanotube characteristics responsible for variations in CNT toxicity could be

identified, this could potentially reduce the testing burden and provide guidance for the design and production of safer CNTs. This is the motivation of various attempts to identify property/toxicity relationships for nanomaterials, including this one.

The development of quantitative structure activity relationships (QSARs) for nanomaterials is also motivated by the desire to transition out of the animal testing paradigm. The goal of QSARs is to predict the activity—in this case, toxicity—of a material from its chemical and physical structure. QSAR studies and toxicity modeling studies for nanomaterials have focused predominately on *in vitro* toxicity indicators, especially those amenable to high throughput screening[4], [15], [16], [17], [18], [19].The only QSAR toxicity study on CNTs[20] to date concluded that the cytotoxicity of multi-walled carbon nanotubes (MWCNT) to bacteria is enhanced by chemical treatments that increase MWCNT-cell contact opportunities (uncapping, debundling, length shortening, and dispersion).

Our work differs from these studies in its focus on quantitative *in vivo* pulmonary toxicity indicators, and those factors varying between different batches of carbon nanotubes, specifically physical dimensions and impurity content. For nearly all of the other traditional QSAR descriptors, such as polarizability, hydrophobicity, or surface charge, there are currently no studies of the *in vivo* toxicity of functionalized or otherwise chemically modified CNTs that could be used to determine the effects of these properties. The major differences that do exist between batches of CNTs used in the literature on pulmonary toxicity studies are variations in dimensions, dispersion or aggregation, impurity content, and CNT configuration (multi-walled versus single-walled).

Only one effort has been made to date towards meta-analysis of CNT toxicity studies, and it did not address physical or chemical variation in the CNT materials themselves[21], focusing instead on the question of whether penetration of cell membranes is likely to occur in *in vitro* experiments. The study concluded that penetration of human cells is "somewhat possible" and could result in cellular damage, recommending the minimization of human exposure.

One challenge of our meta-analysis was the absence of standardized nanomaterial characterization protocols. Investigators typically used procedures generally similar to one another, but with certain specific differences. Processing of CNTs prior to animal exposure varied between studies. Characterization of CNTs used in instillation exposures sometimes occurred before, after, or both before and after combining the CNTs with the suspension fluid. Researchers often only reported the properties of the nanomaterial believed to be important for their research question, leaving a substantial amount of uncertainty regarding the actual form of the material. For example, CNT inhalation studies differ significantly in measurement and reporting of the distribution of aggregate sizes[22], [23], the amount and type of metallic impurities[24], [25], and the presence of carbon impurities[26], [27]. These practices can make employing traditional meta-analysis techniques difficult, because they result in a data set with many missing values.

This paper presents a meta-analysis based on machine-learning-algorithms, specifically regression trees (RT)[6] and the associated ensemble method, random forests (RF)[7]. These methods experience less degradation from missing data than multiple linear regression by ordinary least squares, enabling meaning to be extracted from the relatively thinly populated data set.

RT and RF models have been employed in generating quantitative predictive models to for biological activity of different chemical compounds including toxicity[28] and bioaccumulation[29], applications similar to this study of physical and chemical characteristics associated with CNT toxicity. Investigators have also used RF models to identify important epidemiological and phenotypic differences in salmonella strains[30]. Also, RF models increasingly play a role in identifying important interactions between single gene risk factors in disease association studies[31].

The objectives of this research are three-fold, (1) to rank CNT properties and experimental conditions by their information content, (2) to quantify the influence of different CNT properties on dose-response relationships, and (3) to lay the computational foundation RT and RF based meta-analysis for toxicology studies.

2.3 Methods

2.3.1 Selection of CNT Toxicity Studies

As of October 2012, the archival literature contained 17 carbon nanotube toxicity studies using rodent (rat and mouse) models exposed through inhalation, aspiration, or intratracheal instillation that met our screening criteria. To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study (Table 2-1). We did not include studies whose endpoints were the presence or absence of gross pathologies because there were insufficient data to differentiate between the many variations in CNT characteristics. Limiting the analysis to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

Given the importance of pathologies in understanding the risks of exposure to inhaled nanomaterials, especially the possibility of cancer, this analysis should be extended to incorporate these categorical or class outcomes as soon as sufficient data are available.

Four individual pulmonary toxicity endpoints covering a range of effects were reported in the identified studies: polymorphonuclear neutrophils (PMN), macrophages (MAC), lactate dehydrogenase (LDH) and total protein (TP). These endpoints reflect several dimensions of immune response and cell membrane damage and death. These indicators were all measured in bronchoalveolar lavage (BAL) fluid extracted from the lungs of the mice or rats, and were reported as a counts per subject or fold of control measurements (the average indicator count or concentration in animal test subjects divided by the average count or concentration in control animals).

We converted all toxicity results to fold of control format, a form that many of the studies already reported. We found by experimentation that this step was key for comparing test subjects across species and across modes of exposure. This data set is available at http://nanohub.org/resources/13515[32].

2.3.2 Preparation of CNT Toxicity Data for Analysis

We treated all nanoparticle properties and experimental conditions as independent variables. In Table II there are four categorical variables, CNT configuration, exposure mode, sex, and animal species. The inclusion of these variables is an essential feature of the meta-analysis, allowing the dissimilar experimental systems to be included in a single analysis. Another important experimental design variable was the Post-Exposure Period, the time after exposure that elapsed before the BAL fluid was examined for toxicity

indicators. Table I only shows dose measured in μg CNT per kg of animal. This is just one of five ways dose was reported in the input data (Table 2-2), and it was the dose unit that resulted in the regression trees with the highest explanatory power.

The input for RT and RF analysis is a matrix in which the rows represent individual experimental animals. The 41 columns of the matrix contained 20 experimental conditions, 17 nanoparticle properties (such as CNT length, diameter, specific surface area, and average aggregate size), and the 4 experimental endpoints (toxicity measures). For diagnostic purposes, a series of 2 to 6 uniformly distributed random variables between 0 and 1 were added to the data array as columns at different stages of the analysis. The random variables—by definition information-free and not correlated with any toxicity endpoint values—supplied an independent measure with which to compare the discriminatory performance between information-rich and information-poor attributes of the RT and RF model generation and pruning algorithms. Missing values for any particular nanomaterial property or experimental condition in any published study were left blank.

Several variables utilized in the meta-analysis were calculated or inferred from the data when not reported directly in the individual studies. For example, values for median CNT length and diameter were derived from reported maximum and minimum values by using the midpoint between them. When not reported, total and 24-hour doses for each experiment were calculated from reported concentrations. Specific surface area and total mass dose were combined to produce a value for surface area dose. In no cases were nanomaterial properties or experimental values inserted based on typical values or model-predicted quantities. In the individual studies, the toxicity endpoint results were

all reported as a mean and standard deviation. We assumed the endpoint values were normally distributed.

The number of matrix rows per experiment was 100 times the number of animals in each experimental group. For example, if an experiment involved 12 test animals at one dose and 12 test animals at a different dose, the experiment would occupy 1200 + 1200 = 2400 rows. Each of these rows would contain identical nanomaterial and experimental attributes, the appropriate exposure metrics, but for the experimental endpoints (PMN, MAC, LDH, and TP), each row would have a unique, discrete realization of a normal distribution with mean and standard deviation reported for its group of animals. (Endpoint values deemed impossible—negative protein counts, for instance—were discarded and replaced with the nearest plausible value, usually zero.) Higher rates of sampling at 500 or 1000 samples per experimental animal did not alter the ordinal value of model variables or goodness of fit measures in the resulting RTs.

The probabilistic representation of the experimental results preserves the inherent variability in the measured responses and reduces the likelihood of the models reporting noise or of their oversensitivity to small differences in experimental exposure conditions. The assumption of a normal distribution for the toxicity endpoint measures likely overestimates the amount of uncertainty present given that biological limits restrict the range of certain measures. This procedure also has the effect of weighting the influence of each study on the RT and RF models in relation to the number of animal subjects involved. Furthermore, incorporating experimental uncertainty of the data in this manner enables the models to output the corresponding range of likely values rather than being limited to a single value.

2.3.3 Regression Trees

For each measured toxicity endpoint we created a RT model using the MATLAB™ function "classregtree." The RT algorithm successively divides the population of observations (toxicity endpoint values, also called output variables) into binary groups based on an inequality for quantitative input variables or a categorical grouping for categorical variables. At each branching point, the input variable and the split criterion (a value of the input variable) are chosen to produce the greatest possible information gain between the resulting two populations of observations. In this study we developed and applied a calculation of variance reduction as a measure of information gain. While entropy increase may also be used as a measure of information gain, it does not result in substantially different results, and variance reduction is more consistent with the statistics produced by alternative regression models, facilitating comparisons.

To prevent over-fitting, each RT model is pruned through a process similar to backwards stepwise elimination for linear models. In order to prune the model, the error of the model is calculated through ten-fold cross-validation—a process by which the data set is divided randomly into 10 subsets of rows, the RT model branches and split criteria are set based on a matrix containing 9 of those subsets, and the error is measured against the model predictions for the 10^{th} subset. This process is repeated until each of the 10 subsets has been withheld from the model and used for error calculation, and then the total sum-squared error for the model is calculated. The RT model is then pruned by removing the branches providing the least error reduction until the model reaches the smallest size possible within one standard error of the minimum error. The RT model for neutrophils is shown in Figure 2-1. The other three RTs can be found at http://nanohub.org/resources/15901.

The importance of each variable included in the RT model is calculated based on the total variance reduction achieved by each branch in the final model. For variables appearing on multiple branches, the variance reduction is summed across the entire model.

One reads a tree from the top down, following each branch to its terminal leaf. For example, in Figure 2-1, if the total CNT dose is less than 5150 μg/kg and the dose of Cr is more than 3.13 μg/kg and neutrophils were measured less than 17.5 days after exposure, and the dose of Fe was less than 2600 μg/kg, and the neutrophils were measured more than 4 days after exposure, and the total CNT dose was more than 1300 μg/kg, then the neutrophil counts in the 600 matrix rows representing animals satisfying these conditions would be on average 293 times higher than neutrophils in control animals.

While linear models and other algebraic models are undefined whenever any of the input variables are missing, this is not true of a RT model. Since overall 19% of the characterization or exposure attributes in the data matrix are missing and 74% of the variables have at least one missing value, complete utilization of all available data is a distinct advantage for RTs over multiple linear regression (MLR) models. Further, this RT property of producing a prediction regardless of the completeness of the data record prevents the underestimation of error that is common when regression models are applied to data sets with missing values.

While other machine learning methods such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) can also handle nonlinear relationships and may be more computationally efficient, they become undefined for cases where any of the input data is missing. So, while imputation of missing values would be required for MLR,

ANN, or SVM models to take full advantage of the entire experimental data set, this step and potential source of artificially introduced uncertainty is unnecessary for RTs.

RTs show the best model for a given data set in the sense of parsimony, but there are ways to interrogate the data that are even less sensitive to their biases, such as the Random Forest (RF).

2.3.4 Random Forests

RF models have the advantage of more fully exploring the value of each variable in the data set, compared to RT models, which typically include a small subset of them. They also are less sensitive to weaknesses in the data. With RTs, variables can appear influential due to their over-representation in the data. The randomized variable selection process for the RF model generation procedure ensures that all variables are evaluated.

We generated a RF model for each toxicity endpoint implementing Breiman's algorithm with the MATLAB™ function "treebagger." The RF model is composed of an aggregate collection of RT models, each created from a data set where each branch is selected from a random subset of one third of the available variables (matrix columns). Each RF model contains at least 1000 RT models, whose elements are averaged to produce the RF output. The RFs were grown (i.e. trees added) until the error decreased less than 0.1% with the addition of 100 trees to the RF model. (See Appendix A for full model documentation.)

The variable importance results from the RF model are calculated in the same way as for the RT model; however, the final variance reduction for each variable is calculated from the average across all of the trees in the RF model. This is accomplished by calculating the variance reduction caused by each branch variable in each individual tree in the forest, summing the variance reductions by variable across all of the trees in the

RF, and then dividing the final values by the total number of trees in the RF model. These error reduction values form the measure of variable information content used in ranking the variables by their importance in Figure 2-2.

The RF models used for ranking variables and in the dose-response curve generation (in the next section) were trimmed by first removing all variables with less calculated importance than the average of the uniform random variables. Then, model refinement proceeded in a forward stepwise fashion evaluated on the basis of the R^2 value. In each step, the additional variable producing the highest possible R^2 value was selected.

To investigate the contribution of individual variables to a baseline CNT dose-response curve, we use a reduced version of the RF model containing only the variables accounting for >99% of the variance reduction plus the variable of interest. We ran this model at minimum and maximum levels of the variable of interest in the input matrix, holding all other variables constant at their median values while varying the dose variable. The two resulting curves show the shift in the dose-response associated with the variable of interest (Figure 2-3). Calculating the mean shift across the range of applied CNT doses enables a quantitative evaluation of the relative effect of different variables on toxicity (See Results).

In order to estimate the uncertainty in each of these model results, a RT model was trained to predict the observed experimental variance rather than the observed mean output value. Mean dose response curves plus and minus one estimated standard deviation are plotted in Figure 2-3. The RF standard deviations reflect an expectation of the population standard deviation.

The pixelated appearance of RF dose-response curves is due to the small number of unique experimental data points available in a given sub-section of the attribute space. If

more data were available, the shape of the response curve would be smoothed. Locations of abrupt steps in the RF curves are reflective of fewer data points being available under those conditions. These large steps should be considered indicative of limited knowledge at these dose levels rather than indicative of a significant transition occurring at that specific point.

Of the 37 input variables, the RT and RF models contain only between 3 and 13 variables (Table 2-3), yet the information these models explain approaches the maximum possible information value a model could contain if all differences between experimental groups were fully captured. (See the 'Maximum Model Performance' column in Table 2-3.) An exposure group is defined as a group of animals that experienced the same level of exposure to the same batch of CNTs and the same length of recovery period. When the number of variables in a RF model is much smaller than the number of exposure groups, this gives confidence that the RF model is not over-fit. That the performance of the RF models approaches the limit of explaining all but the inherent experimental variability means that the most significant controllable factors affecting CNT toxicity are explicitly or implicitly included.

2.3.5 Model Validation

Table 2-4 displays a study-by-study validation of the RF model in terms of mean squared error (MSE). A RF model was generated after withholding the target study from the training data set. Then, the specifics of the withheld study were fed into the model as test data inputs. The output of the model is what would be predicted if that experiment had yet to be performed. The experimental results and model predictions (and residuals) display the overall good performance of this modeling technique for anticipating the results of future experiments. The RF model cannot extrapolate based on trends,

however, and can only be used in this manner for combinations of inputs that lie within the limits of the training data.

As seen in Table 2-4, the mean squared error of the overall model to all observations is similar to that resulting when each individual study is withheld and used as a test case. The one outlier, the study by Pauluhn et al., 2010, involved an inhalation exposure with a much longer exposure period than the other studies. The MSE also reflects the relative similarity in results between the selected study and the information contained in the others, as illustrated in particular by the 90-day inhalation study (Pauluhn, 2010), and the LDH results for the Shvedova et al., 2007 study. The overall impression from these results indicates that the RF models are generally reliable when predicting the results of a new study so long as no significant extrapolation of any of the input variables of the training data is required.

A complete model description and validation document including code with model training results is available in Appendix A.

2.4 Results

The RT model for PMN displayed in Figure 2-1 demonstrates the structure and grouping of PMN observations based on differences between the various experiments and exposure conditions in the literature. Five unique variables are included in this model, some appearing in multiple branches. The model variables are a mix of experimental conditions and CNT properties, with the total dose by mass (positively correlated with PMN count) providing the greatest variance reduction, followed by post-exposure time (mostly negatively correlated with PMN count), the dose of chromium

(positively correlated with PMN count), the dose of iron (negatively correlated with PMN count), and the dose of cobalt (positively correlated with PMN count).

The random forest model permits us to put a more comprehensive quantitative value on the information gain or variance reduction provided by each possible model variable, rather than just the few that might appear in the best RT model. Figure 2-2 shows that the total dose by mass, the dose of cobalt, the post exposure time, and mass concentration (an exposure attribute for inhalation experiments, measured in $\mu g/m^3$) contribute most to the RF PMN model's ability to explain variance, a similar but more robust result than the best RT. It would be imprecise to say that the variables with the longest bars in Figure 2-2 are the most important for CNT toxicity. It is better to say that those variables are responsible for most of the performance of the best RF model. If those variables were deleted from the data set, it is likely that other variables that are correlated with them would take their place in the new RF without totally undermining the RF model's ability to explain the variance in the data.

We also used RF models to contrast the effects of different levels of input variables on the output variables controlling for CNT dose. This is analogous to the kind of experiment that would contrast two batches of CNTs made with different catalysts at the same CNT dose, except that it does this for the full range of reported CNT doses. Since RF models are not for extrapolation, we could only vary the tested variables over the range represented in the data set. For the output variable PMN, several variables could increase the dose response relationship by more than 25% (Figure 2-3 and Table 2-5). These were median diameter, mass mode aerodynamic diameter (MMAD), and cobalt content. Shorter median lengths and smaller specific surface areas also increased the dose response relationship. The negative signs in Table 2-5 indicate inverse relationships.

Thus, for example, CNTs with the highest specific surface area (1,040 m^2/g) decreased the neutrophil response by 90%. For Macrophages, the same five variables were also important. For Lactate Dehydrogenase, the most influential variable was short median length, followed by cobalt content; and for Total Protein, the geometric variables, median length and median diameter, and MMAD were salient.

2.5 Discussion

While the literature conveys different conclusions with respect to whether normal levels of impurities dominate[33] or insignificantly affect[34] CNT toxicity, whether increasing aggregation is mitigating[35] or aggravating[36], whether thinner SWCNTs are more toxic[37] or less toxic[22], and whether aspect ratio[38] or surface area[13] could account for differences in biological interactions with CNTs, the results of this study reveal that some of these effects are non-linear (aggregation), others apparently insignificant as currently measured (aspect ratio), and some confirmed aggravators (impurities and increasing CNT diameter).

2.5.1 Effects of CNT Dimensions and Aggregation

Larger aggregate sizes increased PMN and MAC and decreased Total Protein and LDH. Larger aggregate sizes increased the post-exposure time required to recover to levels approaching normal. The largest aggregates resulted in cell counts as much as 30% higher than those observed with the smallest aggregates, for the same post-exposure times. These results are consistent with the idea that larger aggregates are more difficult to clear from the lungs, but cause less damage to cell membranes on an equivalent mass basis. While this result is inconsistent with the findings of some *in vitro* studies that increasing aggregate size leads to greater cell membrane damage[39], it does confirm the

observation from one *in vivo* study where larger aggregates increased neutrophil count and reduced LDH release[24].

Increases in the median length of CNTs produced a marked decrease in the observed toxicity (−30% to −60%), except for in BAL total protein (+30%), contrary to expectations that increasing aspect ratio would increase toxicity. Although, long CNTs may become more readily aggregated and tangled reducing their characteristics as fibers in the lung, the response to increases in CNT length do not mirror those of increasing aggregation. Aspect ratio (length divided by diameter) was tested as a predictive factor in the models, but provided negligible predictive value compared to either length or diameter alone. While CNT lengths varied by several orders of magnitude (from a few hundred nanometers up to 100 micrometers), diameters for all CNTs varied within a much smaller range, of between 1 and 35 nm. It should be noted then that aspect ratio may be comparable between a long, but relatively thin and flexible SWCNT and a much stiffer MWCNT with an even greater length, while their toxicity mechanisms may differ. This significantly reduces the utility of aspect ratio as a CNT toxicity predictor, relative to length and stiffness.

While the categorical attribute, single-walled *vs.* multi-walled CNTs, produced only a limited shift in the dose-response curves (<2%), the diameter of the CNTs (a plausible proxy for the difference between single-walled and multi-walled CNTs) proved to consistently increase toxicity (causing shifts in dose response curves between +13 and +72%). These findings are consistent with the hypothesis that stiffer nanofibers (MWCNTs) produce greater cell damage in the lungs and present more resistance to the body's natural particle breakdown and clearance systems[40], [41].

The model preference for the continuous variable, diameter, over the categorical variable suggests that not all multi-walled CNTs are equally toxic. It is important, however, to realize that this could also reflect other properties correlated in certain parts of the variable space, such as different catalyst recipes for different CNT structures.

Contrary to observations of non-fibrous nanomaterials, increasing the specific surface area of the CNTs had a generally negative effect on toxicity, reducing the responses of most toxicity indicators or having very little effect[42]. CNT surface area measurements were all made by the N_2-BET method, which relates the amount of N_2 adsorbed to a sample to its surface area. Surface area values in our data generally increased with or were unaffected by increasing aggregate sizes. Micrographs of CNT aggregates in the source papers typically showed loose, tangled structures, relatively open and more consistently 3-dimensional from the perspective of N_2 molecules than disaggregated CNTs. Thus the interiors of the tangled aggregates would be shielded from interactions with biological surfaces, even while contributing significantly to the surface area measurement. The N_2-BET measurement methodology appears to act as a proxy for the degree of CNT aggregation as it is positively correlated with a coefficient of 0.88. This interpretation is consistent with the experimental finding that CNTs manipulated to be smaller and better dispersed had a smaller measured specific surface area and induced a more severe toxic response [24].

2.5.2 Effects of Impurities

CNT pulmonary toxicity is significantly increased but not dominated by the amounts of metallic impurities. As these metals are removed from the CNTs, toxicity is reduced, but not eliminated. Of all the metal contaminants, toxicity measures were most negatively impacted by the content of cobalt. Other metals including chromium, nickel,

aluminum, iron, and copper had relatively weak or inconsistent effects on observed toxicity.

Impurity contents are not truly independent variables but reflect a few specific catalyst constituent combinations, so the relative toxicity of one metal over another cannot be definitively addressed with the current data. One of the principal metals co-occurrence groupings is composed of cobalt and nickel (correlation coefficient of 0.992) with another grouping composed of iron, copper, and chromium (correlation coefficient of 0.988 between Fe and Cu or Cr). However, given the greatest RF response to impurities results from cobalt exposure, and the fact that cobalt is a known sensitizer [43], and has been associated together with cobalt carbides in acute lung toxicity [44], it is plausible that the dominant effect of cobalt is real. Though the cobalt content of the CNTs in these experiments only varied between 0% and 0.53%, this was sufficient to produce a large change in the CNT dose response curve.

2.5.3 Effects of Exposure Mode

Contrary to expectations, the exposure mode of the experiment did not have a significant value in explaining the difference in study outcomes, only accounting for a relative variance reduction (in relation to the variance reduction of the total model) of between 0.01% and 0.5%.We expected that the differences in CNT deposition patterns between instillation exposures and inhalation exposures and the uncertainty in translating ambient particle concentration to total dose received would have produced a significant difference in the observed response for a given dose. The fact that the models did not find this factor significant means that either there is a non-obvious proxy for exposure mode contained within the combinations of other experimental variables, or that for

these conditions (relatively high dose and short duration experiments) the mode of exposure is minimally influential to experimental outcomes.

2.6 Conclusions

We have developed a suite of machine-learning-based methods for meta-analysis of nanotoxicity studies that has enabled the extraction of information across studies that was not observable in the individual studies. This work has identified a short list of CNT attributes that contribute to uncoated unfunctionalized CNT pulmonary toxicity, metallic impurities, CNT length and diameter, surface area, and aggregate size (the variables listed in Table V), because they alter the CNT dose-response relationship. Being the result of a meta-analysis, our results are limited by the experimental choices made in the archival literature and do not involve a theory of toxicity. Nevertheless, this work has demonstrated the utility of RT and RF methods for identifying and ranking attributes of nanoparticles that contribute to their toxicity and quantifying those contributions.

Table 2-1. Summary of published data on *in vivo* pulmonary exposures of carbon nanotubes included in the meta-analysis.

Publication Reference	Author(s)	Year	CNT Configuration	Exposure Mode	Animal	Average Aggregate Size (nm)	Purity (%)	Total Dose (μg/kg)	Post Exposure Period (days)
[45]	Warheit, D. et al.	2004	SWCNT	instillation	rats		90	1000 - 5000	1 - 90
[24]	Muller, J. et al.	2005	MWCNT	instillation	rats		97.8	2 - 8	3 - 60
[46]	Shvedova, A. et al.	2005	SWCNT	aspiration	mice		99.7	490 - 1970	1 - 60
[47]	Shvedova, A. et al.	2007	SWCNT	aspiration	mice		99.7	1851	1
[48]	Shvedova, A. et al.	2008	SWCNT/MWCNT	instillation	mice	4200	82	250 - 1000	1 - 7
[49]	Muller, J. et al.	2008	MWCNT	instillation	rats		95 - 99	8890	3
[50]	Elgrabli D. et al.	2008	MWCNT	instillation	rats			5 - 500	1 - 180
[51]	Mercer R. et al.	2008	SWCNT	aspiration	mice	690	98	303	1 - 30
[52]	Inoue K. et al.	2008	SWCNT/MWCNT	instillation	mice			4000	1
[53]	Ryman-Rasmussen J. et al.	2008	MWCNT	inhalation	mice	714	94	12000	1 - 14
[27]	Ma-Hock, L. et al.	2009	MWCNT	inhalation	rats	400	90	190 - 2400	3 - 21
[22]	Nygaard, U.	2009	SWCNT/MWCNT	aspiration	mice		95	5000 - 20000	26
[54]	Ellinger-Ziegelbauer H. et al.	2009	MWCNT	inhalation	rats	2900		180 - 3900	7 - 90
[23]	Pauluhn, J.	2010	MWCNT	inhalation	rats	1670 - 2190	98.6	105 - 6290	1 - 90
[55]	Porter D. et al.	2010	MWCNT	aspiration	mice		99.5	435 - 1740	1 - 56
[56]	Park, E-J et al.	2011	SWCNT	instillation	mice		90	100	1 - 28
[57]	Teeguarden, J.G. et al.	2011	SWCNT	aspiration	mice		99.7	12000	1

Table 2-2. List of the 37 carbon nanotube input attributes included in models and their definitions.

Category	Variable	Units	Definition
Size and Shape	configuration	MWCNT or SWCNT	a categorical variable indicating whether the carbon nanotubes are multi-walled (MWCNTs) or single walled (SWCNTs)
	minimum CNT length	nm	the minimum reported length of the free individual CNT fibers either measured or stated by manufacturer's specifications
	median CNT length	nm	the median length of the free individual CNT fibers either measured or stated by manufacturer's specifications
	maximum CNT length	nm	the maximum reported length of the free individual CNT fibers either measured or stated by manufacturer's specifications
	minimum CNT diameter	nm	the minimum reported diameter of the free individual CNT fibers either measured or stated by manufacturer's specifications
	median CNT diameter	nm	the median diameter of the free individual CNT fibers either measured or stated by manufacturer's specifications
	maximum CNT diameter	nm	the maximum reported diameter of the free individual CNT fibers either measured or stated by manufacturer's specifications
	aggregate diameter (MMAD)	nm	mass mode aerodynamic diameter – the most frequent size of the particle aggregates by the mode of the distribution by mass
	specific surface area	m^2/g	specific surface area as measured by the N2-BET (Nitrogen, Brunauer-Emmett-Teller) gas adsorption method
	purity	%	the fraction by percent mass of the amount of the CNT sample composed of carbon atoms
Impurities	dose cobalt [total]	pg/kg	the total dose received by the animal subject of cobalt impurities present in the CNT particulate
	dose cobalt [24 hr average]	pg/kg	the average daily dose received by the animal subject of cobalt impurities present in the CNT particulate
	dose aluminum [total]	pg/kg	the total dose received by the animal subject of aluminum impurities present in the CNT particulate
	dose aluminum [24 hr average]	pg/kg	the average daily dose received by the animal subject of aluminum impurities present in the CNT particulate
	dose iron [total]	pg/kg	the total dose received by the animal subject of iron impurities present in the CNT particulate
	dose iron [24 hr average]	pg/kg	the average daily dose received by the animal subject of iron impurities present in the CNT particulate
	dose copper [total]	pg/kg	the total dose received by the animal subject of copper impurities present in the CNT particulate

Category	Variable	Units	Definition
	dose copper [24 hr average]	pg/kg	the average daily dose received by the animal subject of copper impurities present in the CNT particulate
	dose chromium [total]	pg/kg	the total dose received by the animal subject of chromium impurities present in the CNT particulate
	dose chromium [24 hr average]	pg/kg	the average daily dose received by the animal subject of chromium impurities present in the CNT particulate
	dose nickel [total]	pg/kg	the total dose received by the animal subject of nickel impurities present in the CNT particulate
	dose nickel [24 hr average]	pg/kg	the average daily dose received by the animal subject of nickel impurities present in the CNT particulate
	dose oxidized carbon [total]	pg/kg	the total dose received by the animal subject of oxidized carbon impurities present in the CNT particulate
	dose oxidized carbon [24-hr average]	pg/kg	the average daily dose received by the animal subject of oxidized carbon impurities present in the CNT particulate
Exposure characteristics	exposure hours	hours	the number of hours that the animal subject was exposed to the CNTs
	first to last exposure period	hours	the time period in hours between the first hour of exposure and the last hour of exposure by the animal subject to CNTs
	animal species	rat or mouse	a categorical variable indicating the type of the animal subject
	animal breed or strain	Sprague-Dawley, Wistar, C57BL/6, ICR, Crl:CD(SD)IGS BR,BALB/cAnNCrl	a categorical variable indicating the specific breed or strain of the animal subject
	sex	male or female	A categorical variable indicating the sex of the animal subject
	mean animal mass	g	the mean mass of the animal subjects in a given experiment
	post exposure	days	the number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period
	exposure mode	inhalation, instillation, aspiration	a categorical variable indicating the mode of exposure
	mass concentration	mg/m^3	the mass concentration of CNTs in the air of the animal subject inhalation chamber (inhalation exposures only)
	total mass dose	µg/kg	the total mass dose of CNTs received over the course of the experiment by the animal subject
	average mass dose	µg/kg/24-hr	the average daily mass dose of CNTs received over the course of the experiment (first to last exposure period) by the animal subject

Category	Variable	Units	Definition
	total surface area dose	m^2/kg	the highest peak hourly surface area dose of CNTs received over the course of the experiment by the animal subject
	average surface area dose	$m^2/kg/24\text{-hr}$	the average daily surface area dose of CNTs received over the course of the experiment (first to last exposure period) by the animal subject

Table 2-3. Model goodness-of-fit performance and the number of model input attributes included.

Output Variable	Regression Tree		Random Forest		Maximum Model Performance	
	R^2	# of variables	R^2	# of variables	R^2	# of exposure groups
Neutrophils	0.89	5	0.83	5	0.90	103
Macrophages	0.62	3	0.84	3	0.88	62
Lactate Dehydrogenase	0.84	7	0.89	5	0.90	84
Total Protein	0.92	13	0.95	5	0.96	67

Table 2-4. The RF model validation results in terms of mean squared error (MSE) for each of the studies included in this analysis and the observed change in outcome when that particular study was excluded from the training data.

First Author, Publication Date	Total Dose (μg/kg)	Recovery Period (days)	Exposure Mode	RF Model MSE (fold of control squared)
Neutrophils				
Pauluhn, 2010	105 - 6290	1 - 90	inhalation	38,500
Shvedova, 2008	250 - 1000	1	instillation	6,700
Muller, 2005	2 - 8	3	instillation	6,300
Shvedova, 2007	1851	1	aspiration	2,900
Inoue, 2008	4000	1	instillation	2,100
Porter, 2010	435 - 1740	1 - 28	aspiration	1,600
Shvedova, 2005	490 - 1970	1 - 60	aspiration	260
MSE for all points in full model	--	--	--	2,700
Macrophages				
Pauluhn, 2010	105 - 6,290	1 - 90	inhalation	46,700
Shvedova, 2008	250 - 1,000	1 - 7	instillation	7,000
Mercer, 2008	303	1 - 30	aspiration	1,200
Shvedova, 2005	490 - 1,970	1 - 60	aspiration	740
Nygaard, 2009	5,000 - 20,000	26	aspiration	41
Ma-Hock, 2009	190 - 2,400	3 - 21	inhalation	12
MSE for all points in full model	--	--	--	7,900
Lactate Dehydrogenase				
Pauluhn, 2010	105 - 6,290	1 - 90	inhalation	46,700
Shvedova, 2007	1,851	1	aspiration	28,800
Warheit, 2004	1,000 - 5,000	1 - 30	Instillation	10,000
Muller, 2005	2 - 8	3	instillation	6,200
Shvedova, 2008	250 - 1,000	1	instillation	4,900
Shvedova, 2008	250 - 1,000	1 - 7	instillation	4,100
Muller, 2008	8,890	3	Instillation	2,300
Ellinger-Ziegelbauer, 2009	180 - 3,900	7 - 90	inhalation	1,500
Porter, 2010	435 - 1,740	1 - 28	aspiration	300
Ma-Hock, 2009	190 - 2,400	3 - 21	inhalation	104
MSE for all points in full model	--	--	--	8,400
Total Protein				
Pauluhn, 2010	105 - 6,290	1 - 90	inhalation	46,000
Ma-Hock, 2009	190 - 2,400	3 - 21	inhalation	8,100
Shvedova, 2007	1,851	1	aspiration	2,800
Ellinger-Ziegelbauer, 2009	180 - 3,900	7 - 90	inhalation	2,700
Inoue, 2008	4,000	1	instillation	2,400
Muller, 2008	8,890	3	instillation	1,000
Muller, 2005	2 - 8	3	instillation	37
Shvedova, 2005	490 - 1,970	1 - 60	aspiration	22
MSE for all points in full model	--	--	--	8,500

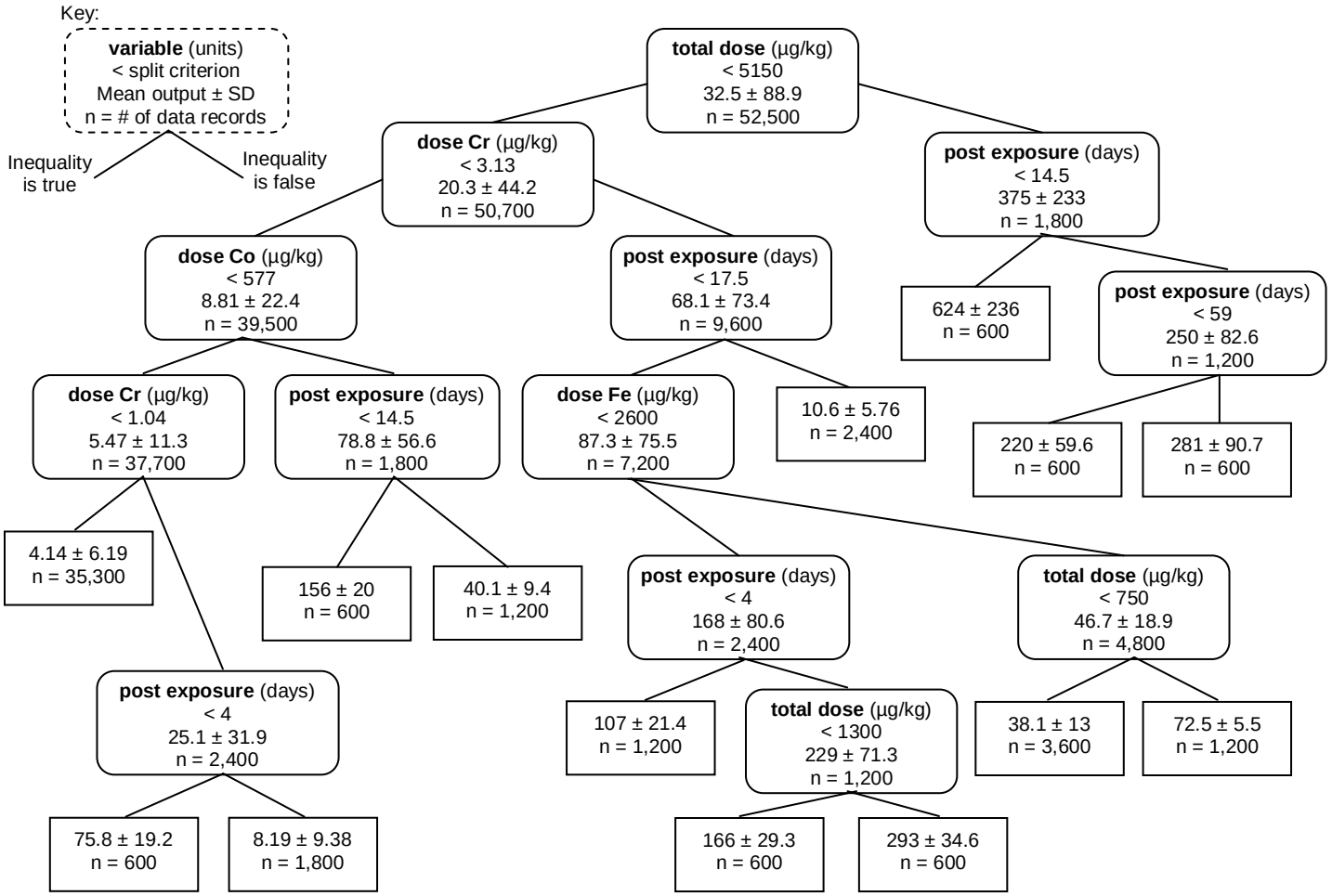

Figure 2-1. Regression tree model for the effect of CNT exposure on BAL neutrophils. The mean output values in the leaf nodes (rectangular terminal nodes) are the model's predictions.

