Taylor & Francis
Taylor & Francis Group

Inhalation Toxicology

International Forum for Respiratory Research

ISSN: 0895-8378 (Print) 1091-7691 (Online) Journal homepage: http://www.tandfonline.com/loi/iiht20

Prediction of lung function response for populations exposed to a wide range of ozone conditions

William F. McDonnell, Paul W. Stewart, Marjo V. Smith, Chong S. Kim & Edward S. Schelegle

To cite this article: William F. McDonnell, Paul W. Stewart, Marjo V. Smith, Chong S. Kim & Edward S. Schelegle (2012) Prediction of lung function response for populations exposed to a wide range of ozone conditions, Inhalation Toxicology, 24:10, 619-633, DOI: 10.3109/08958378.2012.705919

To link to this article: https://doi.org/10.3109/08958378.2012.705919

Inhalation Toxicology, 2012; 24(10): 619–633

ISSN 0895-8378 print/ISSN 1091-7691 online
DOI: 10.3109/08958378.2012.705919

informa
healthcare

RESEARCH ARTICLE

Prediction of lung function response for populations exposed to a wide range of ozone conditions

William F. McDonnell[1], Paul W. Stewart[2], Marjo V. Smith[3], Chong S. Kim[4], and Edward S. Schelegle[5]

[1]William F. McDonnell Consulting, Chapel Hill, NC, USA, [2]Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA, [3]SRA International, Durham, NC, USA, [4]NHEERL, U.S. Environmental Protection Agency, RTP, NC, USA, and [5]Department of Anatomy, Physiology, and Cell Biology, University of California, Davis, CA, USA

Abstract

Context: A human exposure–response (E–R) model previously demonstrated to accurately predict population mean FEV_1 response to ozone exposure has been proposed as the foundation for future risk assessments for ambient ozone.*Objective*: Fit the original and related models to a larger data set with a wider range of exposure conditions and assess agreement between observed and population mean predicted values.

Materials and methods: Existing individual E–R data for 23 human controlled ozone exposure studies with a wide range of concentrations, activity levels, and exposure patterns have been obtained. The data were fit to the original model and to a version of the model that contains a threshold below which no response occurs using a statistical program for fitting nonlinear mixed models.

Results: Mean predicted FEV_1 responses and the predicted proportions of individuals experiencing FEV_1 responses greater than 10, 15, and 20% were found to be in agreement with observed responses across a wide range of exposure conditions for both models. The threshold model, however, provided a better fit to the data than the original, particularly at the lowest levels of exposure.

Conclusion: The models identified in this manuscript predict population FEV_1 response characteristics for 18–35-year-old individuals exposed to ozone over a wide range of conditions and represent a substantial improvement over earlier E–R models. Because of its better fit to the data, particularly at low levels of exposure, the threshold model is likely to provide more accurate estimates of risk in future risk assessments of ozone-induced FEV_1 effects.

Keywords: Ozone, lung function, exposure–response, air pollution, risk assessment, model evaluation

Introduction

Ozone inhalation is known to induce reversible decrements in forced expiratory volume in one second (FEV_1) in humans, and increasingly sophisticated attempts to quantify the relationship between ozone exposure and response have been made over the past several decades (Silverman et al., 1976; Folinsbee et al., 1978; Adams et al., 1981; Colucci, 1983; Hazucha, 1987; Kriebel & Smith, 1990; Larsen et al., 1991; McDonnell & Smith, 1994; Smith & McDonnell, 1999). We recently identified a dynamic two-compartment model (the original model) that describes the relationship between FEV_1 response and inhaled ozone concentration (C), minute ventilation ($\dot{V}_E$), and duration of exposure while also accounting for the wide range of individual variability in response and the known individual risk factors (McDonnell et al., 2007, 2010). Using cross-validation techniques, we found that this model accurately predicted the mean observed responses to a wide range of exposure conditions for an existing dataset of 540 males for which we had individual data, and that it also predicted the mean responses for an independent set of seven published studies for which we had mean values for the exposure conditions (McDonnell et al., 2010).

More recently, the individual participant data from a set of seven published studies conducted at the University of California, Davis, which includes low as well as variable

Address for Correspondence: William F. McDonnell MD PhD, 1207 Hillview Road, Chapel Hill, NC 27514, USA. Tel.: 919–967-1219. E-mail: McDonnell.William@Earthlink.Net

(Received 27 May 2012; revised 20 June 2012; accepted 20 June 2012)

C exposures, and from a recently published study conducted at the U.S. Environmental Protection Agency of low *C* ozone exposures have been made available to us. These data will be more fully described in the methods. Furthermore, Schelegle et al. (2012), working from data that substantially overlaps the above mentioned data, have recently identified a model that shares some characteristics of our recent model and that also includes a threshold dose of ozone below which no response is observed.

The purposes of this study were to (1) merge the individual exposure and response measures from multiple studies into a single analysis dataset; (2) develop several new models derived from our original model form that incorporate a threshold; (3) fit the original model as well as related threshold models to the combined dataset; (4) evaluate the agreement between mean observed responses and mean population predicted responses for the original model as well as threshold models; and (5) evaluate the agreement between the observed and predicted proportions of individuals experiencing a 10, 15, and 20% FEV_1 decrement.

By estimating model coefficients from a large dataset with a wide range of exposure conditions our intent was to identify one or more exposure–response (E–R) models that will be useful as the foundation for improved future ozone risk assessment activities and for comparing lung function responses under different regulatory scenarios and ambient conditions.

Methods

Data

The data for this study comprise measures of FEV_1 before, during, and after ozone exposure, measures of ozone *C* and $\dot{V}_E$ during exposure, and measures of physical characteristics of participants from 23 previously conducted human controlled ozone exposure studies (20 published individually). Data were originally organized into three subsets, two of which were compiled from multiple studies conducted at two different laboratories, and one of which comprises data from a recently published study.

The original subset (EPA data) contains exposure and response information from 15 studies conducted at the U.S. Environmental Protection Agency Human Research Facility in Chapel Hill, NC, over the period from 1980 to 1993 and has been previously described in detail (McDonnell et al., 2007). Exposures fell into one of six categories: short-duration (2–2.5 h) exposures either at rest, with alternating periods of heavy exercise and rest, or with alternating periods of moderate exercise and rest; and long-duration (6.6–7.6 h) exposures either at rest, with moderate exercise interrupted by short periods of rest, or with light exercise interrupted by short periods of rest. Ozone concentrations ranged from 0.08 to 0.40 part per million (ppm). All exposures were conducted in an environmentally controlled chamber. A total of 540 males, ages 18–35 years participated in these studies and contributed 3483 measures of lung function response during or after exposure.

The UCD dataset contains exposure and response data from seven studies conducted at the Human Performance Laboratory at the University of California, Davis, and published between the years 1997 and 2009. The published exposure protocols of the UCD data have been described in detail (McDonnell et al., 2010) and these, as well as several previously unpublished exposure protocols conducted within these studies, have been more recently summarized in tabular form (Schelegle et al., 2012). The most prevalent exposure protocol involved long-duration (6.6 h) exposures with exercise interrupted by short periods of rest. A second protocol involved an 8-h exposure with alternating 30 min periods of rest and exercise. With a few exceptions, all protocols utilized exercise of moderate intensity with a target $\dot{V}_E$ of ~20 $L \cdot min^{-1} \cdot m^{-2}$ body surface area (BSA). Hourly ozone concentrations ranged from 0.03 to 0.24 ppm, with some exposures being at a constant square-wave concentration (SW) and others using changing levels of ozone either continuously (RAMP) or in hourly steps (STEP). Some exposures were conducted via facemask while others were conducted in an environmentally controlled chamber. Adams (2002, 2003) previously demonstrated that responses for the two methods of exposure are similar. The data from one subject were excluded from the UCD dataset due to obvious transcription errors from the original data which were no longer available. A total of 142 individuals (70 males, 72 females) ages 18–33 years contributed 7340 measures of response.

The EPA2 dataset comprised data from a single study conducted in a new controlled exposure chamber of the US EPA Human Studies Facility during the winters of 2008–2009 and 2009–2010 (Kim et al., 2011). Participants underwent long-duration (6.6 h) exposure with moderate exercise interrupted by short periods of rest. The first 30 volunteers were exposed to 0.06 and 0.08 ppm, and the remaining 29 were exposed to 0.06 ppm. In contrast to other studies, lung function was not measured after the first or second hour of exposure. A total of 59 individuals (27 M, 32 F) ages 19–35 years contributed 592 measures of lung function response.