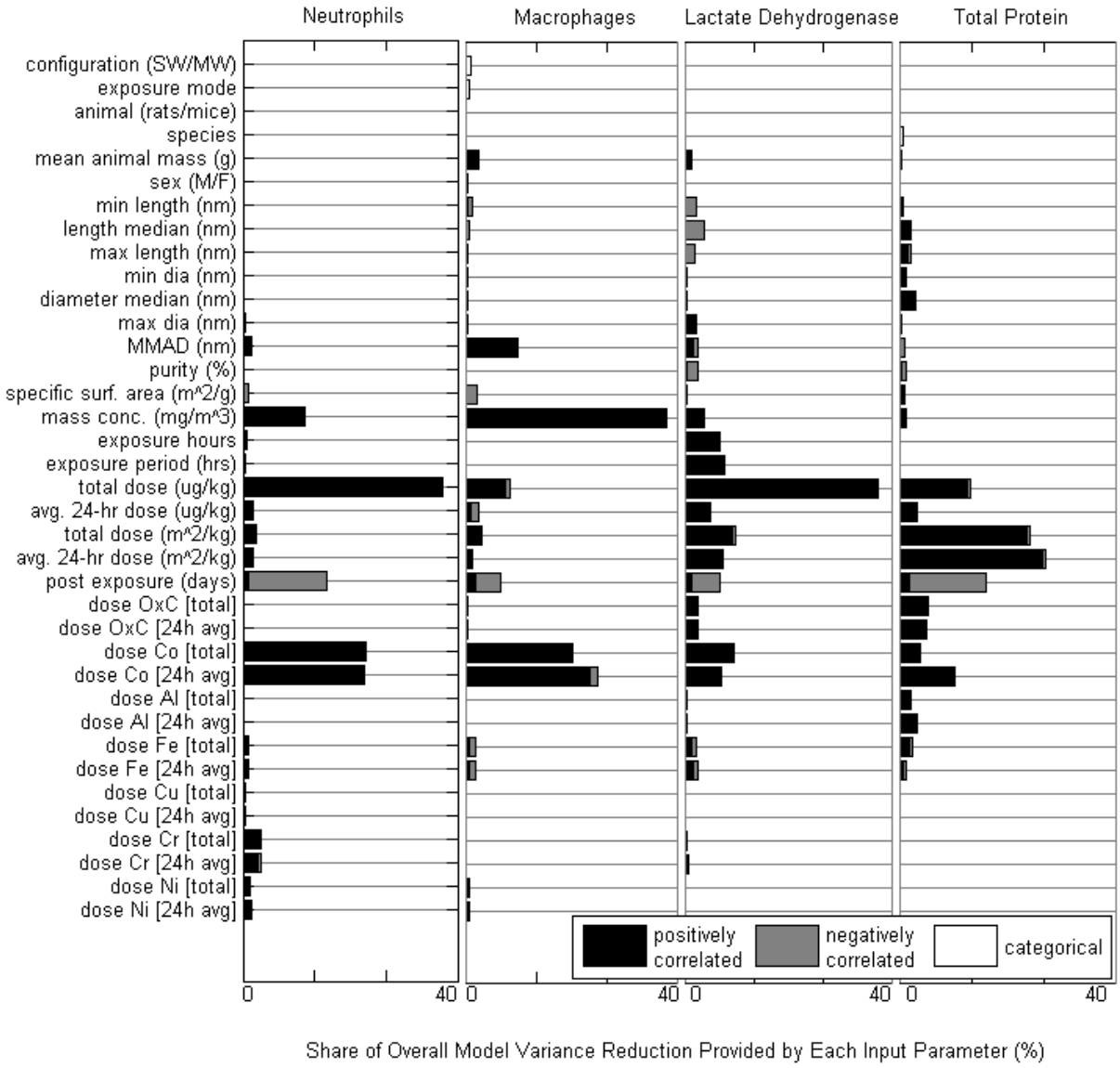

Figure 2-2. Random forest model variable importance comparison as measured by mean variance reduction (shown as share of total model variance reduction) for all four toxicity indicators.

Bar shading indicates whether the variance reduction occurred with a positive or negative correlation between the input and output, or with a branch based on a non-numeric categorical variable.

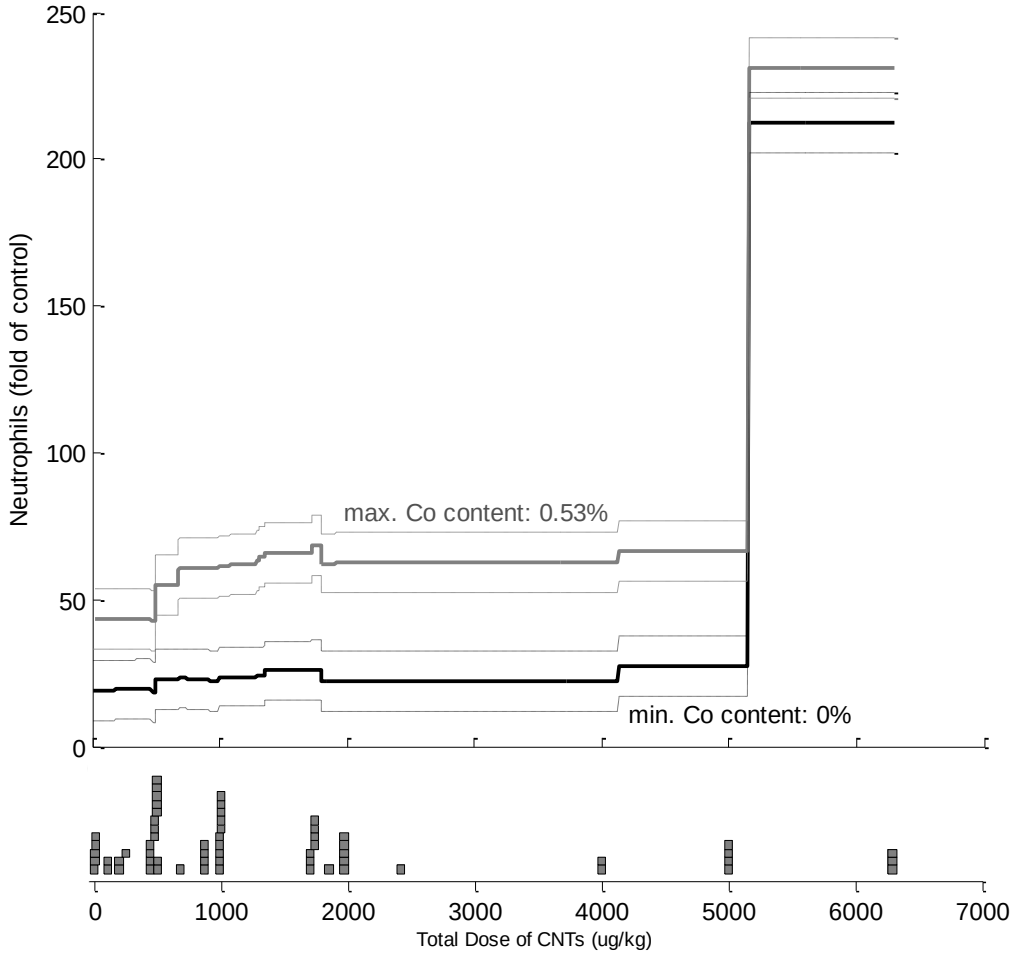

Figure 2-3. RF dose-response graphs of the effects of CNT exposure on PMNs at 0% and 0.53% cobalt impurity by mass.

At CNT doses below 5000 μg/kg, CNTs with 0.5% cobalt double the neutrophil response. Dashed and dotted lines indicate the response ± the experimental standard deviation. All other model input variables are held at their median values. The gray blocks in the sub-x-axis indicate dose levels at which experimental data exists.

Table 2-5. Mean relative shift (%) observed in CNT total mass dose-response curves in four BAL inflammation indicators from changes in select CNT attributes based on RF model results.

In each case the reported shift reflects the effect of changing the input variable from its reported minimum numerical value to its maximum in the input data set.

		Output Variables			
Input Variable	**Minimum/Maximum Value (units)**	**Neutrophils (%)**	**Macrophages (%)**	**Lactate Dehydrogenase (%)**	**Total Protein (%)**
Median Length	320/100,000 (nm)	−51	−36	−60	+30
Median Diameter	0.8/49 (nm)	+46	+32	+13	+72
MMAD	400/4,200 (nm)	+62	+29	−23	−64
Specific Surface Area	109/1,040 (m^2/g)	−90	−28	−19	+2
Aluminum Content	0/9.6(%)	+0	+16	+15	+10
Cobalt Content	0/0.53 (%)	+57	+58	+28	+1
Iron Content	0/17.7 (%)	+14	−20	+0	+16
Copper Content	0/0.16 (%)	−4	−8	−1	−16
Nickel Content	0/5.53 (%)	+3	+0	−1	+0
Chromium Content	0/0.05 (%)	+21	−8	+3	−25

3 Nano-Titanium Dioxide Particle Size, Aggregation, and Toxicity – Results from a Machine Learning Based Meta-Analysis

3.1 Abstract

Titanium dioxide nanoparticles (nano-TiO_2) are one of the nanomaterials produced in the largest quantity, and inhalation is a potential exposure route for both workers and the general public. NIOSH has proposed a new exposure limit based on one particular type of nano-TiO_2 and fixed surface to mass ratio while cautioning that different nano-TiO_2 may have various degrees of toxic potency. To quantitatively determine the influence of nano-TiO_2 physicochemical properties and exposure attributes on observed pulmonary toxicity, we conducted a meta-analysis of 25 *in vivo* studies in which rodents were exposed to nano-TiO_2 through respiratory track (e.g., inhalation or instillation). Machine learning models, including regression trees and random forests, revealed that primary particle size had the most influence. Compared with 100 nm nano-TiO_2, 3.5 nm TiO_2 nanoparticles (the smallest in our analysis) had 5% more total cells and 45% higher concentrations of lactate dehydrogenase, an indicator of cytotoxicity, in bronchoalveolar lavage (BAL) fluid. Most other particle characteristics including crystalline phase (anatase fraction), aggregation, surface coating, shape, and purity had negligible effect on toxicity. We conclude that, given the present information, it is sufficient to treat nano-titanium dioxide particles as one class.

3.2 Introduction

Titanium dioxide particles are used in a variety of industrial and consumer applications, e.g, as a white pigment, photocatalyst, or UV-absorbing sunscreen ingredient. While titanium dioxide is relatively inert at micrometer and larger sizes, it can become reactive at the nanometer scale (roughly between 1 nm and 100 nm). A special case of nano-TiO_2 toxicity is associated with its photocatalytic activity. In the presence of ultraviolent light, TiO_2 generates reactive oxygen species (ROS), which have been shown to be highly toxic to cells *in vitro* [58]. Nano-TiO_2 can create reactive oxygen even without light stimulation [59] to a lesser extent.

While dermal exposure can cause sensitivity [60], the most worrisome exposure scenario for humans is pulmonary, due to workplace exposure or inhalation of aerosol personal care products.

The literature on nano-TiO_2 pulmonary toxicity is somewhat inconsistent. Inhaled nano-TiO_2 were reported to both produce lesions similar to emphysema in mice [61], and result in only limited toxicity under similar conditions [62]. Significant differences in pulmonary toxicity have been attributed to the proportional makeup of anatase or rutile crystalline forms of the nano-TiO_2 [63].

The Occupational Health and Safety Administration (OSHA) rules permit a maximum airborne concentration of 15 mg/m^3 in the workplace, which is 3 times higher than the upper limit allowed for uncharacterized non-toxic nuisance dust (limited to 5 mg/m^3) [64]. More recently, the National Institutes of Occupational Safety and Health (NIOSH) published a draft recommended exposure limit for nano-titanium dioxide of 0.3 mg/m^3—a reduction by a factor of 50 [65]. This recommended modification reveals

the general trend of current knowledge that anticipates nanoparticle or ultrafine particulates to be potentially more toxic at lower doses than their larger counterparts.

The question is, can all nanoscale TiO_2 particles be treated as a nearly identical class, or do variations between different types of particles in dimensions, shape, or surface properties mean that the regulatory definition of nano-TiO_2 particles should include certain particle properties?

An efficient though preliminary approach to addressing this question is through meta-analysis. No published meta-analysis of *in vivo* exposures to titanium dioxide nanoparticles currently exists. Three nano-SAR studies using *in vitro* toxicity from different metal oxide nanoparticles concluded that TiO_2 nanoparticles are less toxic than other metal oxide nanoparticles of similar sizes [4], [15], [17]. Another effort modeled the effects of particle properties on *in vitro* toxicity of titanium dioxide and zinc oxide nanoparticles. It found that aggregation and surface charge affected cellular membrane damage [66]. However, in the literature, no model yet explains the differences in effects observed from pulmonary exposures to titanium dioxide nanoparticles.

This study compares the factors affecting pulmonary nano-titanium dioxide toxicity, including changes in crystalline phase composition (rutile versus anatase), surface treatments (hydrophobic or hydrophilic), aggregation, and primary particle sizes through a meta-analysis of the extant published literature on rodent pulmonary exposure studies.

One of the challenges in interpreting these data is that due to differences in investigator approaches and specific study objectives, the papers do not report all relevant nanoparticle characteristics. The TiO_2 database we assembled for this meta-analysis contained approximately 40% missing values for the nanoparticle and experimental characteristics.

The usual meta-analysis technique is multiple linear regression by ordinary least squares or maximum likelihood estimation. However, given the frequency of missing data, MLR models perform poorly. MLR also makes certain strong assumptions about the non-relatedness of the independent variables for which do not hold for our data. Imputation of missing values can address part of this problem, but it introduces another source of uncertainty in that the unknown values may actually lie outside the range of the prescribed distribution or not follow the presumed distribution shape or relationship. Careful selection of variable distributions and multiple imputation data sets can mitigate some issues, but they remain substantial drawbacks.

This paper instead proposes a meta-analysis based on machine learning algorithms: regression trees (RTs) [6] and random forests (RFs) [7] that experience less degradation from missing values than other regression techniques and so do not necessitate imputed values, but remain relatively simple and force few assumptions onto the nature of the underlying relationships between input variables. These techniques have demonstrated the capability to reveal important determinants of carbon nanotube toxicity even when there were substantial amounts of missing data [67].

RT models are made up of a hierarchical series of decision rules (in branch nodes) based on the values of specific input variables that eventually terminate in a predicted value (in a leaf node). Each branch of a RT represents a set of conditions resulting in a specific outcome (level of toxicity response).

RF models have the advantage of more fully exploring the value of each variable in the data set, compared to RT models, which typically include a small subset of them. They also are less sensitive to weaknesses in the data. With RTs, variables can appear

influential due to their over-representation in the data. The randomized variable selection process for the RF model generation procedure ensures that all variables are evaluated.

3.3 Methods

3.3.1 Data Selection

Studies of *in vivo* exposure to nano-TiO_2 via respiratory track exposure (inhalation, instillation, and aspiration) in animals were identified in 2010 and 2011 via literature search of peer-reviewed journal publication in PubMed and references within. Due to the scarcity of reports on toxicity outside the lung, we focused our analysis on pulmonary toxicity. Studies published prior to and in 2010 were included. 25 studies (Table 3-1) met our inclusion criteria, i.e., reporting physico-chemical characterization measurements of the particles, well described exposure (method, dose/concentration, duration, frequency, etc), and quantitative toxicity endpoints.

Via carefully reading, we captured initially 120 parameters pertaining to particle (physicochemical properties of neat or as received TiO_2), animal (e.g., species, breed, weight and age at the beginning of the experiment), exposure (e.g., media for suspension, aerosol size, exposure method, concentration, duration, frequency), toxicity (e.g., weight after exposure, organ weight, cell counts in BAL, pathology), dosimetry (e.g.,, TiO2 distribution, retention), and others (e.g., funding source, publication year). No single study provided information in all 120 parameters. For papers missing some commonly reported particle characterization results, we asked the corresponding and/or first author for the information. When data were reported in only graphs, software was used to convert the results in line chart/bar chart into numbers. Few studies reported both animal age and weight at the beginning of the study, but one of the two was reported in

all the rest studies. The unreported weight or age was estimated based on age (for weight) or weight (for age) of the specific breed historical data from the animal supplier identified in the paper. Cumulative particle exposure, if not reported, was estimated using Multiple Path Particle Dosimetry (MPPD) model 2.1, which includes nanoparticle calculation.

For particle characteristics, the parameters were further selected into our analysis with consideration of various recommended minimal characterization and the number of studies reporting the measurement results or vendor provided information. In general, parameters on the minimal characterization lists were also reported more often than ones not on the list. In addition parameters on the minimal characterization list and reported often, we purposely included a few parameters not on the list to compare the contribution of these parameters.

For toxicity and dosimetry, 5 pulmonary toxicity endpoints measured in bronchoalveolar lavage (BAL) fluid were reported in sufficient numbers of studies across the published data for analysis. These include total cell count (TCC), polymorphonuclear neutrophils (PMN), macrophage count (MAC), lactate dehydrogenase concentration (LDH), and total protein concentration (TP). Toxicity endpoints were either reported as or converted by us to the relative change from the negative control group ("fold of control"). Six other reported toxicity endpoints, including pathologies and cytokines, as well as all dosimetry endpoints were reported in only few studies and therefore excluded from our analysis.

3.3.2 Data Preparation

To enable the meta-analysis of dissimilar studies, we treated all nanoparticle properties and experimental conditions as independent variables (Table 3-2). In addition

to numeric variables we included the categorical variables, material name, manufacturer, shape, hydrophobic/hydrophilic, surface coating, suspension media, animal species, animal breed/strain, animal sex, exposure route, and exposure route-detail. The other 12 numeric independent variables were primary particle size, aggregate size in aerosol or suspension, anatase fraction, purity, specific surface area, aerosol or instillation exposure concentrations, exposure period, exposure frequency, recovery period, and total cumulative dose.

The input for this analysis is a matrix in which the rows represent individual experimental animals and the 41 columns of the matrix contained 24 experimental conditions, 12 nanoparticle properties (such as primary particle size, specific surface area, and average aggregate size), and the 5 experimental endpoints (toxicity measures). For diagnostic purposes, a series of 2 to 4 uniformly distributed random variables between 0 and 1 were added to the data array as columns at different stages of the analysis. The random variables—by definition information-free and not correlated with any toxicity endpoint values—supplied an independent measure with which to compare the discriminatory performance between information-rich and information-poor attributes. Missing values for any particular nanomaterial property or experimental condition in any published study were left blank.

The number of matrix rows per experiment was 20 times the number of animals in each experimental group. For example, if an experiment involved 12 test animals at one dose and 15 test animals at a different dose, the experiment would occupy 240 + 300 = 540 rows. Each of these rows would contain identical nanomaterial and experimental attributes, the appropriate exposure metrics, but for the experimental endpoints (PMN, MAC, TCC, LDH, and TP), each row would have a unique, discrete realization of a

normal distribution with the mean and standard deviation reported for its group of animals. Endpoint values deemed impossible—negative protein concentrations, for instance—were discarded and replaced with the nearest plausible value, usually zero. We assumed endpoint measures were normally distributed. Higher rates of sampling at 50 or 100 samples per experimental animal did not alter the ordinal value of model variables or goodness of fit measures in the resulting RT's.

The probabilistic representation of the experimental results preserves the inherent variability in the measured responses and reduces the likelihood of the models reporting noise or of their oversensitivity to small differences in experimental exposure conditions. The assumption of a normal distribution for the toxicity endpoint measures is reasonable based on the actual observed distributions in similar toxicity studies. This procedure weights the influence of each study on the model structure in relation to the number of animal subjects involved. Furthermore, incorporating experimental uncertainty of the data in this manner enables the models to output the corresponding range of likely values rather than being limited to a single value.

3.3.3 Regression Trees

For each measured toxicity endpoint we created a RT model using the MATLAB™ function "classregtree." The RT algorithm successively divides the population of observations (toxicity endpoint values, also called output variables) into two groups based on an inequality for quantitative input variables or a categorical grouping for categorical variables. At each branching point, the inequality test variable and value are chosen to produce the highest possible information gain between the resulting two groups of observations. In this study we used a calculation of variance reduction as a measure of information gain. While entropy increase may also be used as a measure of

information gain, it does not produce substantially different results, and variance reduction is more consistent with the statistics produced by alternative regression models, facilitating comparisons. The process of pruning the model to prevent overfitting is described in Appendix B.

The importance of each variable included in the RT model is calculated based on the total variance reduction achieved by each branch in the final model. For variables appearing on multiple branches, the variance reduction is summed across the entire model.

While RTs show the best model for a given data set in the sense of parsimony, there are ways to interrogate the data that are even less sensitive to its biases, such as the Random Forest (RF).

3.3.4 Random Forests

We generated a RF model for each toxicity endpoint implementing Breiman's algorithm with the MATLAB™ function "treebagger". The RF model is composed of an aggregate collection of RT models, each created from a data set where each branch is selected from a random subset of one third of the available variables (matrix columns). Each RF model contains at least 500 RT models, whose elements are averaged to produce the RF output. The RFs were grown (i.e. trees added) until the error decreased less than 0.1% with the addition of 100 trees to the RF model.

The variable importance results from the RF model are calculated in the same way as for the RT model; however, the final variance reduction for each variable is calculated from the average across all of the trees in the RF model. This is accomplished by calculating the variance reduction caused by each branch variable in each individual tree in the forest, summing the variance reductions by variable across all of the trees in the

RF, and then dividing the final values by the total number of trees in the RF model. These error reduction values form the measure of variable information content used in ranking the variables by their importance in Figure 3-1.

To examine effects of specific particle characteristics on the overall dose-response, we first removed all variables with less calculated importance than the average of the uniform random variables. Then, model refinement proceeded in a forward stepwise fashion evaluated on the basis of the R^2 value. In each step, the additional variable producing the highest possible R^2 value was selected. Once the model achieved more than 95% of its potential information gain or the model achieved a peak greater than the comprehensive model, variable addition was halted. The goodness of fit statistics for the RFs and RTs are found in Appendix B.

3.3.5 Multiple Imputation

In order to apply linear regression models to these data, we needed to fill in the missing measurements through imputation. Based on the imputation distributions defined in Table 3-2, we independently generated 5 separate data sets via sampling with new randomized seeds for each. Each of these data sets was the same size, a factor of 20 times the number of animal subjects in the reported studies. The output data was not resampled from the distributions but remained identical for each imputed array.

3.3.6 Linear Regression

A series of stepwise multiple linear regression models were each trained on one of the imputed data sets to predict a specific quantitative toxic endpoint measure using the MATLAB™ function "stepwise". We standardized numeric input variables on the basis of their z-scores. Categorical or class variables were converted to a series of binary

dummy variables. Variables included in the linear regression models are shown in Table 3-4, based on an inclusion criterion of a p-value less than 0.05.

3.4 Results

3.4.1 Random Forest Model Variable Importance with No Imputed Data

The importance of particle, exposure, and animal characteristics to toxicity was modeled in five RF models, each for one specific toxicological endpoint, namely one type of measurement in BAL (Figure 3-1).

Dosimetry characteristics along with aerosol concentration (mg/m^3) and aerosol concentration x day ($mg\text{-}day/m^3$) were among the most important parameters, and they consistently appear important in four out of five toxicological endpoints. This is consistent with the toxicology principle of the dose makes the poison.

Among particle characteristics, in general over all five models, TiO_2, increasing particle size and increasing aggregation are generally associated with decreased toxicity for both inhaled and instilled particles. This agrees with the general expectation that TiO_2 nanoparticles will be more toxic than larger particles. The chemical impurity of the TiO_2 particulate had only a limited effect on LDH, and lacked significant explanatory power for all of the cell count measures.

Smaller average aerosol size and mass median aerodynamic size (MMAD) were associated with increased toxicity in four of the five endpoints. These two parameters also appear to have more influence than any single primary particle characteristics in most toxicological endpoints (except for LDH). This suggests that the size of aerosol/particle at the time of exposure is more important than the primary particle size.

3.4.2 Linear Regression Variable Importance with Imputed Data

No statistically significant linear models could be fit to the un-filled data so the imputation-filled data sets were used. All linear regression models performed poorly, with R^2 values < 0.10 (Table 3-3). The linear models primarily employed categorical variables such as material name, particle manufacturer, animal breed/strain, and suspension medium (Table 3-4).

3.4.3 RF Model Validation

We validated the RF models by generating a series of models, each one leaving out the data from a single paper in the set of publications (Table 3-1) reporting that endpoint. Then, the error each of the models make when predicting the data from their omitted publication is calculated. These results provide two insights into the data. The first is confidence that the model is not overly trained on the results of one particular study over the results of the others. The other insight is that a quantitative comparison is provided for the similarity of each study to the data provided by the others.

The model validation results generally show that the overall model performance is not adversely affected with the removal of most individual studies' data, except in the case of Bermudez, 2002. This study stands out for its very high inhaled particle concentrations, and long recovery times, as well as the inclusion of 3 species of animals, and a relatively large amount of experimental groups in comparison to other studies. Given the exposure characteristics are outside the range of most other experiments, it is not surprising that a model trained without any of these data would poorly predict the results of that study.

The results of this validation are included in Appendix B.

3.4.4 Effects on Dose-Response

The RF models can be used to produce dose-response curves (see Figure 3-2) based on the aggregated information created from the published data by holding all input variables constant (in this case, at their median values) while varying the total dose from the minimum to maximum value. These shifts of the dose-response curves can be quantified by integrating and comparing the areas under the curves for changes in the secondary variable of interest. Results of this exercise are shown in Table 3-5, for the most influential variables and those of greatest current interest.

The shape of the dose-response curves is governed by a collected family of stepwise functions contained within the RF model. In areas of the experimental space where fewer data are available, large steps or flat areas can occur. These should be considered indicative of model uncertainty and limited information rather than signals of important effects at specific dose thresholds.

The effect of a secondary input variable such as particle size on the mass-dose-response can be observed by contrasting the dose-response curve calculated at different values for the secondary variable.

3.4.5 Model Comparison

Table 3-3 contrasts the explanatory power of the three types of regression models: linear (using the imputation-filled data set) RT and RF (using the unfilled data set) for the 5 experimental endpoints.

Model cross-validation results are found in Appendix B. Also in the appendix are the results of using the filled data set on the RF models. The use of these data had little effect on the RF results.

3.5 Discussion

It should be noted that overall, TiO_2 particles appear to produce lower toxic responses than many of the other tested nanoparticles such as carbon nanotubes [67] or nano-copper [68], even at relatively large doses. Of all the dose-response curves derived from the RF models in this study, most only predict changes of a factor of 2 to 3 from the control group response (see supplemental information). This is not unexpected given the relatively consistent inert behavior of TiO_2 in the absence of light, which is always the case in rodent lungs. What damage does occur may be predominately mechanical in nature rather than a unique chemical property-induced effect. This has implications for the kinds of TiO_2 particle characteristics that actually could significantly affect toxic responses in rodents.

3.5.1 Performance of Linear and Machine Learning Models and Multiple Imputation

Given the overall poor performance of the linear models on this data set (Table 3-3, $R^2 < 0.10$), we are skeptical of the significance of the variables that were included in the linear models. The combination of the large number of variables, the relatively low number of experimental groups for this number of independent variables, and variability in the reported outcomes degraded the capability of the linear models to explain these data. These factors result in the overreliance by the models on the binary dummy variables created for the categorical exposure characteristics.

The highly non-linear machine learning models perform much better statistically, which is to be expected, and also are able to reproduce the anticipated exponential dose-response curve shape (Figure 3-2) for many of the measured endpoints. Since, the

important variables in the machine learning models including particle size, aggregation, lung burden, dose, and recovery period are also supported by individual studies and current understanding, and the models experience little change whether using the imputed or un-filled datasets, we have more confidence in their results.

3.5.2 The Importance of Particle Size and Aggregation

The influence of TiO_2 particle size on pulmonary toxicity is not large given the range of sizes tested (Table 3-5), but its toxicity consistently decreases with increasing particle size. While the variance reduction attributed to primary particle size in the complete RT model (Figure 3-1) is minimal, its value increases as dosimetry measurements and other plausibly correlated variables are removed. In the reduced model, increasing particle size, which does vary over a wide range, significantly reduces measures of cellular damage (LDH and TP), while moderately reduces total cell count (TCC) (Table 3-5). This finding is consistent with that of individual studies testing the effects of changes in TiO_2 particle sizes on toxicity, and is likely due to the fact that smaller particles reach locations deeper in the lung inducing damage throughout a greater proportion of the lung's internal surface area.

Aggregation on the other hand does not seem to produce a large effect either in terms of variance explanation or effect on the dose-response curves (shifting the total dose-response between +8% and −5%). We considered whether this could have been due to a relatively consistent ratio of particle size to aggregate size among the published experiments, but in fact, the ratio of MMAD to primary particle size varied between 0.4 and 69. Information on the distribution of particle and aggregate sizes was lacking from the available data, however even in the case of large MMAD aggregate size there can be expected to exist a considerable quantity of smaller particles, which could account for

the discrepancy between primary particle size displaying a significant effect, while aggregate size does not. Here we see one of the important differences between *in vivo* and *in vitro* toxicity tests for nano-titanium dioxide, where aggregate size has proven an important predictor of cellular toxicity in well plates [69]. Measured distributions of particle sizes and their relative contribution to the overall particle mass would permit the cause of this discrepancy to be tested.

3.5.3 The Influence of Crystalline Phase and Impurities

The relatively muted toxic response to nano-TiO_2 particles combined with the relatively high proportion of missing characterization values reduces the capability to observe significant effects caused by changes in crystalline structure, geometry, or surface treatments.

The anatase crystalline phase of the TiO_2 particles does seem to be associated with a small increase in cellular responses, but also a moderate decrease in LDH and TP (Table 3-5). *In vitro* studies have shown an effect of anatase-phase induced cellular membrane damage greater than that of the rutile phase [70], and evidence from single pulmonary exposures seems to confirm a relatively smaller aggravating effect of the anatase phase [63]. This effect does not seem consistent across studies, however. Given a wider range of anatase fractions and a greater variety of exposure scenarios and toxicity endpoints measured, we cannot validate the same observation that the anatase phase of TiO_2 is more toxic in pulmonary exposures.

Impurities delivered with the titanium dioxide nanoparticles were not characterized, so their exact makeup is generally unknown, but the data show a generally positive correlation between purity and measured LDH and TP in BAL fluid. This means that

whatever the constituents of the impurities are, they are generally less toxic than the nano-TiO_2.

3.6 Conclusions

From meta-analysis of the 25 *in vivo* studies, we can confirm that decreasing TiO_2 particle size is associated with increased toxicity. The hypothesized greater toxicity of the anatase phase could not be confirmed. The increased toxicity of increasing purity suggests that the observed effects are in fact products of exposure to titanium dioxide and not some unknown substance. Aggregation, at least as it is currently reported, does not appear to significantly affect measured toxicity. Based on this current information, the treatment of all TiO_2 nanoparticles as members of the same class is justified.

This work has also demonstrated the utility of this analytical framework in extracting variable importance information of different nanoparticle characteristics even in a data set missing 40% of the potential measurements, a data set so porous that MLR was impossible without resorting to imputation methods. Although these techniques are necessarily limited by the experimental choices made in the published studies, they produced a quantitative ranking of the magnitude of effects of different particle attributes in explaining differences in observed toxicity of TiO_2 nanoparticles.

Table 3-1: Published nano-TiO_2 pulmonary toxicity experiments that were used in this meta-analysis.

Reference	Year	Primary Size (nm)	Exposure Route	Animal Species	Phase Composition (% anatase)	Instilled Exposure (mg)	Aerosol Concentration (mg/m^3)	Post Exposure Period (weeks)
Oberdorster G et al. [71]	1992	20, 250	instillation	rats	100	0.1 – 0.5	-	1 day
Osier M et al.[72]	1997	21, 250	inhalation, instillation	rats	100	-	125	0 – 1
Bermudez E et al.[73]	2002	unknown	inhalation	mice, rats, hamsters	0	-	9.5 – 240.3	0 – 52
Rehn B et al.[74]	2003	20	instillation	rats	unknown	0.15 – 1.2	-	0 – 13
Renwick L et al.[75]	2004	29, 250	instillation	rats	unknown	0.125 – 0.5	-	1 day
Warheit D et al.[62]	2006	10, 200×35	instillation	rats	0, 100	4.05 – 20.25	-	0 – 12
Grassian V. et al.[76]	2007	3.5, 17.8	inhalation	mice	100	-	0.77 – 7.35	0 – 1 day
Warheit D et al.[63]	2007	130 – 380	instillation	rats	0, 80	0.245 – 1.225	-	0 - 12
Nemmar A. et al.[77]	2008	6	instillation	rats	0	0.36 – 1.8	-	1 day
Kobayashi et al.[78]	2009	18 – 180	instillation	rats	100	0.25 – 1.25	-	0 – 4
Warheit D et al.[79]	2010	135 – 164	instillation	rats	15	1.25	-	0 – 12

Table 3-2: List of nano-TiO_2 characteristics and exposure variables included in meta-analysis data set.

Variable names	No. of entries	Variable types	Imputation distribution (if applicable)
Particle characteristics			
Material name	422	class	NONE
Manufacturer/supplier	422	class	Discrete (based on proportions of manufacturers existing in data)
Anatase%	422	numeric	Discrete (based on proportional distribution of existing values)
Purity (% TiO2)	422	numeric	Normal distribution based on mean and standard deviation of existing purity data
Shape	65	class	Discrete (based on proportions of shapes existing in data)
Average size (nm)	422	numeric	Normal distribution based on mean and standard deviation of existing size data
Average Surface Area (m^2/g)	161	numeric	Normal distribution based on mean and standard deviation of existing surface area data
Hydrophilic or hydrophobic	422	class	Discrete (based on proportions of classes existing in data)
Animal characteristics			
Animal species	422	class	not necessary
Animal breed/strain	422	class	not necessary
Animal sex	422	class	not necessary
Animal age (weeks)	422	numeric	not necessary
Animal weight (g)	422	numeric	Normal distribution based on mean and standard deviation of measured values for a particular species
Exposure characteristics			
Suspension media	195	class	NONE
Average size in suspension (nm)	29	numeric	NONE
Exposure route- general	418	class	NONE
Exposure route- detail	418		NONE
Exposure length	318	numeric	NONE

Variable names	No. of entries	Variable types	Imputation distribution (if applicable)
Exposure frequency	422		NONE
Exposure duration (days)	237		NONE
Exposure time post instillation (days)	422	numeric	NONE
Exposure cumulative particle exposure (mg)	258	numeric	NONE
Number of Animals for this concentration and exposure duration (and time postexposure)	391	weighting variable	N/A, weighting variable
Aerosol size (nm)	143	numeric	Normal distribution from mean and standard deviation of existing measures (inhalation only)
Aerosol concentration (mg/m^3)	197	numeric	NONE
Toxicity characteristics			
Health, mortality	146	numeric	Output 1
Health, weight	147	numeric	Output 2
Dosimetry # animals used for dosimetry at this dose, duration, and time postexposure	207	weighting variable	N/A, weighting variable
# animals used for BAL at this dose, duration, and time postexposure	314	weighting variable	N/A, weighting variable
BAL total cell count	259	numeric	Output 3
BAL polymorphonuclear neutrophil cell count	247	numeric	Output 4
BAL eosinophil cell count	120	numeric	Output 5
BAL lymphocyte cell count	131	numeric	Output 6
BAL macrophage cell count	195	numeric	Output 7
BAL MTP (micro total protein) level (mg/mL)	172	numeric	Output 8
BAL LDH (lactate-dehydrogenase) level (U/L)	132	numeric	Output 9
Lung absolute weight (g)	55	numeric	Output 10
Lung histopathology/morphology	212	class	Output 11

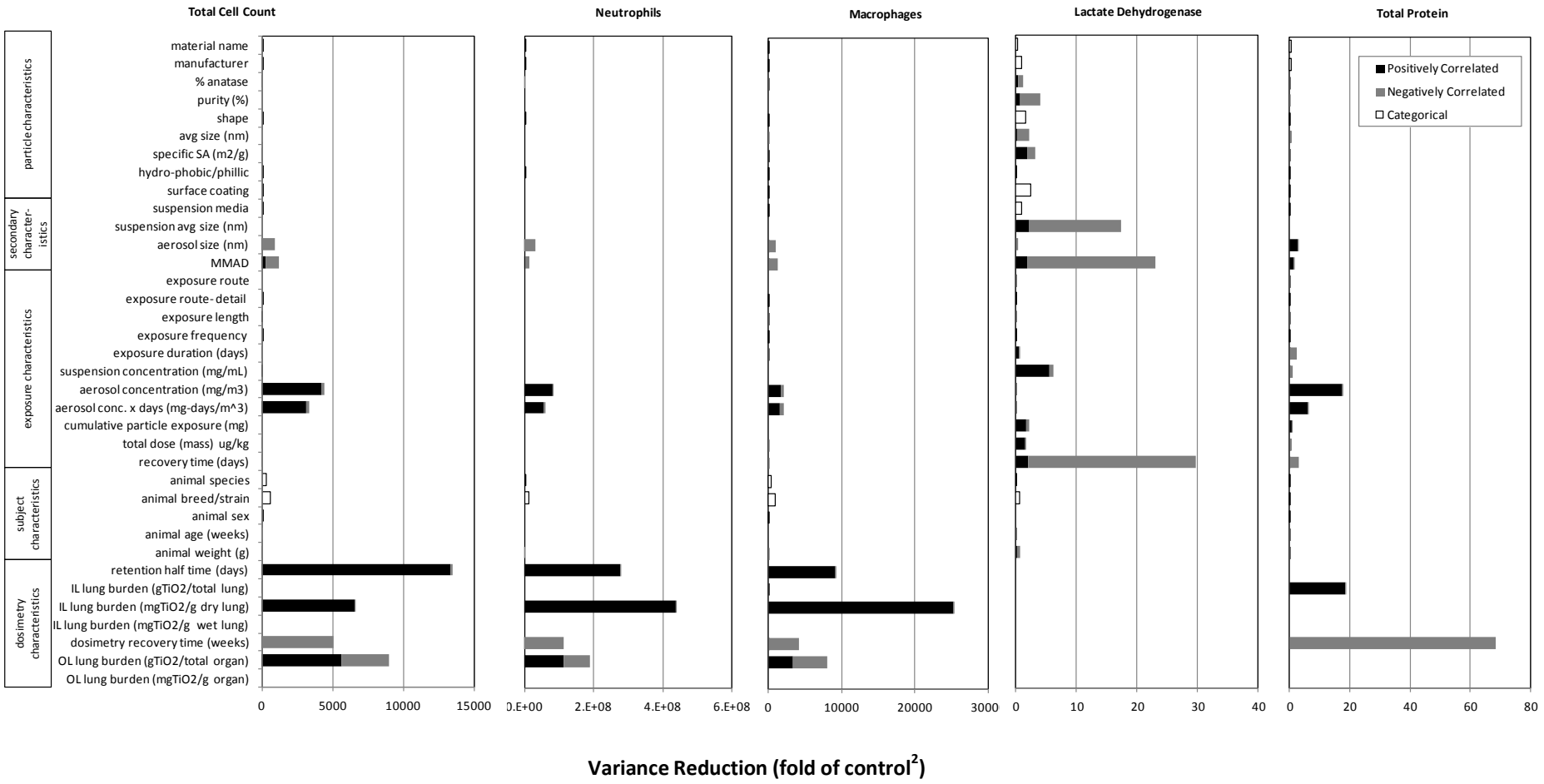

Figure 3-1: Variable importance results as measured by average variance reduction for the Random Forest models predicting total cell count (TCC), polymorphonuclear neutrophils (PMN), macrophages (MAC), lactate dehydrogenase (LDH) and total protein (TP) in BAL fluid.

The parameters with high variance reduction have more influence on the toxicological endpoint specified than the parameters with low variance reduction.

Table 3-3: RT, RF, and MLR models' goodness of fit statistics.

Modeled Endpoint	Regression Tree		Random Forest		Multiple Linear Regression	
	R^2	# of variables	R^2	# of variables	R^2	# of variables
Neutrophils	0.95	3	0.95	3	0.07	11
Macrophages	0.95	1	0.96	3	0.06	3
Total Cell Count	0.68	1	0.72	5	0.05	7
Total Protein	0.31	1	0.41	6	0.08	12
Lactate Dehydrogenase	0.49	9	0.54	5	0.07	6

Table 3-4: Variables included in linear regression models. Asterisks (*) indicate the number of times the given variable appeared in models generated separately for the 5 independently imputed data sets.

	Neutrophils	Macrophages	Total Cell Count	Lactate Dehydrogenase	Total Protein
Particle characteristics					
material name	*****	*****	*****	*****	*****
manufacturer	*****	*****			*****
shape	*****		*****	*****	*****
hydro-phobic/philic			*		*
specific surface area	***				*
surface coating	*		*****	*****	****
Exposure characteristics					
suspension media	****			*****	**
exposure route					*****
exposure route - detail	*				*****
cumulative particle exposure				*****	
Subject characteristics					
animal species	*		*****		
animal sex	*				
animal breed/strain	*****	*****	*****	*****	***
animal weight	****		*****		*
animal age					*

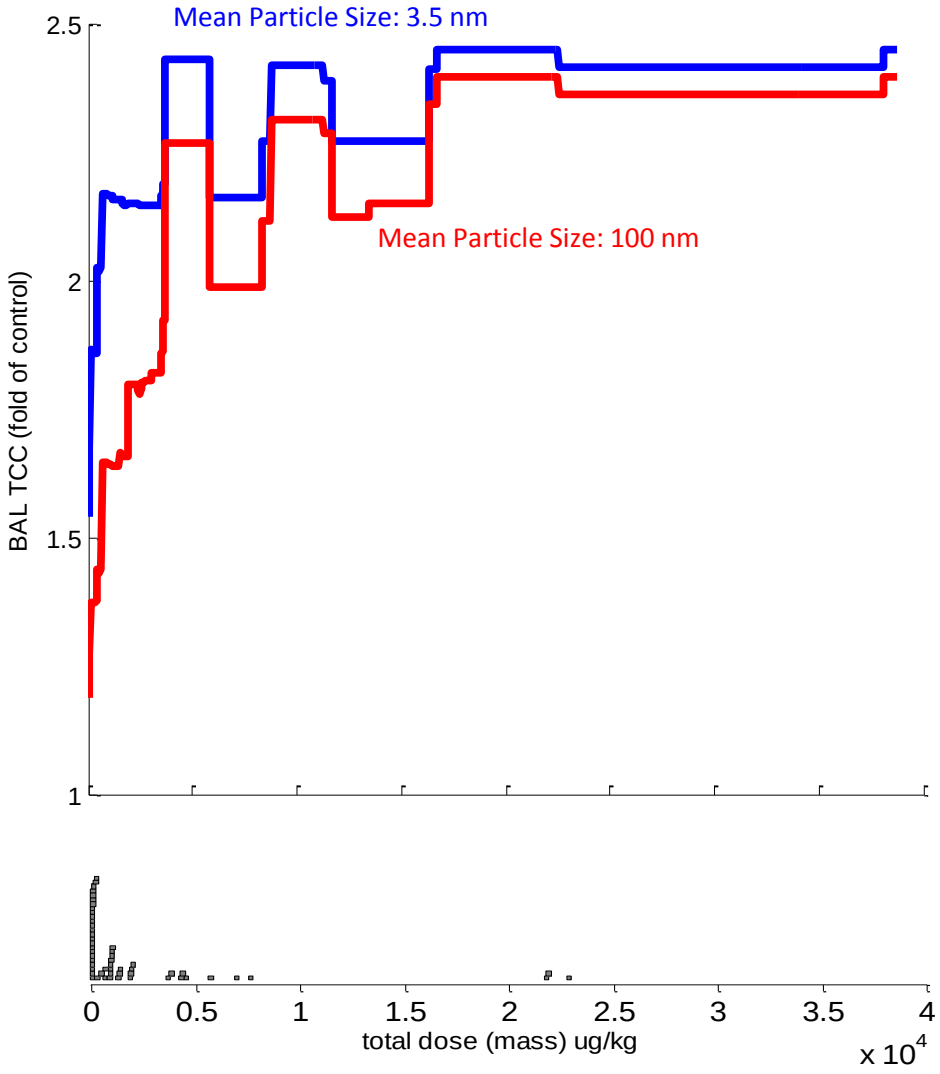

Figure 3-2: Dose-response curves as generated by the RF model for change in BAL total cell count (TCC) following exposure to different sized titanium dioxide particles.

Over the entire range of doses, the smallest TiO_2 particles (3.5 nm) demonstrate a 5% greater mean change in TCC across all dose levels than the largest nanoparticles (100 nm). The gray blocks in the sub-x-axis indicate dose levels at which experimental data exists in terms of total dose by mass (other dose-equivalent effects are inferred from concentration and time of exposure to produce the final model).

Table 3-5: Mean relative shift (%) observed in nano-TiO_2 dose-response curves for three BAL inflammation indicators resulting from changes in select nanoparticle attributes.

Mean relative shift is calculated as the relative change in integrated area under the dose-response curve with the secondary input variable and comparing that to the integrated area under the curve created with the secondary input at its maximum value. The result marked "n/a" did not have sufficient data with which to determine an effect. MAC and PMN results lacked sufficient input variable data with which to evaluate shift effects of these variables on dose-response.

		Output Variables		
Input Variable	**Minimum/Maximum Value (units)**	**TCC (%)**	**LDH (%)**	**TP (%)**
Aggregate Size, MMAD	18/1,450 (nm)	+8	+7	−5
Primary Particle Size	3.5/1,000 (nm)	−12	−47	−45
Purity	88/100 (%)	n/a	+13	+21
Anatase Fraction	0/100 (%)	+9	−18	−13

4 Selecting Nanoparticle Properties to Mitigate Risks to Workers and the Public – A Machine Learning Modeling Framework to Compare Pulmonary Toxicity Risks of Nanomaterials

4.1 Abstract

Due to their size and unique chemical properties, nanomaterials have the potential to interact with living organisms in novel ways, leading to a spectrum of negative consequences. Though a relatively new materials science, already nanomaterial variants in the process of becoming too numerous to be screened for toxicity individually by traditional and expensive animal testing. As with conventional pollutants, the resulting backlog of untested new materials means that interim industry and regulatory risk management measures may be mismatched to the actual risk. The ability to minimize toxicity risk from a nanomaterial during the product or system design phase would simplify the risk assessment process and contribute to increased worker and consumer safety.

Some attempts to address this problem have been made, primarily analyzing data from in vitro experiments, which are of limited predictive value for the effects on whole organisms. The existing data on the toxicity of inhaled nanomaterials in animal models is sparse in comparison to the number of potential factors that may contribute to or

aggravate nanomaterial toxicity, limiting the power of conventional statistical analysis to detect property/toxicity relationships. This situation is exacerbated by the fact that exhaustive chemical and physical characterization of all nanomaterial attributes in these studies is rare, due to resource or equipment constraints and dissimilar investigator priorities.

This paper presents risk assessment models developed through a meta-analysis of *in vivo* nanomaterial rodent-inhalational toxicity studies. We apply machine learning techniques including regression trees and the related ensemble method, random forests in order to determine the relative contribution of different physical and chemical attributes on observed toxicity. These methods permit the use of data records with missing information without substituting presumed values and can reveal complex data relationships even in nonlinear contexts or conditional situations.

Based on this analysis, we present a predictive risk model for the severity of inhaled nanomaterial toxicity based on a given set of nanomaterial attributes. This model reveals the anticipated change in the expected toxic response to choices of nanomaterial design (such as physical dimensions or chemical makeup). This methodology is intended to aid nanomaterial designers in identifying nanomaterial attributes that contribute to toxicity, giving them the opportunity to substitute safer variants while continuing to meet functional objectives.

Findings from this analysis indicate that carbon nanotube (CNT) impurities explain at most 30% of the variance pulmonary toxicity as measured by polymorphonuclear neutrophils (PMN) count. Titanium dioxide nanoparticle size and aggregation affected the observed toxic response by less than ±10%. Difference in observed effects for a group of metal oxide nanoparticle associated with differences in Gibbs Free Energy on

lactate dehydrogenase (LDH) concentrations amount to only 4% to the total variance. Other chemical descriptors of metal oxides were unimportant.

4.2 Introduction

Nanoparticles or ultrafine particles, sometimes designated PM0.1 (i.e. particulate matter 0.1 micrometers or less), are solid particles where at least one dimension is in the range of 1-100 nanometers. A nanometer is one billionth of a meter, and is just slightly longer than 3 molecules of H_2O would stretch if lined up end to end. Since many properties of solids including magnetism, electrical conductivity, and tensile strength arise only after sufficient numbers of atoms have aggregated together, the properties exhibited by nanoparticles can either be in gentle transition from solute to solid behavior or reveal steep threshold transitions to some combination of characteristics as particle size increases. Conversely, at small sizes, particles can exhibit an entirely new property like semi-conductivity that is unexpected based on studies of the material bulk properties.

These peculiar aspects of nanoparticles have led to the conjecture that some particular substance in nano-form could prove to be anomalously toxic or more toxic [80]. Since many biological structures and molecules are nano-sized themselves (e.g., like the protein immunoglobulin-G that measures 33 nm across), there is a potential for unique interactions between these small solids and biological processes.

That exposure to nanomaterials in the workplace may result in dangerous pathologies associated with their unique properties is reflected in recent National Institutes of Occupational Safety and Health (NIOSH) recommendations. Existing regulations by the Occupational Safety and Health Administration (OSHA) limit

titanium dioxide particulate to a concentration in the workplace of 15 mg/m^3, and limit carbon particulate to a concentration of 5 mg/m^3 [64]. But, upon review of the available research, NIOSH published a recommendation that titanium dioxide nanoparticles be limited to no more than 0.3 mg/m^3 [65], and also proposed that carbon nanotubes (CNTs) be limited to concentrations no greater than 7 $\mu g/m^3$ [81]. The recommended maximum allowable quantities of nano-particulate in the workplace are 50 and 700 times smaller than for carbon soot and micron and larger diameter titanium dioxide respectively. No new OSHA regulations for nanoparticles have yet been formally adopted.

The new proposed exposure limits apply to all particles with a primary size smaller than 100 nm equally, even though substantial differences in toxicity may exist between 20 nm and 40 nm particles, between thinner or thicker nanotubes, or between lightly aggregated or significantly aggregated nanoparticles. Small differences in chemistry including coatings, functionalizations, or contaminants may produce divergent toxic responses. Thus, a limit based on all shapes and varieties for a certain chemical compound sized on average below a single threshold may be inadequate.

Developing specific standards for nanoparticle variants would require a lot more data than the size-based standard and, animal studies, while the most applicable to human risks, are expensive. Since, there are potentially dozens of variable characteristics between different nanoparticles, such studies can only offer limited conclusions on the importance of specific characteristics (for example, that multi-walled carbon nanotubes may be more toxic than single-walled carbon nanotubes [22]).

Reaching conclusions on the interactions of specific characteristics and dose on observed toxic effects is further complicated by the inconsistent or incomplete

measurement of nanoparticle properties among the published studies. So, in many published nanotoxicology studies there is uncertainty about the properties of the substance that was tested.

Since, there is limited capability of cellular *in vitro* studies to predict the outcome of *in vivo* mammalian studies [82], and individual animal studies have not adequately studied the variations in nanoparticle properties, understanding the contributions of several nanoparticle properties at once must rely on a virtual experiment assembled from combinations of individual studies across the literature, a meta-analysis [67].

The identification of properties responsible for toxicity could conceivable allow product and process designers to design safer nanoparticles while achieving the same functional objectives. For example, if the length of a carbon nanotube was critical to a design's functionality, but its diameter was not, and diameter proved to be a critical determinant of the carbon nanotube's toxicity risk, careful selection of the CNT's diameter could mitigate that risk without compromising functionality.

There has been limited research to date on predictively modeling the toxicity of nanomaterials, and those studies have focused entirely on cell culture toxicity *in vitro* [15], [18], [66]. Summarizing the knowledge gained to date from *in vivo* mammalian pulmonary nanoparticle toxicity studies of relevance to workplace safety is the intent of this paper.

This work seeks to quantify and visualize the degree to which changes in certain nanoparticle characteristics change the overall magnitude of the toxic response to a given dose of nanoparticles. To accomplish this, our study utilizes a machine learning algorithm called the random forest (RF), which has unique capabilities for quantitatively learning from data with a high proportion of missing values, revealing relationships that may be conditional or only applicable after a certain threshold has been passed, and

without having to assume statistical independence between each of the inputs. These strengths make RF models especially suited to risk assessment activities in the early stages of implementing a new technology.

4.3 Methods

We perform a meta-analysis of pulmonary nanoparticle toxicity studies in order to determine the degree to which design variables such as chemical composition, dimensions, shape, and surface treatments affect the magnitude of the toxic dose response.

4.3.1 Data Sources

We collected data from published peer-reviewed literature describing experiments where rodents were exposed to nanoparticles through inhalation (dry aerosols), aspiration (a small volume of saline fluid with suspended nanoparticles positioned just beyond the trachea and naturally inhaled by the animal), or instillation (a small volume of saline fluid with suspended nanoparticles injected into the bronchial tubes).

All of the included studies reported quantitative toxicity measures for either the concentration of lactate dehydrogenase (LDH) or the number of polymorphonuclear neutrophils (PMN) in bronchoalveolar lavage (BAL) fluid. BAL is a procedure where saline fluid is used to rinse out the lungs of the rodent some time following exposure. The fluid is collected along with dislodged cells, particles, and biomolecules and analyzed for indicators of toxicity, like PMN, LDH, and Total Protein.

LDH is a cellular protein and as such is an indicator of cytotoxicity or cell membrane damage in the lungs. This is typically measured as a concentration on a picogram per milliliter basis, but we translate all data for this analysis as a multiple change from control

basis (i.e. fold of control). The mean and standard deviation of LDH concentration for a group of animals exposed to nanoparticulate are normalized to the mean measurement for the control group. Control animals are either exposed to only air for inhalation experiments, or an instillation of saline fluid for instillation experiments.

The PMN cell count is an indicator of inflammation and the early stages of an immune response. PMNs are measured in terms of absolute cell counts per milliliter of BAL fluid. We translate these values into "fold of control" in the same way as the LDH values to reflect the change from the control group response.

Studies included in this meta-analysis had to include characterization of the nanoparticles used in the experiment as well as control groups and a quantitatively measured output for PMN, LDH, or Total Protein including reported uncertainty. A complete listing of all data sources utilized in this study is provided in Annex C.

4.3.2 Data Preparation

The input for the data analysis is a matrix with rows representing data specific to individual animals in the selected studies and columns containing experimental and output variables.

Before analysis, the output data for this study were expanded using a Monte Carlo resampling technique. At the rate of 100 samples per animal subject, a given set of experimental inputs including dose levels and nanoparticle characteristics were associated with 100 discrete realizations of the reported distribution of measured output responses. Distributions for measured endpoints were assumed to be normal with values deemed to be impossible (cell counts less than zero, for example) excluded. For example, if the PMN average of 6 animals with a given exposure to nanoparticles was measured to be 10 $\pm$ 2, the data set would contain 600 rows (6 $\times$ 100) with identical input characteristics,

but each row having a discrete sample from a normal distribution with a mean of 10 and a standard deviation of 2. We found that higher sampling rates (e.g., 500 rows per animal or 1,000 rows per animal) did not alter the results.

This procedure accomplishes two important tasks for this analysis: (1.) it reduces the likelihood of overfitting, since the model must measure its error against the entire range of experimental outcomes and not just the mean values; and (2.) it permits the uncertainty of the multi-dimensional model with respect to the actual measured results to be traced through and evaluated at any desired point or sub-region of interest.

The final data set contained one column for each experimental variable including measured nanoparticle characteristics and attribute of exposure (e.g., total dose, length of recovery, mode of exposure, etc.), and 100 rows for each animal subject utilized in the experiments.

4.3.3 Random Forest Models

We employ random forest (RF) models [7]—an unsupervised machine learning method—to discover and quantify the relationships in the existing data set. RF models are made up of ensembles of regression trees (RT) [6], which are hierarchical structures of decision rules that divide observations into two groups on the basis of a specific criterion (see Figure 4-1). The decision rules are automatically selected by the algorithm on the basis of those with produce the greatest possible information gain. Random forests extend this process by creating large numbers of regression trees using subsets of the data randomly omitting a fraction of the experimental variables. The results are averaged and considered to be more robust than a single regression tree constructed from the complete data set.

We have implemented Breiman's random forest algorithm via the MATLAB™ function treebagger, creating RF models of 1,000 trees each, each branch being established from a randomly selected sub-set of one third of the available input variables. The learning progression diagrams shown in Annex B indicate that the models have already reached their maximum performance with several hundred fewer trees. One may download the final model objects for this study and instructions on implementation here: http://nanohub.org/resources/17539. For further detail on the model implementation, internal structure and validation results, see Appendix C.

4.3.4 Visualizing Interactions

In order to visualize the interactions between multiple nanoparticle and exposure characteristics at the same time, we record the RF predicted output while changing 2 or 3 of the input parameters of interest from their minimum to maximum values in 20 steps. For 2 variables, this creates a matrix of 400 values, which we represent as a filled contour plot (see Figure 4-2). Changes in 3 variables are represented by multiples of two dimensional contour plots, for example showing the relationship between toxicity and changes in length and diameter for different doses of carbon nanotubes as shown in Figure 4-3.

4.4 Results

Examining the dose-response effects of exposure to nanomaterials (see Figure 4-2), one can see that for carbon nanotubes the total dose dominates the effects from the length of recovery time in their influence on PMNs, while the recovery from exposure to titanium dioxide nanoparticles dominates the expected LDH and total protein

concentrations, while total dose explains most of the LDH observations for metal oxides (Figure C-9).

To consider the effects of nanoparticle design tradeoffs, one can see, for example, how length and diameter of CNTs appear to affect toxicity across a range of dose levels (Figure 4-3). With the highest observed responses in PMN and LDH occuring when the diameter of the CNTs is large and the length of the CNTs is short. These effects are consistent proportionally across several dose levels, even as the total magnitude of the observed response increases.

For titanium dioxide we display the effects of chemical purity and aggregation (MMAD, mass mode aerodynamic diameter—a metric for the average size of aggregated particles) as total dose changes (Figure 4-4). Aggregation appears to have a limited effect on pulmonary inflammation as compared to changes in purity.

For a broader set of metal oxides, we find that the total dose for an animal subject is a much more important predictor of measured BAL LDH than any other physical or chemical attribute of the nanoparticles (Figure 4-5). The total dose of metal oxide nanoparticles appears to explain almost all of the variation (Figure 4-5 and C-9)

Additional contour plots generated by these models are included in Annex A. These include the relationship between carbon nanotube dose, cobalt impurity dose and PMN and LDH (Figures C-1 and C-2); the relationship between titanium dioxide dose and aggregate diameter for LDH and total protein (Figures A3 and A4); the relationship between aggregate diameter and recovery time for titanium dioxide nanoparticles and LDH and total protein (Figures C-5 and C-6); and the relationship between aggregate diameter and purity (Figure C-7) and aggregate diameter and Gibbs Free Energy for BAL LDH following exposure to metal oxide nanoparticles.

4.5 Discussion

In terms of an individual designing or specifying a nanomaterial for a particular application and wanting to minimize risks from toxicity at the same time, certain factors including particle size, shape, and chemical makeup would be at the forefront of easily manipulatable design characteristics. If a designer can reduce the toxic risk through careful selection of these factors while continuing to meet functional objectives, they would likely do so.

However, the effects of changes in particle size on toxicity are still a matter of some debate. Aggregation of nanoparticles into larger particles is also debated as whether it may exacerbate toxicity or alternatively to not have a significant effect [83]. It is also unclear from the published literature whether impurities should be considered important or unimportant contributors to toxicity [23], or further, to what actual extent differences in chemical makeup account for differences in toxicity between different nanoparticles.

4.5.1 Effects of Particle Size and Aggregation

Particle size and aggregation were thought to be important determinants of toxicity for nanomaterials, especially the idea that as the particles became smaller and potentially more highly reactive, their toxicity could markedly increase [84]. Although the experimental data including cellular-level in vitro experiments is mixed, larger particle sizes and aggregate do at least sometimes increase the resulting toxicity [85].

For carbon nanotubes, we see an overall effect of both length and diameter. Increasing diameter, which may also be an indicator of carbon nanotube stiffness, was associated with increasing toxicity (Figure 4-3), with two thresholds of 5nm and 30nm. CNTs with a diameter less than 5nm are single-walled nanotubes, while those with larger

diameters are multi-walled. The greatest toxicity is exhibited by CNTs with large diameters and short lengths. By way of comparison, asbestos fibers are on average longer (by 2-5 times) and have much larger diameters (by about two orders of magnitude) than the typical carbon nanotube.

For titanium dioxide, very small particles do in fact seem to generally produce higher dose-response effects than larger titanium dioxide nanoparticles (Figure C-10). This only occurs for very small particles, and fewer data points are available in this size range making this conclusion more uncertain than the general observation that dose-response is not influenced much by particle size over most of the wide range of tested sizes.

When considering the metal oxides as a group, particle size does affect the BAL concentration of LDH to some extent, with larger particles causing higher LDH concentrations. Particle size contributes to model variance reduction (Figure C-15), but the magnitude of the difference in LDH is dwarfed by the change associated with higher doses (Figure 4-5).

This is opposite of the effect observed for titanium dioxide data analyzed alone, where titanium dioxide nanoparticles smaller than 40 nm caused at most a 2-fold increase in LDH. It must be noted that the entire metal oxide data set did not include any particles that small except for titanium dioxide nanoparticles, so the effect of very small diameter metal oxide nanoparticles should not be considered to be well defined.

4.5.2 Effects of Impurities

The importance of impurities in explaining the toxicity of nanoparticles has long been debated for carbon nanotubes, and these data appear to clearly indicate that the cobalt content of CNTs (see Figure C-1) has the effect of increasing the immune response, whether sensitizing the system to the effects of CNTs exposure, or causing

such an effect independently. The metallic impurities that exist together with the CNTs are remants of the metallic catalysts used in the manufacturing process.

For titanium dioxide nanoparticles, on the other hand, the impurities (or purity) of the particles is not a significant contributor (see Figure 4-4) to inhalational toxicity. This is most likely due to the fact that impurities in titanium dioxide nanoparticle manufacturing include much more inert materials than the metals associated with carbon nanotubes. The impurities were not often characterized in the nano-TiO2 toxicology studies.

4.5.3 Effects of Chemical Differences

A variety of different quantitative chemical descriptors have been proposed and tested with models to predict *in vitro* toxicity of metal oxide nanoparticles [4], [15]. But, as seen in Figure 4-5, the results of pulmonary exposure studies on rodents appear to indicate that the total mass of metal oxide nanoparticles is a much more important predictor than any chemical or physical descriptors. In fact, if all chemical descriptors were excluded from the model, the fraction of explained variance (or R^2 value) only decreased from 0.97 to 0.93.

The magnitude of the change in LDH as shown in Figure 4-5 due to particle size, aggregation, or Gibb's Free Energy are dwarfed by the magnitude of change due to simply increasing the total mass of metal oxide nanoparticles the animals are exposed to. While this analysis only includes a few different metal oxides, these oxides do differ significantly in terms of solubility, thermodynamic stability, and reactivity. While other quantitative chemical descriptors were tested including metal group or period from the periodic table, the mean isoelectric point, the surface charge, the enthalpy of formation,

and crystalline structure, the Gibb's Free Energy proved to have the greatest apparent effect, but only a slight one.

4.6 Conclusions

Random forest models even when trained on an incomplete data set can provide useful risk assessment of the benefits or costs of possible design tradeoffs in the area of nanoparticle toxicity. Using these models to quantitatively summarize the current knowledge and visualize the relationships between particle design parameters contributes to understanding the risks of a new technology. This is especially true during the early stages of implementation when the science may not have developed mechanistic explanations for why one material may pose a higher risk than another.

The pulmonary toxicity measured by LDH release of metal oxide nanoparticles as a group including titanium dioxide, magnesium oxide, silicon dioxide, and zinc oxide does not appear to be highly dependent on physical characteristics of the particles, and depends only slightly on chemical characteristics, at least within the ranges that have been tested to date in animals. This leads to the conclusion that for these materials, the first and best risk mitigation may be only to minimize exposure.

Design characteristics for carbon nanotubes are much more important, relatively, to pulmonary toxicity, at least for the relatively short term exposures that have been examined so far. These characteristics include the proportion of metallic impurities like cobalt, and the nanotube length and diameter. CNT diameter is important over a wide range of doses and combinations of other variables and should be minimized to mitigate toxicity.

Meta-analysis of toxicity studies such as this one have the ability to quantitatively compare the claims of single studies against the larger field of study and to quantify the relative contributions of a large number of factors. Those individual studies form the basis for this analysis, their conclusions are re-evaluated in light of other findings and minor effects can be distinguished from major ones. Such information could be taken into account in future product and process design decisions that utilize nanoparticles in order to mitigate risks to workers, consumers, and businesses. Meta-analyses could also play a role in determining future regulatory decisions regarding these materials, by helping distinguish significant from insignificant effects on toxicity.

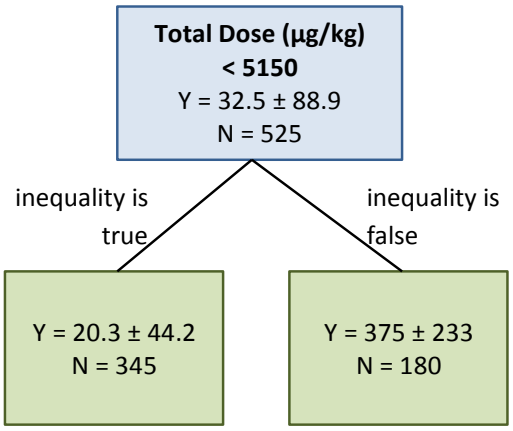

Figure 4-1: Example of a single-branch regression tree with two leaf nodes

Table 4-1: Listing of all rodent pulmonary toxicity studies included in the characteristic interaction risk analysis models.

Ref.	First Author	Year	Nano-particle	Total Dose (µg/kg)	Recovery Period (days)	Exposure Mode	End-point(s) Measured
[34]	Pauluhn J.	2010	CNT*	105 – 6,290	1 – 90	Inhalation	PMN, LDH
[24]	Muller J.	2005	CNT*	2 – 8	3	Instillation	PMN, LDH
[86]	Shvedova A.	2008	CNT*	250 – 1,000	1 – 7	Instillation	PMN, LDH
[55]	Porter	2010	CNT*	435 – 1,740	1 – 28	Aspiration	PMN, LDH
[87]	Inoue	2008	CNT*	4,000	1	Instillation	PMN
[46]	Shvedova A.	2005	CNT*	490 – 1,970	1 – 60	Aspiration	PMN
[47]	Shvedova A.	2007	CNT*	1,851	1	Aspiration	PMN, LDH
[45]	Warheit D.	2004	CNT*	1,000 – 5,000	1 – 30	Instillation	LDH
[49]	Muller J.	2008	CNT*	8,890	3	Instillation	LDH
[88]	Ellinger-Ziegelbauer H.	2009	CNT*	180 – 3,900	7 – 90	Inhalation	LDH
[89]	Bermudez E.	2002	TiO_2	10,000 – 90,000	0 – 365	Inhalation	PMN
[76]	Grassian V.	2007	TiO_2	35 – 3,300	0 – 14	Inhalation	PMN, LDH
[77]	Nemmar A.	2007	TiO_2	1,000 – 5,000	1	Instillation	PMN
[90]	Oberdorster G.	1994	TiO_2	295 – 2300	0	Instillation	PMN
[91]	Warheit D.	2006	TiO_2	300 – 10,000	0 – 84	Instillation	PMN, LDH
[75]	Renwick L.	2004	TiO_2	300 – 1,200	1	Instillation	PMN
[74]	Rehn B.	2003	TiO_2	750 – 6,000	0 – 90	Instillation	PMN
[72]	Osier M.	1997	TiO_2	750 – 3,750	0 – 7	Inhalation	PMN
[63], [92]	Warheit D.	2007	TiO_2	1,000 – 5,000	0 – 84	Instillation	PMN, LDH
[93]	Warheit D.	2010	TiO_2	1,000 – 5,000	0 – 30	Instillation	LDH
[78]	Kobayashi	2009	TiO_2	5,000	0 – 10	Instillation	LDH

	N.						
[94]	Sayes C.	2007	SiO_2, ZnO	1,000 – 5,000	1 – 90	Instillation	LDH
[83]	Gosens I.	2010	SiO_2	1,600	1	Instillation	LDH
[82]	Warheit D.	2009	ZnO, MgO	1,000 – 5,000	1 – 90	Instillation, Inhalation	LDH

*The carbon nanotube (CNT) portion of this pulmonary toxicity data set is available for download at http://nanohub.org/resources/13515 [32]

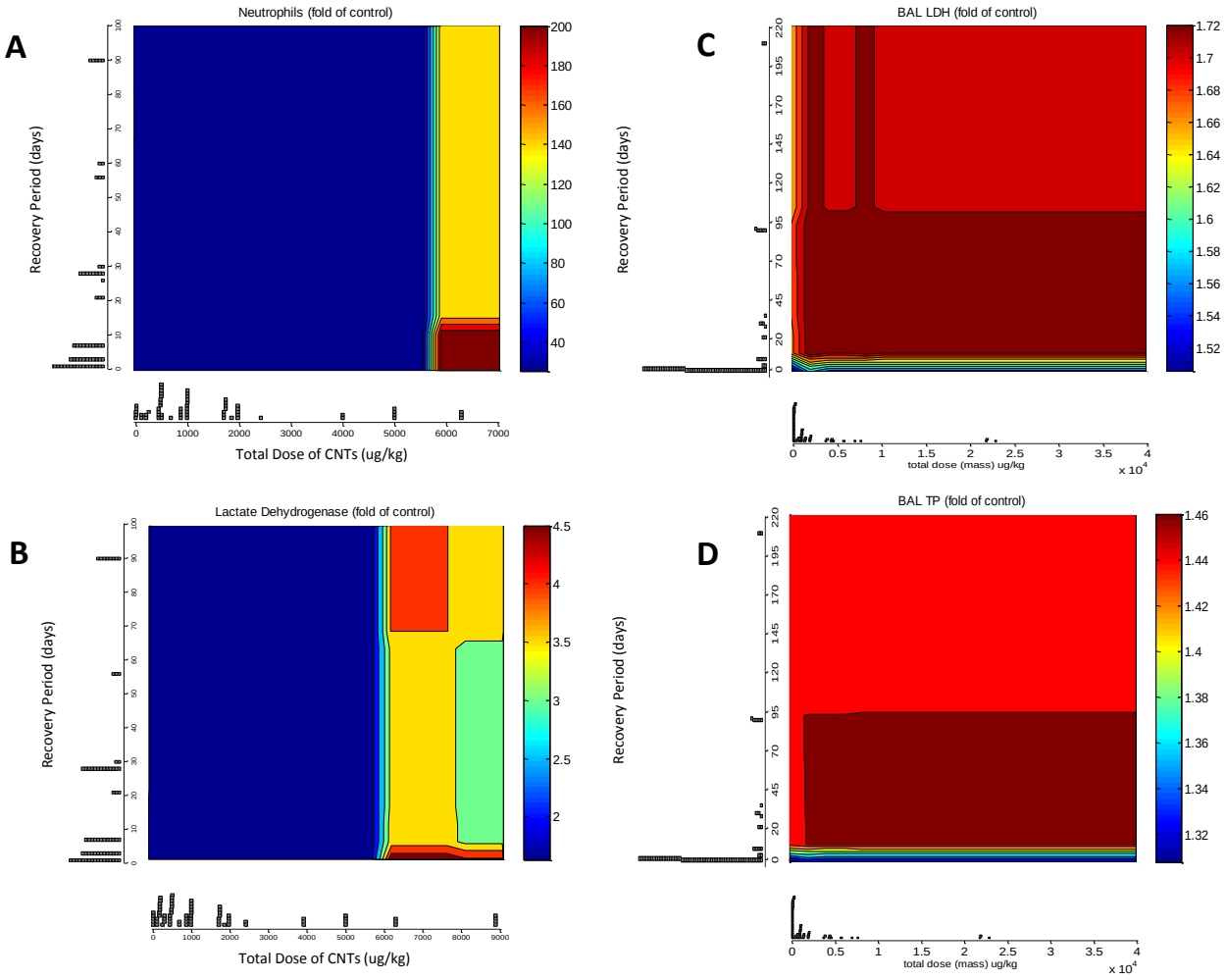

Figure 4-2: Dose- and Recovery-response contour risk plots for CNT and TiO_2 nanoparticle toxicity.

[A] Change in Neutrophil count in BAL fluid following pulmonary exposure to carbon nanotubes. [B] Change in lactate dehydrogenase in BAL fluid following pulmonary exposure to carbon nanotubes. [C] Change in lactate dehydrogenase in BAL fluid following exposure to titanium dioxide nanoparticles. [D] Change in total protein in BAL fluid following exposure to titanium dioxide nanoparticles. The gray blocks along the x- and -y-axis indicate dose and recovery time levels at which experimental data exists. In the case of TiO_2 nanoparticles other dose-equivalent effects are inferred from concentration and time of exposure to produce the final model).

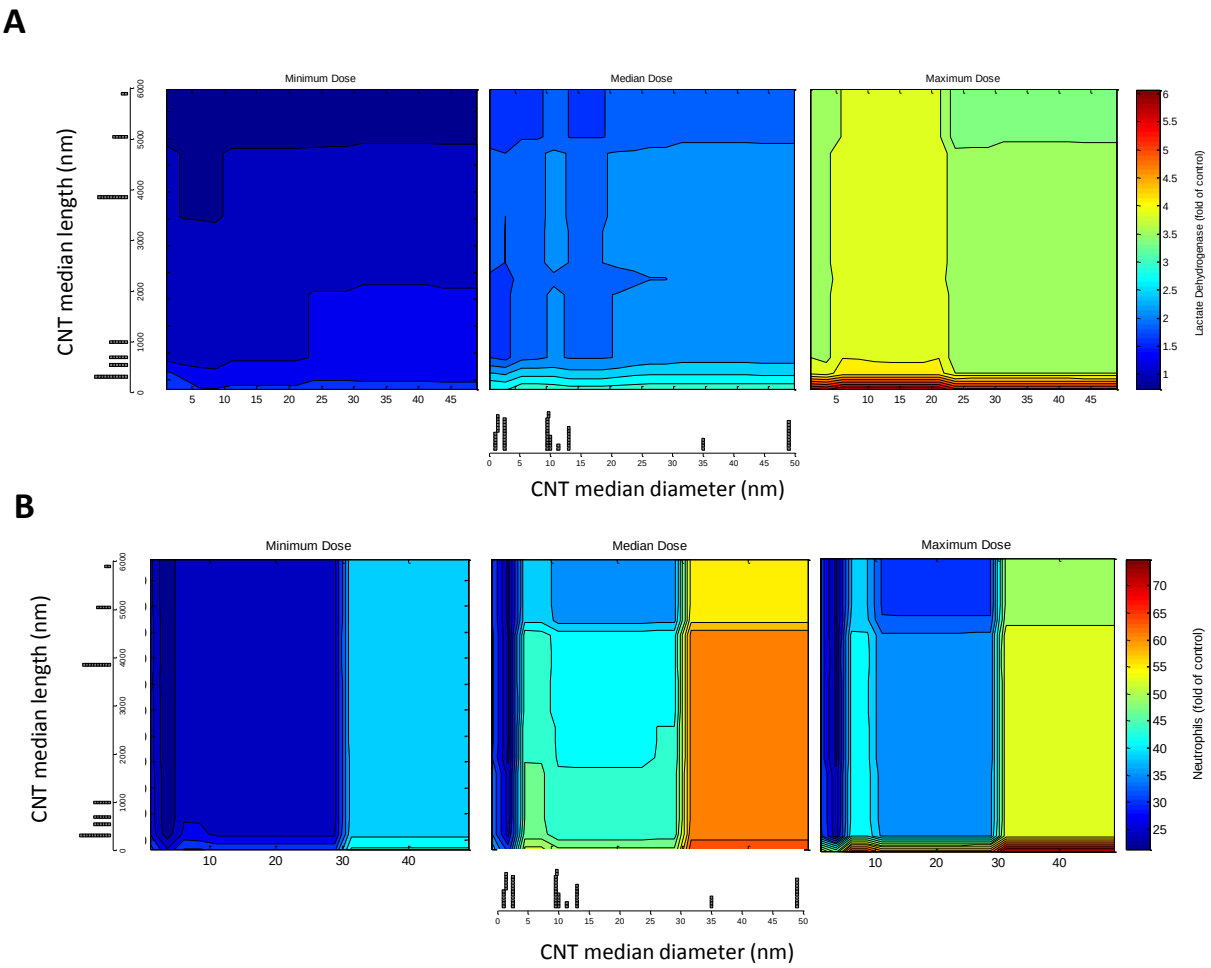

Figure 4-3: Effects of pulmonary exposure to carbon nanotubes at three dose levels, and all values of nanotube length and diameter:

Minimum dose is 2 μg/kg; median dose is 3250 μg/kg; maximum dose is 6500 μg/kg. [A] Change in Lactate dehydrogenase (LDH) in BAL fluid following exposure. [B] Change in Neutrophils count in BAL fluid following exposure. Values other than dose, length, and diameter, such as recovery period, and % cobalt impurity are held constant at their median reported values. These results suggest that larger diameter CNTs (multi-walled CNTs) produce a significantly increased immune response (PMN counts), but only a mildly increased LDH concentration. The small gray blocks along the median x-axis and y-axis for the group indicate the CNT length and diameter measures for which experimental data exist.

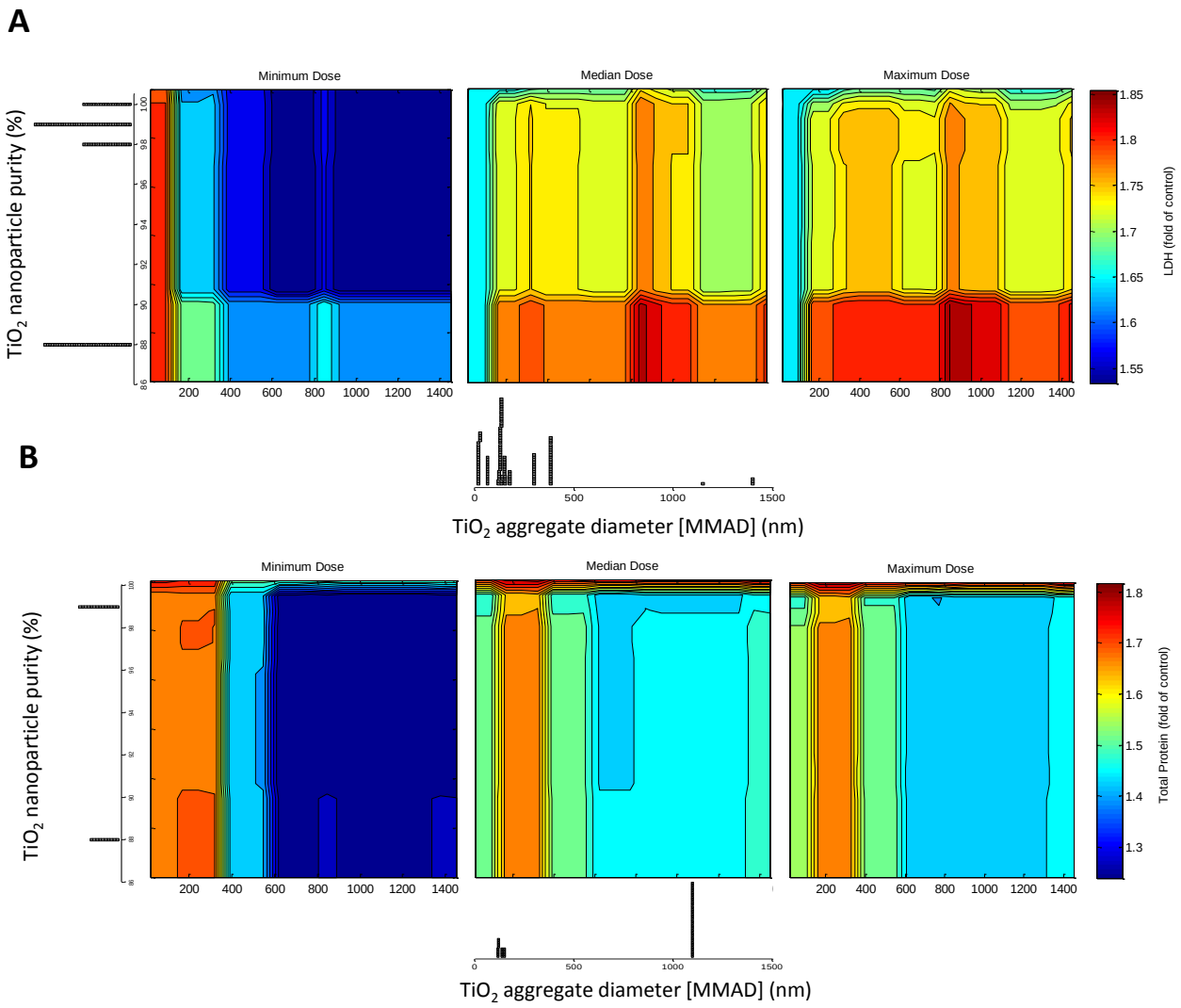

Figure 4-4: Effects of pulmonary exposure to titanium dioxide nanoparticles based on changes in dose, aggregate diameter (MMAD), and purity.