All studies in the EPA, UCD, and EPA2 data subsets included filtered air (FA) control exposures, and all original studies were approved by the governing Institutional Review Boards. All participants were informed of the risks of participation and provided informed consent prior to participation. Participants were nonsmoking, generally healthy males (EPA data) or males and females (UCD and EPA2 data) between the ages of 18 and 35 years. Individuals were excluded for a history of asthma or other chronic respiratory disease with the exception of one study of individuals with mild seasonal allergic rhinitis who were currently asymptomatic and who were exposed outside of their allergy season (McDonnell et al., 1987). Overall a total of 741 individuals (637 male, 104 female) contributed 11,415

measures of lung function (8477 for ozone exposures and 2938 for FA exposures) for this analysis.

For the exposures conducted via facemask in several of the UCD studies, $\dot{V}_E$ was continuously measured during the periods of both rest and exercise, and the average of all measures for each period of rest or exercise was used to represent $\dot{V}_E$ for that individual for that period. For chamber exposures in the EPA, EPA2, and UCD studies, $\dot{V}_E$ during exercise was usually measured on one or two occasions during each exercise period, and the single measure or the average of two measures was used to represent $\dot{V}_E$ during the entire exercise period. For the chamber studies conducted completely at rest, a resting $\dot{V}_E$ /BSA of 5 L·min^{-1}·m^{-2} body surface area (BSA) was assigned to each individual (McGregor et al., 1961; Chadha et al., 1987). For chamber studies in which rest and exercise alternated, a resting $\dot{V}_E$ /BSA of 6 L·min^{-1}·m^{-2} BSA was assigned to each individual to reflect a higher level of resting metabolism consequent to recent exercise. This differentiation of resting $\dot{V}_E$ between the two groups is a change from our previous efforts (McDonnell et al., 2007, 2010).

The outcome variable of interest was the percent decline in FEV_1 from baseline (preexposure) due to ozone exposure. For each time point, for each individual, the largest of the multiple (generally three) FEV_1 trials was selected, and the percent decrement in FEV_1 from the preexposure baseline was calculated as 100% × [preexposure FEV_1−current FEV_1]/preexposure FEV_1.

Because many procedures inherent in the studies (e.g. exercise, repeated measurement of lung function, and the passage of time) may have independent effects upon FEV_1, the observed decrements in FEV_1 during ozone exposures represent the sum of these study effects and the effects of ozone itself. We wished to model only the effects that could be ascribed to ozone and therefore subtracted an unbiased estimate of the magnitude of the extraneous study effects observed during filtered air (FA) control exposures from the observed responses during ozone exposure as follows. For each study we calculated the mean (across participants) FEV_1 % decrement for each time point during the FA control exposures, and then, for each exposure to ozone in that study, subtracted this mean FA response from the observed FEV_1 response of each individual at the corresponding time point. It is these FA-adjusted decrements ($\%\Delta FEV_1$) that we model as the observed response to ozone exposure. Because the resulting $\%\Delta FEV_1$ had a mean value of zero for each time point of each FA exposure after subtracting the mean FA response we did not include the data from the FA exposures when the model was fit. This is also a change from our previous efforts.

Model specification

The original statistical model used to characterize the relationship between ozone exposure and response is described in the following two equations. More detailed descriptions and interpretations of this model and earlier versions have been previously published (Smith & McDonnell, 1999, McDonnell et al., 1994, 2007).

$$\frac{d(X_{ijk})}{dt} = C_{ijk} \times [(\dot{V}_E/\mathrm{BSA})^{\beta 6}]_{ijk} - \left(\beta 5 \times X_{ijk}\right) \qquad (1)$$

Here, X_{ijk} is the value of an intervening variable X for the i^{th} person, at the j^{th} time, in the k^{th} exposure protocol. The value of X increases over time as a function of the dose rate ($C \times \dot{V}_E$) and decreases over time according to first order reaction kinetics. One could interpret X to represent the level of oxidant stress resulting from accumulation and removal of ozone or its reactive byproducts, although the validity of the model is not dependent upon this interpretation. The initial ($T = 0$) value of $X_{ijk} = 0$. Time is in minutes, C is the instantaneous concentration in ppm, $\dot{V}_E$ is the instantaneous minute ventilation in L·min^{-1}, BSA is body surface area in m^2, the coefficient β5 is the inverse of the time constant for this differential equation, and β6 allows response to differ in sensitivity to C and $\dot{V}_E$. Note that for any given temporal profile of C and $\dot{V}_E$ /BSA this differential equation can be integrated over time and the temporal profile of X can be calculated if β5 and β6 are known.

$$\%\Delta FEV_{1ijk} = e^{(Ui)} \times M_{ijk} + E_{ijk} \qquad (2)$$

in which

$$M_{ijk} = \begin{Bmatrix} \left(\beta 1 + \beta 2\left[\mathrm{Age}_{ijk} - 23.8\right]\right) \Big/ \left[1 + \beta 4\, e^{(-\beta 3 Xijk)}\right] \\ -\left(\beta 1 + \beta 2\left[\mathrm{Age}_{ijk} - 23.8\right]\right) \Big/ \left[1 + \beta 4\right] \end{Bmatrix}$$

and is equal to the median of the population response distribution.

The right hand side of Equation 2 is an algebraic function of X that is modified by Age (centered on 23.8 years, the mean age of the sample) and U, and consists of 3 parts: the inter-subject variation in response (e^{Ui}); a sigmoid-shaped function (M_{ijk}) that describes the central tendency of response to ozone via the variable X_{ijk} with a mathematical adjustment to allow predictions of zero ozone-induced effect when $X = 0$ (e.g. at time = 0 and during FA exposures); and an error term (E_{ijk}) representing intra-subject variation. One could interpret this second equation as representing the neurally-mediated lung function response to the oxidant stress described in Equation 1, although again, the validity of the model is not dependent upon this interpretation. The random-effect variable U_i describes the individual responsiveness to ozone and is normally distributed with a mean assumed to equal zero. Equations 1 and 2 comprise a parametric, nonlinear statistical model for clustered data and completely specify a well-defined functional form for central tendency and for variance and correlation.

Exponentiation of the estimated population distribution of U values multiplied by the sigmoid function

results in a log normal distribution of possible individual responses with a lower limit of zero for any given age and exposure. Note that for any given temporal profile of X values calculated from Equation 1, one can calculate the temporal profile of predicted $\%\Delta FEV_1$ for a volunteer with average ozone responsiveness (i.e., $U_i = 0$). Once the variance of U is estimated from the data, the entire frequency distribution of individual responses (for given age and exposure conditions) can be described as well as any summary measures of the population response such as the mean, the median, the proportion of individuals with response greater than a given magnitude, and percentiles of response. The population mean of the distribution of predicted individual responses for those of given age and exposure conditions can be calculated as exp(variance $[U]/2) \times M_{ijk}$ based on Equation 2.

We also developed and fit a modification of the original model that assumes no FEV_1 response occurs unless X in Equation 1 exceeds a threshold (β9). This threshold model is identical to the initial model with the exception that the value of $(X_{ijk})_{Th}$ defined in Equation 3 is substituted into Equation 2 in place of X_{ijk}. The value of $(X_{ijk})_{Th}$ does not become nonzero until X_{ijk} in Equation 1 exceeds the threshold β9. Other versions of the threshold model allowing β9 to vary with age or to be treated as a random effect were also fit.

$$\left(X_{ijk}\right)_{\mathrm{Th}} = \left(X_{ijk} - \beta 9\right) \times I \qquad (3)$$

in which I has a value of 1 if $X_{ijk} > \beta 9$; otherwise, I has a value of 0.

In order to characterize the association between body mass index (BMI) calculated as wt (kg)/ht^2 (m) and FEV_1 response, we replaced Equation 2 with Equation 4 for the original model as well as for the threshold model. The value of BMI is centered around the sample mean BMI of 23.1 kg·m^{-2}.

$$\%\Delta FEV_{1ijk} = e^{(Ui)} \times M_{ijk} + E_{ijk} \qquad (4)$$

in which

$$M_{ijk} = \left\{ \begin{array}{l} \left(\beta 1 + \beta 2\left[\mathrm{Age}_{ijk} - 23.8\right] + \beta 8\left[\mathrm{BMI} - 23.1\right]\right) \Big/ \left[1 + \beta 4\, e^{(-\beta 3 X ijk)}\right] \\ -\left(\beta 1 + \beta 2\left[\mathrm{Age}_{ijk} - 23.8\right] + \beta 8\left[\mathrm{BMI} - 23.1\right]\right) \Big/ \left[1 + \beta 4\right] \end{array} \right\}$$

and is the median of the population response distribution.

Model fitting

The iterative model-fitting algorithms used were implemented using PROC NLMIXED (SAS 9.2, SAS Institute, Inc.), a procedure specially designed for fitting nonlinear random-effects models. Extensive grid searching was employed. The optimization algorithms used for maximizing the likelihood function included "Newton-Raphson with Ridging", "Trust Region", and "Dual Quasi-Newton". The integration method used within this algorithm was the adaptive Gaussian quadrature method. Model fitting provided statistical estimates of the primary model parameters (β1 . . . β9, variance of U, and variance of E). These maximum likelihood estimators (MLEs) of the primary parameters were used to compute confidence intervals and to compute MLEs of secondary parameters such as predicted values and the probability that an individual would experience an effect greater than a given magnitude.