[A] Changes in lactate dehydrogenase (LDH) in BAL fluid [B] Changes in total protein concentration in BAL fluid. Other variables in the model are held constant at their median values. The minimum dose is 35 μg/kg. The median dose is 1.8×10^6 μg/kg. The maximum dose is 3.5×10^6 μg/kg. These results indicate that increasing purity is associated with a mildly decreasing LDH concentration, but has little impact on total protein concentration. The size of particle aggregates appears to have negligible effect for either measure. The small gray blocks along the median x-axis and y-axis for the group indicate the aggregate diameter (MMAD) and purity measures for which experimental data exist.

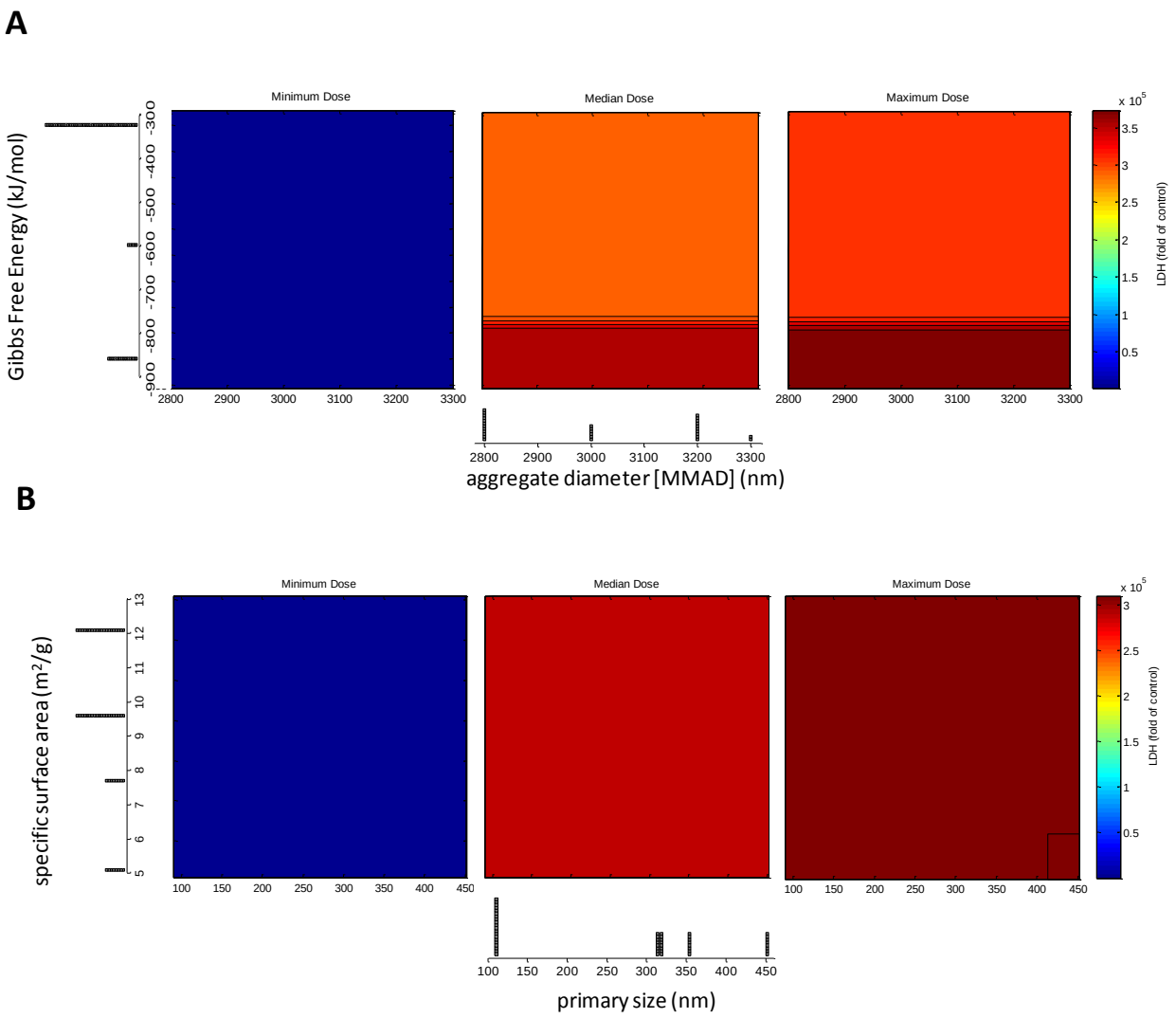

Figure 4-5: Effects of pulmonary exposure to metal oxide nanoparticles including titanium dioxide, zinc oxide, magnesium oxide, and silicon dioxide, based on changes in LDH based on variations in [A] aggregation (MMAD) and the Gibbs free energy, and primary particle size and specific surface area [B].

The minimum dose is 300 μg/kg. The median dose is 8,000 μg/kg. And, the maximum dose is 16,000 μg/kg. These plots indicate that changes in total dose by mass affect the observed toxicity to a much greater degree than any effects from size or chemical factors. The small gray blocks along the median x-axis and y-axis for the group indicate the nanoparticle diameter and specific surface area or Gibbs Free Energy measures for which experimental data exist.

5 Conclusions

This thesis has contributed to the risk assessment of pulmonary exposures to nanomaterials, and quantitative risk modeling methods appropriate for uncertain and incomplete data sets. Although traditional meta-analysis methods do not resolve to usable models for this data set, the machine learning methods employed herein reveal the relative importance of changes in specific nanomaterial characteristics and their effects on toxicity.

Chapter 2 presented a risk modeling methodology using regression tree and random forest machine learning algorithms and analyzed the current state of knowledge regarding the contributing characteristics of carbon nanotubes to pulmonary toxicity in rodents. These models identify important contributors such as carbon nanotube diameter and length, metallic impurity content, and aggregation contrasted against specific surface area and aspect ratio.

Chapter 3 presented an analysis of titanium dioxide nanoparticle characteristics contributing to toxicity. With a more uncertain data set featuring more missing values and more experimental variability, multiple linear regression models were not even possible without imputation of missing values, and even with imputation the linear models fell short of performance demonstrating plausibility. This paper also showed how the random forest method is not sensitive to the effects of multiple imputation.

Chapter 4 described how random forest models could be used to examine the interactions between multiple nanomaterial characteristics and toxicity at the same time allowing material designers to meet function objectives while minimizing toxicity risk.

This thesis explored the use of machine learning modeling methodologies to ascribe the observed toxicity in pulmonary exposure animal studies to specific nanomaterial characteristics. In addition to quantifying existing knowledge on specific questions on the pulmonary toxicity of nanomaterials, this research as made the data sets and models publically available to other investigators for the purposes of facilitating further analysis and improvements in nanomaterial risk assessment.

This research pioneered the use of RT and RF models for meta-analysis and risk assessment in the early stages of technology development. While the particular questions here focused on connecting nanomaterial characteristics to pulmonary toxicity outcomes in rodents, the need for estimating risks while data are incomplete or uncertain is a common occurrence. These data-driven analytical methodologies should in the future complement the more traditional approach of expert elicitation in similar situations.

The analysis in this dissertation investigated only one category of toxic responses measured in bronchoalveolar lavage fluid: inflammation, cell damage, and early stage immune response. The current data are heavily weighted towards short term acute exposure scenarios. Continuing toxicological research on these materials is necessary to more fully explore the consequences of exposure. As more data become available on other toxicological endpoints such as pathologies from chronic exposures, these models should contribute to evaluating our evolving knowledge on those risks and reconciling differences between test methodologies. The modeling methods demonstrated herein can be used to predict pathologies on a categorical or classification basis once sufficient

data become available. It is recommended that experiments exploring the carcinogenicity of these and other nanomaterials vary nanomaterial characteristics to the greatest possible extent so that efforts similar to this one can evaluate the contribution of controllable parameters to the most significant risks.

Organizations such as the Occupational Safety and Health Administration (OSHA) and the Environmental Protection Agency (EPA) may soon be making regulatory decisions regarding exposure and emission limits for nanomaterials. The modeling methods exemplified in this thesis should complement and inform those policy decisions by evaluating whether categories of nanomaterials like carbon nanotubes can be sufficiently addressed by single regulatory limits, or whether other important characteristics of the materials need to be accounted for.

Future research in this area will enable nanomaterial designers and regulators to limit risks while achieving the maximum functional benefit from these unique materials. Demonstrating the use of machine learning models for meta-analysis-based risk assessment should encourage the application of these methods to similar problems, and spur the development of enhanced capabilities to more quickly identify and mitigate risks as they arise.

References

[1] C. O. Robichaud, A. E. Uyar, M. R. Darby, L. G. Zucker, and M. R. Wiesner, "Estimates of upper bounds and trends in nano-TiO2 production as a basis for exposure assessment.," *Environ. Sci. Technol.*, vol. 43, no. 12, pp. 4227–33, Jun. 2009.

[2] N. C. Mueller and B. Nowack, "Exposure modeling of engineered nanoparticles in the environment," *Environ. Sci. Technol*, vol. 42, pp. 4447–4453, 2008.

[3] T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H.-M. Hwang, A. Toropov, D. Leszczynska, and J. Leszczynski, "Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles.," *Nature nanotechnology*, vol. 6, no. 3, pp. 175–8, Mar. 2011.

[4] D. Fourches, D. Pu, C. Tassa, R. Weissleder, S. Y. Shaw, R. J. Mumper, and A. Tropsha, "Quantitative nanostructure-activity relationship modeling.," *ACS nano*, vol. 4, no. 10, pp. 5703–5712, Oct. 2010.

[5] C. M. Sayes, A. a Marchione, K. L. Reed, and D. B. Warheit, "Comparative pulmonary toxicity assessments of C60 water suspensions in rats: few differences in fullerene toxicity in vivo in contrast to in vitro profiles.," *Nano letters*, vol. 7, no. 8, pp. 2399–406, Aug. 2007.

[6] L. Breiman, J. Friedman, C. Stone, and R. A. Olshen, *Classification and Regression Trees*. New York: Chapman and Hall/CRC, 1984, p. 368.

[7] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, May 2001.

[8] B. Kane and R. H. Hurt, "The asbestos analogy revisited," *Nature Nanotechnology*, vol. 3, no. July, pp. 378–379, 2008.

[9] C. Varga and K. Szendi, "Carbon nanotubes induce granulomas but not mesotheliomas.," *In Vivo*, vol. 24, no. 2, pp. 153–156, 2010.

[10] J. Muller, M. Delos, N. Panin, V. Rabolli, F. Huaux, and D. Lison, "Absence of carcinogenic response to multiwall carbon nanotubes in a 2-year bioassay in the peritoneal cavity of the rat.," *Toxicol. Sci.*, vol. 110, no. 2, pp. 442–448, Aug. 2009.

[11] A. Takagi, A. Hirose, T. Nishimura, N. Fukumori, A. Ogata, N. Ohashi, S. Kitajima, and J. Kanno, "Induction of mesothelioma in p53+/- mouse by intraperitoneal

application of multi-wall carbon nanotube.," *J. of Tox. Sci.*, vol. 33, no. 1, pp. 105–116, Feb. 2008.

[12] Y. Sakamoto, D. Nakae, N. Fukumori, K. Tayama, A. Maekawa, K. Imai, A. Hirose, T. Nishimura, N. Ohashi, and A. Ogata, "Induction of mesothelioma by a single intrascrotal administration of multi-wall carbon nanotube in intact Fischer 344 rats," *J. of Tox. Sci.*, vol. 34, pp. 65–76, 2009.

[13] A. F. Hubbs, R. R. Mercer, S. A. Benkovic, J. Harkema, K. Sriram, D. Schwegler-Berry, M. P. Goravanahally, T. R. Nurkiewicz, V. Castranova, and L. M. Sargent, "Nanotoxicology--A Pathologist's Perspective.," *Toxicol. Path.*, vol. 39, no. 2, pp. 301–324, Jan. 2011.

[14] L. M. Sargent, D. W. Porter, D. Lowry, L. Battelli, M. Siegrist, M. L. Kashon, B. T. Chen, D. G. Frazer, L. M. Staska, A. F. Hubbs, W. McKinney, W. Andrew, S. Tsuruoke, M. Endo, V. Castranova, and S. H. Reynolds, "Multi-Walled Carbon Nanotube-induced Lung Tumors," in *Proc. of the Annual Mtg. of the Soc. of Toxicology*, 2013, p. 130:A457.

[15] T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H.-M. Hwang, A. Toropov, D. Leszczynska, and J. Leszczynski, "Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles.," *Nature nanotechnology*, vol. 6, no. February, pp. 175–178, Feb. 2011.

[16] C. Sayes and I. Ivanov, "Comparative study of predictive computational models for nanoparticle-induced cytotoxicity.," *Risk Analysis*, vol. 30, no. 11, pp. 1723–1734, Nov. 2010.

[17] R. Liu, R. Rallo, S. George, Z. Ji, S. Nair, A. Nel, and Y. Cohen, "Classification nanoSAR development for cytotoxicity of metal oxide nanoparticles," *Small*, vol. 7, no. 8, pp. 1118–1126, 2011.

[18] D. Fourches, D. Pu, and A. Tropsha, "Exploring Quantitative Nanostructure-Activity Relationships (QNAR) Modeling as a Tool for Predicting Biological Effects of Manufactured Nanoparticles," *Comb. Chem. High T. Scr.*, vol. 14, pp. 217–225, 2011.

[19] S. Kang, M. S. Mauter, and M. Elimelech, "Physicochemical Determinants of Multiwalled Carbon Nanotube Bacterial Cytotoxicity," *Environ. Sci. Technol.*, vol. 42, no. 19, pp. 7528–7534, Oct. 2008.

[20] S. Kang and M. S. Mauter, "Physicochemical Determinants of Multiwalled Carbon Nanotube Bacterial Cytotoxicity," *Environ. Sci. Technol.*, vol. 42, no. 19, pp. 7528–7534, 2008.

[21] A. Genaidy, T. Tolaymat, R. Sequeira, M. Rinder, and D. Dionysiou, "Health effects of exposure to carbon nanofibers: systematic review, critical appraisal, meta analysis

and research to practice perspectives.," *Science of the Total Environment*, vol. 407, no. 12, pp. 3686–701, Jun. 2009.

[22] U. C. Nygaard, J. S. Hansen, M. Samuelsen, T. Alberg, C. D. Marioara, and M. Løvik, "Single-walled and multi-walled carbon nanotubes promote allergic immune responses in mice.," *Toxicol. Sci.*, vol. 109, no. 1, pp. 113–123, May 2009.

[23] J. Pauluhn, "Subchronic 13-week inhalation exposure of rats to multiwalled carbon nanotubes: toxic effects are determined by density of agglomerate structures, not fibrillar structures," *Toxicol. Sci.*, vol. 113, no. 1, pp. 226–242, 2010.

[24] J. Muller, F. Huaux, N. Moreau, P. Misson, J.-F. Heilier, M. Delos, M. Arras, A. Fonseca, J. B. Nagy, and D. Lison, "Respiratory toxicity of multi-wall carbon nanotubes.," *Tox. App. Pharma.*, vol. 207, no. 3, pp. 221–231, 2005.

[25] D. B. Warheit, T. R. Webb, V. L. Colvin, and C. M. Sayes, "Pulmonary bioassay studies with nanoscale and fine-quartz particles in rats: toxicity is not dependent upon particle size but on surface characteristics," *Toxicol. Sci.*, vol. 95, no. 1, pp. 270–280, 2007.

[26] L. A. Mitchell, J. Gao, R. Vander Wal, A. Gigliotti, S. W. Burchiel, and J. D. McDonald, "Pulmonary and systemic immune response to inhaled multiwalled carbon nanotubes.," *Toxicol. Sci.*, vol. 100, no. 1, pp. 203–214, Nov. 2007.

[27] L. Ma-Hock, S. Treumann, V. Strauss, S. Brill, F. Luizi, M. Mertler, K. Wiench, A. O. Gamer, B. van Ravenzwaay, and R. Landsiedel, "Inhalation toxicity of multiwall carbon nanotubes in rats exposed for 3 months.," *Toxicol. Sci.*, vol. 112, no. 2, pp. 468–481, 2009.

[28] R. Solimeo, J. Zhang, M. Kim, A. Sedykh, and H. Zhu, "Predicting chemical ocular toxicity using a combinatorial QSAR approach.," *Chem. Res. Toxicol.*, vol. 25, no. 12, pp. 2763–9, Dec. 2012.

[29] P. Gong, P.-R. Loh, N. D. Barker, G. Tucker, N. Wang, C. Zhang, B. L. Escalon, B. Berger, and E. J. Perkins, "Building quantitative prediction models for tissue residue of two explosives compounds in earthworms from microarray gene expression data.," *Environ. Sci. Technol.*, vol. 46, no. 1, pp. 19–26, Jan. 2012.

[30] L. Barco, M. Mancin, M. Ruffa, C. Saccardin, C. Minorello, P. Zavagnin, A. A. Lettini, J. E. Olsen, and A. Ricci, "Application of the Random Forest method to analyse epidemiological and phenotypic characteristics of Salmonella 4,[5],12:i:- and Salmonella Typhimurium strains.," *Zoonoses Public Health*, vol. 59, no. 7, pp. 505–12, Nov. 2012.

[31] K. L. Lunetta, L. B. Hayward, J. Segal, and P. Van Eerdewegh, "Screening large-scale association study data: exploiting interactions using random forests.," *BMC genetics*, vol. 5, p. 32, Jan. 2004.

[32] J. Gernand, "Carbon Nanotube (CNT) Pulmonary Toxicity Data Set," 2012. [Online]. Available: http://nanohub.org/resources/13515.

[33] M. S. Hull, A. J. Kennedy, J. a Steevens, A. J. Bednar, C. a Weiss, and P. J. Vikesland, "Release of metal impurities from carbon nanomaterials influences aquatic toxicity.," *Environ. Sci. Technol.*, vol. 43, no. 11, pp. 4169–4174, Jun. 2009.

[34] J. Pauluhn, "Multi-walled carbon nanotubes (Baytubes): approach for derivation of occupational exposure limit.," *Regul. Toxicol. Pharm.*, vol. 57, no. 1, pp. 78–89, Jun. 2010.

[35] G. Oberdörster, E. Oberdörster, and J. Oberdörster, "Nanotoxicology: An Emerging Discipline Evolving from Studies of Ultrafine Particles," *Environ. Health*, vol. 113, no. 7, pp. 823–839, 2005.

[36] A. Helland, P. Wick, A. Koehler, K. Schmid, and C. Som, "Reviewing the environmental and human health knowledge base of carbon nanotubes.," *Environ. Health Perspect.*, vol. 115, no. 8, pp. 1125–1131, 2007.

[37] S. Kang, M. Herzberg, D. F. Rodrigues, and M. Elimelech, "Antibacterial effects of carbon nanotubes: size does matter!," *Langmuir*, vol. 24, no. 13, pp. 6409–6413, Jun. 2008.

[38] S. Kang, M. S. Mauter, and M. Elimelech, "Physicochemical determinants of multiwalled carbon nanotube bacterial cytotoxicity.," *Environ. Sci. Technol.*, vol. 42, no. 19, pp. 7528–7534, Oct. 2008.

[39] P. Wick, P. Manser, L. K. Limbach, U. Dettlaff-Weglikowska, F. Krumeich, S. Roth, W. J. Stark, and A. Bruinink, "The degree and kind of agglomeration affect carbon nanotube cytotoxicity.," *Toxicol. Lett.*, vol. 168, no. 2, pp. 121–131, Jan. 2007.

[40] K. Kostarelos, "The long and short of carbon nanotube toxicity.," *Nature biotechnology*, vol. 26, no. 7, pp. 774–776, Jul. 2008.

[41] K. Donaldson, R. Aitken, L. Tran, V. Stone, R. Duffin, G. Forrest, and A. Alexander, "Carbon nanotubes: a review of their properties in relation to pulmonary toxicology and workplace safety.," *Toxicol. Sci.*, vol. 92, no. 1, pp. 5–22, 2006.

[42] F. Watari, N. Takashi, A. Yokoyama, M. Uo, T. Akasaka, Y. Sato, S. Abe, Y. Totsuka, and K. Tohji, "Material nanosizing effect on living organisms: non-specific, biointeractive, physical size effects.," *Journal of the Royal Society*, vol. 6 Suppl 3, no. April, pp. S371–88, Jun. 2009.

[43] D. Lison, "Human toxicity of cobalt-containing dust and experimental studies on the mechanism of interstitial lung disease (hard metal disease)," *Crit. Rev. Toxicol.*, vol. 26, no. 6, pp. 585–616, 1996.

[44] D. Lison, M. De Boeck, V. Verougstraete, and M. Kirsch-Volders, "Update on the genotoxicity and carcinogenicity of cobalt compounds," *J. Occup. Environ. Med.*, vol. 58, no. 10, pp. 619–625, 2001.

[45] D. B. Warheit, B. R. Laurence, K. L. Reed, D. H. Roach, G. A. M. Reynolds, and T. R. Webb, "Comparative pulmonary toxicity assessment of single-wall carbon nanotubes in rats," *Toxicol. Sci.*, vol. 77, pp. 117–125, 2004.

[46] A. A. Shvedova, E. R. Kisin, R. Mercer, A. R. Murray, V. J. Johnson, A. I. Potapovich, Y. Y. Tyurina, O. Gorelik, S. Arepalli, D. Schwegler-Berry, A. F. Hubbs, J. Antonini, D. E. Evans, B.-K. Ku, D. Ramsey, A. Maynard, V. E. Kagan, V. Castranova, and P. Baron, "Unusual inflammatory and fibrogenic pulmonary responses to single-walled carbon nanotubes in mice.," *Am J. Physiol.-Lung C.*, vol. 289, no. 5, pp. 698–708, 2005.

[47] A. A. Shvedova, E. R. Kisin, A. R. Murray, O. Gorelik, S. Arepalli, V. Castranova, S. Young, F. Gao, Y. Y. Tyurina, T. D. Oury, and V. E. Kagan, "Vitamin E deficiency enhances pulmonary inflammatory response and oxidative stress induced by single walled carbon nanotubes in C57BL/6 mice," *Toxicol. Appl. Pharmacol.*, vol. 221, no. 3, pp. 339–348, 2007.

[48] A. A. Shvedova, E. Kisin, A. R. Murray, V. J. Johnson, O. Gorelik, S. Arepalli, A. F. Hubbs, R. R. Mercer, P. Keohavong, N. Sussman, J. Jin, J. Yin, S. Stone, B. T. Chen, G. Deye, A. Maynard, V. Castranova, P. A. Baron, and V. E. Kagan, "Inhalation vs. aspiration of single-walled carbon nanotubes in C57BL/6 mice: inflammation, fibrosis, oxidative stress, and mutagenesis.," *Am J. Physiol.-Lung C.*, vol. 295, no. 4, pp. 552–565, 2008.

[49] J. Muller, F. Huaux, A. Fonseca, J. B. Nagy, N. Moreau, M. Delos, E. Raymundo-Pinero, F. Beguin, M. Kirsch-Volders, I. Fenoglio, B. Fubini, and D. Lison, "Structural Defects Play a Major Role in the Acute Lung Toxicity of Multiwall Carbon Nanotubes: Toxicological Aspects," *Chem. Res. Toxicol.*, vol. 21, pp. 1698–1705, 2008.

[50] D. Elgrabli, S. Abella-Gallart, F. Robidel, R. Rogerieux, J. Boczkowski, and G. Lacroix, "Induction of apoptosis and absence of inflammation in rat lung after intratracheal instillation of multiwalled carbon nanotubes," *Toxicology*, vol. 253, pp. 131–136, 2008.

[51] R. R. Mercer, J. Scabilloni, L. Wang, E. Kisin, a R. Murray, D. Schwegler-Berry, a a Shvedova, and V. Castranova, "Alteration of deposition pattern and pulmonary response as a result of improved dispersion of aspirated single-walled carbon nanotubes in a mouse model.," *Am J. Physiol.-Lung C.*, vol. 294, no. 1, pp. L87–97, Jan. 2008.

[52] K.-I. Inoue, H. Takano, E. Koike, R. Yanagisawa, M. Sakurai, S. Tasaka, A. Ishizaka, and A. Shimada, "Effects of pulmonary exposure to carbon nanotubes on lung and

systemic inflammation with coagulatory disturbance induced by lipopolysaccharide in mice.," *Exp. Biol. Med.*, vol. 233, no. 12, pp. 1583–1590, Dec. 2008.

[53] J. P. Ryman-Rasmussen, E. W. Tewksbury, O. R. Moss, M. F. Cesta, B. a Wong, and J. C. Bonner, "Inhaled multiwalled carbon nanotubes potentiate airway fibrosis in murine allergic asthma.," *Am. J. Respir. Cell Mol. Biol.*, vol. 40, no. 3, pp. 349–358, Mar. 2009.

[54] H. Ellinger-Ziegelbauer and J. Pauluhn, "Pulmonary toxicity of multi-walled carbon nanotubes (Baytubes) relative to alpha-quartz following a single 6h inhalation exposure of rats and a 3 months post-exposure period.," *Toxicology*, vol. 266, no. 1–3, pp. 16–29, Dec. 2009.

[55] D. W. Porter, A. F. Hubbs, R. R. Mercer, N. Wu, M. G. Wolfarth, K. Sriram, S. Leonard, L. Battelli, D. Schwegler-Berry, S. Friend, M. Andrew, B. T. Chen, S. Tsuruoka, M. Endo, and V. Castranova, "Mouse pulmonary dose- and time course-responses induced by exposure to multi-walled carbon nanotubes," *Toxicology*, vol. 269, pp. 136–147, 2010.

[56] E.-J. Park, J. Roh, S.-N. Kim, M.-S. Kang, Y.-A. Han, Y. Kim, J. T. Hong, and K. Choi, "A single intratracheal instillation of single-walled carbon nanotubes induced early lung fibrosis and subchronic tissue damage in mice.," *Archives of toxicology*, vol. 85, no. 9, pp. 1121–31, Sep. 2011.

[57] J. Teeguarden, B. Webb-Robertson, K. Waters, A. Murray, E. Kisin, S. Varnum, J. Jacobs, J. Pounds, R. Zanger, and A. Shvedova, "Comparative proteomics and pulmonary toxicity of instilled single-walled carbon nanotubes, crocidolite adbestos, and ultrafine carbon black in mice," *Toxicol. Sci.*, vol. 120, no. 1, pp. 123–135, 2011.

[58] C. Xue, J. Wu, F. Lan, W. Liu, X. Yang, F. Zeng, and H. Xu, "Nano Titanium Dioxide Induces the Generation of ROS and Potential Damage in HaCaT Cells Under UVA Irradiation," *Journal of Nanoscience and Nanotechnology*, vol. 10, no. 12, p. 8, 2010.

[59] J.-R. Gurr, A. S. S. Wang, C.-H. Chen, and K.-Y. Jan, "Ultrafine titanium dioxide particles in the absence of photoactivation can induce oxidative damage to human bronchial epithelial cells.," *Toxicology*, vol. 213, no. 1–2, pp. 66–73, Sep. 2005.

[60] N. Lu, Z. Zhu, X. Zhao, R. Tao, X. Yang, and Z. Gao, "Nano titanium dioxide photocatalytic protein tyrosine nitration: a potential hazard of TiO2 on skin.," *Biochemical and biophysical research communications*, vol. 370, no. 4, pp. 675–80, Jun. 2008.

[61] H.-W. Chen, S.-F. Su, C.-T. Chien, W.-H. Lin, S.-L. Yu, C.-C. Chou, J. J. W. Chen, and P.-C. Yang, "Titanium dioxide nanoparticles induce emphysema-like lung injury in mice.," *The FASEB journal*, vol. 20, no. 13, pp. 2393–2395, Nov. 2006.

[62] D. B. Warheit, T. R. Webb, C. M. Sayes, V. L. Colvin, and K. L. Reed, "Pulmonary instillation studies with nanoscale TiO2 rods and dots in rats: toxicity is not dependent upon particle size and surface area.," *Toxicol. Sci.*, vol. 91, no. 1, pp. 227–236, May 2006.

[63] D. B. Warheit, T. R. Webb, K. L. Reed, S. Frerichs, and C. M. Sayes, "Pulmonary toxicity study in rats with three forms of ultrafine-TiO2 particles: differential responses related to surface properties," *Toxicology*, vol. 230, pp. 90–104, 2007.

[64] *Limits for Air Contaminants. Code of Federal Regulations. Title 29. 1910.1000*. United States, p. 1910.1000.

[65] NIOSH, "Current Intelligence Bulletin 63: Occupational Exposure to Titanium Dioxide," DHHS (NIOSH) Publication No. 2011-160, 2011.

[66] C. Sayes and I. Ivanov, "Comparative study of predictive computational models for nanoparticle-induced cytotoxicity.," *Risk analysis : an official publication of the Society for Risk Analysis*, vol. 30, no. 11, pp. 1723–34, Nov. 2010.

[67] J. Gernand and E. Casman, "A meta-analysis of carbon nanotube pulmonary toxicity studies - How physical dimensions and impurities afffect the toxicity of carbon nanotubes," *Risk Analysis*, vol. XX, no. XX, pp. XXX–XXX, 2013.

[68] L. V Stebounova, A. Adamcakova-Dodd, J. S. Kim, H. Park, P. T. O'Shaughnessy, V. H. Grassian, and P. S. Thorne, "Nanosilver induces minimal lung toxicity or inflammation in a subacute murine inhalation model.," *Particle and fibre toxicology*, vol. 8, no. 1, p. 5, Jan. 2011.

[69] C. Sayes and I. Ivanov, "Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity," vol. 30, no. 11, 2010.

[70] L. K. Braydich-Stolle, N. M. Schaeublin, R. C. Murdock, J. Jiang, P. Biswas, J. J. Schlager, and S. M. Hussain, "Crystal structure mediates mode of cell death in TiO2 nanotoxicity," *Journal of Nanoparticle Research*, vol. 11, no. 6, pp. 1361–1374, Oct. 2008.

[71] G. Oberdorster, J. Ferin, R. Gelein, S. C. Soderholm, and J. Finkelstein, "Role of the Alveolar Macrophage in Lung In jury : Studies with Ultrafine Particles," *Environ. Health Perspect.*, vol. 97, pp. 193–199, 1992.

[72] M. Osier and R. et al. Baggs, "Intratracheal instillation versus intratracheal inhalation: influence of cytokines on inflammatory response," *Environ. Health Perspect.*, vol. 105, no. S5, pp. 1265–1271, 1997.

[73] E. Bermudez, J. B. Mangum, B. Asgharian, B. a Wong, E. E. Reverdy, D. B. Janszen, P. M. Hext, D. B. Warheit, and J. I. Everitt, "Long-term pulmonary responses of three laboratory rodent species to subchronic inhalation of pigmentary titanium dioxide particles.," *Toxicol. Sci.*, vol. 70, no. 1, pp. 86–97, Nov. 2002.

[74] B. Rehn and F. et al Selier, "Investigations on the inflammatory and genotoxic lung effects of two types of titanium dioxide: untreated and surface treated," *Tox. App. Pharma.*, vol. 189, no. 2, pp. 84–95, 2003.

[75] L. C. Renwick, D. Brown, A. Clouter, and K. Donaldson, "Increased inflammation and altered macrophage chemotactic responses caused by two ultrafine particle types," *Occupational and Environmental Medicine*, vol. 61, no. 5, pp. 442–447, May 2004.

[76] V. H. Grassian, P. T. O'shaughnessy, A. Adamcakova-Dodd, J. M. Pettibone, and P. S. Thorne, "Inhalation exposure study of titanium dioxide nanoparticles with a primary particle size of 2 to 5 nm.," *Environmental health perspectives*, vol. 115, no. 3, pp. 397–402, Mar. 2007.

[77] A. Nemmar, K. et al. Melghit, and B. H. Ali, "The acute proinflammatory and prothrombotic effects of pulmonary exposure to rutile TiO2 nanorods in rats," *Exp. Biol. Med.*, vol. 233, no. 5, pp. 610–619, May 2008.

[78] N. Kobayashi, M. Naya, S. Endoh, J. Maru, K. Yamamoto, and J. Nakanishi, "Comparative pulmonary toxicity study of nano-TiO2 particles of different sizes and agglomerations in rats: different short- and long-term post-instillation results," *Toxicology*, vol. 264, pp. 110–118, 2009.

[79] D. B. Warheit, C. M. Sayes, S. R. Frame, and K. L. Reed, "Pulmonary exposures to Sepiolite nanoclay particulates in rats: resolution following multinucleate giant cell formation.," *Toxicology letters*, vol. 192, no. 3, pp. 286–93, Feb. 2010.

[80] M. Auffan, J. Rose, J.-Y. Bottero, G. V Lowry, J.-P. Jolivet, and M. R. Wiesner, "Towards a definition of inorganic nanoparticles from an environmental, health and safety perspective.," *Nature nanotechnology*, vol. 4, no. 10, pp. 634–41, 2009.

[81] NIOSH, "NIOSH Current Intelligence Bulletin 65: Occupational Exposure to Carbon Nanotubes and Nanofibers," DHHS (NIOSH) Publication No. 2013-145, 2013.

[82] D. B. Warheit, C. M. Sayes, and K. L. Reed, "Nanoscale and fine zinc oxide particles: can in vitro assays accurately forecast lung hazards following inhalation exposures?," *Environ. Sci. Technol.*, vol. 43, no. 20, pp. 7939–45, Oct. 2009.

[83] I. Gosens, J. A. Post, L. J. J. de la Fonteyne, E. H. J. M. Jansen, J. W. Geus, F. R. Cassee, and W. H. de Jong, "Impact of agglomeration state of nano- and submicron sized gold particles on pulmonary inflammation.," *Particle and fibre toxicology*, vol. 7, no. 1, p. 37, Jan. 2010.

[84] L. E. Yu, L.-Y. L. Yung, C. Ong, Y.-L. Tan, K. S. Balasubramaniam, D. Hartono, G. Shui, M. R. Wenk, and W. Ong, "Translocation and effects of gold nanoparticles after inhalation exposure in rats," *Nanotoxicology*, vol. 1, no. 1–4, pp. 234–241, 2007.

[85] C. Sayes and I. Ivanov, "Comparative study of predictive computational models for nanoparticle-induced cytotoxicity.," *Risk Analysis*, vol. 30, no. 11, pp. 1723–1734, Nov. 2010.

[86] A. A. Shvedova, E. Kisin, A. R. Murray, V. J. Johnson, O. Gorelik, S. Arepalli, A. F. Hubbs, R. R. Mercer, P. Keohavong, N. Sussman, J. Jin, J. Yin, S. Stone, B. T. Chen, G. Deye, A. Maynard, V. Castranova, P. A. Baron, and V. E. Kagan, "Inhalation vs. aspiration of single-walled carbon nanotubes in C57BL/6 mice: inflammation, fibrosis, oxidative stress, and mutagenesis.," *Am J. Physiol.-Lung C.*, vol. 295, no. 4, pp. 552–565, 2008.

[87] K.-I. Inoue, H. Takano, E. Koike, R. Yanagisawa, M. Sakurai, S. Tasaka, A. Ishizaka, and A. Shimada, "Effects of pulmonary exposure to carbon nanotubes on lung and systemic inflammation with coagulatory disturbance induced by lipopolysaccharide in mice.," *Exp. Biol. Med.*, vol. 233, no. 12, pp. 1583–1590, Dec. 2008.

[88] H. Ellinger-Ziegelbauer and J. Pauluhn, "Pulmonary toxicity of multi-walled carbon nanotubes (Baytubes) relative to a-quartz following a single 6 h inhalation exposure of rats and a 3 months post-exposure period," *Toxicology*, vol. 266, pp. 16–29, 2009.

[89] E. Bermudez, J. B. Mangum, B. Asgharian, B. a Wong, E. E. Reverdy, D. B. Janszen, P. M. Hext, D. B. Warheit, and J. I. Everitt, "Long-term pulmonary responses of three laboratory rodent species to subchronic inhalation of pigmentary titanium dioxide particles.," *Toxicol. Sci.*, vol. 70, no. 1, pp. 86–97, Nov. 2002.

[90] G. Oberdorster, J. Ferin, and B. E. Lehnert, "Correlation between Particle Size, In Vivo Particle Persistence, and Lung Injury," *Environmental Health Perspectives*, vol. 102, p. 173, Oct. 1994.

[91] D. B. Warheit, T. R. Webb, C. M. Sayes, V. L. Colvin, and K. L. Reed, "Pulmonary instillation studies with nanoscale TiO2 rods and dots in rats: toxicity is not dependent upon particle size and surface area.," *Toxicol Sci*, vol. 91, no. 1, pp. 227–236, May 2006.

[92] D. B. Warheit, R. A. Hoke, C. Finlay, E. M. Donner, K. L. Reed, and C. M. Sayes, "Development of a base set of toxicity tests using ultrafine TiO2 particles as a component of nanoparticle risk management," *Toxicology Letters*, vol. 171, pp. 99–110, 2007.

[93] D. B. Warheit, C. M. Sayes, S. R. Frame, and K. L. Reed, "Pulmonary exposures to Sepiolite nanoclay particulates in rats: resolution following multinucleate giant cell formation.," *Toxicology letters*, vol. 192, no. 3, pp. 286–93, Feb. 2010.

[94] C. M. Sayes, K. L. Reed, and D. B. Warheit, "Assessing toxicity of fine and nanoparticles: comparing in vitro measurements to in vivo pulmonary toxicity profiles.," *Toxicol. Sci.*, vol. 97, no. 1, pp. 163–80, May 2007.

Appendix A: Carbon Nanotube Regression Tree and Random Forest Model Details

1 Summary

This document contains model learning statistics, and structure of the models utilized in the paper "A meta-analysis of carbon nanotube pulmonary toxicity studies – How physical dimensions and impurities affect the toxicity of carbon nanotubes." This information is meant to supplement and support the explanations and conclusions reached in that paper.

This document includes the detailed structure of the pruned regression tree models (Figures A-1 through A-3) as well as the tree's error performance as a function of model growth (Figures A-4 through A-7). The random forest model performance versus model growth for each of the 4 output measures is also included (Figures A-8 through A-11).

The stepwise random forest models and their performance as a function of included variables are displayed in Figures A-12 through A-15. The random forest generated dose-response profiles and the effects of cobalt impurities are shown in Section 6 (Figures A-16 through A-18).

The MATLAB™ code that creates these regression tree and random forest model objects including the stepwise random forest model object is included in Section 7. The data used to train the models can be found at https://nanohub.org/resources/13515.

2 *Pruned Regression Tree Models*

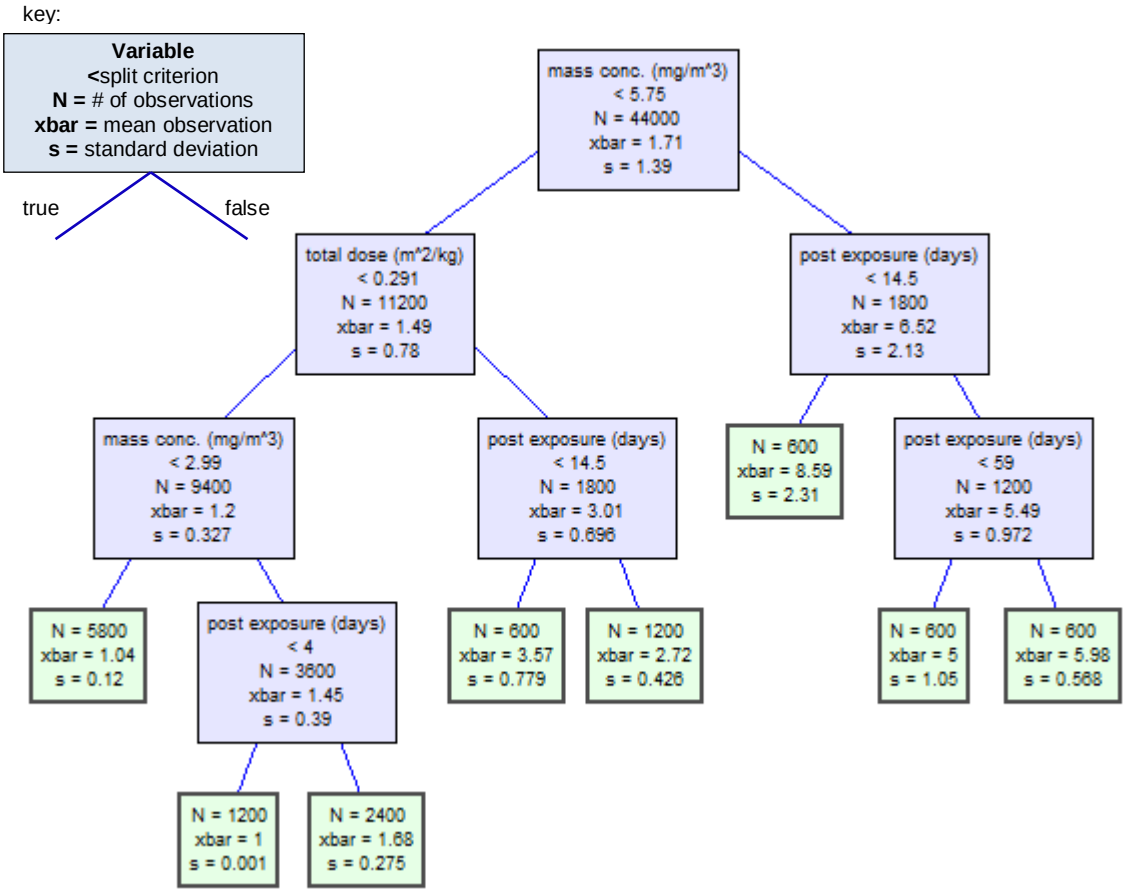

Figure A-1: RT model for BAL macrophages as measured by fold of control. Each branch divides the population of observations into two child populations based on an inequality in one variable. The mean values in the leaf nodes (terminal nodes) are the model's predictions. Characteristics about the BAL macrophage values including number of observations (N), mean (xbar), and standard deviation (s) are provided at each leaf and branch.

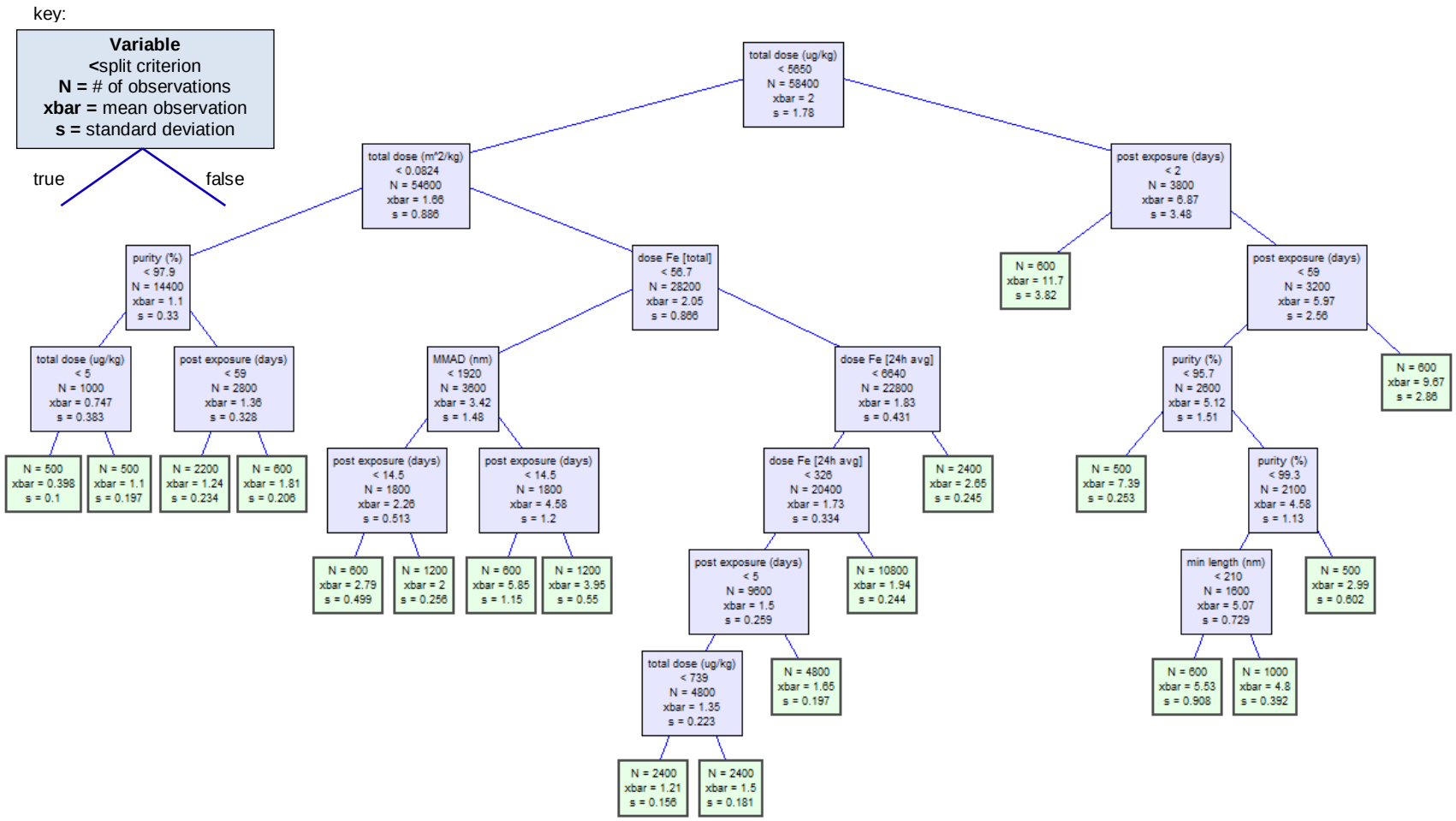

Figure A-2: RT model for BAL Lactate Dehydrogenase (LDH). Each branch divides the population of observations into two child populations based on an inequality in one variable. The mean values in the leaf nodes (terminal nodes) are the model's predictions. Characteristics about the BAL LDH values including number of observations (N), mean (xbar), and standard deviation (s) are provided at each leaf and branch.

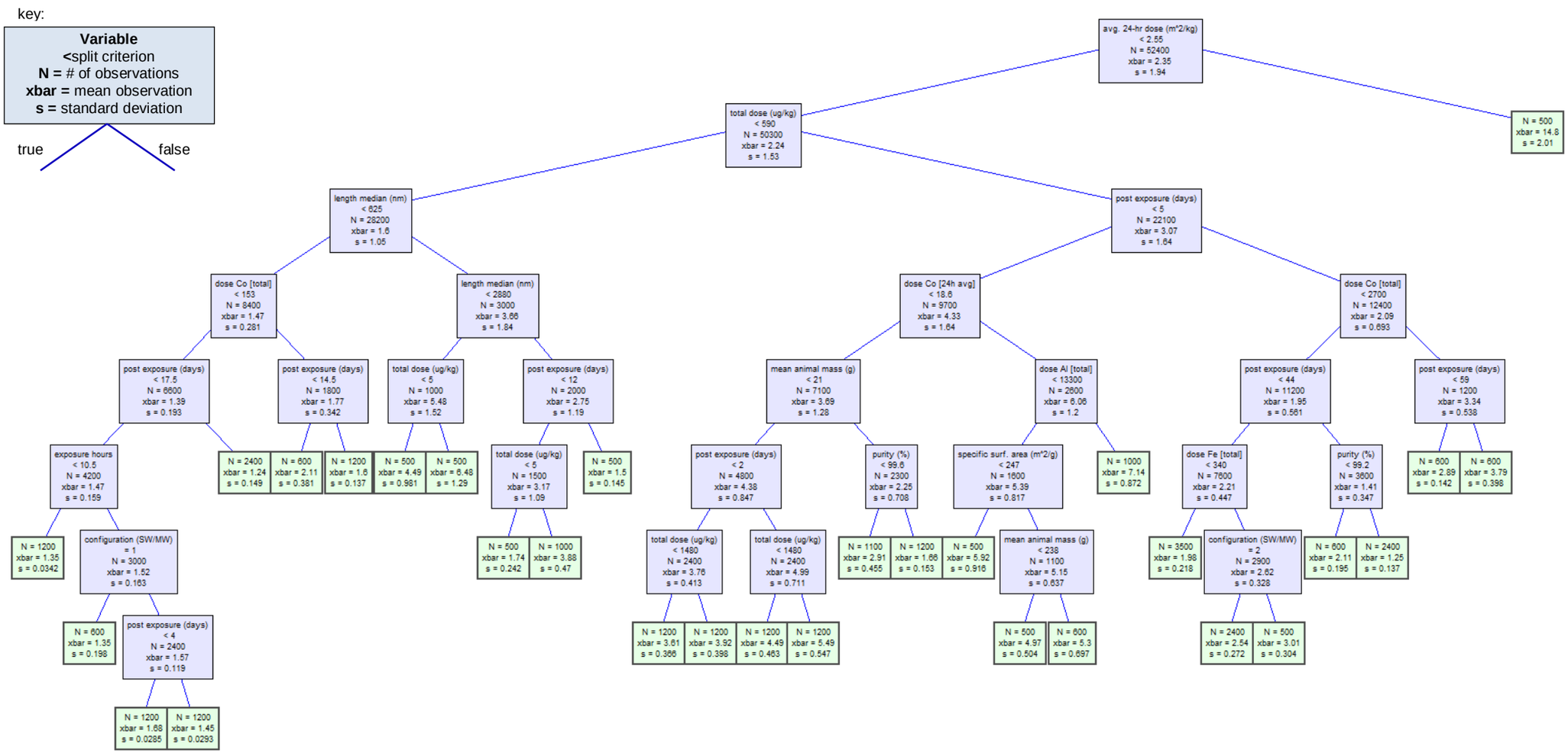

Figure A-3: RT model for BAL Total Protein. Each branch divides the population of observations into two child populations based on an inequality in one variable. The mean values in the leaf nodes (terminal nodes) are the model's predictions. Characteristics about the BAL total protein values including number of observations (N), mean (xbar), and standard deviation (s) are provided at each leaf and branch.

3 Regression Tree Variable Importance

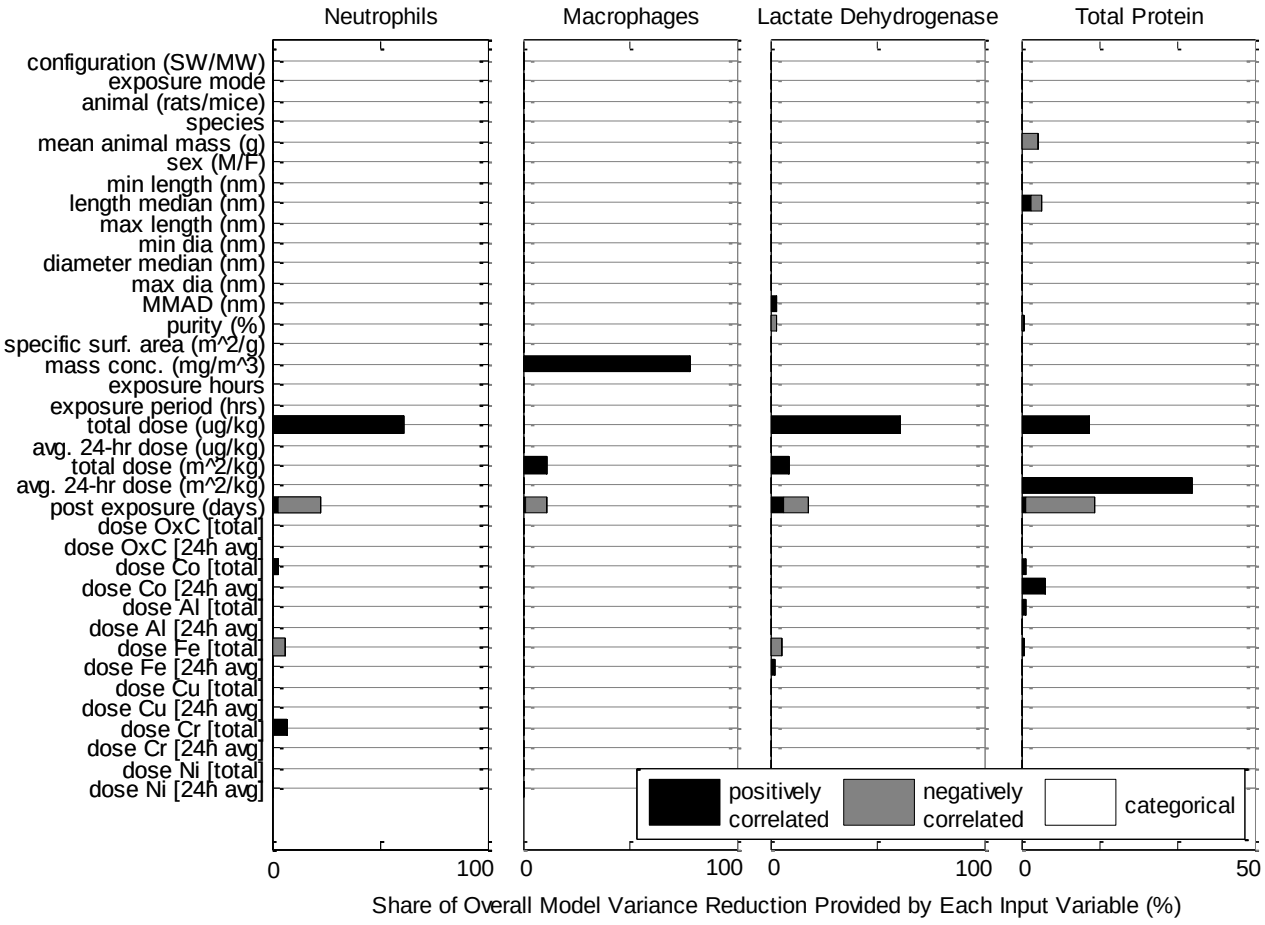

Figure A-4: Variable importance as determined by the relative variance reduction for each of the 4 regression tree (RT) models. Bar shading indicates whether the variance reduction occurred with a positive or negative correlation between the input and output, or with a branch based on a non-numeric categorical variable.

4 Regression Tree Model Growth Curves

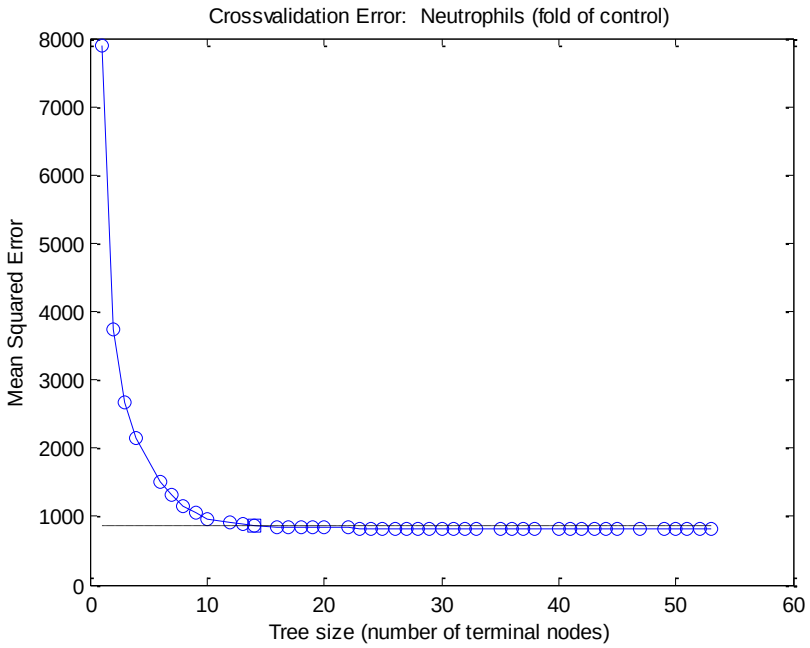

Figure A-5: The regression tree model cross-validation error for BAL neutrophils. The dashed line indicates the level of one standard error above the minimum potential error. Based on these results the regression tree model is pruned to 12 branches.

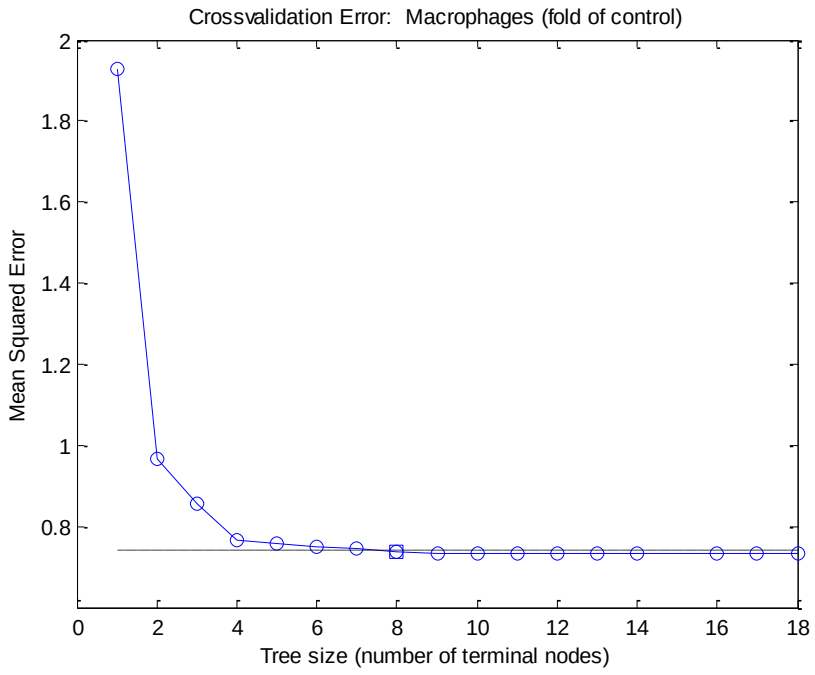

Figure A-6: The regression tree model cross-validation error for BAL macrophages. The dashed line indicates the level of one standard error above the minimum potential error. Based on these results the regression tree model is pruned to 8 branches.

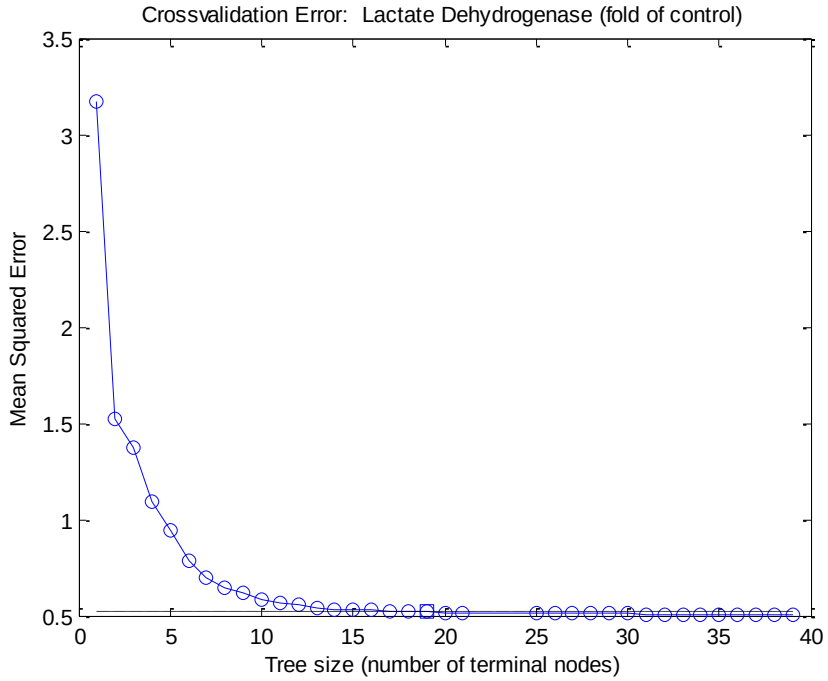

Figure A-7: The regression tree model cross-validation error for BAL lactate dehydrogenase. The dashed line indicates the level of one standard error above the minimum potential error. Based on these results the regression tree model is pruned to 19 branches.

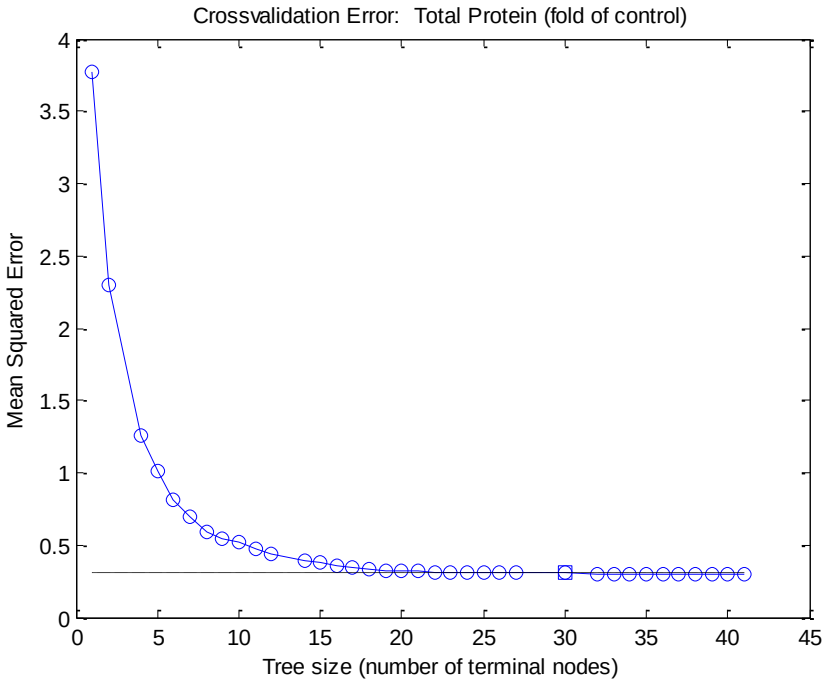

Figure A-8: The regression tree model cross-validation error for BAL total protein. The dashed line indicates the level of one standard error above the minimum potential error. Based on these results the regression tree model is pruned to 30 branches.

5 Random Forest Model Growth Curves

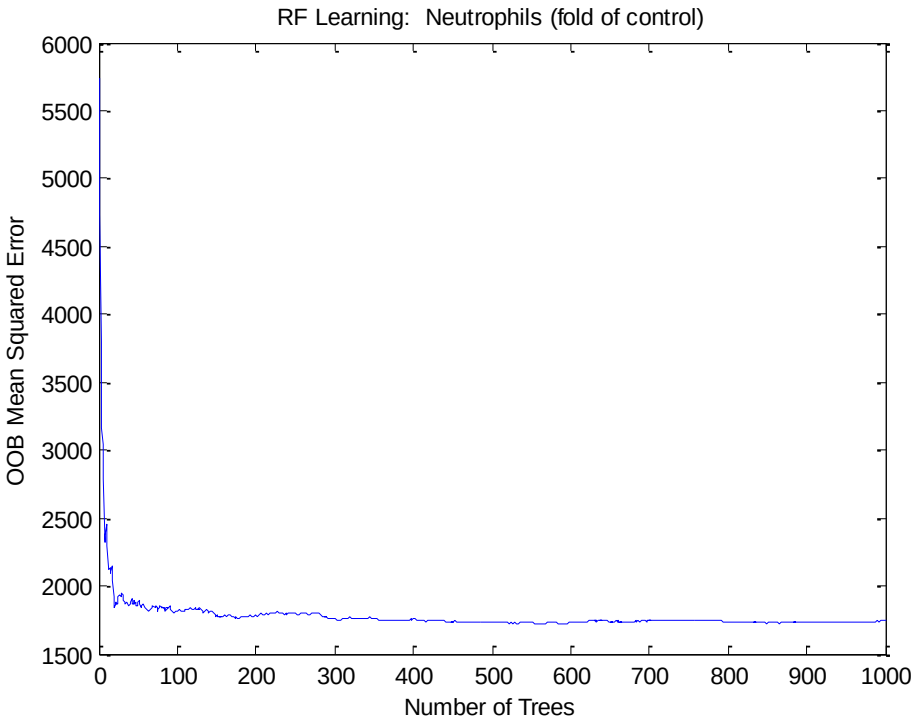

Figure A-9: The random forest model out-of-bag (out-of-bag describes data samples withheld from the training of individual tree models making up the forest) mean squared error as a function of tree models included in forest for BAL neutrophils. These results reflect that nearly all potential information gain has been achieved prior to 1000 trees being included.

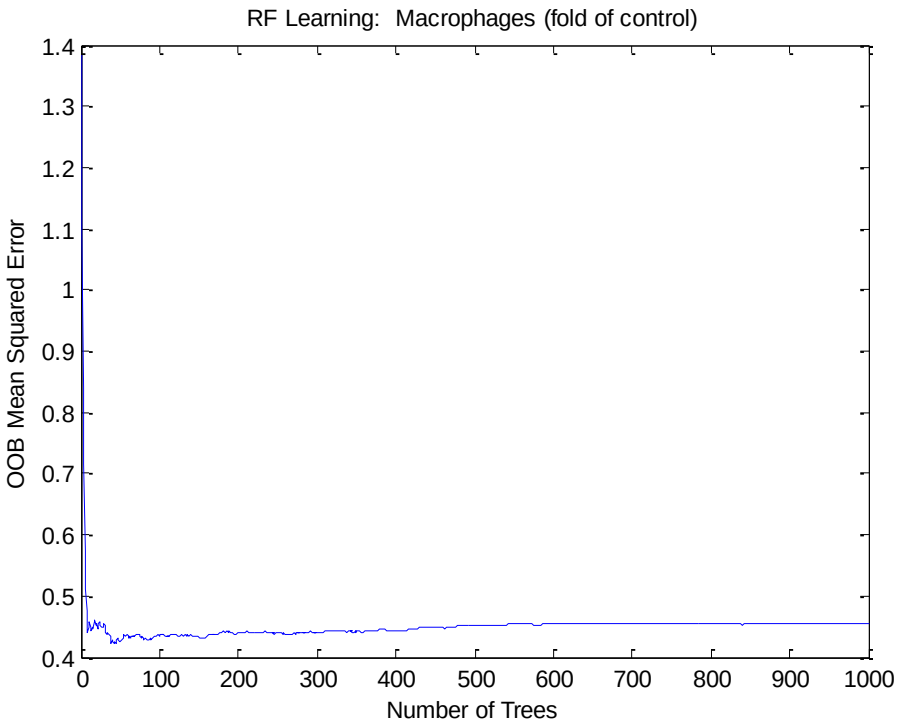

Figure A-10: The random forest model out-of-bag (out-of-bag describes data samples withheld from the training of individual tree models making up the forest) mean squared error as a function of tree models included in forest for BAL macrophages. These results reflect that nearly all potential information gain has been achieved prior to 1000 trees being included.

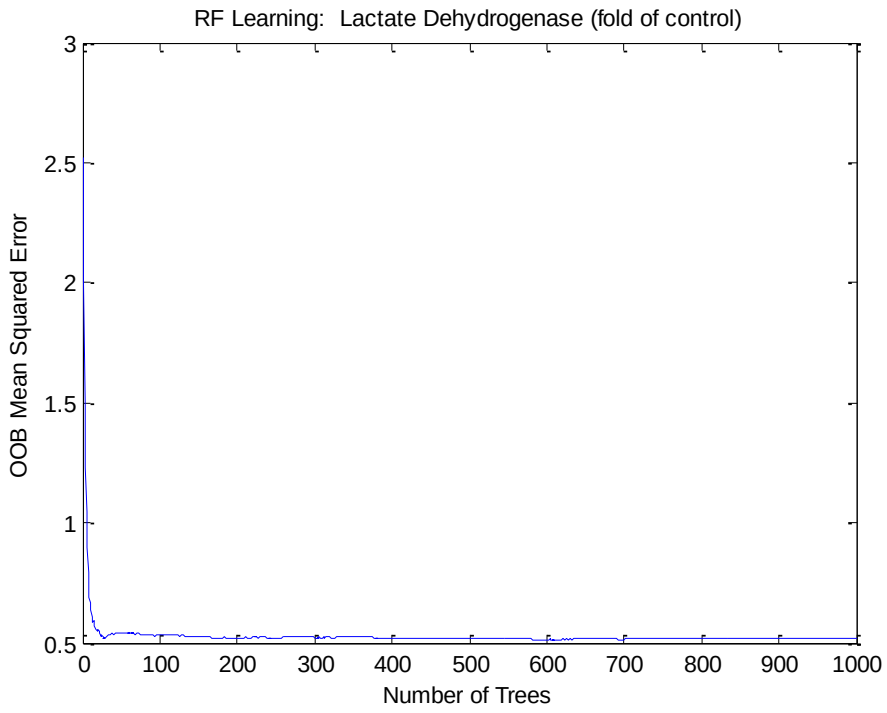

Figure A-11: The random forest model out-of-bag (out-of-bag describes data samples withheld from the training of individual tree models making up the forest) mean squared error as a function of tree models included in forest for BAL lactate dehydrogenase. These results reflect that nearly all potential information gain has been achieved prior to 1000 trees being included.

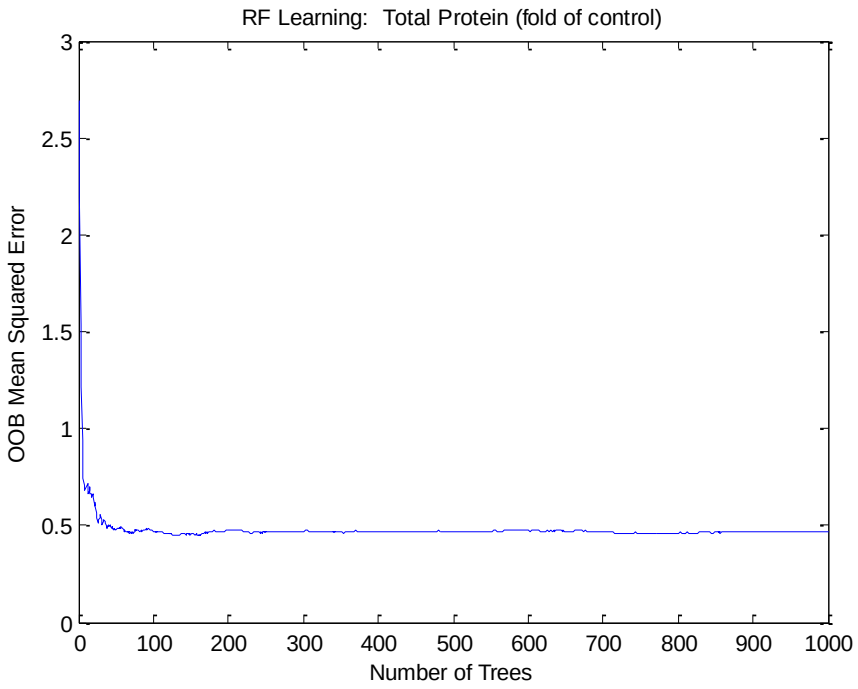

Figure A-12: The random forest model out-of-bag (out-of-bag describes data samples withheld from the training of individual tree models making up the forest) mean squared error as a function of tree models included in forest for BAL total protein. These results reflect that nearly all potential information gain has been achieved prior to 1000 trees being included.

6 Stepwise Random Forest Model Growth Curves

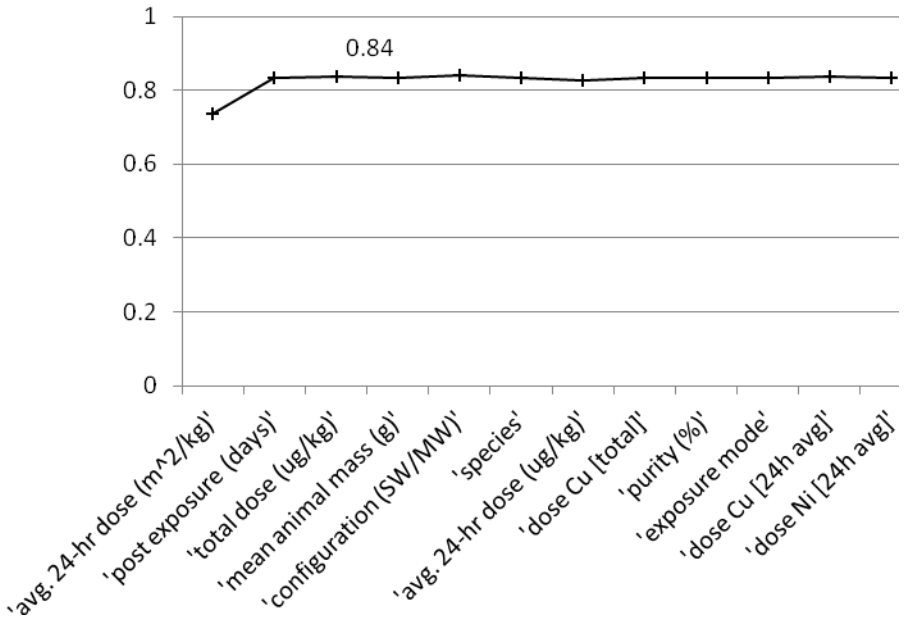

Figure A-13: Stepwise RF model performance by added variable for BAL macrophages. The maximum performance of 0.84 is reached at the addition of the 3rd model parameter, total dose.

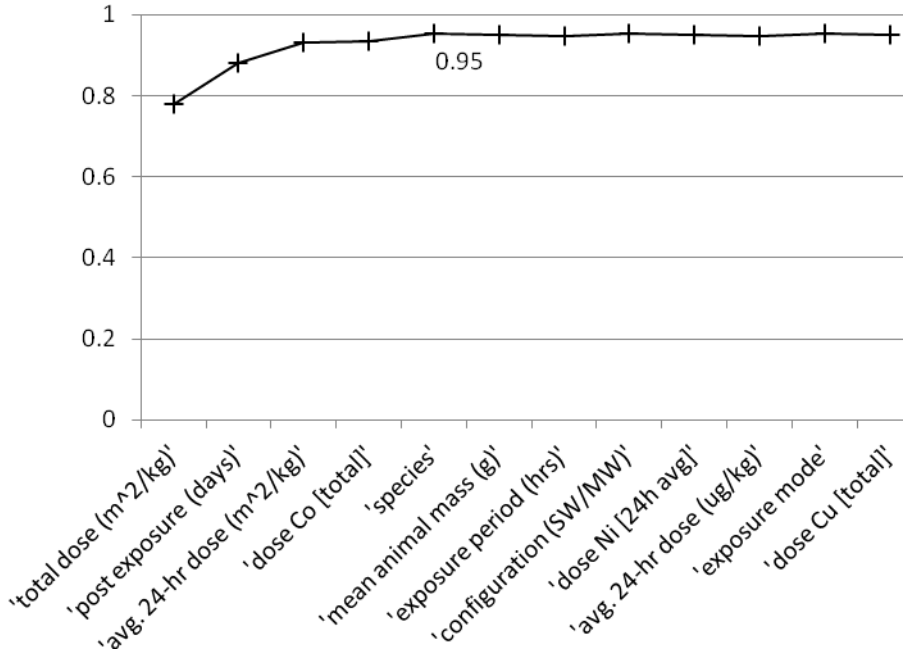

Figure A-14 Stepwise RF model performance by added variable for BAL Total Protein. The maximum performance of 0.95 is reached at the addition of the 5th model parameter, species.

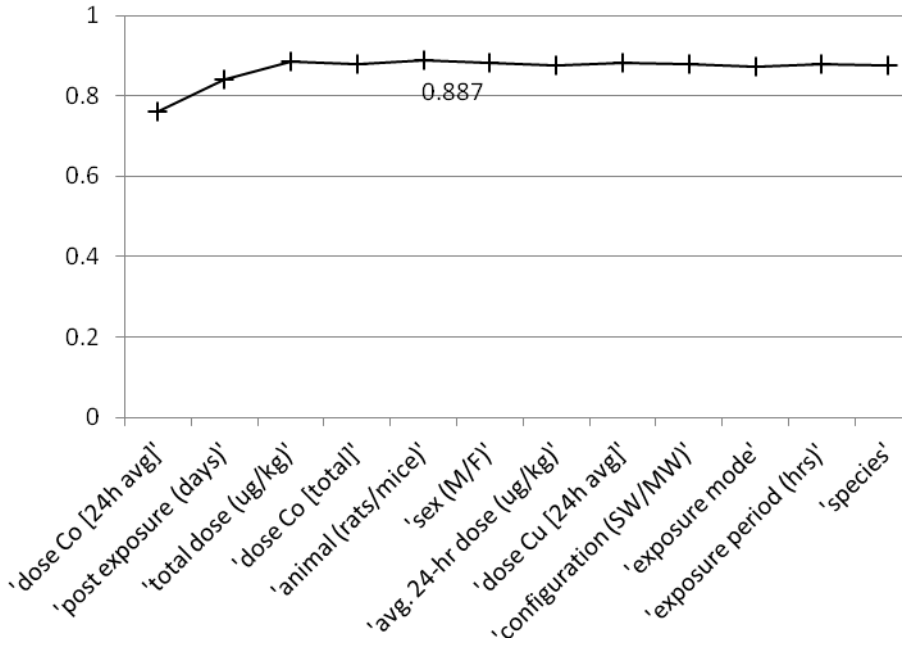

Figure A-15: Stepwise RF model performance by added variable for BAL Lactate Dehydrogenase (LDH). The maximum performance of 0.89 is reached at the addition of the 5th model parameter, animal type.