Calculation of predicted values

Predicted values for comparison with observed values at each data point were calculated in two ways: (1) For each person for whom exposure and response data were available as model input, PROC NLMIXED provided an empirical best linear unbiased predictor (eBLUP) of U_i (U_{eBLUP}) that was used to compute an individual-specific predicted value for each data point for that individual: $\exp(U_{\mathrm{eBLUP}}) \times M_{ijk}$. These estimates of the individual's central tendency provide the most optimistic description of the fit of the model to the data, but are not useful for predicting responses of individuals who were not members of the input data set. (2) We calculated the predicted value for each individual at each data point as the population mean of the distribution of expected responses for all individuals of the same age with the same exposure conditions: exp(Variance(U)/2) × M_{ijk}. This population mean is useful for predicting responses of individuals regardless of whether they are members of the input dataset.

The probabilities that a randomly selected individual of given age and exposure conditions will experience a $\%\Delta FEV_1$ change greater than 10, 15, and 20%, respectively, at any time point were calculated from the estimated primary model parameters in two ways: (1) by direct computation of an exact double-integral expression for the probability of interest; and (2) by repeated sampling from the estimated distributions of U and E (both assumed to be normally distributed with mean zero and with variance estimated from the data) as follows. For each time point of each exposure for each participant (n = 8477), the median (M_{ijk}) response was calculated from Equation 2. Up to 100,000 values of U were then randomly drawn from the estimated frequency distribution of U, and for each value of U drawn, a value of E was randomly selected from the estimated distribution of E. For each selection, the predicted $\%\Delta FEV_1$ was calculated from Equation 2. The proportion of sampled values with $\%\Delta FEV_1$ greater than 10, 15, or 20% thus provided a Monte Carlo estimate of the probability that an arbitrary new individual of given age and exposure will experience an effect greater than the given magnitude. Note that when the standard error of the estimated proportions reached 10^{-6} and the proportions no longer changed with increasing selections of U, the simulation was terminated and the last calculated values used. When applied to these data the direct method and the simulation method provided identical

answers to at least four decimal places. The simulation method was used to estimate such probabilities for each study participant for each data point. The average of the estimated probabilities of all participants at a given data point provided a predicted value for the proportion of the sample that will experience such changes at that data point (total data points = 365).

Model evaluation

Models were evaluated by plotting observed responses against predicted responses for (1) individual data points; (2) the means (across participants) of individual responses at each time point of each exposure of each study; and (3) the proportions of individuals experiencing FEV_1 decrements greater than 10, 15, and 20% at these time points. Unless otherwise noted, the predicted value for population mean was used in all comparisons with observed values as they are more relevant to ultimate use of this model for predicting response in independent samples. Unfortunately, there is no statistically valid, simple statistic (such as an R^2) for evaluating the fit of the model to the observed data for nonlinear mixed effects models. For visual assessment of goodness-of-fit we plotted observed values versus predicted values ignoring any aspects of repeated measures or unequal weights among the groups. Simple linear regression for these plots provided estimates of slopes, intercepts, and a crude pseudo R^2. At best, these are imperfect descriptors of the goodness-of-fit of the nonlinear mixed model. We also plotted average observed responses (averaging across participants) and average predicted responses against time for each exposure of each study in order to look for areas of the E–R surface where the model might not predict well. Similar plots for the observed and predicted proportions with response greater than 10, 15, and 20% versus time were also examined. We examined plots of model residuals against predicted values, independent variables, and other subject characteristics such as height. Akaike's Information Criteria (AIC) was computed for comparing the various models for goodness-of-fit.

Comparisons among predictions of the various models were made by plotting the predicted values from one model against the predicted values from another model, both at the individual and group levels. Linear regression was performed, again ignoring any aspects of repeated measures or unequal weights. Bland-Altman plots of agreement between predictions were also examined.

Results

Personal characteristics of the participants in each of the three subsets of studies are presented in Table 1. The volunteers in the EPA and EPA2 data were, on average, older than those in the UCD data. The EPA males had lower BMI and weight than males in EPA2 and UCD although BSA was equal for the three groups. The UCD females were taller and lighter with lower BMI than the EPA2 females although BSA was equal for the two groups. Males overall were older than females although ages were equal for males and females in the UCD data, and as expected males were taller and heavier with larger BSA and BMI than females.

For each exposure within the UCD and EPA2 studies, each of which included approximately equal numbers of males and females, activity level during exercise was controlled to produce equal values of $\dot{V}_E$ /BSA for both males and females with a target value of 20 $L \cdot min^{-1} \cdot m^{-2}$ BSA for most exposures. The means of the measured values across all exposures were 20.0 and 19.8 for the males and females, respectively, while the measured values of $\dot{V}_E$ were 39.2 and 33.0 $L \cdot min^{-1}$ reflecting the larger BSA of the males. For each time point of each exposure condition of each of these studies (n = 229 data points), the mean response for the males and for the females was calculated. A scatter plot of these observed mean male responses versus the corresponding observed mean female responses

Table 1. Characteristics of volunteers exposed to ozone for three subsets and for the combined data.

	EPA	UCD		EPA2		Combined		
Data subsets	Males	Males	Females	Males	Females	Males	Females	Total
Participants (n)	540[a]	70	72	27	32	637	104	741
%ΔFEV_1 obs. (n) (all exposures)	3483	3696	3644	276	316	7455	3960	11,415
%ΔFEV_1 obs (n) (ozone exposures)	2445	2856	2860	168	188	5469	3008	8477
Age (year)[b]	24.1 (18–36)	22.3 (18–33)	22.2 (18–31)	26.1 (19–35)	24.0 (19–31)	24.0 (18–36)	22.7 (18–31)	23.8 (18–36)
Ht (cm)[b]	180.6 (158–208)	179.1 (161–191)	166.7 (147–180)	178.9 (167–193)	164.7 (153–173)	180.4 (158–208)	166.1 (147–180)	178.3 (147–208)
Wt (kg)[b]	75.4 (53–116)	78.6 (55–105)	60.8 (44–83)	78.8 (56–105)	63.4 (47–97)	75.9 (53–116)	61.6 (44–97)	73.8 (44–116)
BSA ($L \cdot min^{-1} \cdot m^{-2}$)[b]	2.0 (1.6–2.5)	2.0 (1.6–2.3)	1.7 (1.3–2.0)	2.0 (1.7–2.3)	1.7 (1.4–2.1)	2.0 (1.6–2.5)	1.7 (1.3–2.1)	1.9 (1.3–2.5)
BMI ($kg \cdot m^{-2}$)[b]	23.1 (17–35)	24.5 (19–32)	21.8 (18–27)	24.6 (17–32)	23.3 (19–33)	23.3 (17–35)	22.3 (18–33)	23.1 (17–35)

%ΔFEV_1 obs., number of observations of FEV_1 decrements; Ht, height; Wt, weight; BSA, body surface area; BMI, body mass index; FEV_1, forced expiratory volume in 1 s.

[a]55 of these participants had only filtered air exposures.

[b]Values are means (range). Calculation of means and ranges utilizes only one measure per volunteer from his/her earliest participation.

demonstrates that, in general, the male responses were nearly equal to the female responses for equivalent exposures (with $\dot{V}_E$ adjusted for BSA) with most data clustered along the line of identity (Figure 1). This indicates that use of the dose rate metric ($C \times \dot{V}_E$ /BSA) provides adequate control for sex differences in response in these data. Most of the studies in Figure 1 included approximately 15 males and 15 females at each time point. As expected, the data points for which the male and female responses were most discordant tended to be those from studies or time points with small ($n < 8$) numbers of each sex (not identified on the graph).

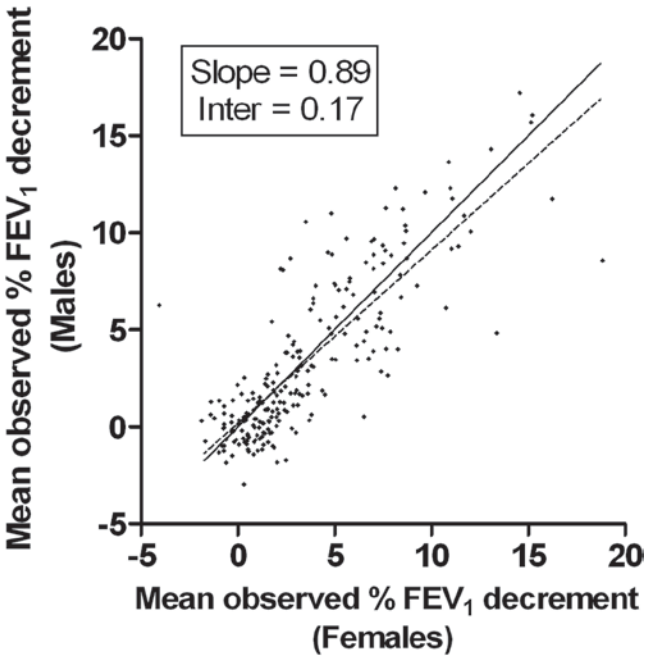

Figure 1. Mean observed ozone-induced FEV_1 decrements for males versus those for females for each time point and exposure condition of each of the UCD and EPA2 studies (n = 229 data points). Solid line: identity; dashed line: regression. (See colour version of this figure online at www.informahealthcare.com/iht)

The parameter estimates of the original and threshold models fit to the combined data (Rows 1 and 2 of Table 2, respectively) were similar with the exception of a difference in β4 due to a relative shift of the sigmoid curve induced by the introduction of a threshold. With the exception of the age coefficient (β2) all regression coefficients including β9 (the threshold in the threshold model) and the variance of U were statistically significantly different from zero in both models. The approximate two-sided 95% confidence interval for the exponent β6 barely included 1 in both models. Akaike's Information Criteria (AIC) was lower (indicating better fit) for the threshold model (49,594) than for the original model (49,614).

As noted in the methods, predicted values for each time point of each exposure of each study were calculated for each individual in two different ways for comparison with observed values. Using the eBLUP method, which provides estimates of each individual's response for the given time point, and which demonstrates the most optimistic fit of the model to the current data, one sees a wide range of individual calculated responses that are in agreement with the observed responses (n = 8477) for the original model (Figure 2A). The averages (n = 365) of these individual calculated and observed values at each time point of each exposure of each study are also in agreement (Figure 2B).

The second method, which can be used to generate predictions for individuals not in the dataset, estimates the frequency distribution of response for the population of individuals of same age and exposure conditions. In this instance, for each individual we calculate the mean of his/her response distribution (the population mean) at each data point. As expected, this results in a narrower range of the individual predicted responses (n = 8477), and although there is less correlation between individual observed and predicted responses, overall the data follow the line of identity (Figure 3A). The average (n = 365) of the individual observed versus the average of the individual population mean predicted response at each time

Table 2. Estimated coefficients and statistical parameters for various models fit to the combined data[a].

Model	β1	β2	β3	β4	β5	β6	β8	β9	Var (U)	Var (E)	AIC
Original	9.8057	−0.1907	0.01839	65.826	0.003191	0.8753	–	–	0.9449	17.120	49,614
	(0.7422)	(0.2807)	(0.00512)	(12.469)	(0.000214)	(0.0863)			(0.0833)	(1.177)	
Threshold	10.916	−0.2104	0.01506	13.497	0.003221	0.8839	–	59.284	0.9373	17.0816	49,594
	(0.8446)	(0.3100)	(0.00333)	(4.734)	(0.000207)	(0.0647)		(10.192)	(0.0824)	(1.1506)	
Original	9.942	−0.2593	0.01823	65.862	0.003193	0.8776	0.4814	–	0.9251	17.114	49,604
(+BMI)	(0.753)	(0.2316)	(0.00506)	(12.470)	(0.000215)	(0.0860)	(0.3251)		(0.0828)	(1.164)	
Threshold	11.092	−0.2873	0.01486	13.442	0.003224	0.8867	0.5467	59.963	0.9166	17.076	49,583
(+BMI)	(0.866)	(0.2542)	(0.00326)	(4.710)	(0.000207)	(0.0644)	(0.3687)	(10.286)	(0.0805)	(1.139)	
Original	10.679	−0.4278	0.01805	46.635	0.003724	0.8653	–	–	0.8754	15.069	14,543
(EPA data)[a]	(0.826)	(0.1194)	(0.00435)	(11.185)	(0.000392)	(0.0743)			(0.0855)	(1.258)	

Original model is as defined in Equations 1 and 2. Threshold model is defined in Equations 1, 2, and 3, and models with BMI (body mass index) in Equation 4.

AIC, Akaike's Information Criteria; Var, estimated variance of parameter.

Values in parenthesis are standard errors of the estimates. Confidence limits (95%) for each parameter are approximately equal to the estimate ± 2SE.

[a]All models are fit to the combined EPA, UCD, and EPA2 data with the exception of Row 5 which is fit only to the EPA data. Filtered air responses are not included in the data fit by the models.

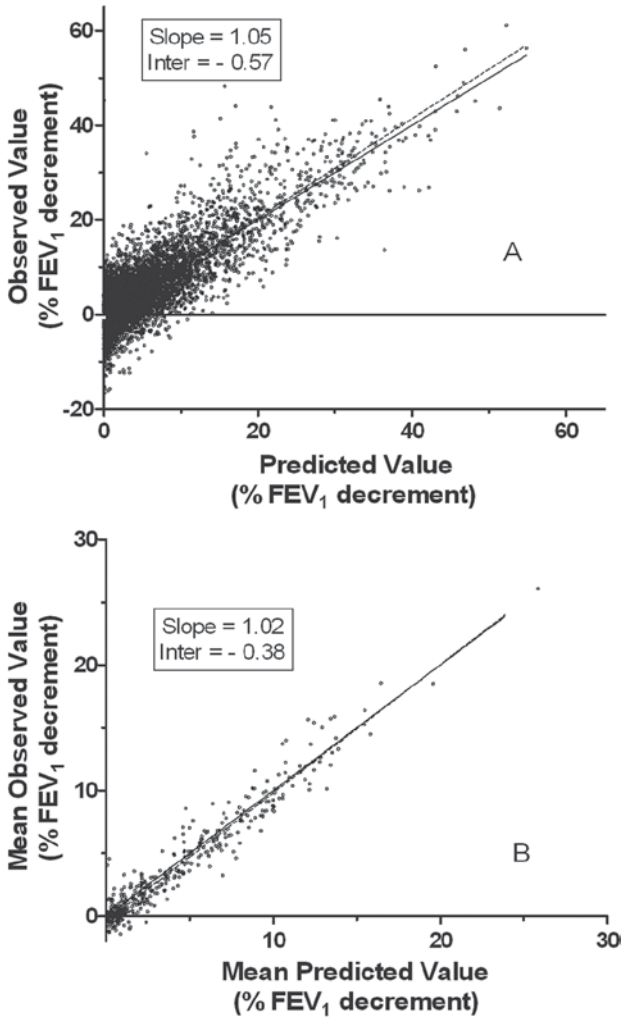

Figure 2. (A) Observed versus predicted individual FEV_1 decrements (n = 8477). (B) Mean observed versus mean predicted FEV_1 decrements. Means (across subjects) are calculated for each time point of each exposure condition of each study (n = 365 data points). Both figures use the empirical best linear unbiased predictor of U for each individual for calculation of predicted values. Solid line: identity; dashed line: regression.

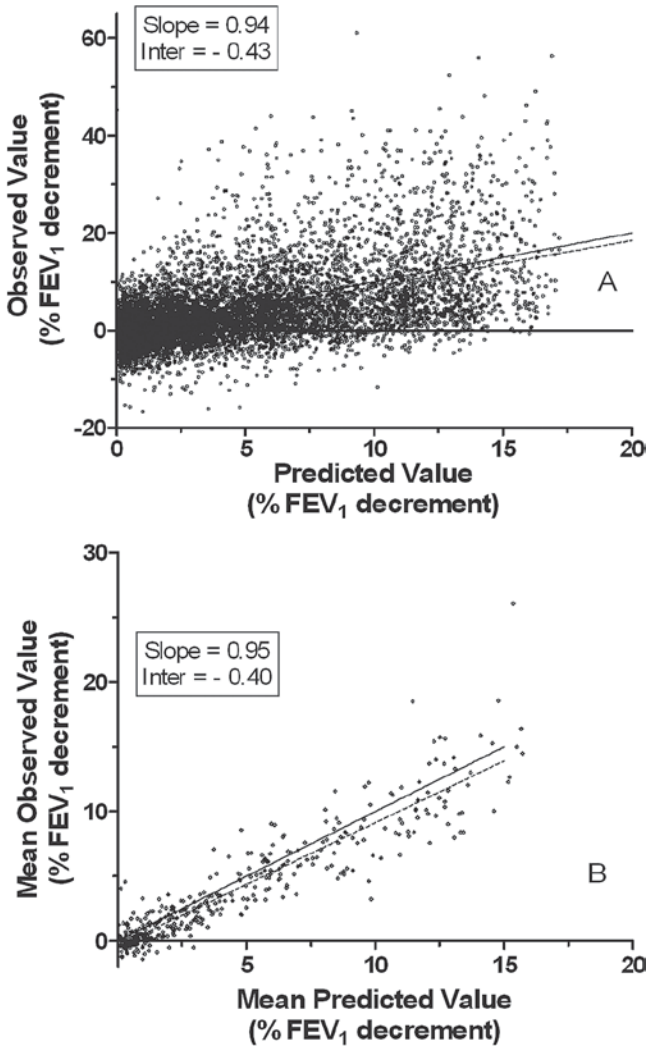

Figure 3. (A) Observed versus predicted individual FEV_1 decrements (n = 8477). (B) Mean observed versus mean predicted FEV_1 decrements. Means are calculated for each time point of each exposure condition of each study (n = 365). Both figures use the population mean of U for calculation of predicted values. Solid line: identity; dashed line: regression.

point for each exposure of each study are tightly clustered around the line of identity (Figure 3B). In general, there appears to be a slight overprediction of individual and mean responses for the original model using the population mean method. The observed and predicted proportions of individuals at each data point (n = 365) with greater than 10, 15, and 20% decrements in $\%\Delta FEV_1$ (Figure 4A, 4B and 4C) are generally in agreement.