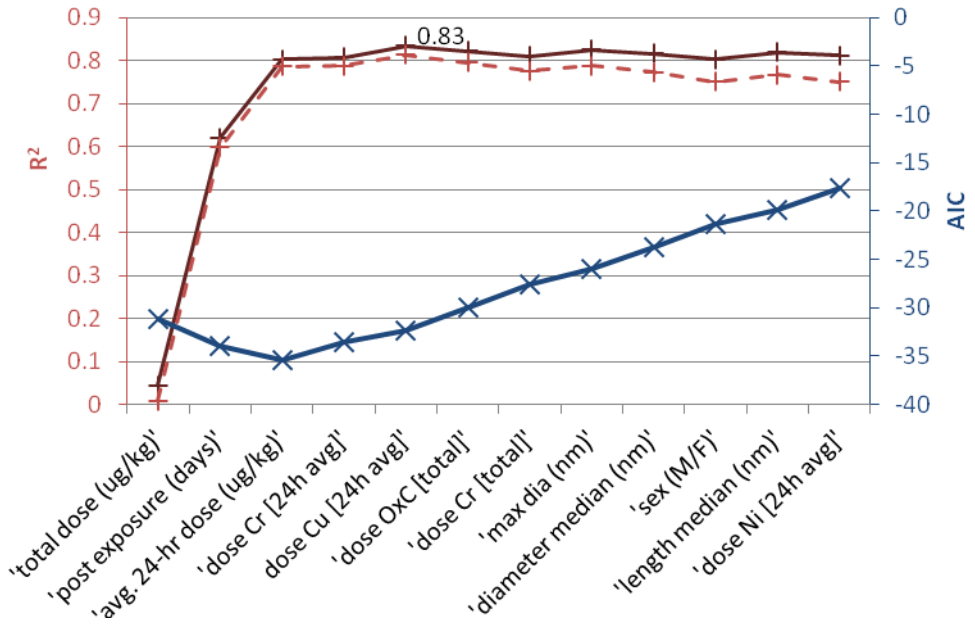

Figure A-16: Stepwise RF model performance by added variable for BAL neutrophils. The maximum performance of 0.83 is reached at the addition of the 5th model parameter, 24 hour average dose of copper. The Akaike Information Criterion (AIC) is plotted on the right axis.

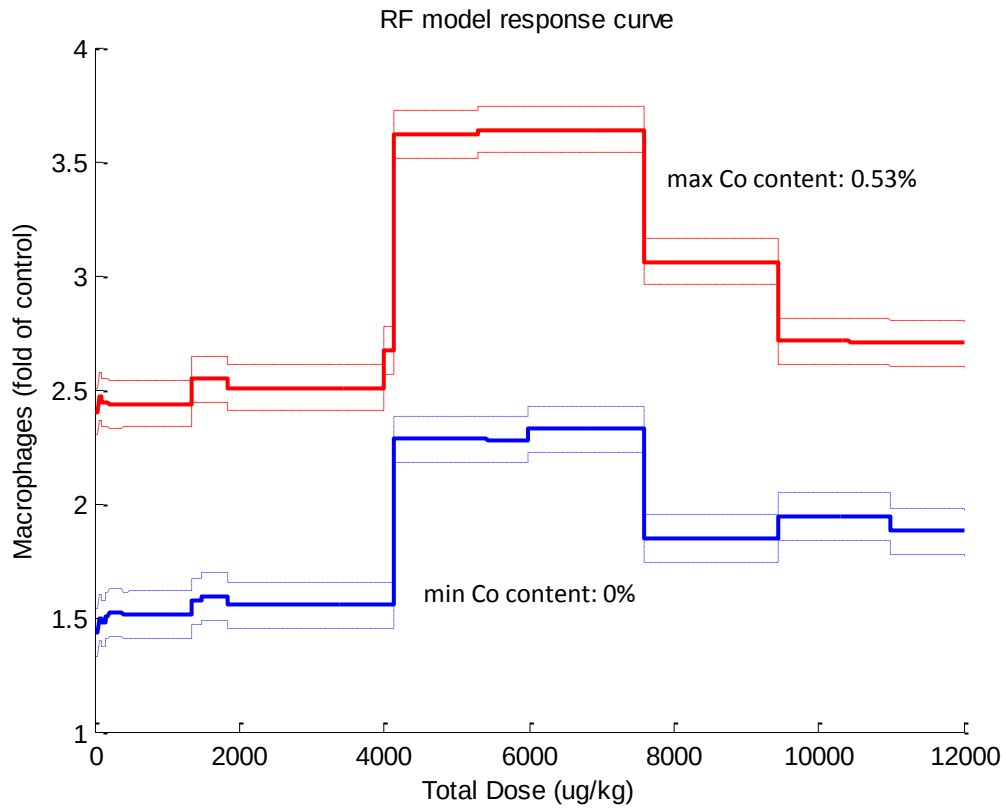

Figure A-17: RF model dose-response response curve for macrophages as modified by cobalt content. Dashed and dotted lines indicate the RT model predicted experimental standard deviation. All other model inputs are held constant at their median value.

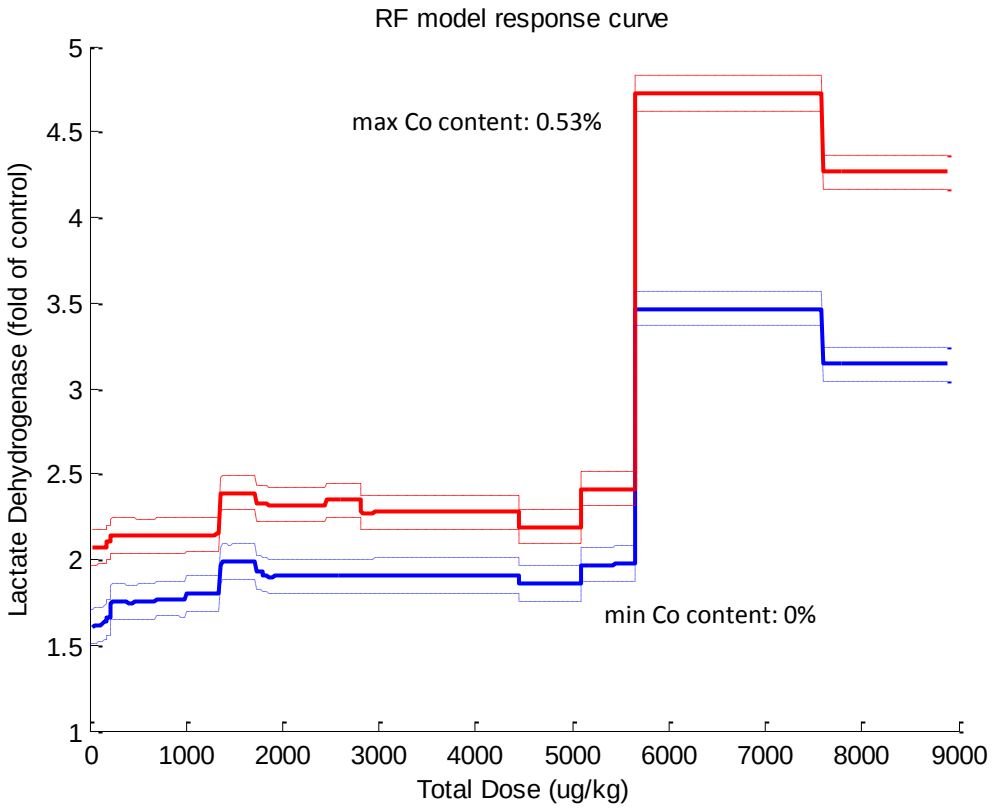

Figure A-18: RF model dose-response response curve for lactate dehydrogenase as modified by cobalt content. Dashed and dotted lines indicate the RT model predicted experimental standard deviation. All other model inputs are held constant at their median value.

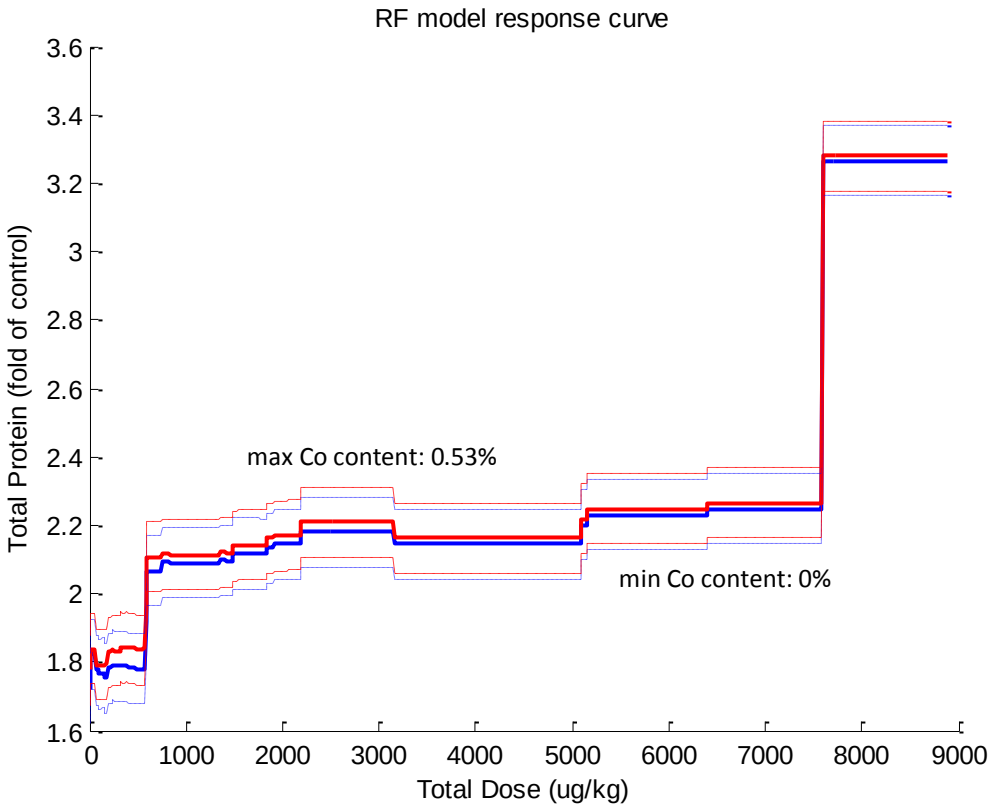

Figure A-19: RF model dose-response response curve for BAL total protein as modified by cobalt content. Dashed and dotted lines indicate the RT model predicted experimental standard deviation. All other model inputs are held constant at their median value.

8 MATLAB Code for Regression Tree and Random Forest Model Generation

Regression Tree Model:

```
catcol = [1 2 3 4 6];
%numerical designations of 'Inputs' columns that are categorical

t = classregtree(Inputs,Outputs,'names',InputNames,'categorical',catcol);

[crscost,crserr,crsnodes,crsbstlvl] = ...
    test(t,'crossvalidate',InputsNew,OutputsNew,'nsamples',10);

tmin = prune(t,'level',crsbstlvl);              %crsbstlvl for auto pruning


fig(1) = view(tmin,outputcol,OutputNames{outputcol});


[mincrscost,mincrsloc] = min(crscost);
figure('Name',['Crossvalidation Error:  ',OutputNames{outputcol}]);
    plot(crsnodes,crscost,'b-o',...
    crsnodes(crsbstlvl+1),crscost(crsbstlvl+1),'bs',...
    crsnodes,(mincrscost+crserr(mincrsloc))*ones(size(crsnodes)),'k--');
    xlabel('Tree size (number of terminal nodes)');
    ylabel('Mean Squared Error');
    title(['Crossvalidation Error:  ',OutputNames{outputcol}]);
```

Random Forest Model:

```
numtrees = 1000;     %numer of trees in forest
catcol = [1 2 3 4 6];        %numerical designations of 'Inputs' columns that
are categorical

inputcol = size(Inputs,2);

b = TreeBagger(numtrees,Inputs,Outputs,'method','r',...
      'oobpred','on','oobvarimp','on','NVarToSample',round(inputcol/3),...
      'categorical',catcol);

figure('Name',['RF Learning:  ',OutputNames{outputcol}]);  % Plot of error as
a function of trees in model
    plot(oobError(b));
    xlabel('Number of Trees');
    ylabel('OOB Mean Squared Error');
    title(['RF Learning:  ',OutputNames{outputcol}]);
```

Stepwise Random Forest Model:

```
%   This script creates a stepwise random forest model.
%
%   24 January 2012
%   (c) Jeremy M. Gernand
%
%

first = tic;        %start CPU timer
disp('   Initializing...')
%   Designate output variable and process parameters **********************
%
%   >>>>>>>>>>>>>>>>>>>>>>>>>>> USER INPUT <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
outputcol = 23;      %designate output column
numtrees = 1000;            %numer of trees in forest
catcol = [1 2 3 4 6];       %numerical designations of 'Inputs' columns
                            %   that are categorical
%   >>>>>>>>>>>>>>>>>>>>>>>>>>> END USER INPUT <<<<<<<<<<<<<<<<<<<<<<<<<<<<

%   Capture matrix parameters ***********************************************
%
inputcol = size(Inputs,2);

%   Trim inputs and outputs to only those with numerical output value *****
%
InputsNew = Inputs;
OutputsNew = Outputs(:,outputcol);
for i = 1:size(Inputs,1);
    if isnan(OutputsNew(size(Inputs,1)-i+1));

        OutputsNew(size(Inputs,1)-i+1) = [];
        InputsNew(size(Inputs,1)-i+1,:) = [];

    end
end

%   Set up and record results of stepwise random forest models ************
%
disp('   Generating stepwise random forest model...')

DataArray = [];
NewDataArray = [];
UsedVars = [];
CatCols = [];
OutputMean = mean(OutputsNew);
rfTSS = 0;
er_count = size(InputsNew,1);
TSSarray = OutputsNew(1:er_count) - OutputMean;
TSSarray = TSSarray.^2;
rfTSS = sum(TSSarray);
RFstepwiseR2 = zeros(round(inputcol/3),inputcol);
RFstepwiseStats = cell(round(inputcol/3),4);
```

Continued…

```
for step = 1:round(inputcol/3);
    second = tic;

    %   display progress on screen
    stepstr = ['      step ',num2str(step),'... '];
    disp(stepstr)

    %   define data array for RF model generate (loop)
    for trial = 1:inputcol;
        %   check if variable has already been used (skip calc if true)
        if ~isempty(find(UsedVars==trial,1));
            RFstepwiseR2(step,trial) = NaN;
        else

            %   define new total data array
            NewDataArray = InputsNew(:,trial);
            InputArray = [DataArray NewDataArray];

            %   set up list of categorical variables
            if ~isempty(find(catcol,trial));
                CatCols = [CatCols trial];
            end

            NumColInMatrix = size(InputArray,2);

            %   define NumVarsToSample
            if NumColInMatrix == 1;
                NumVarsToSample = 1;
            elseif NumColInMatrix == 2;
                NumVarsToSample = 1;
            elseif NumColInMatrix > 2;
                NumVarsToSample = round(NumColInMatrix/3);
            end

            %   generate RF model and compact
            trialstr = ['          trial ',num2str(trial),'... '];
            disp(trialstr)

            b = TreeBagger(numtrees,InputArray,OutputsNew,'method','r',...

'oobpred','on','oobvarimp','on','NVarToSample',NumVarsToSample,...
                'categorical',CatCols);
            bcompact = compact(b);

            %   calculate and record R-squared value for RF model
            rfSSE = 0;
            [RFpredict,RFsd] = predict(bcompact,InputArray);
            SSEarray = OutputsNew(1:er_count) - RFpredict(1:er_count);
            SSEarray = SSEarray.^2;
            rfSSE = sum(SSEarray);
            RFstepwiseR2(step,trial) = 1 - (rfSSE / rfTSS);

        end

    end
```