The predicted values generated by the original and threshold models across the range of exposure conditions are very similar for the various metrics of response and methods of calculating predicted values described above. Linear regression of the predicted values of the original model against the predicted values of the threshold model yielded slopes that vary between 0.99 and 1.01, intercepts that vary between 0.00 and 0.09, and a pseudo R^2 of 1.00 for all comparisons. Consistent with that finding, the levels of agreement between observed and predicted responses, reflected by the slopes, intercepts, and pseudo R^2 values of the regression lines for observed versus predicted responses, for the threshold model are similar to those of the original model for the various metrics of response (Table 3). In general, however, for individual and for mean responses, the intercepts of the threshold model are somewhat closer to zero than those of the original model which suggests slightly better agreement between observed and predicted responses, including at low levels of exposure. This is consistent with the lower AIC of the threshold model compared with the original model.

Plots of averages (across participants) of observed and averages of population mean predicted responses versus time for each exposure of each study were examined and found to be similar for both models with the agreement between observed and predicted responses being somewhat better for the threshold model at the

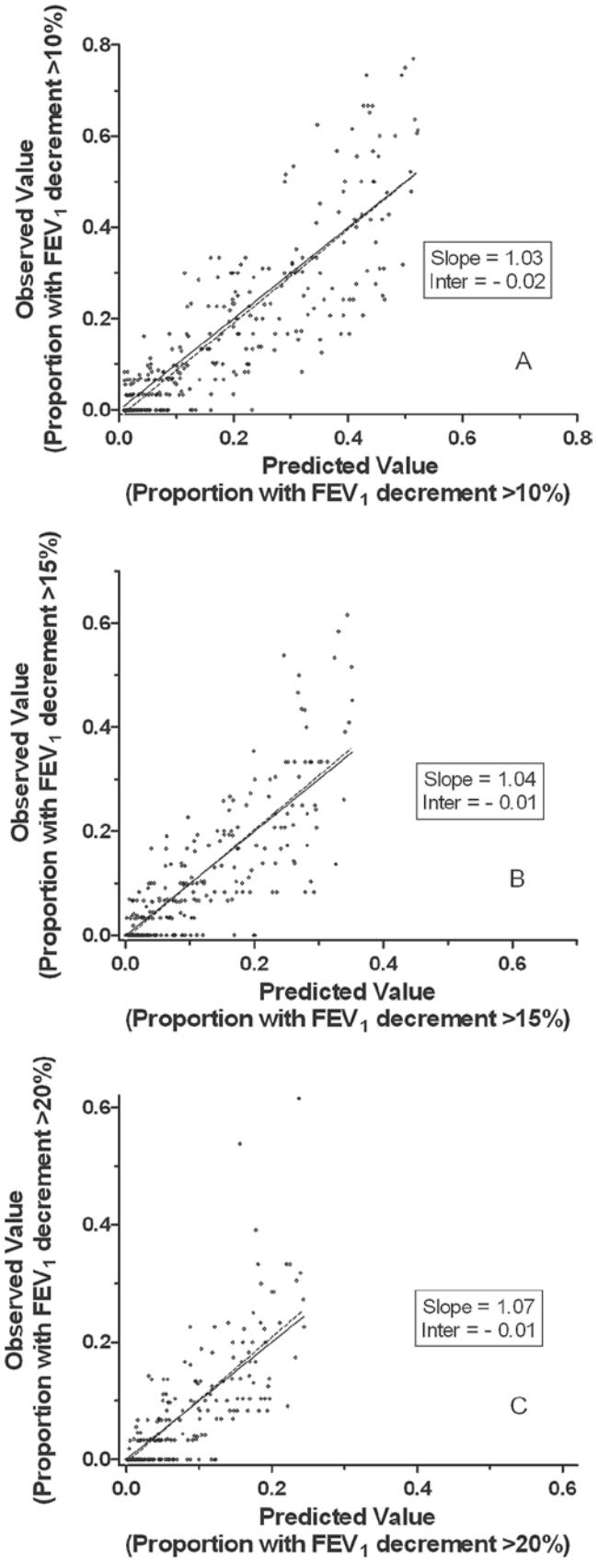

Figure 4. Observed versus predicted proportions of individuals experiencing an FEV_1 decrement greater than 10% (A), 15% (B), or 20% (C). Proportions are calculated for each time point of each exposure condition of each study (n = 365). Solid line: identity; dashed line: regression.

earliest time points of the low concentration exposures due to lower predictions at these time points. The plots of observed and predicted values versus time for this current analysis are very similar to those presented for the EPA studies and for the UCD studies in the McDonnell et al. (2007, 2010) manuscripts, respectively, and are not shown here. With the following exceptions the agreement between observed and predicted responses was generally very good across the range of exposures. Both models overpredicted the responses of the 8-h square-wave (SW) and variable *C* (RAMP) 0.12 ppm exposures of the Adams (2006b) study (see Figure 2 of McDonnell et al., 2010). The models captured the temporal dynamics of response in those exposures, however, and both either accurately predicted or underpredicted responses of other exposures with similar magnitudes of observed response. These include the responses to several 6.6-h 0.12 ppm SW and variable *C* (STEP) exposures and to a 0.87 ppm 6.6-h STEP exposure. Both models tended to slightly overpredict response of the intermediate time points of the SW and STEP 0.06 ppm UCD studies (see Figures 4, 5, and 6 of McDonnell et al., 2010), but accurately predicted the earliest and latest time points of these exposures and most responses of the 0.07 and 0.08 ppm 6.6-h SW and STEP exposures. These differences are partially due to the greater recovery that appears to occur during the lunch period of the 6.6-hr exposures in the Adams studies compared to the EPA studies. Both models also slightly overpredicted the responses for 6.6-hr resting exposures to a high concentration of ozone (0.24 ppm). It is not clear whether these few discrepancies represent areas of the E–R surface where the model is deficient or simply reflect the variability in response observed among subjects and studies. Similar plots of response versus time for the EPA2 data (Kim et al., 2011) which were not available for the earlier manuscripts but are of interest to the standard setting process are presented in Figure 5A and 5B. The first 30 study participants (Figure 5A) were exposed to FA, 0.06 ppm, and 0.08 ppm and the last 29 (Figure 5B) were exposed to FA and 0.06 ppm only. Observed responses were smaller than predicted responses for both exposure conditions for the first 30 participants although the patterns of observed responses with C and T were consistent with predicted. Observed responses to 0.08 ppm in this group of participants were also generally smaller than those observed for 0.08 ppm exposures in prior studies (Horstman et al., 1990; McDonnell et al., 1991; Adams 2002, 2003, 2006a). There was relatively good agreement between the magnitudes of the observed and predicted responses for the latter 29 volunteers, although the expected temporal pattern in observed responses was not evident.

The estimated coefficients of the original model and the simple threshold model that included BMI (Equation 4) are presented in Rows 3 and 4 of Table 2, respectively. Inclusion of BMI resulted in a slightly more negative age coefficient ($\beta 2$) with little change in the

other parameters. The BMI coefficient (β8) was not statistically significantly different from zero in either model, but its inclusion resulted in a reduction in the AIC for both models. Although addition of BMI to the models did change some individual predicted responses, it did not meaningfully change the overall relationships between observed and predicted responses (Table 3). The threshold coefficient (β9) remained statistically significantly different from zero when BMI was included, and the AIC of the threshold model containing BMI was lower than that for the corresponding non-threshold model (compare lines 3 and 4 of Table 2) indicating a better fit of the threshold model to the data.

Several other models that allowed the threshold (β9) to be a random effect or to be a function of age were successfully fit. In neither case were additional terms statistically significant, and we concluded that the existing data were not adequate to support these additional terms.

Because the threshold model provides a better fit to the data than the original model, we present additional figures demonstrating the performance of this model. In Figure 6, we plot both observed and predicted responses versus time for the exposures in the dataset with the lowest ozone concentrations which are of current regulatory significance. The Adams (2006a) study included a 0.06 ppm SW exposure and STEP exposures averaging 0.04 and 0.06 ppm. Target hourly *C*s were 0.03, 0.04, 0.05, 0.05, 0.04, and 0.03 ppm for the 0.04 STEP exposure and 0.04, 0.07, 0.09, 0.7, 0.05, and 0.04 ppm for the 0.06 STEP

Table 3. Linear regression descriptors for the relationships between observed and predicted responses for various models, levels of outcome, and prediction methods.

Outcome Level	Individual	Mean	Individual	Mean	P > 10%	P > 15%	P > 20%
Prediction Method	eBLUP	eBLUP	Population Mean	Population Mean	Simulation	Simulation	Simulation
Model							
Original							
Slope	1.05	1.02	0.94	0.95	1.03	1.04	1.07
Intercept	−0.57	−0.38	−0.43	−0.40	−0.02	−0.01	−0.01
Psuedo R^2	0.72	0.93	0.28	0.85	0.78	0.73	0.67
Threshold							
Slope	1.04	1.00	0.94	0.94	1.03	1.05	1.08
Intercept	−0.47	−0.25	−0.30	−0.26	−0.02	−0.01	−0.01
Psuedo R^2	0.71	0.93	0.28	0.86	0.78	0.74	0.68
Original (+BMI)							
Slope	1.05	1.02	0.93	0.95	1.03	1.04	1.06
Intercept	−0.57	−0.38	−0.41	−0.42	−0.02	−0.01	−0.01
Psuedo R^2	0.72	0.93	0.28	0.85	0.77	0.73	0.66
Threshold (+BMI)							
Slope	1.04	1.00	0.93	0.94	1.03	1.05	1.08
Intercept	−0.47	−0.25	−0.28	−0.28	−0.02	−0.01	−0.01
Psuedo R^2	0.71	0.93	0.28	0.85	0.78	0.74	0.67

Individual refers to individual $\%\Delta FEV_1$ decrements at each time point for each exposure. Mean refers to the mean (across participants) of all individual decrements at each point. eBLUP, empirical best linear unbiased predictor. Models are as described in Table 2. A "true" R^2 that represents the proportion of variance explained by the model does not exist for nonlinear mixed effect models. For descriptive purposes only, slopes, intercepts and the psuedo R^2 are calculated for the linear regression of observed versus predicted responses ignoring aspects of repeated measures. p >10, >15, and >20% are the proportions of individuals with $\%\Delta FEV_1$ values greater than 10, 15, and 20%, respectively, for each time point for each exposure. See Figures 2–4 for graphical representation of the Original model results.

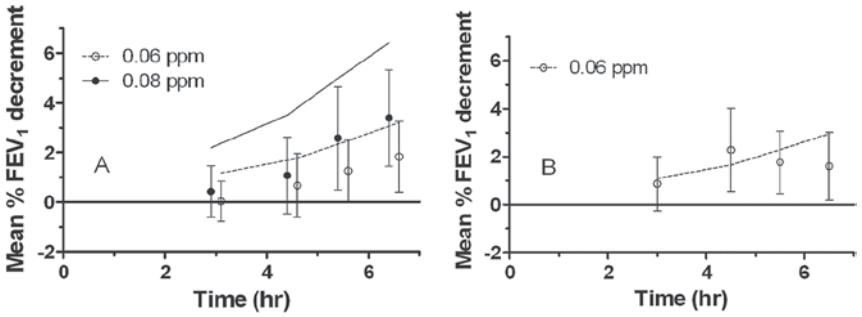

Figure 5. Mean observed and population mean predicted FEV_1 decrements (percent from baseline) versus time for the first 30 participants (A) and for the last 29 participants (B) of the Kim et al. (2011) study. Predicted values are invisible points connected by line segments, and observed values are symbols with 95th percentile confidence interval bars. Overlapping data at nominal times are offset to improve clarity.

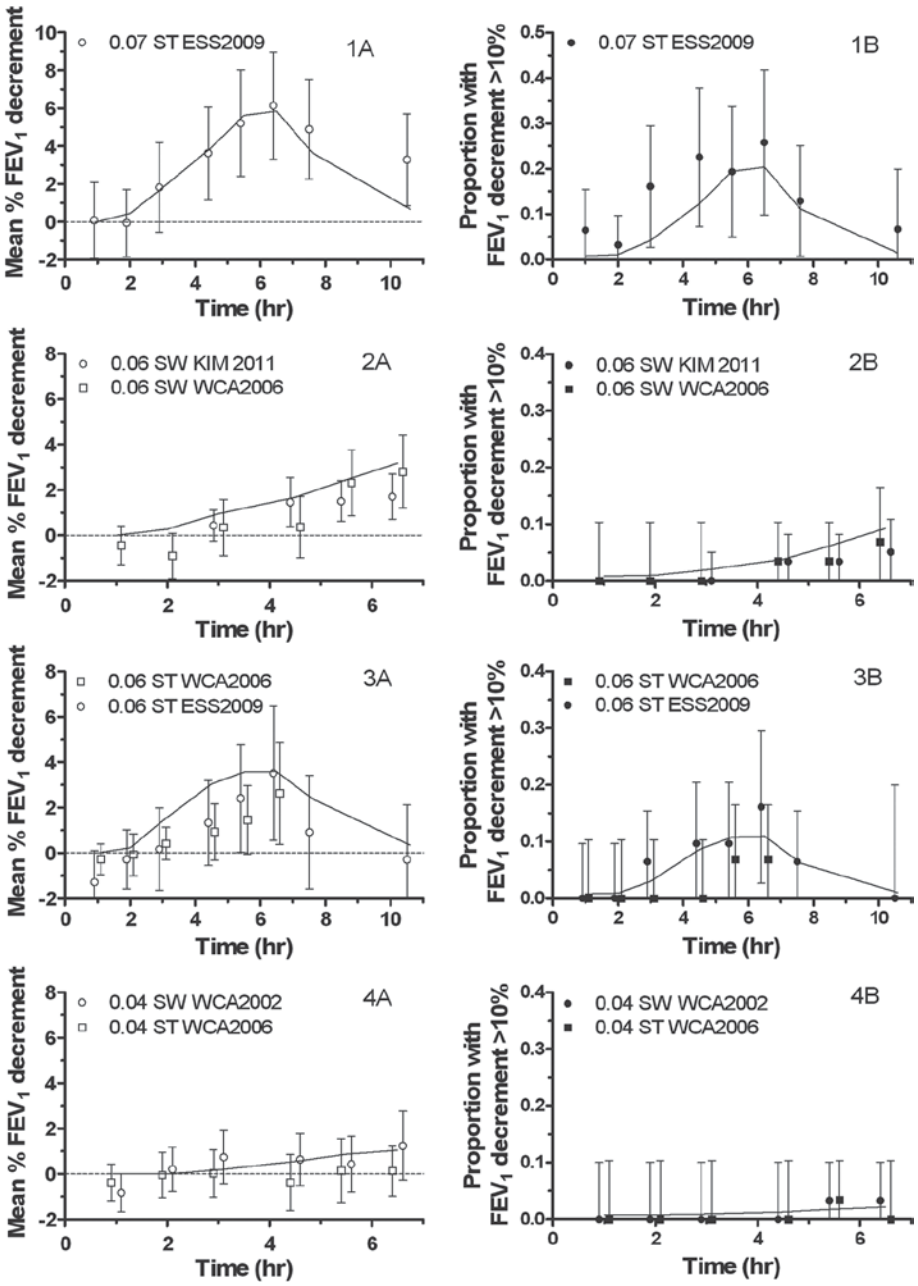

Figure 6. Observed (symbols with 95th percentile confidence interval bars) and predicted (invisible points connected by line segments) responses versus time for low concentration exposures. Population mean predicted values were generated by the threshold model. Mean responses are plotted in the A panels, and the proportions with an FEV_1 decrement > 10% are plotted in B panels. Studies are grouped by mean ozone concentration (in part per million) in panels 1–4. SW, constant concentration exposure; ST, variable concentration exposure; ESS2009, Schelegle et al. (2009); WCA2002, Adams (2002); WCA2006, Adams (2006a); Kim2011, Kim et al. (2011).

exposure. A 0.04 ppm SW exposure was included in the Adams (2002) study, and a 0.06 ppm SW exposure was conducted by Kim et al. (2011). Schelegle et al. (2009) conducted 0.06 and 0.07 ppm STEP exposures with target hourly *C*s of 0.04, 0.07, 0.07, 0.09, 0.05, and 0.04 ppm and 0.05, 0.07, 0.08, 0.09, 0.08 and 0.05 ppm, respectively. In the Figure 6A panels, mean observed $\%\Delta FEV_1$ for individual studies are plotted as symbols with 95% confidence interval bars while the weighted averages of predicted values for the studies in each panel are plotted as invisible points connected by line segments. There was generally good agreement between observed and predicted mean response with some overprediction for the 0.06 ppm exposures. In the Figure 6B panels, there was generally good agreement between the observed and predicted proportions of individuals with FEV_1 decrements greater than 10% with some underprediction of response to the 0.07 ppm exposures.