Continued…

```
%   select highest performing variable
    [BestTrial, BestLocation] = max(RFstepwiseR2(step,:));

    %   record elapsed time
    elapsedtime(step) = round(toc(second)/60);

    %   record performance stats
    RFstepwiseStats(step,1) = {step};               %stepwise row number
    RFstepwiseStats(step,2) = {BestLocation};       %variable number
    RFstepwiseStats(step,3) = InputNames(BestLocation);     %variable name
    RFstepwiseStats(step,4) = {RFstepwiseR2(step,BestLocation)}; %R^2 value
    RFstepwiseStats(step,5) = {elapsedtime(step)};          %step time(min)

    %   define new input array
    DataArray = [DataArray InputsNew(:,BestLocation)];
    UsedVars = [UsedVars BestLocation];

    %   display progress on screen
    progstr = ['       R2 progress ',num2str(BestTrial),'... '];
    disp(progstr)
    timestr = ['       time elapsed (min): ',num2str(elapsedtime(step))];
    disp(timestr)

end

%   Graph plot of R-squared versus variable addition
%       include order of inclusion on graph...

figure('Name',['RF Stepwise Model Growth:  ',OutputNames{outputcol}]);
hold on;
R2fig = plot(RFstepwiseStats{:,1},RFstepwiseStats{:,4},'-');
title(['RF Stepwise Model Growth:  ',OutputNames{outputcol}]);
xlabel('Variable Names in Order of Addition');
ylabel('Model Performance (R^2)');
hold off;
        %here, need to add axis labels, put variable names on x-axis
        %change formatting to show point markers with line

toc(first);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

9 Stepwise Linear Regression Models

Table A-1: Coefficients and model performance statistics for stepwise linear regression models. These results provide another perspective on input variable importance, however the amount of data excluded from these models reduces the confidence as compared to the RT and RF models.

Output Variable	Input Variable	Coefficient	P-Value
Neutrophils	Mass Concentration	170.09	0
	Post Exposure Period	-38.75	5.23E-175
	MMAD	-18.86	1.69E-9
$R^2 = 0.66$ 86% of the records excluded from this model due to missing values			
Macrophages	Mass Concentration	2.47	0
	Configuration (SW/MW)	0.91	1.10E-39
	MMAD	-0.55	3.62E-4
	Post Exposure Period	-0.20	5.10E-9
$R^2 = 0.80$ 75% of the records excluded from this model due to missing values			
Lactate Dehydrogenase	MMAD	3.94	2.37E-47
	Configuration (SW/MW)	3.36	9.02E-166
	Mass Concentration	2.47	0
	Post Exposure Period	-0.21	7.12E-33
$R^2 = 0.71$ 82% of the record excluded from this model due to missing values			
Total Protein	Mean Animal Mass	-14.84	1.62E-95
	Configuration (SW/MW)	8.56	1.49E-81
	MMAD	2.73	3.75E-55
	24h Avg Dose Cobalt	1.90	3.07E-95
	Mass Concentration	-1.16	5.57E-42
	Post Exposure Period	-0.31	1.23E-294
$R^2 = 0.66$ 74% of the records excluded from this model due to missing values			

Appendix B: Titanium Dioxide Nanoparticle Toxicity Regression Tree and Random Forest Model Details

Regression Tree Models

The following figures contain the pruned regression tree (RT) models predicting toxic responses in BAL fluid following exposure to TiO_2 nanoparticles.

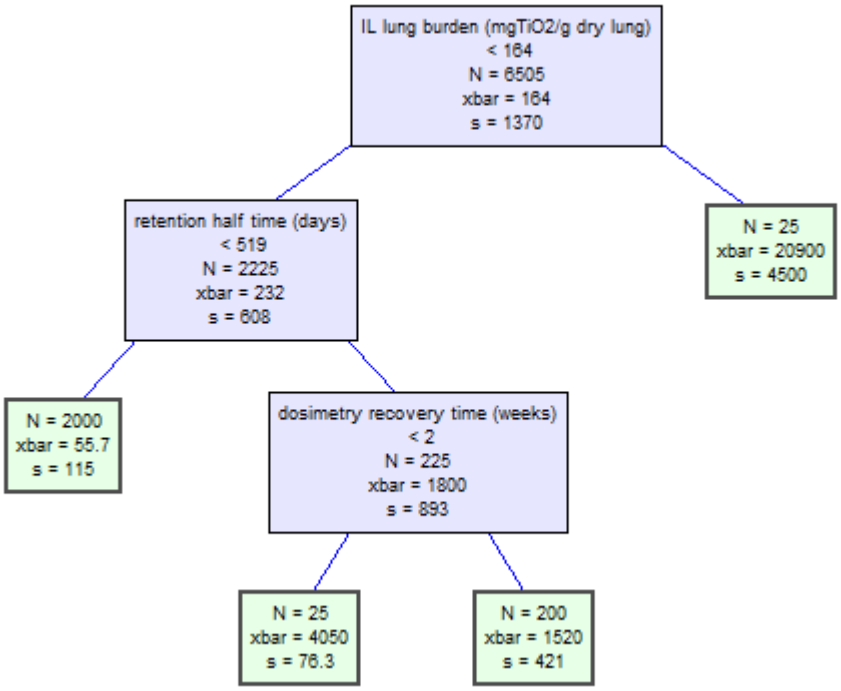

Figure B-1: Regression tree model predicting the change in BAL neutrophils following exposure to TiO_2 nanoparticles.

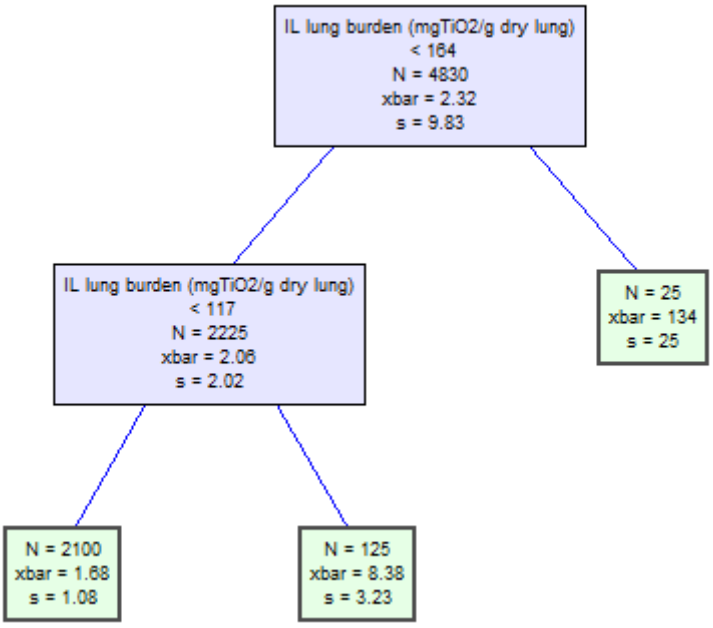

Figure B-2: Regression tree model predicting the change in BAL macrophages following exposure to TiO_2 nanoparticles.

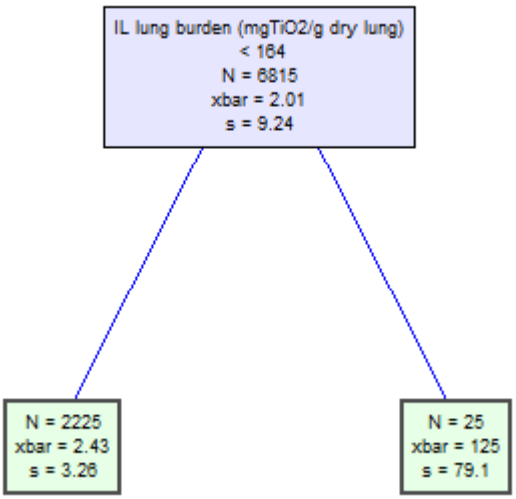

Figure B-3: Regression tree model predicting the change in BAL total cell count following exposure to TiO_2 nanoparticles.

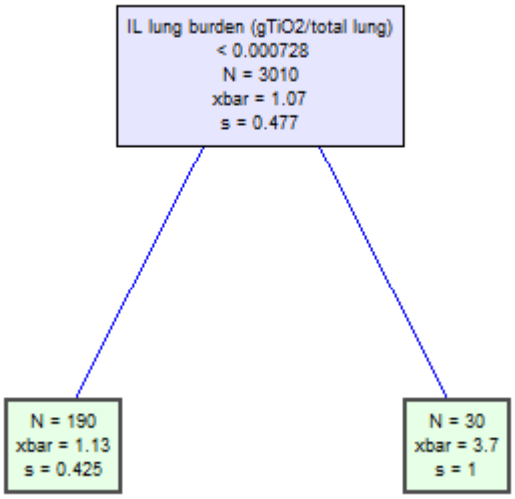

Figure B-4: Regression tree model predicting the change in BAL total protein following exposure to TiO_2 nanoparticles.

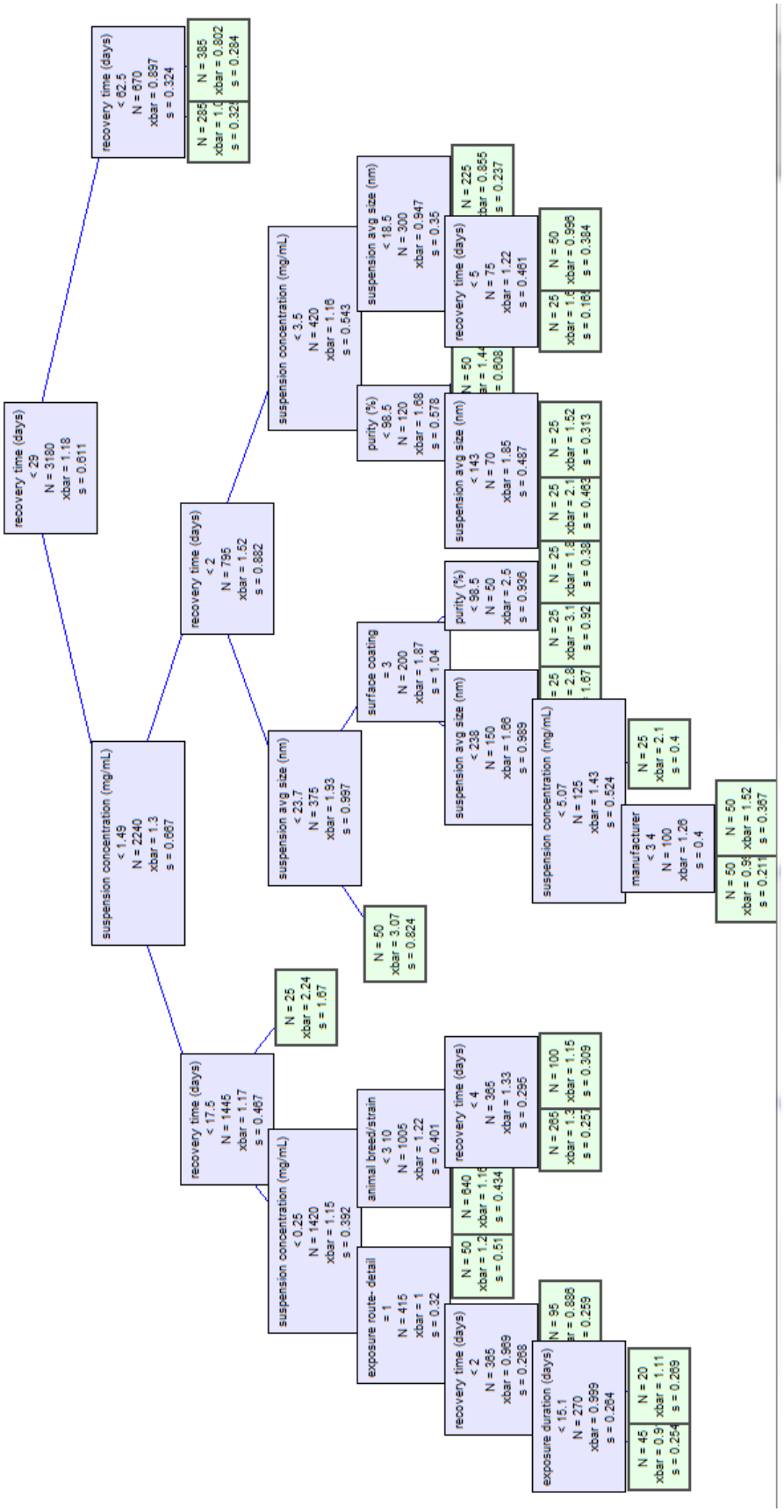

Figure B-5: Regression tree model predicting the change in BAL lactate dehydrogenase concentration following exposure to TiO_2 nanoparticles.

Regression Tree Model Learning Graphs

These graphs display the error reduction achieved by each additional branch added to the regression tree (RT) models. To prevent over-fitting, each RT model is pruned through a process similar to backwards stepwise elimination for linear models. In order to prune the model, the error of the model is calculated through ten-fold cross-validation—a process by which the data set is divided randomly into 10 subsets of rows, the RT model branches and split criteria are set based on a matrix containing 9 of those subsets, and the error is measured against the model predictions for the 10th subset. This process is repeated until each of the 10 subsets has been withheld from the model and used for error calculation, and then the total sum-squared error for the model is calculated. The RT model is then pruned by removing the branches providing the least error reduction until the model reaches the smallest size possible within one standard error of the minimum error. The pruned RT model for neutrophils is shown in Figure 3-1.

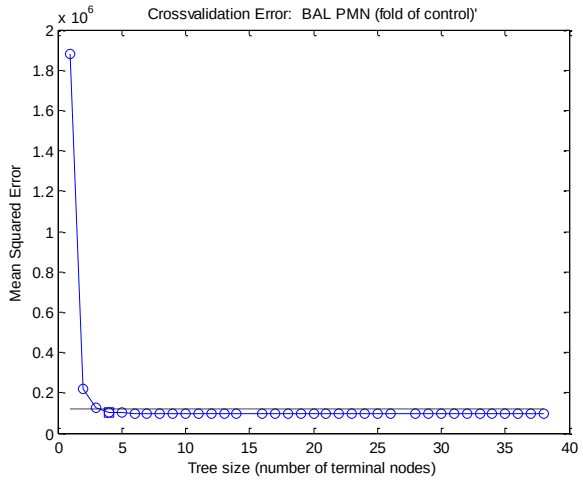

Figure B-6: **Crossvalidation error for RT prediction of change in BAL PMN**

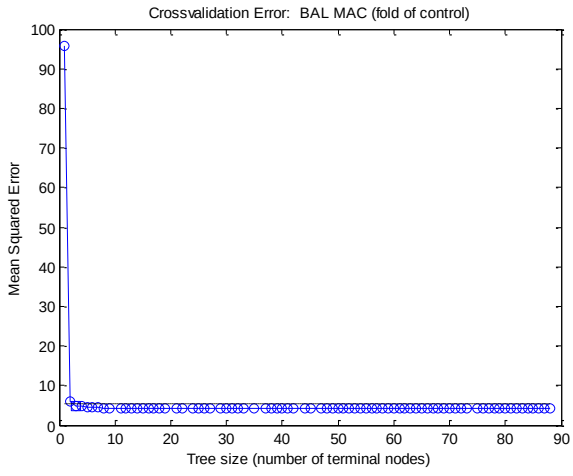

Figure B-7: **Crossvalidation error for RT prediction of change in BAL MAC**

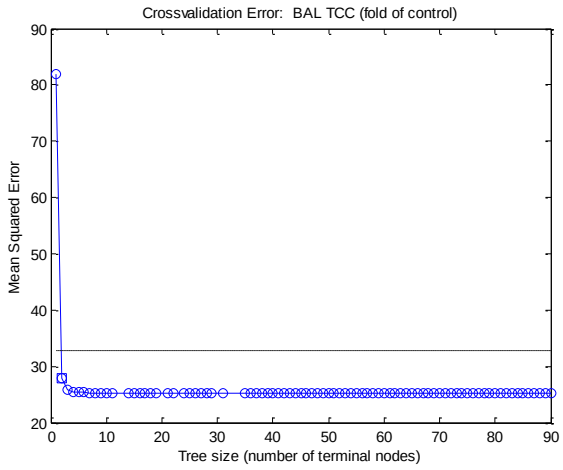

Figure B-8: **Crossvalidation error for RT prediction of change in BAL TCC.**

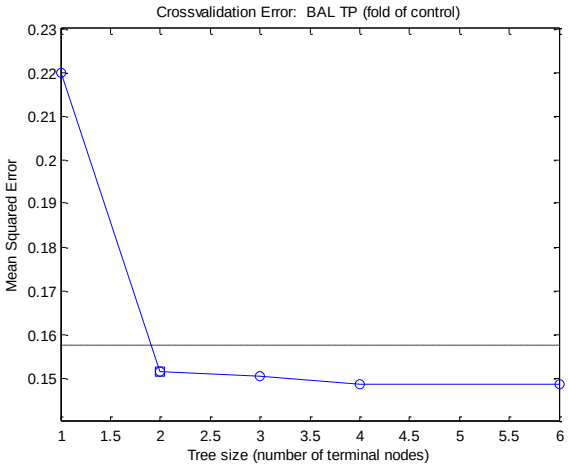

Figure B-9: **Crossvalidation error for RT prediction of change in BAL TP.**

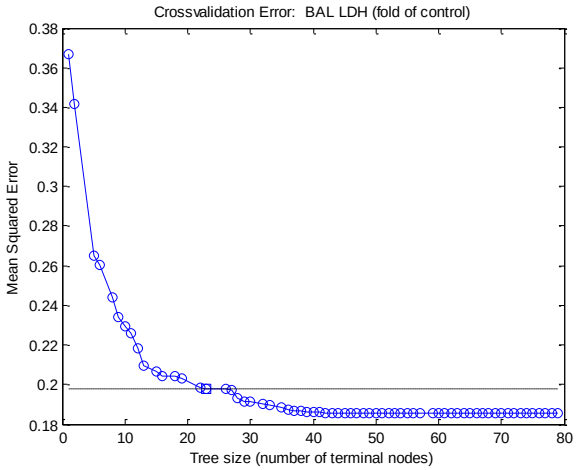

Figure B-10: **Crossvalidation error for RT prediction of change in BAL LDH.**

Random Forest Model Learning Graphs

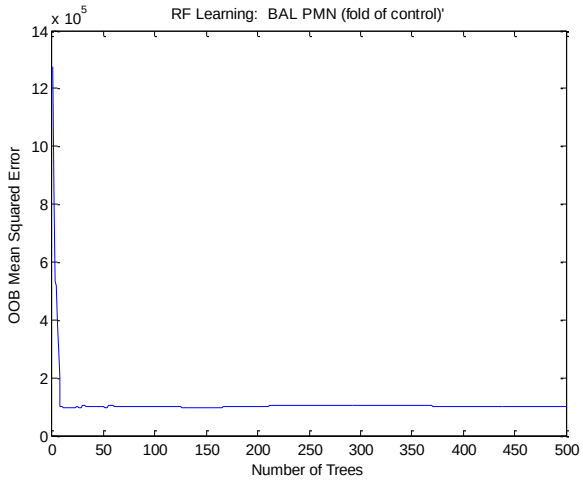

Figure B-11: **Mean squared error per RF model growth for BAL PMN.**

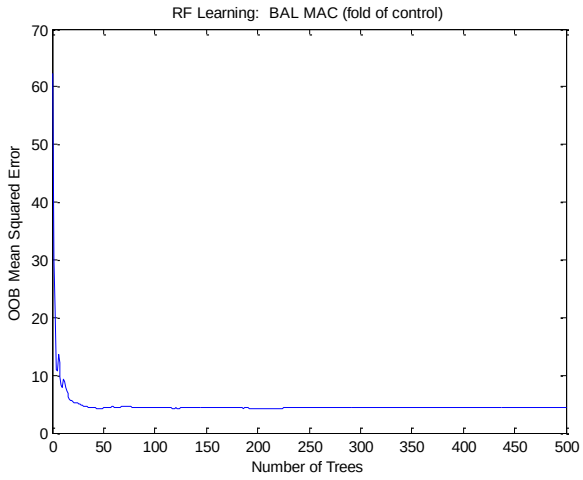

Figure B-12: **Mean squared error per RF model growth for BAL MAC.**

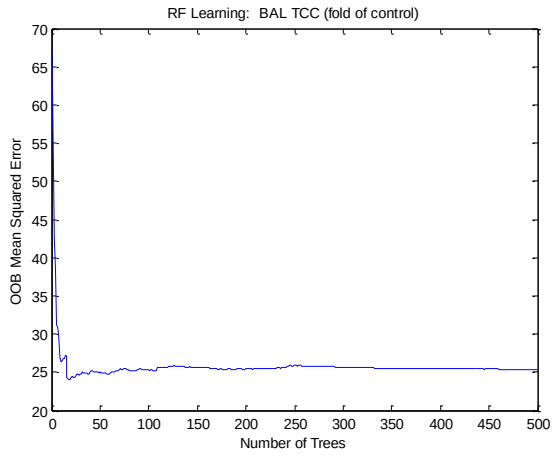

Figure B-13: **Mean squared error per RF model growth for BAL TCC.**

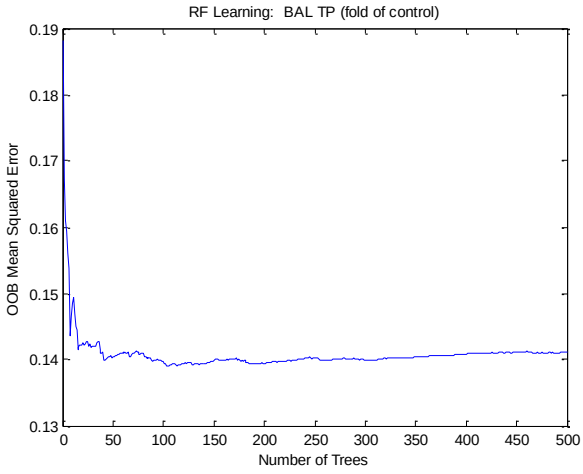

Figure B-14: **Mean squared error per RF model growth for BAL TP.**

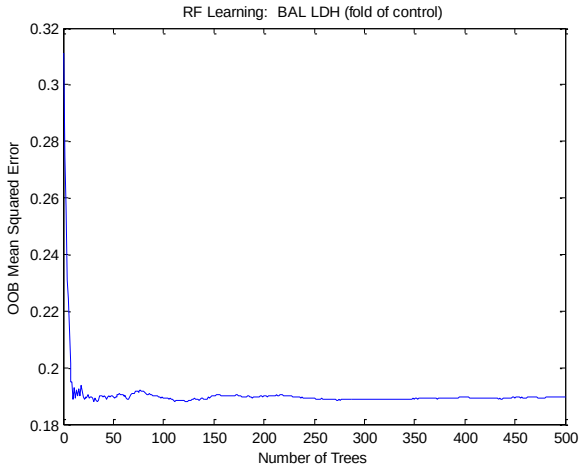

Figure B-15: **Mean squared error per RF model growth for BAL LDH.**

Stepwise Random Forest Model Structures

A stepwise RF model is generated in the same manner as a stepwise linear regression model, except using the RF algorithm—one variable at a time is added so as to produce the highest possible variance reduction (or R^2 value).

In conjunction with the RF variable importance results for all input variables together, these results provide an indication of what smaller set of variables is providing the most information value in terms of improving the prediction of the output. A general cutoff can be applied here such that the model is limited to those variables necessary to achieve 95% or 99% of the total possible variance reduction. For the 5 outputs shown here, the number of necessary variables is generally 3 to 5.

It should also be noted here that for steps 1 and 2, true RF models do not exist. Step 1 is a single RT model, while step 2 is a "small forest" (usually less than 10 RT models). Once 3 or more variables are available, more complete RF models are produced.

A better metric for defining the "best" set of variables for the a model is something like the AIC (Akaike Information Criterion) that takes into account both model error as well as complexity. This factor isn't completely defined for models like a Random Forest, but with a few assumptions, the best model size usually occurs between 3 and 6 variables for all of these outputs, even when the model performance continues to improve. Adjusted R-squared can be calculated too, and results in similar model sizes. In some cases, interestingly, the model performance actually degrades with the addition of more variables (I am still investigating what this really means).

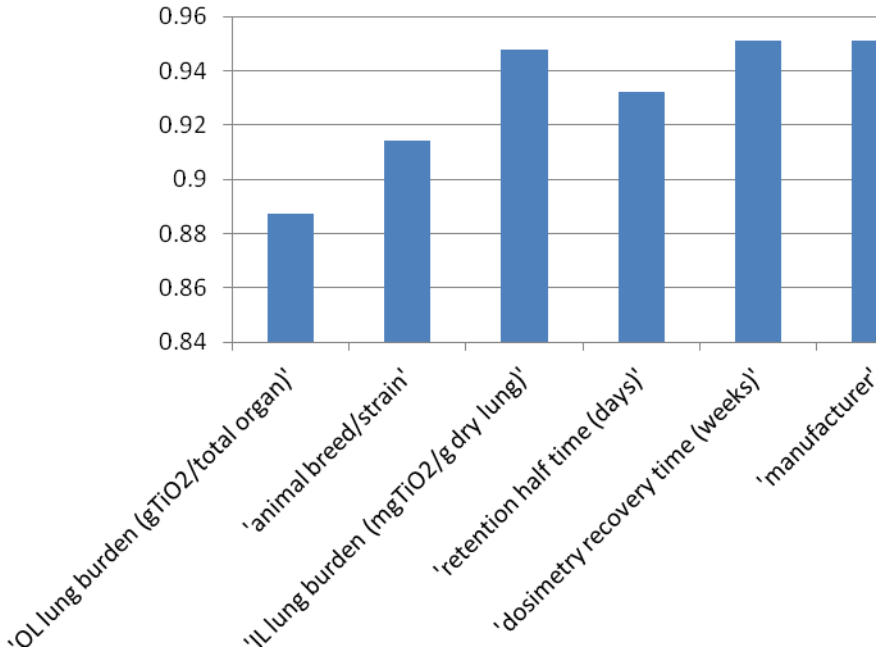

Figure B-16: PMN Stepwise RF Model Growth: R-Squared Performance by Variable Added

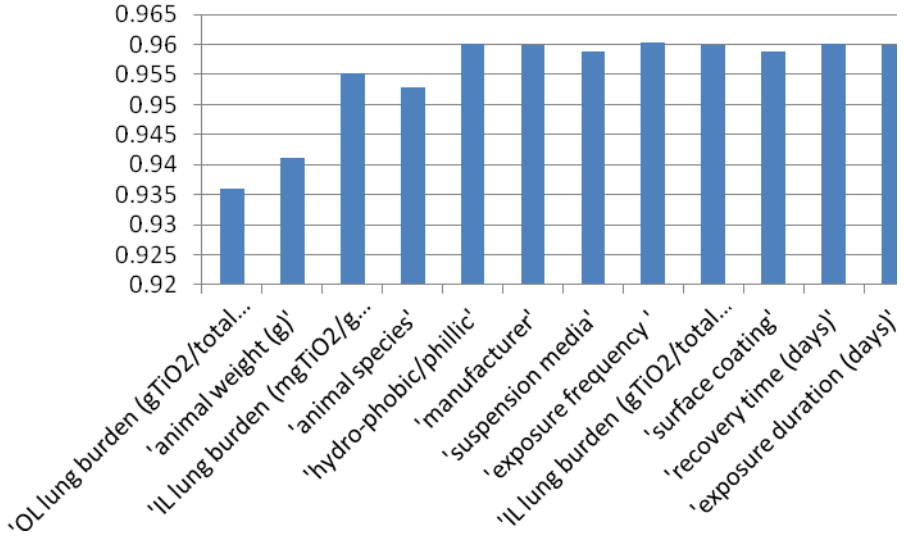

Figure B-17: MAC Stepwise RF Model Growth: R-Squared Performance by Variable Added

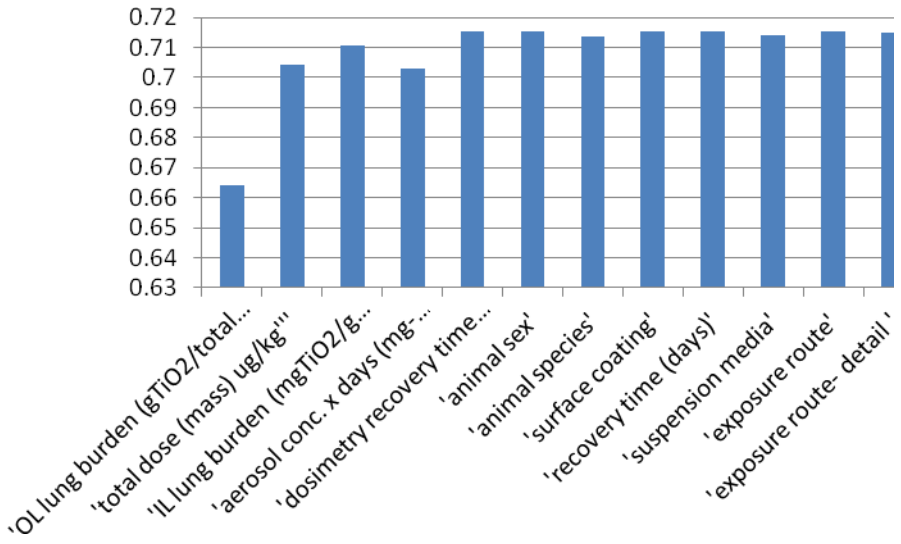

Figure B-18: TCC Stepwise RF Model Growth: R-Squared Performance by Variable Added

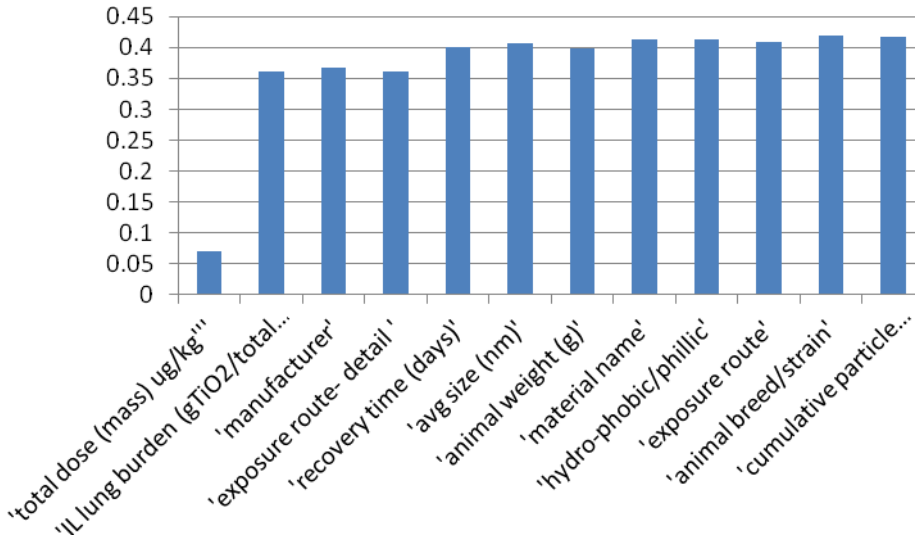

Figure B-19: MTP Stepwise RF Model Growth: R-Squared Performance by Variable Added

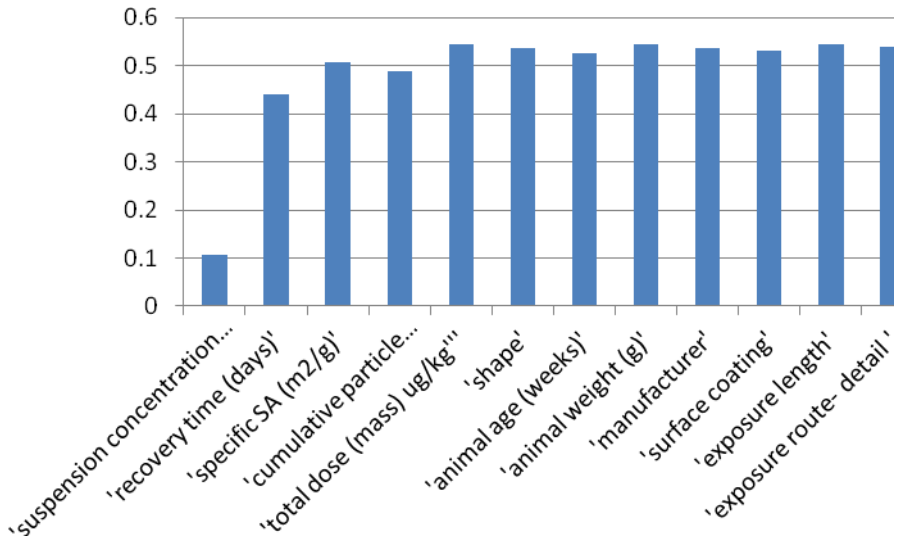

Figure B-20: LDH Stepwise RF Model Growth: R-Squared Performance by Variable Added

Comparing Random Forests from Imputed Data to Those from Non-Imputed Data

These following charts compare the median variable importance results from the RFs trained on the imputed data sets (the median response for each variable from the 5 imputed data-derived models, and the results from the RFs trained on the non-imputed data (with missing values).

In general, the imputed variable importance results did not differ significantly from the non-imputed results, with the exception of aerosol concentration × days, which displayed a higher importance in the non-imputed data than in the imputed data. This indicates that the higher importance is probably due to the fact that so many other experiment characterization values are missing. This situation is reversed for LDH, however, where the imputed values happen to have reduced the relative importance of other variables as compared to aerosol concentration, resulting in the RF model becoming more reliant on that value.

Total Cell Count

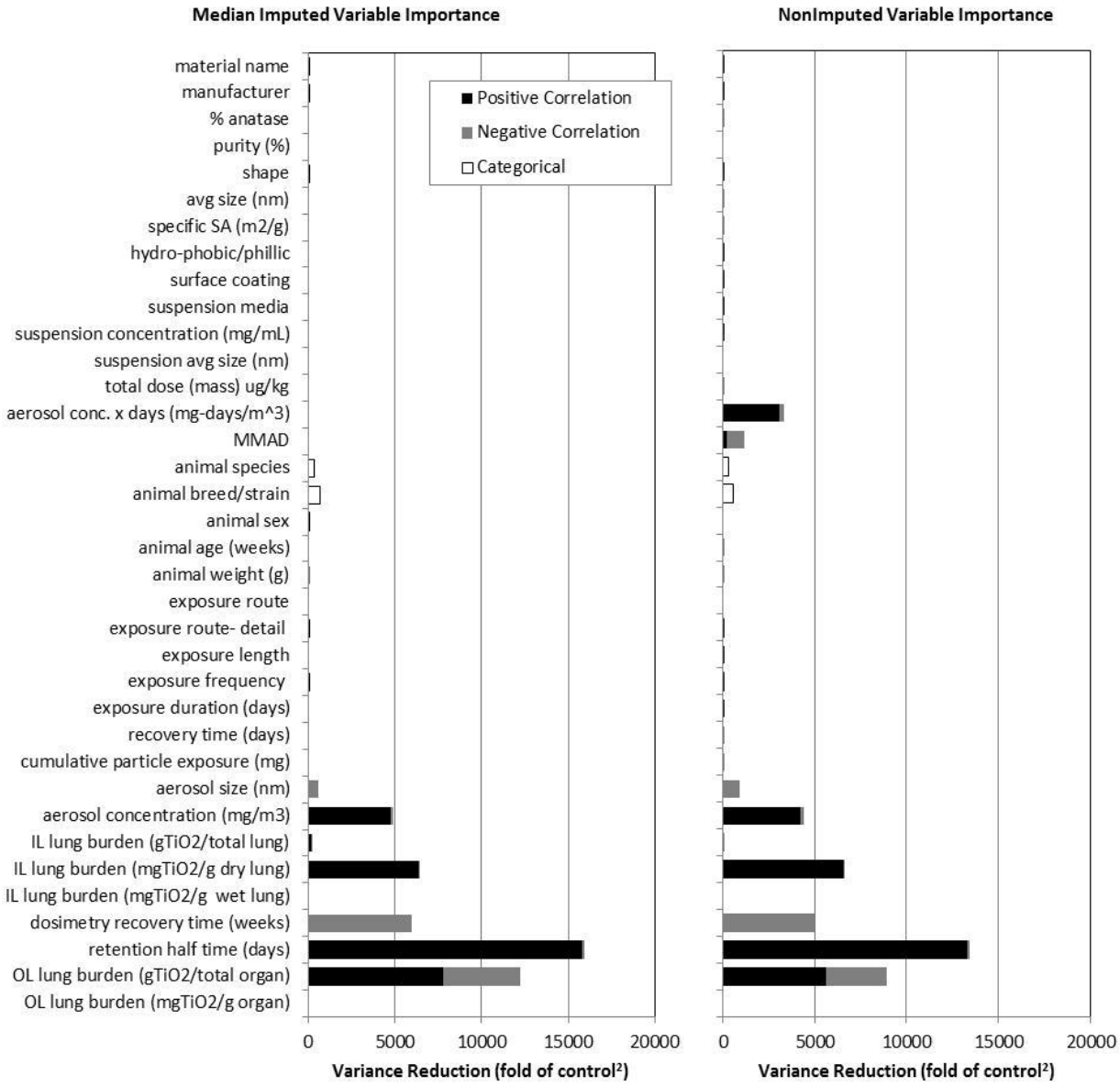

Figure B-21: **Comparison of RF variable importance between imputed and non-imputed data sets for the prediction of BAL total cell count.**

Neutrophils

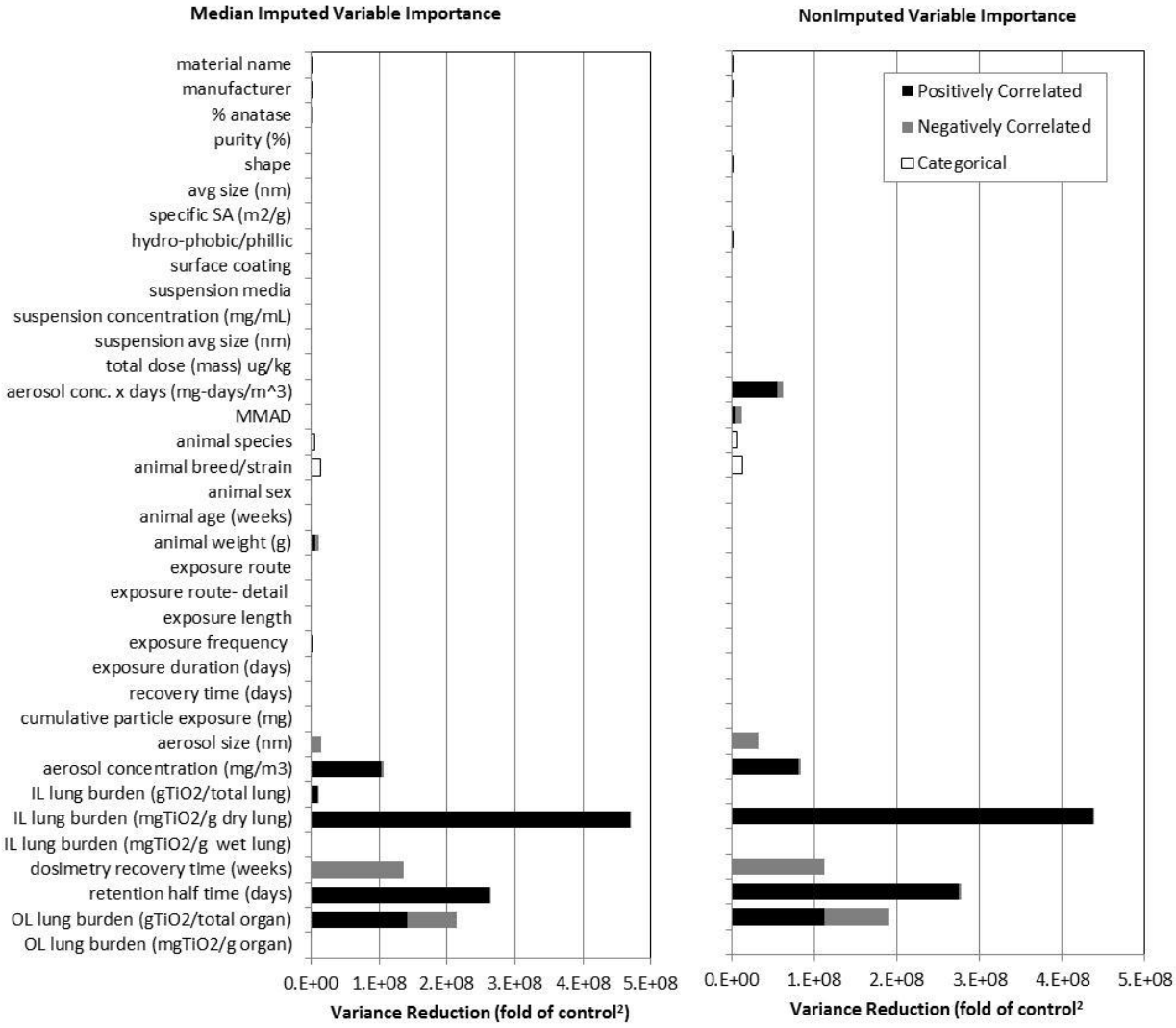

Figure B-22: **Comparison of RF variable importance between imputed and non-imputed data sets for the prediction of BAL neutrophils count.**

Macrophages

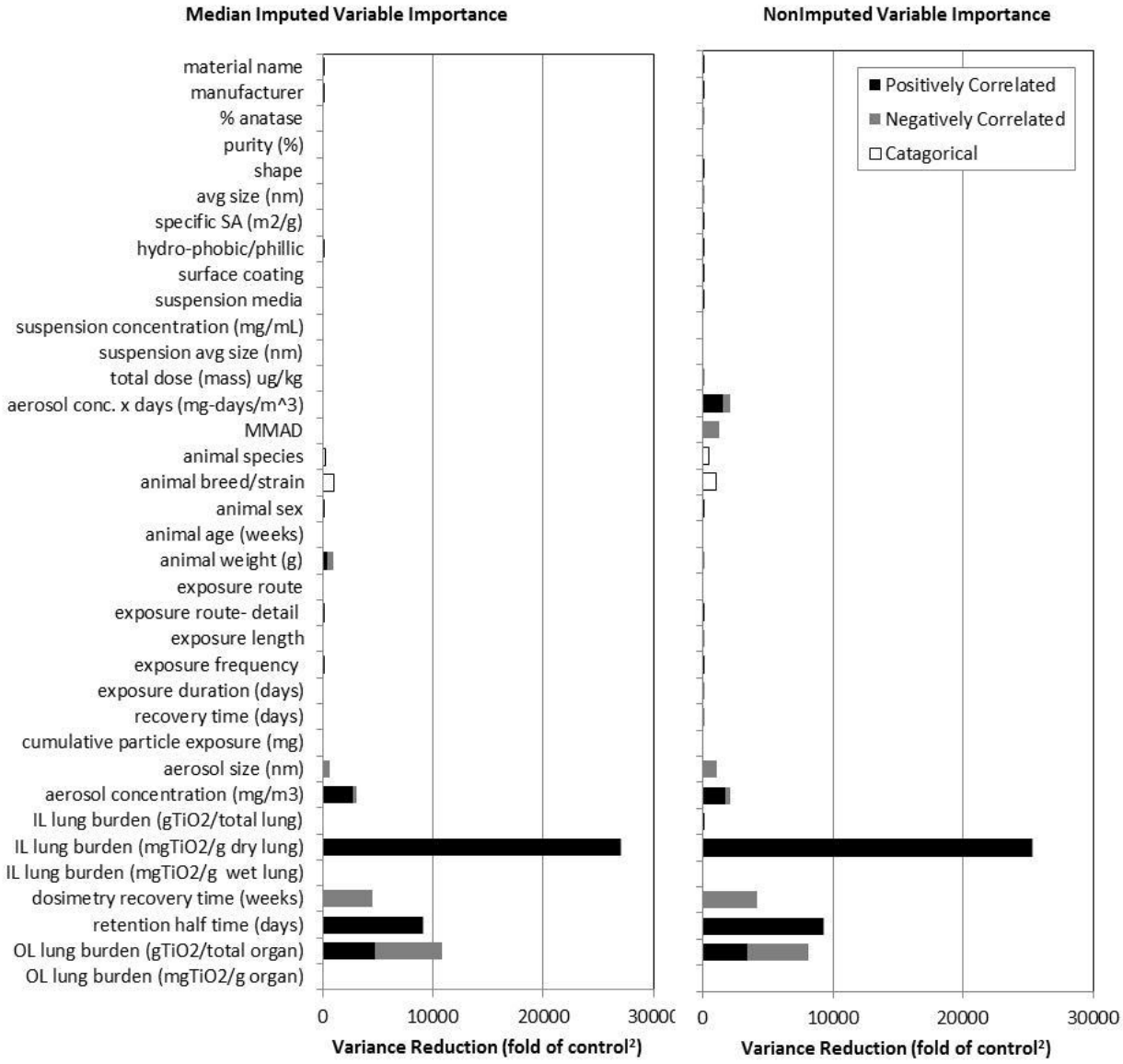

Figure B-23: **Comparison of RF variable importance between imputed and non-imputed data sets for the prediction of BAL macrophages count.**

Lactate Dehydrogenase

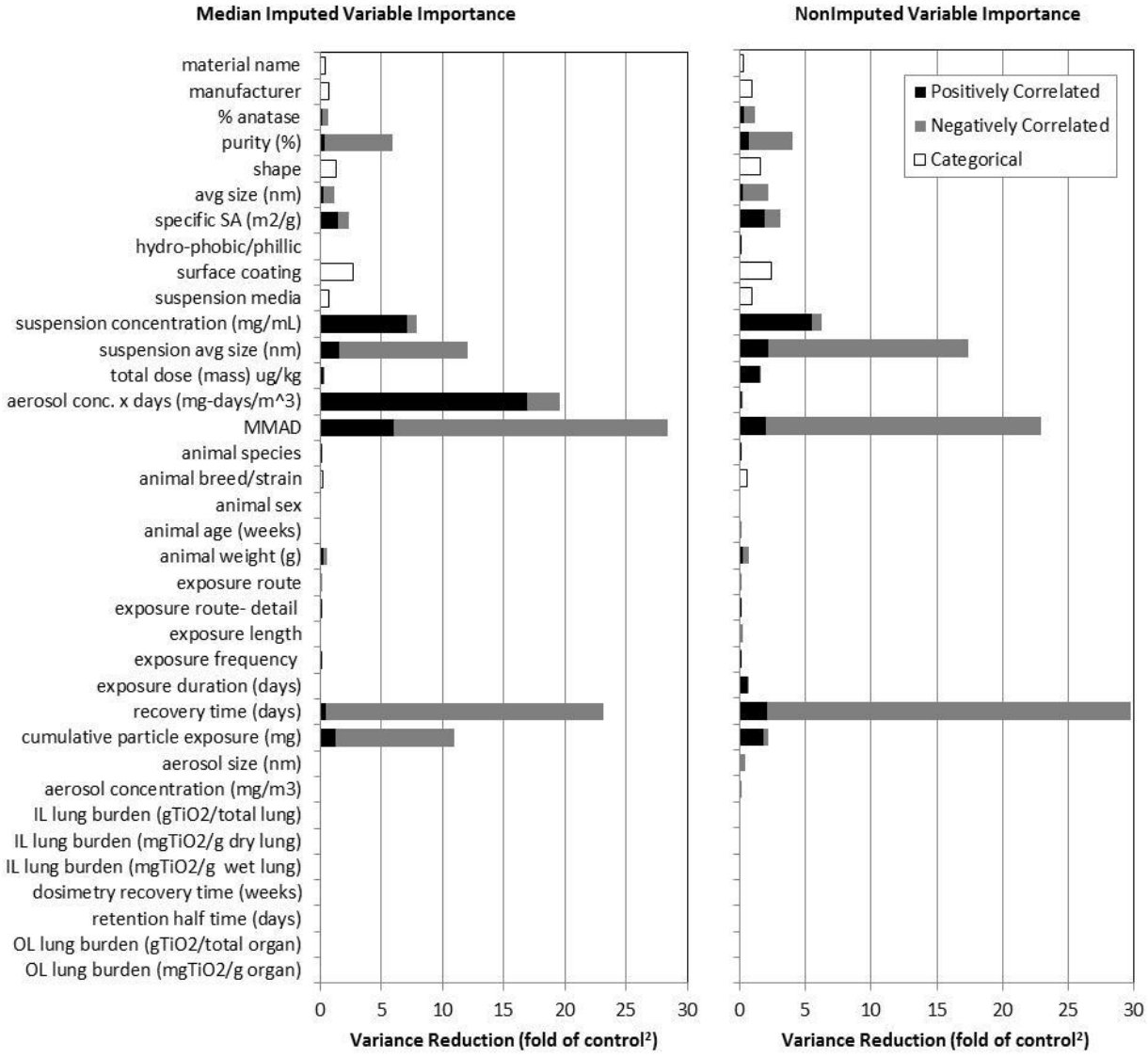

Figure B-24: **Comparison of RF variable importance between imputed and non-imputed data sets for the prediction of BAL LDH concentration.**

Total Protein

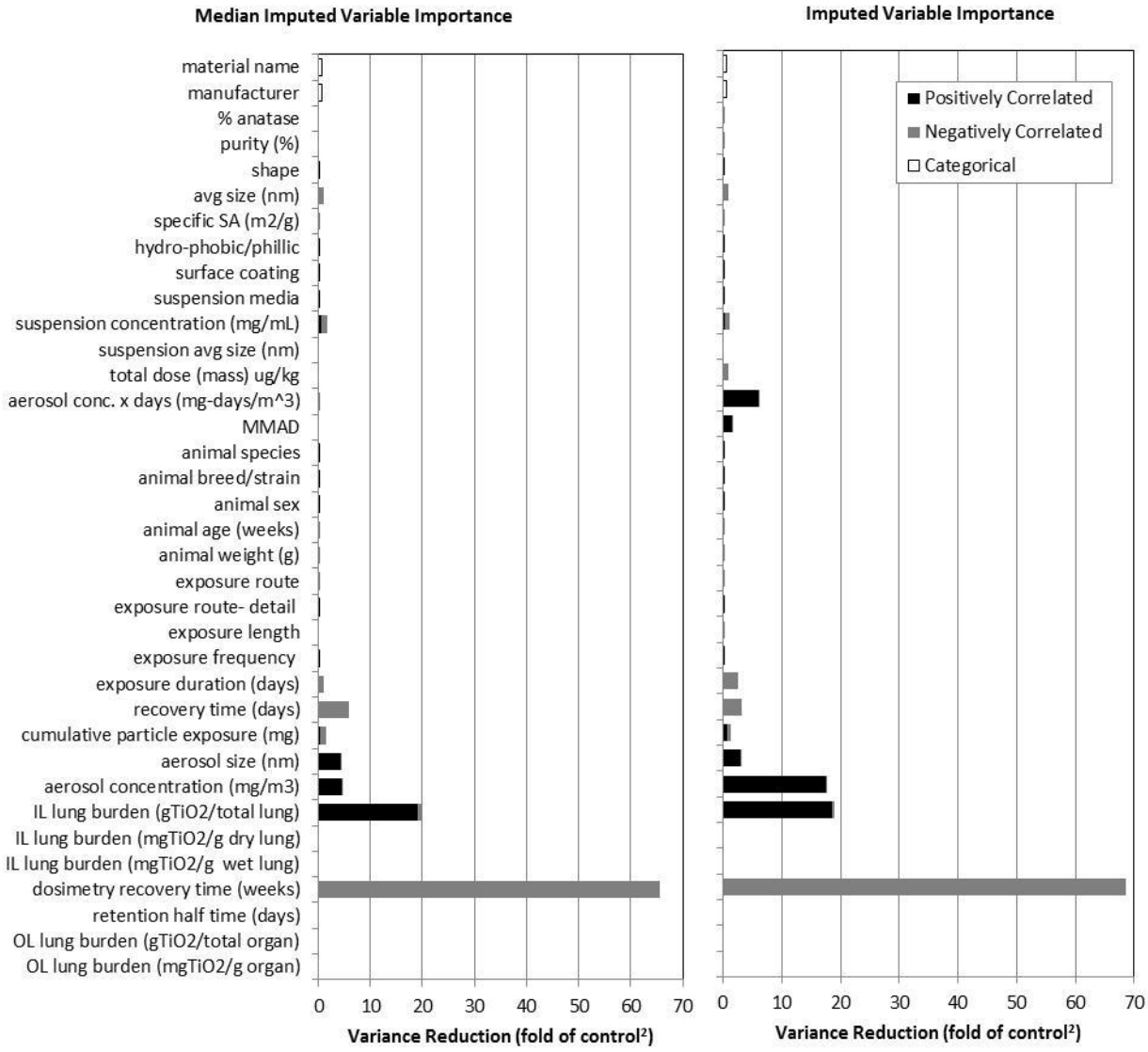

Figure B-25: **Comparison of RF variable importance between imputed and non-imputed data sets for the prediction of BAL total protein concentration.**

Variable Importance without Inclusion of Dosimetry Measurements

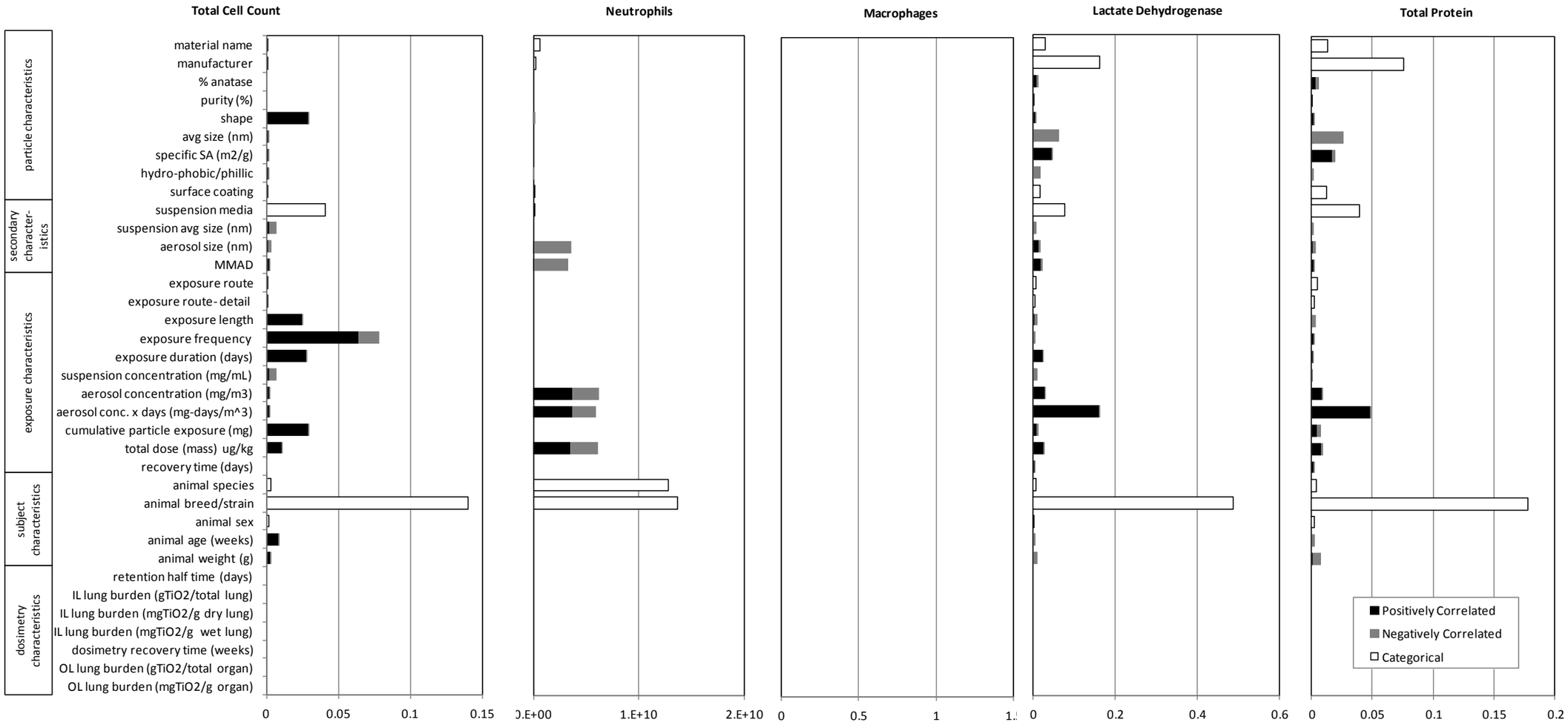

Figure B-25: Variable importance results from the RF models when all dosimetry characteristics are excluded from the data set. No viable model was possible for Macrophages without the dosimetry data.

RF Model Dose-Response and Recovery Curves

These figures contain the RF model summarized dose-response curves for exposures to nano-TiO_2 particles.

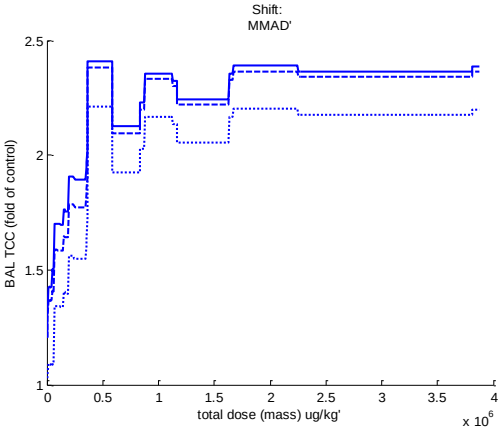

Figure B-26: **Dose-response curve from RF model for BAL TCC shifted in response to changes in MMAD.**

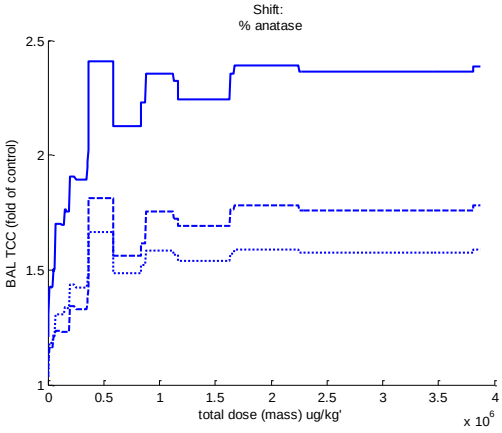

Figure B-28: **Dose-response curve from RF model for BAL TCC shifted in response to changes in percent anatase.**

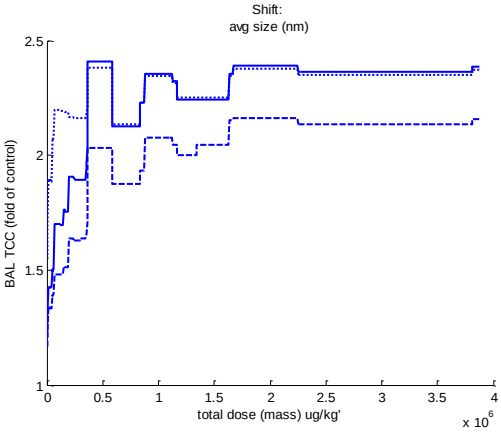

Figure B-27: **Dose-response curve from RF model for BAL TCC shifted in response to changes in average size.**

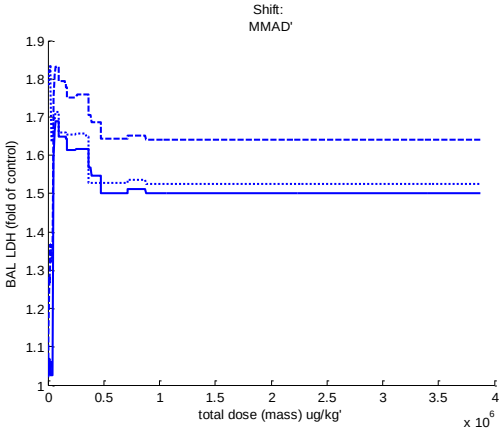

Figure B-29: **Dose-response curve from RF model for BAL LDH shifted in response to changes in MMAD.**

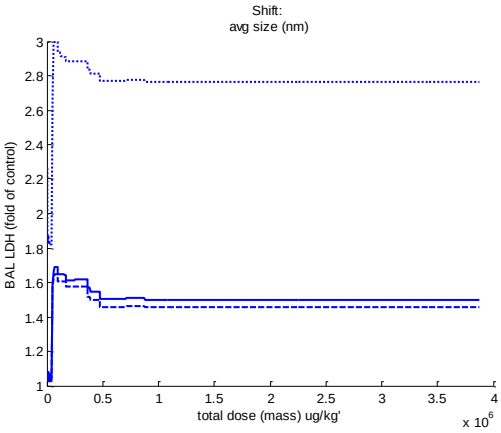

Figure B-30: **Dose-response curve from RF model for BAL LDH shifted in response to changes in average size.**

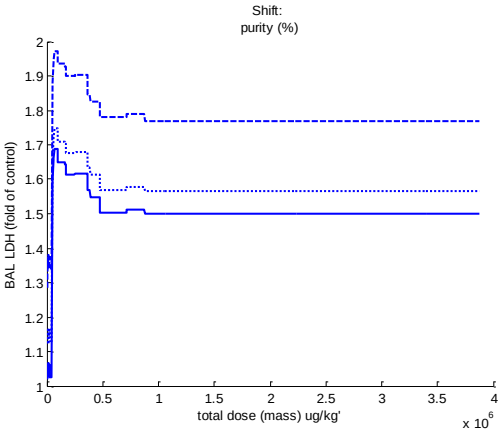

Figure B-31: **Dose-response curve from RF model for BAL LDH shifted in response to changes in purity.**

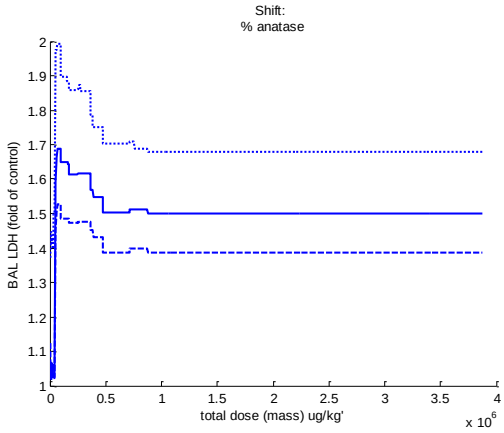

Figure B-32: **Dose-response curve from RF model for BAL LDH shifted in response to changes in anatase fraction.**

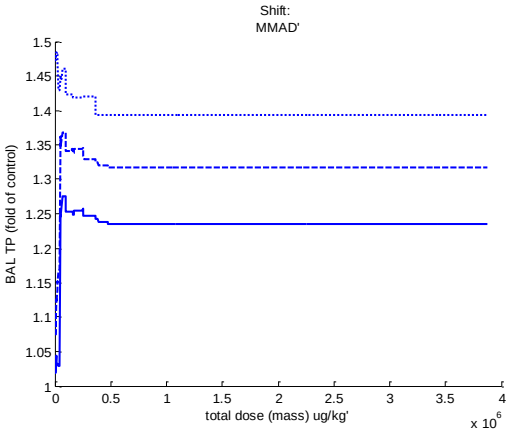

Figure B-33: **Dose-response curve from RF model for BAL TP shifted in response to changes in MMAD.**

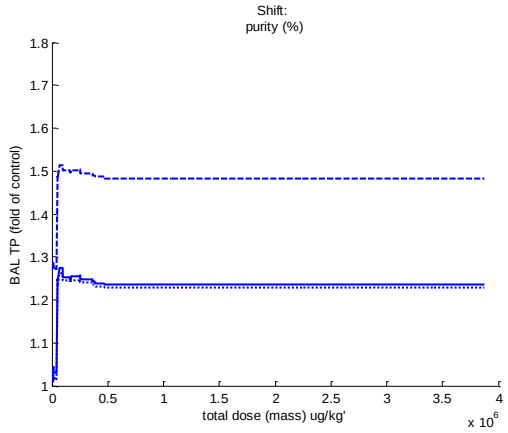

Figure B-34: **Dose-response curve from RF model for BAL TP shifted in response to changes in purity.**

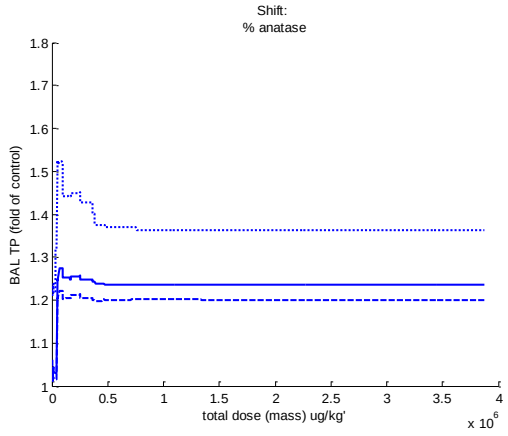

Figure B-35: **Dose-response curve from RF model for BAL TP shifted in response to changes in anatase fraction.**

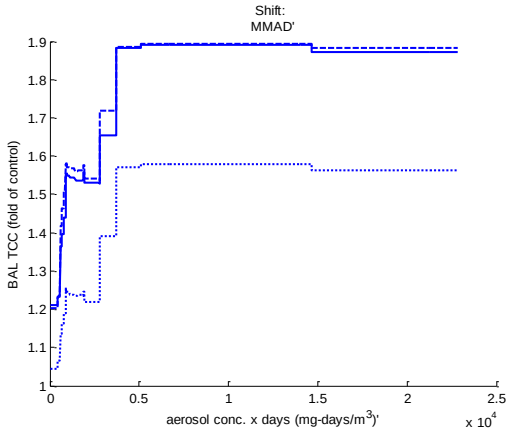

Figure B-36: **Dose-response curve from RF model for BAL TCC in terms of aerosol concentration shifted in response to changes in MMAD.**

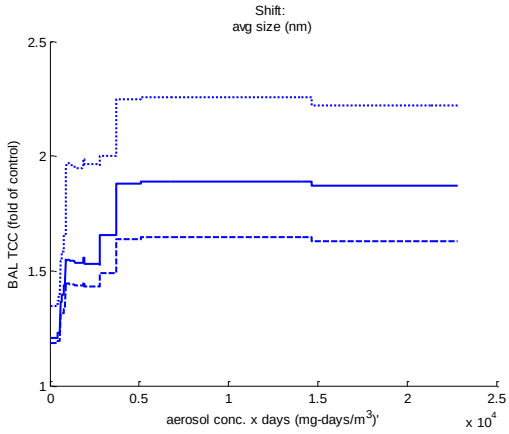

Figure B-37: **Dose-response curve from RF model for BAL TCC in terms of aerosol concentration shifted in response to changes in average size.**

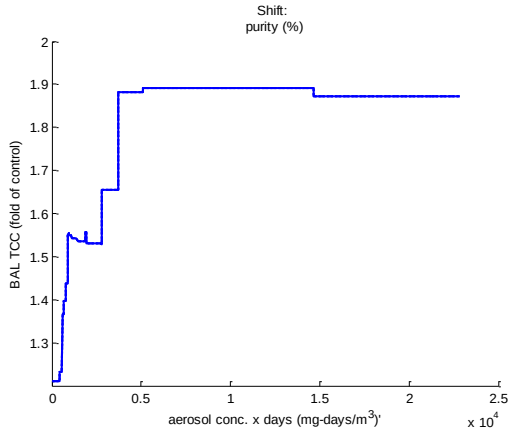

Figure B-38: **Dose-response curve from RF model for BAL TCC in terms of aerosol concentration shifted in response to changes in purity.**

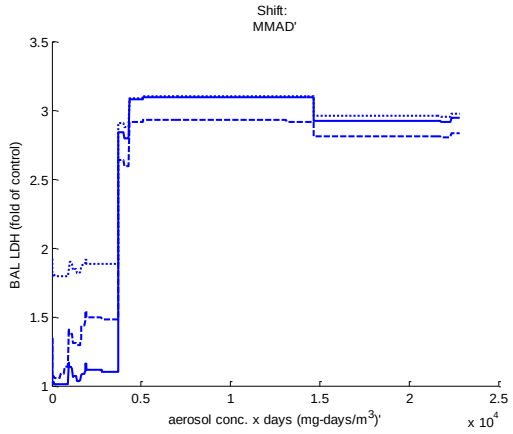

Figure B-39: **Dose-response curve from RF model for BAL LDH in terms of aerosol concentration shifted in response to changes in MMAD.**

Random Forest Model Validation

While linear models and other algebraic models are undefined whenever any of the input variables are missing, a RT model produces a prediction even if no input variables are provided—in this specific case the model produces an output equivalent to the mean of all observations in the training data set. If some, but not all, of the input variables are missing, the model returns the weighted average of all possible resulting leaf nodes. Since 40% of the characterization or exposure attributes are missing and 74% of the input variables in this data set have at least one missing value, complete utilization of all available data is a distinct advantage for RTs over multiple linear regression (MLR) models. Further, this RT property of producing a prediction regardless of the completeness of the data record prevents the underestimation of error that is common when ordinary regression models are applied to data sets with missing values. By not discarding incomplete records, as is the practice in most off-the-shelf MLR algorithms, the RT model's error against every single experimental result is fully captured.