Figure 7 demonstrates the behavior of the threshold model across a range of concentrations. Model predictions are plotted versus *C* for individuals of age 23.8 years (mean age for the dataset) exposed for 6.6 h to a constant concentration of ozone with exercise for 50 min of each hour and an additional 30 min rest period after the 3rd

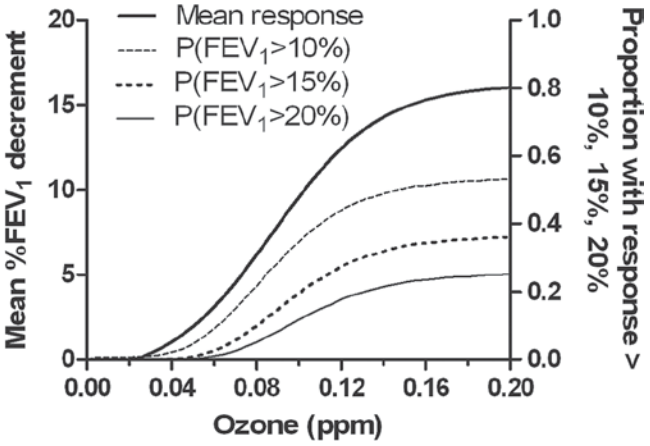

Figure 7. Predicted responses versus concentration for a hypothetical prolonged exposure using the threshold model. Predictions assume a 6.6-h constant concentration exposure of 23.8-year-old individuals exercising for 50 min of each hour to produce a V_E/BSA of 20 L·min^{-1}·m^{-2} during exercise. An additional 30-min rest period occurs after the 3rd hour. Population mean predicted response is plotted on the left *Y*-axis, and predicted proportions with FEV_1 decrements > 10, 15, and 20% are plotted on the right *Y*-axis.

hour. Exercise $\dot{V}_E$/BSA is assumed to be 20 L·min^{-1}·m^{-2} BSA, the target level for most of the prolonged exposures in this dataset. The predicted mean response is plotted on the left *Y* axis while the proportions of individuals predicted to experience FEV_1 decrements in excess of 10, 15, and 20% are plotted on the right *Y* axis. As expected predicted responses increase with increasing *C* and approach a plateau at high *C*. Using $Pr(FEV_1 > 10\%)$ to represent the predicted proportion of individuals with FEV_1 decrements > 10%, Figure 7 shows that $Pr(FEV_1 > 10\%) \geq Pr(FEV_1 > 15\%) \geq Pr(FEV_1 > 20\%)$. For individuals of this age with this exposure protocol, mean response and the proportion with a greater than 10% FEV_1 response are predicted to be barely discernable at around 0.04 ppm with the larger responses only apparent at higher *C*. Points on the curves would be shifted proportionally higher for younger individuals and proportionally lower for older individuals. The curves would be shifted to the right for lower levels of exercise and shorter durations of exposure and to the left for higher levels of exercise and longer durations of exposure.

Discussion

We previously identified a dynamic model that accurately described the relationship between ozone exposure and FEV_1 response in 540, 18–35-year-old males exposed under a wide range of conditions (McDonnell et al., 2007). We later demonstrated that this model was capable of not only describing data to which the model was fit but also capable of predicting the responses of independent data (McDonnell et al., 2010). Because the original EPA data did not include any exposures to concentrations below 0.08 ppm, we had some concern about applying it to lower concentrations of ozone that are much more frequently encountered in the ambient air. Two purposes of the current effort were to refine our estimates of model coefficients by refitting the model to an expanded data set that included individual responses to exposures with average *C* of 0.04, 0.06, and 0.07 ppm ozone from the UCD studies and with C of 0.06 ppm in the EPA2 study and to evaluate the agreement between observed and predicted responses. These additional studies also included exposures with variable ozone concentrations, some of which were intended to mimic diurnal patterns of outdoor ozone *C*. The agreement between observed and predicted responses in the expanded dataset was similar to the earlier results for the EPA data (McDonnell et al., 2007) and was slightly improved for the UCD data (McDonnell et al., 2010) with what appeared to be a slight overprediction for individual and mean $\%\Delta FEV_1$. In addition, we compared the proportions of individuals predicted to experience an FEV_1 change greater than 10, 15, and 20% with the observed data and generally found agreement between observed and predicted responses.

Schelegle et al. (2012) recently explored individual differences in the ozone–FEV_1 relationship using data which overlap to a large extent with the data used here. Rather than fitting a model to the data as a whole, they fit a conceptually similar but simpler model for each of the approximately 220 individual participants with enough data to allow estimation of three model parameters. Their model assumes a threshold number of ozone molecules that must be inhaled following which oxidant stress accumulates in a manner similar to Equation 1 in this manuscript. Change in lung function then increases linearly with increasing oxidant stress. Given the assumptions of their model, their results suggest that the threshold dose for observing responses is nonzero and that all three model parameters (threshold, timing constant, and slope) vary among individuals. Although the inter-subject variability in our original model is described by the distribution of a single variable *U*, the product of e^U and the logistic function in Equation 2 results in a skewed distribution of individual predicted responses for all nonzero levels of exposure. Although the sigmoid has no nonzero threshold below which zero response occurs, there is potentially a region of low values of *X* where the predicted FEV_1 responses from Equation 2 would be undetectable and indistinguishable from a threshold. Because the logistic function is symmetrical, however, the lower end of the curve is constrained by the pattern of responses to higher levels of exposure and may not necessarily provide the best fit to the lowest level exposures. As a result of Schelegle's observations, we modified the original model to include a threshold level of *X* (from Equations 1 and 3) below which the value of the sigmoid would be zero. This provides additional flexibility for the model to better describe response to low-level exposures. As noted in the results, this threshold (β9) in our model was found to be statistically significantly different from zero, and the AIC indicated an improvement in the overall fit of the threshold model compared to the

original model. The threshold model also tended to fit the earliest time point of the low level exposures better than the original model indicating that these two models diverge somewhat at very low exposures. Otherwise the predicted values of the two models were similar.

There are multiple approaches for specifying a threshold. Ours differs somewhat from the method of Schelegle et al., (2012), and we did not attempt to determine whether one was better than another. The estimated value for the threshold in this model, β9 = 59 (SE = 10), is in the same units as the variable *X* defined in Equation 1 which approximate accumulated ozone dose (ppm × liters of inhaled air per m^2 BSA) if one ignores the effect of the exponent β6. Exposures predicted to not reach threshold in this dataset were those with moderate, near-continuous exercise for 1 h to 0.06 and 0.08 ppm ozone and for 2 h to 0.04 ppm and those at rest for 1 h to 0.18 and 0.24 ppm ozone and for 2 h to 0.12 ppm. There are also exposures above threshold with generally small predicted responses because of the sigmoid shape of Equation 2 (e.g. see Figure 6). Our ability to ascertain whether a threshold exists and to estimate the value of a threshold was likely dependent upon the specified form of the threshold, the form of Equation 2 and other equations in the model, and the availability of very low level exposure data.

As noted earlier, we also fit several additional models in which the threshold was allowed to be a function of age and in which the threshold was allowed to vary as a random effect among individuals. Although the models converged, additional model parameters were not statistically significant, and it appeared to us that the data were not adequate to precisely estimate the parameters of these more complex models. It is possible that if a large number of volunteers were each exposed across a wide range of conditions, one might observe greater variability in an exposure threshold using our model, similar to the observation of Schelegle et al., (2012). This was not possible with the currently available data.

In our earlier study (McDonnell et al., 2007), we observed a subtle effect of body size (reflected by either BSA or height) in males on magnitude of effect and found that utilizing $\dot{V}_E$ /BSA rather than $\dot{V}_E$ alone in the model provided a small, but statistically significant improvement in the fit of the model. This is consistent with the idea that for a given dose of inhaled ozone a larger individual will distribute that dose over a larger airway surface area and a larger tissue volume than a smaller person and might therefore be expected to experience a smaller effect. Inclusion of females in the UCD and EPA2 data required that we not only consider the effect of body size within a sex, but that we also determine whether there exist any sex effects not explained by differences in body size. As Figure 1 demonstrates, when activity level is controlled such that $\dot{V}_E$ /BSA is equal for both sexes, the responses of males and females are, on average, approximately equal. This suggests that any effects of sex upon response can be accounted for by utilizing $\dot{V}_E$/BSA as a measure of ventilation and is consistent with several other studies in which both males and females exposed with equal levels of $\dot{V}_E$ /BSA experienced approximately equal FEV_1 responses (Hazucha et al., 2003; Weinmann et al., 1995; Seal et al., 1993). Studies have also indicated that sex differences in ozone uptake in the proximal airways can be accounted for by differences in anatomic dead space which is related to BSA (Bush et al., 1996).