While other machine learning methods such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) can also handle nonlinear relationships and may be more computationally efficient, they become undefined for cases where any of the input data is missing. So, while imputation of missing values would be required for MLR, ANN, or SVM models to take full advantage of the entire experimental data set, this step and potential source of artificially introduced uncertainty is unnecessary for RTs.

Maximum model performance (Table A1) is the maximum R-squared value possible achieved with a model that uniquely identifies each individual experimental exposure group (e.g., each grouping of rodents exposed to the same amount of nano-titania in the same lab with the same amount of recovery time). It is an indicator of the inherent experimental variance, and a ceiling on the performance of any mathematical model.

For each of the cell count indicators the RT and RF models approach the maximum model performance with only a few variables. This is not true for the LDH and Total Protein concentration. The Total Protein RF model only achieved a variance reduction of 0.41. (It is noteworthy that no MLR model could be fit to this data at all, primarily due to the quantity of missing characterization measures). The LDH and TP endpoints in general tend to display a greater degree of experimental variance. As seen in the table below, the maximum RF model performance does not differ significantly from the performance of the minimally sized stepwise RF models, and in the case of TP and LDH, RF model variance explanation performance actually declines a bit.

Table B-1: **Complete RT and RF models' goodness of fit statistics employing all available variables. Maximum model performance is the explained variance possible by assigning each experimental group a unique ID number, and predicting the mean response for each animal in the group.**

Modeled Endpoint	**Regression Tree**		**Random Forest**		**Maximum Model Performance**	
	R^2	# of variables	R^2	# of variables	R^2	# of exposure groups
Neutrophils	0.95	3	0.95	32	0.96	250
Macrophages	0.95	1	0.96	32	0.97	186
Total Cell Count	0.68	1	0.72	34	0.72	261
Total Protein	0.31	1	0.37	30	0.68	112
Lactate Dehydrogenase	0.49	9	0.50	28	0.63	120

Table B-2: **Results of RF model validation whereby each study is consecutively excluded from the training data set and used as a test data set. Error reported as root mean squared error (RMSE). Studies with the highest error are difficult to predict using only the data found in the other experiments.**

First Author, Publication Date	Exposure (mg for instillation, mg/m^3 for inhalation)	Recovery Period (weeks)	Exposure Mode	RF Model RMSE (fold of control)
		Total Cells		
Bermudez, 2002	9.5 – 240 mg/m^3	0 - 52	inhalation	29.9
Oberdorster, 1992	0.1 – 0.5 mg	1 day	instillation	8.9
Rehn, 2003	0.15 – 1.2 mg	0 - 13	instillation	4.5
Warheit, 2006	4.05 – 20.25 mg	0 - 12	instillation	3.6
Warheit, 2007	0.245 – 1.225 mg	0 - 12	instillation	3.1
Osier, 1997	125 mg/m^3	0 - 1	inhalation	2.7
Grassian, 2007	0.77 – 7.22 mg/m^3	0 - 2	inhalation	1.6
Renwick, 2004	0.125 – 0.5 mg	1 day	instillation	1.5
MSE for all points in full model	--	--	--	4.8
		Neutrophils		
Bermudez, 2002	9.5 – 240 mg/m^3	0 - 52	inhalation	6000
Oberdorster, 1992	0.1 – 0.5 mg	1 day	instillation	2400
Grassian, 2007	0.77 – 7.22 mg/m^3	0 - 2	inhalation	1700
Nemmar, 2008	0.36 – 1.8 mg	1 day	instillation	1200
Warheit, 2006	4.05 – 20.25 mg	0 - 12	instillation	570
Rehn, 2003	0.15 – 1.2 mg	0 - 13	instillation	560
Warheit, 2007	0.245 – 1.225 mg	0 - 12	instillation	520
Osier, 1997	125 mg/m^3	0 - 1	inhalation	330
Renwick, 2004	0.125 – 0.5 mg	1 day	instillation	260
MSE for all points in full model	--	--	--	300
		Macrophages		
Bermudez, 2002	9.5 – 240 mg/m^3	0 - 52	inhalation	32.6
Grassian, 2007	0.77 – 7.22 mg/m^3	0 - 2	inhalation	12.7
Nemmar, 2008	0.36 – 1.8 mg	1 day	instillation	8.3
Oberdorster, 1992	0.1 – 0.5 mg	1 day	instillation	5.1
Osier, 1997	125 mg/m^3	0 - 1	inhalation	4.4
Rehn, 2003	0.15 – 1.2 mg	0 - 13	instillation	4.3
MSE for all points in full model	--	--	--	2.0
		Lactate Dehydrogenase		
Warheit, 2006	4.05 – 20.25 mg	0 - 12	instillation	2.0
Kobayashi, 2009	0.25 – 1.25 mg	0 – 4	instillation	1.6
Warheit, 2007	0.245 – 1.225 mg	0 - 12	instillation	1.3
Grassian, 2007	0.77 – 7.22 mg/m^3	0 - 2	inhalation	0.9
Warheit, 2010	1.25 mg	0 – 12	instillation	0.5

First Author, Publication Date	Exposure (mg for instillation, mg/m^3 for inhalation)	Recovery Period (weeks)	Exposure Mode	RF Model RMSE (fold of control)
MSE for all points in full model	--	--	--	0.4
		Total Protein		
Osier, 1997	125 mg/m^3	0 - 1	inhalation	0.9
Grassian, 2007	0.77 – 7.22 mg/m^3	0 - 2	inhalation	0.7
Warheit, 2006	4.05 – 20.25 mg	0 - 12	instillation	0.4
Oberdorster, 1992	0.1 – 0.5 mg	0	instillation	0.3
Rehn, 2003	0.15 – 1.2 mg	0 - 13	instillation	0.2
Renwick, 2004	0.125 – 0.5 mg	1 day	instillation	0.1
Warheit, 2010	1.25 mg	0 – 12	instillation	0.1
MSE for all points in full model	--	--	--	0.4

Appendix C: Nanoparticulate Risk Contour Plots and RF Model Details

This appendix contains additional toxicity risk contour plots generated by the random forest models.

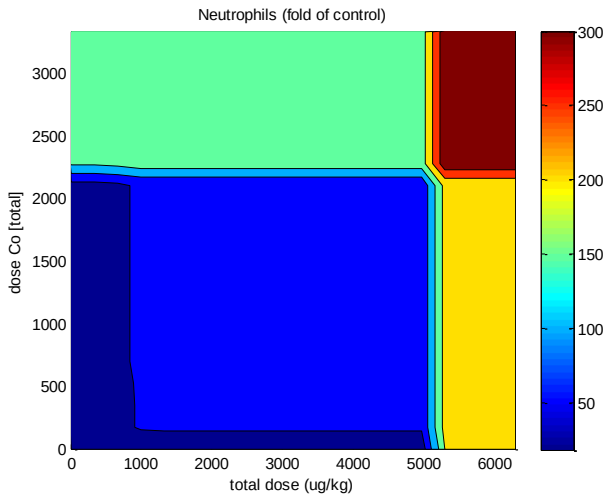

Figure C-1: Changes in RF model predicted BAL neutrophils count following exposure to carbon nanotubes as a function of changes in total dose and the dose of cobalt, a common toxic impurity (up to 0.53% by weight of total CNTs). This suggests that Co and total CNTs both independently contribute to higher neutrophils count.

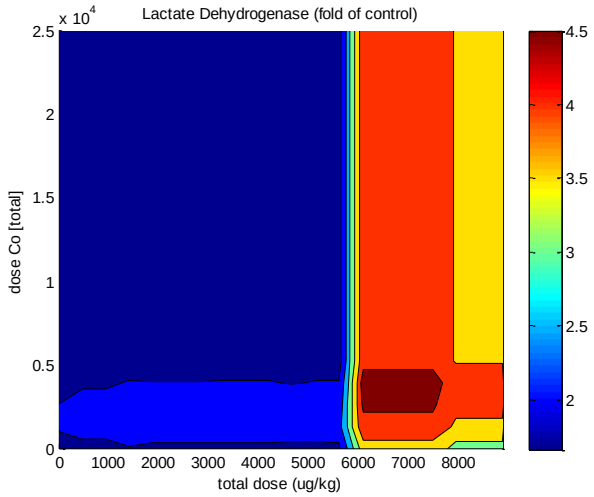

Figure C-2: Changes in RF model predicted BAL LDH following exposure to carbon nanotubes as a function of changes in total dose and the dose of cobalt, a common toxic impurity (up to 0.53% by weight of total CNTs). This suggests that total dose is much more important than Co content for increasing LDH concentration.

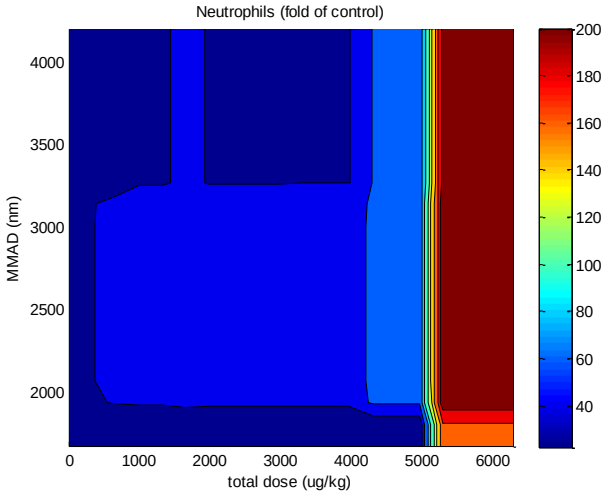

Figure C-3: Changes in RF model predicted BAL neutrophils count following exposure to carbon nanotubes as a function of changes in total dose and aggregation (MMAD, mass mode aerodynamic diameter). This suggests that aggregation only has a small effect on neutrophils count as compared to total dose, and also that low to moderate doses are relatively similar in response.

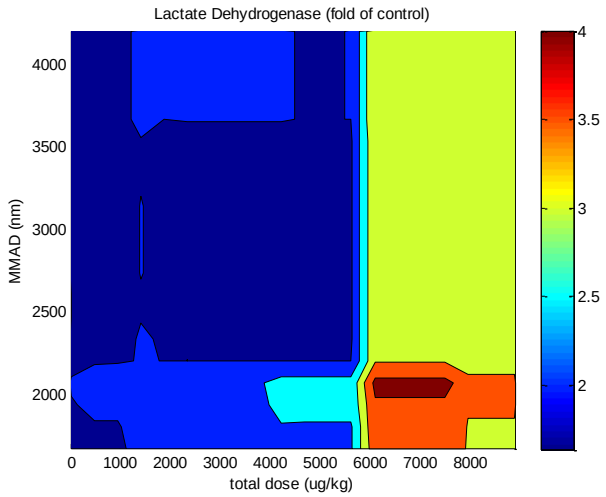

Figure C-4: Changes in RF model predicted BAL LDH following exposure to carbon nanotubes as a function of changes in total dose and aggregation (MMAD, mass mode aerodynamic diameter). This suggests both that total dose is a more important predictor of LDH than aggregation, but also that less aggregation can increase LDH as well.

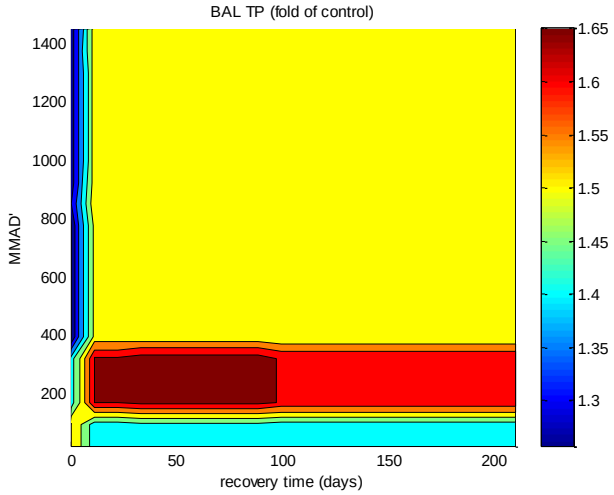

Figure C-5: Changes in RF model predicted BAL total protein following exposure to titanium dioxide nanoparticles as a function of changes in total dose and aggregation (MMAD, mass mode aerodynamic diameter). This suggests that low aggregation levels are more toxic than higher ones, and that recover time is not an important factor, but the scale of these differences in small overall.

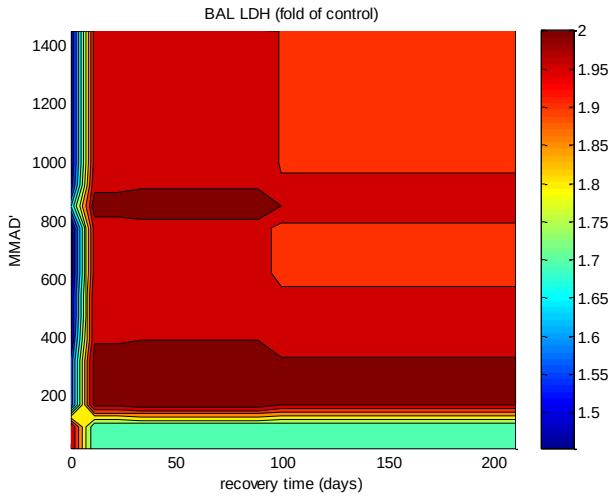

Figure C-6: Changes in RF model predicted BAL LDH following exposure to titanium dioxide nanoparticles as a function of changes in total dose and aggregation (MMAD, mass mode aerodynamic diameter). This suggests that in terms of predicting LDH response, neither recovery time nor aggregation is consistently detrimental or beneficial, so more data would be required.

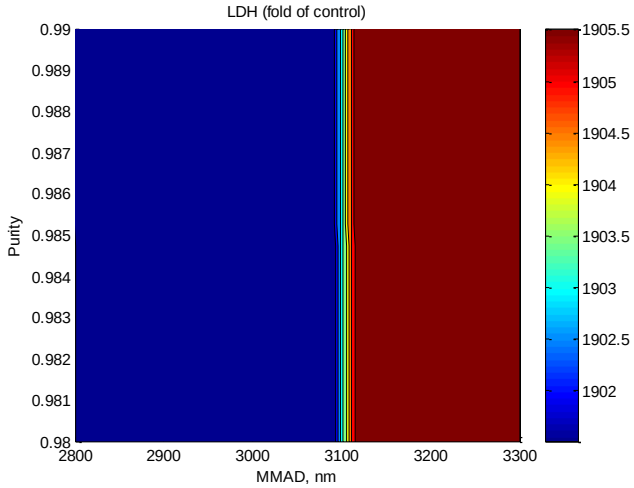

Figure C-7: Changes in RF model predicted BAL LDH following exposure to metal oxide nanoparticles (TiO_2, MgO, ZnO, SiO_2) as a function of changes in aggregation (MMAD, mass mode aerodynamic diameter), and purity. Based on the scale, there is little difference in toxicity across this range of variables indicating these factors play little role in toxicity.

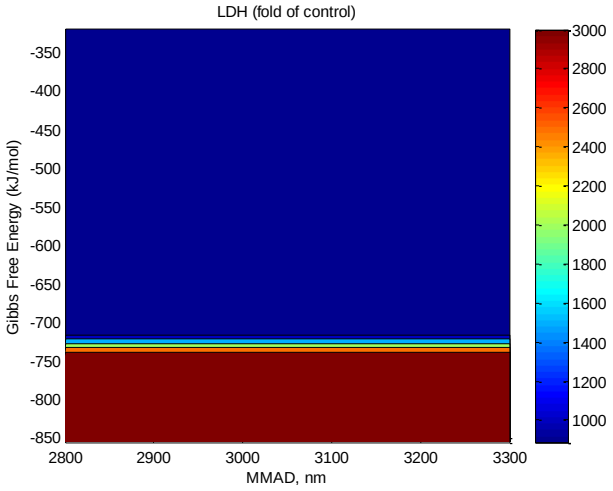

Figure C-8: Changes in RF model predicted BAL LDH following exposure to metal oxide nanoparticles (TiO_2, MgO, ZnO, SiO_2) as a function of changes in aggregation (MMAD, mass mode aerodynamic diameter), and Gibbs Free Energy, a descriptor of the chemical energy available in the metal oxide compound. This suggests that aggregation is much less important than differences in chemical makeup [other results, see Figure 5, indicate that Gibbs Free Energy is much less important than total dose].

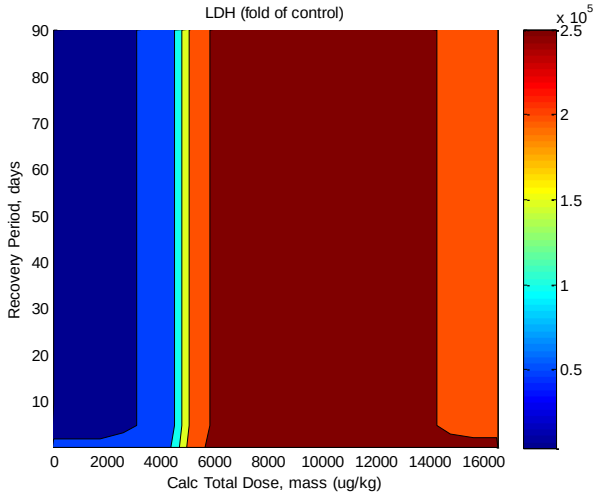

Figure C-9: Changes in RF model predicted BAL LDH following exposure to metal oxide nanoparticles (TiO_2, MgO, ZnO, SiO_2) as a function of total dose and recovery period. This indicates that total dose dominates the change in LDH due to longer recovery periods.

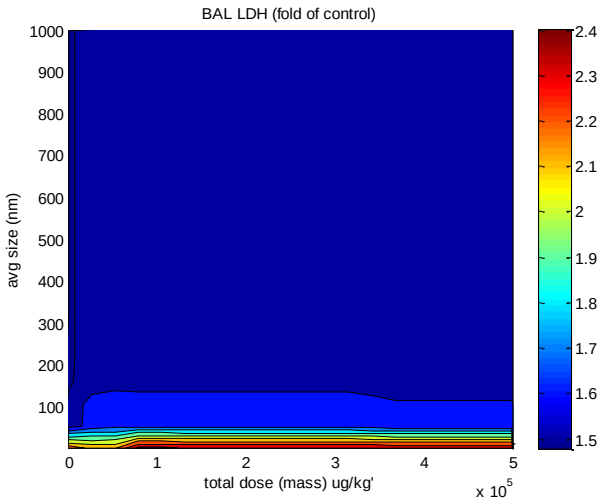

Figure C-10: Changes in RF model predicted BAL LDH following exposure to titanium dioxide nanoparticles as a function of total dose and average particle size. This indicates that very small TiO_2 nanoparticles are more toxic than those in most of the possible range of sizes.

This appendix contains details on the random forest model structures employed in this analysis, learning statistics, error, and goodness-of-fit metrics.

Random Forest Model Variable Importance

There are many different ways to represent the importance of variables in a random forest models. All of these methods consider the information gain achieved by each branch node summed by input variable and averaged across all of the regression trees in the forest. Some methods of calculating information gain include entropy, standardized mean difference, Gini coefficient, and variance reduction. Generally the results as calculated by these methods are very similar. We have chosen to utilize variance reduction as the primary information gain metric primarily due to comparability to other methods of evaluating different kinds of models.

The following figures (B1 through B6) display the internal structure of the RF models and their relative reliance on different input variables to reduce the error of the

models. The height of the columns is not directly analogous to magnitude of the change in outcome associated with a unit change in input—the magnitude of change in toxicity is better observed in the contour plots.

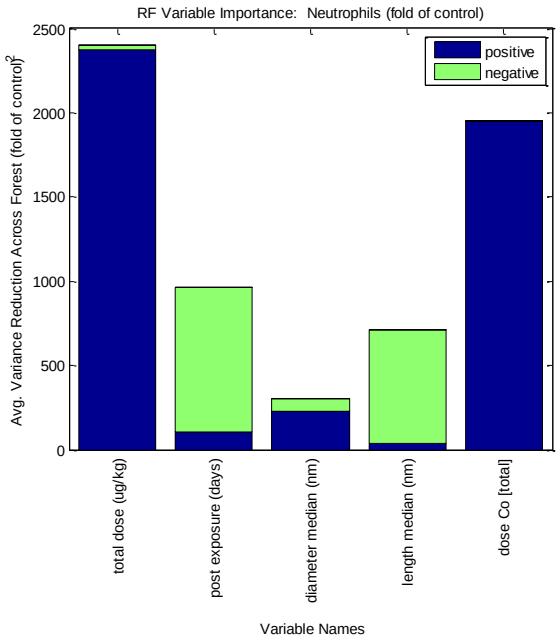

Figure C-11: RF model variable importance as measured by variance reduction attributable to each variable for the prediction of BAL neutrophils count following exposure to carbon nanotubes.

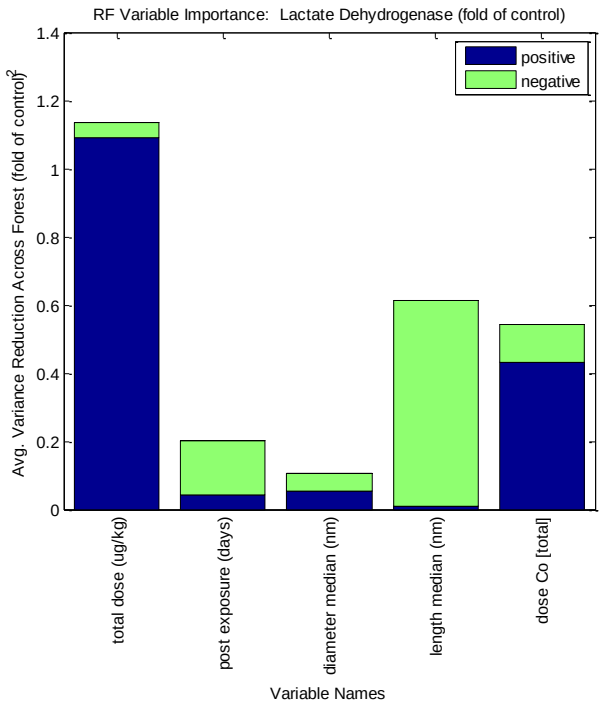

Figure C-12: RF model variable importance as measured by variance reduction attributable to each variable for the prediction of BAL LDH count following exposure to carbon nanotubes.

Column colors reflect whether changes in a given variable when applied to a branch split in the RF model were associated with a positive or negative change in the model output. For example, increases total dose is usually associated with increasing toxic responses, and increasing recovery time is usually associated with decreasing recovery time. Sometimes, these relationships can be complex or non-linear and result in a variable having different effects in different parts of the variable space.

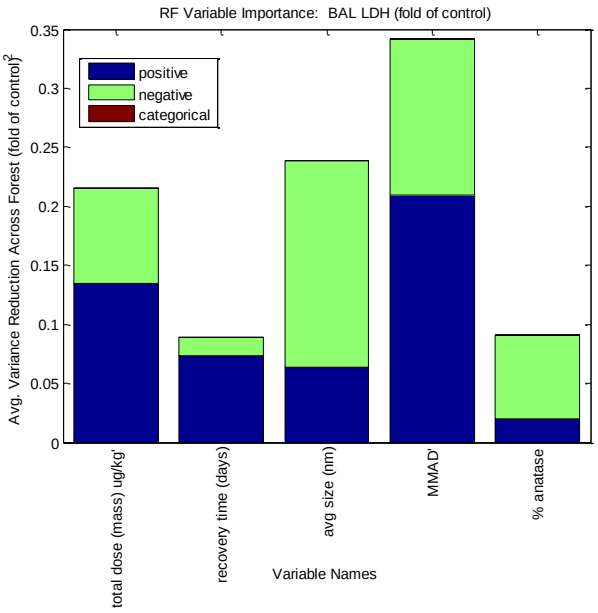

Figure C-13: RF model variable importance as measured by variance reduction attributable to each variable for the prediction of BAL LDH count following exposure to titanium dioxide nanoparticles.

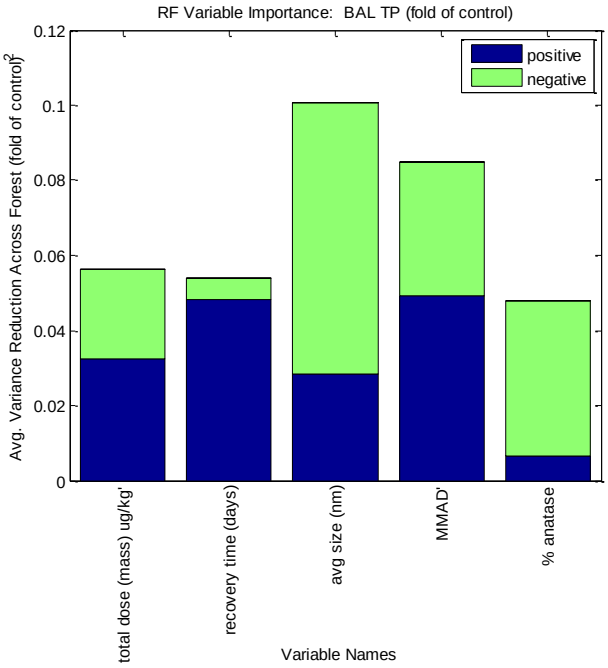

Figure C-14: RF model variable importance as measured by variance reduction attributable to each variable for the prediction of BAL total protein count following exposure to titanium dioxide nanoparticles.

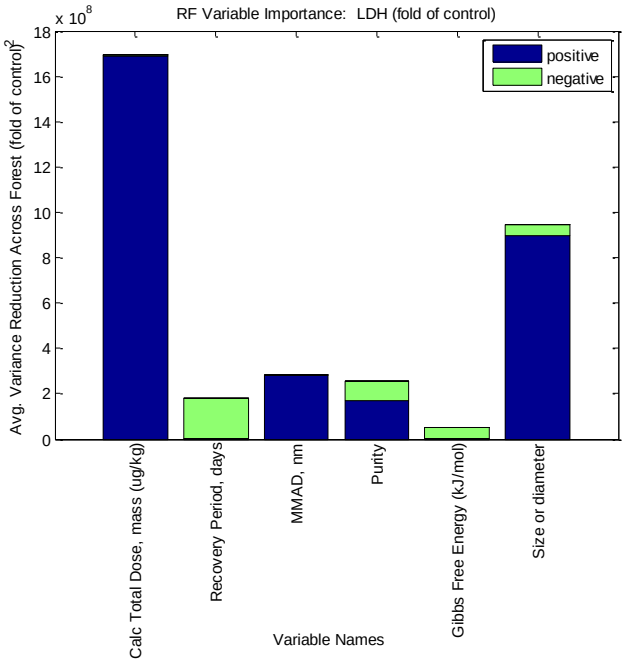

Figure C-15: RF model variable importance as measured by variance reduction attributable to each variable for the prediction of BAL LDH count following exposure to metal oxide nanoparticles include titanium dioxide, magnesium oxide, silicon dioxide,

Random Forest Model Learning Progression

These figures display the error of the RF model in predicting the actual observed values as additional regression trees are added to the forest. Each RF model contains 1,000 trees. As these graphs show, the minimal error state is usually realized by the model once the model has achieved a size of 100-200 trees.

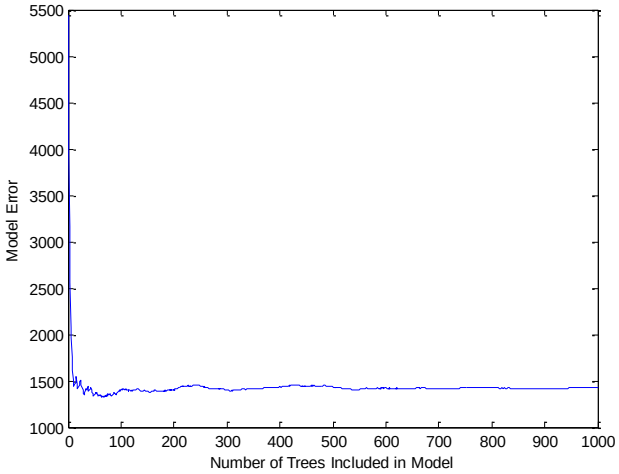

Figure C-16: RF model error as a function of trees included in model for prediction of BAL neutrophils following pulmonary exposure to carbon nanotubes.

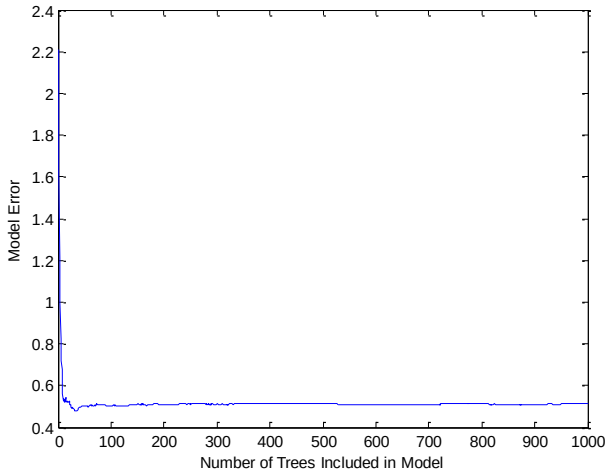

Figure C-17: RF model error as a function of trees included in model for prediction of BAL LDH following pulmonary exposure to carbon nanotubes.

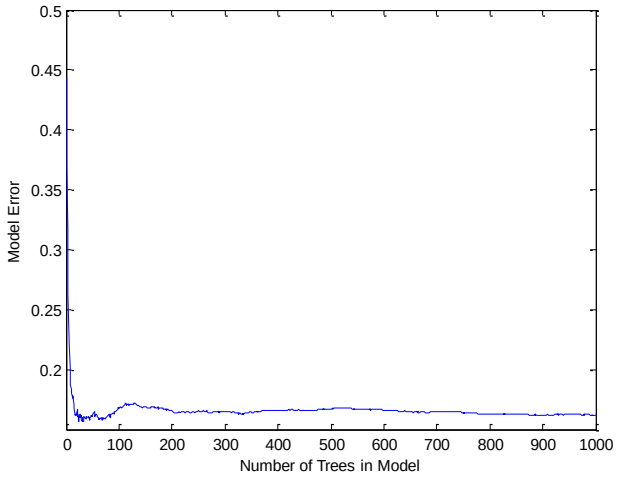

Figure C-18: RF model error as a function of trees included in model for prediction of BAL total protein following pulmonary exposure to titanium dioxide nanoparticles.

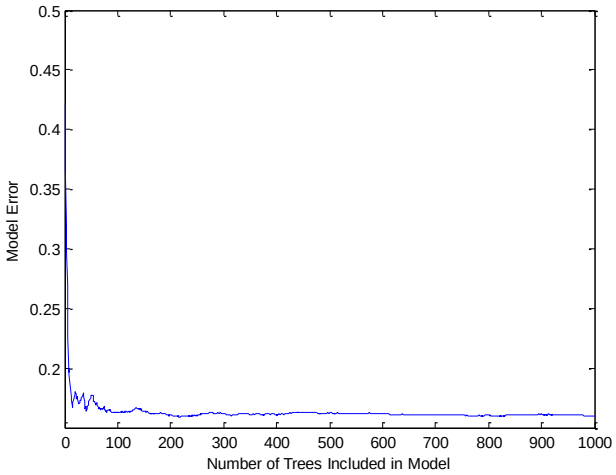

Figure C-19: RF model error as a function of trees included in model for prediction of BAL LDH following pulmonary exposure to titanium dioxide nanoparticles.

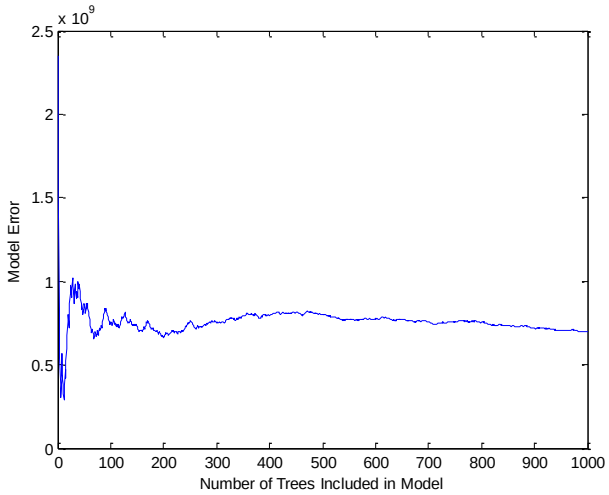

Figure C-20: RF model error as a function of trees included in model for prediction of BAL LDH following pulmonary exposure to metal oxide nanoparticles including titanium dioxide, magnesium oxide, silicon dioxide, and zinc oxide.