The within-sex effect of BSA upon response differed for the males and females of the UCD data, both of which differed from that of the males in the EPA data. There was little apparent relationship between BSA and response for the UCD males while for females, there appeared to be a relationship between response and a higher power of BSA than for the EPA males. There is support in the literature (Bush et al., 1996) for a relationship between ozone uptake and body size, although no relationship between response and body size was observed for females in one study (Messineo and Adams, 1990). We suspect that the inconsistency between males and females in the UCD data and between the females in the UCD data and those of the Messineo and Adams study is the result of BSA being only a surrogate for a more directly related but unmeasured characteristic such as anatomic dead space, the relatively small proportion of the within-sex variance in response explained by the effect, the relatively narrow range of $\dot{V}_E$/BSA in the UCD data, and the small numbers of subjects in individual published studies and in the UCD data relative to the EPA data. We therefore elected to retain the measure of $\dot{V}_E$/BSA that we had used in our earlier work as a means of adjusting for both sex effects and within-sex body-size effects.

We previously demonstrated in the EPA data that BMI was an independent predictor of FEV_1 responsiveness to ozone (McDonnell et al., 2010). In the current expanded dataset, addition of BMI (Equation 4) to both the original and threshold models (Rows 3 and 4, Table 2) did not substantially change the value of coefficients other than the age coefficient (β2) in their respective models. Although the coefficient of BMI was not statistically significantly different from zero in these models, its magnitude was similar to that of our earlier study of the EPA data, albeit with larger standard error, and the AIC was lower when BMI was included. This is consistent with Bennett et al. (2007) that differences in BMI do account for a small proportion of the differences in FEV_1 with those with higher BMI experiencing larger responses. Inclusion of BMI in either model did not result in any meaningful change in the overall relationships between observed and predicted responses (Table 3) although it did result in changes in predicted response for some individuals.

When the original model (Equations 1, 2) was fit to the EPA data alone (Row 5, Table 2), the age coefficient was significantly negative (β2 = –0.43 with SE = 0.12) indicating that response decreases with increasing age. This finding is consistent with the results of Hazucha et al. (2003) who found an effect of age in a sample of young men and women, and it is also consistent with

studies of older adults (>50 year) who experienced very small responses to ozone (Drechsler-Parks et al., 1987). The age coefficient (β2) of the original model fit to the combined EPA, UCD, and EPA2 data (Row 1, Table 2) is about 45% of the magnitude of the coefficient when the model was fit to the EPA data alone, however, and is not statistically significant. When the EPA and UCD studies were allowed to have laboratory-specific age and BMI coefficients with other coefficients of the original model form (+BMI) in common, the EPA and UCD estimated age coefficients (SE) were –0.45 (0.10) and 0.19 (0.60), respectively, and the EPA and UCD estimated BMI coefficients (SE) were 0.45 (0.22) and 0.29 (0.90), respectively. The reasons for these substantial laboratory differences in age and, to a lesser extent, BMI coefficients are not entirely clear. We did observe that there was a statistically significant association between age and BMI in the UCD data ($r = 0.26$) that was absent in the EPA data ($r = 0.03$). Because the effects of age and BMI are additive with opposite signs in this model, it is possible that confounding of the age and BMI effects in the UCD data resulted in a smaller value of the BMI coefficient compared to the EPA data and a more positive age coefficient compared to the EPA data as well as much larger standard errors of the estimates. There were also fewer individuals at the youngest and oldest ages in the UCD studies compared to the EPA studies. Recruitment of volunteers rather than individuals randomly selected from the population in all of these studies leaves open the possibility of selection bias as an explanation of different estimates of the age and BMI effects for the two laboratories.

A further analysis by Bennett et al. (2007) of 197 18–35-year-old males and females from the Hazucha et al. (2003) study found that, for 1.5 h exposures to 0.42 ppm ozone with moderate exercise, the mean $\%FEV_1$ decrement was approximately 16.1%, and that, on average, response decreased by 2.5% of the mean response for each year of age and increased by 3.6% of the mean response for each unit of BMI. Use of the EPA-specific coefficients for age and BMI given in the text above predicts that, on average, response should decrease by 4.4% ($100\% \times \beta2/\beta1$) of the mean response for each year of age and increase by 4.4% ($100\% \times \beta8/\beta1$) for each unit of BMI. The original model (+BMI) using the combined data coefficients (Row 3, Table 2) predicts that, on average, response should decrease by 2.6% of the mean response for each year of age and increase by 4.8% for each unit of BMI. Although laboratory differences in the age coefficient raise uncertainty about the most representative value for this coefficient, it appears that the results from the combined data are broadly consistent with those of Hazucha et al. (2003) and Bennett et al. (2007), a large, independent study specifically designed to examine the effect of age on response. Inclusion of BMI improves the fit of the model according to the AIC and may be useful for controlling any confounding of the age effect by BMI.

Although these models are relatively simple given the complexity of the mechanisms underlying the reductions in FEV_1 that occur in response to inhaled ozone, they predict behavior that is consistent with a number of the widely-accepted characteristics of the ozone-FEV_1 relationship. In addition to quantifying the relationships between response and C and $\dot{V}_E$ at any time point, they offer substantial improvements over earlier models in describing the effects of duration of exposure through the use of a differential equation which allows for both changing exposure conditions and a recovery process, and they include a between-subjects variability structure that is consistent with the observed skewed distributions of individual response. The model appears to predict both mean responses as well as the proportions of individuals with responses with more extreme values. Because the models are applicable to a wide range of exposure conditions and patterns, they allow the use of a large amount of data for estimation of parameters and quantification of uncertainty and also allow comparison of results among dissimilar studies that could not be previously compared.

The complexity of the models, however, is currently limited by the available data including a fairly narrow range of exposure conditions for any one participant. A study of individuals undergoing a series of exposures that covered the full range of exposure conditions would likely lead to a more complete description of the variability in the E–R relationship among individuals perhaps by treating other model parameters as random effects as suggested by the analysis of Schelegle et al. (2012). More data collected during recovery following exposure would probably lead to a more complex understanding of the dynamics of response, and study of individuals across a broader range of ages would allow applicability of the model to a wider age range. Measures of resting $\dot{V}_E$ are difficult to make and were estimated for these studies. For all but the facemask studies, exercise $\dot{V}_E$ was measured at only one or two time points, and we have assumed that changes in $\dot{V}_E$ occur in a square wave fashion ignoring the transients known to occur during changes in activity level. This model currently utilizes only C in inhaled air and $\dot{V}_E$ in calculations of dose rate. Better definition of the site in the respiratory tract where the FEV_1 response is initiated as well as better estimates of how ozone uptake at that site is influenced by physical characteristics other than BSA and by the changes in respiratory pattern (e.g. frequency, tidal volume, inspiratory and expiratory flows, switching between nasal and oro-nasal breathing, etc.) that accompany changes in $\dot{V}_E$ could result in more precise predictions of response. All within-subject variability is currently lumped into a single term E as a result of limitations of the model fitting program. It is likely that some of the within-subject variability is due to true changes in responsiveness to ozone over time while much is simply noise. Improvements in the model fitting software might allow for partition of the within-subject variability into aspects of interest such as daily variation in responsiveness to ozone which are distinct from

nuisance aspects such as measurement error, etc. This might allow better interpretation of the data at the level of the individual which, at present, are noisy.

In conclusion, the original model described in Equations 1 and 2 predicts the exposure–response characteristics of the population of healthy, 18–35-year-old, nonsmoking adults over a wide range of exposure conditions and patterns. This includes prediction of the mean of the population responses as well as the proportions of individuals expected to experience decrements in FEV_1 greater than 10, 15, and 20%. The same model modified to include a threshold in accumulated dose below which no response is observed better fits the observed data particularly at the early time points of the lowest level exposures. Since most ambient exposures are low, the threshold model will likely provide more accurate estimates of risk for the population. Inclusion of BMI in both models was also found to improve the fit to the data and may help control for confounding of the effects of age and BMI on response.

Acknowledgements

The authors thank Dr. William C. Adams of the University of California, Davis, for providing the individual exposure and response data for studies conducted by him.

Declaration of interest

Dr. McDonnell and Dr. Stewart conducted this modeling research under contract to American Petroleum Institute, and Dr. Schelegle has received funding from American Petroleum Institute to conduct other research to assess the effects of ozone on lung function in humans and to develop ozone exposure–response models in the recent past. Dr. Smith and Dr. Kim have no potential conflicts of interest. American Petroleum Institute, Washington, D.C. provided funding for this modeling effort. Although the research described in this article has been supported in part by the United States Environmental Protection Agency, it has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Adams WC, Savin WM, Christo AE. 1981. Detection of ozone toxicity during continuous exercise via the effective dose concept. J Appl Physiol 51:415–422.

Adams WC. 2002. Comparison of chamber and face-mask 6.6-hour exposures to ozone on pulmonary function and symptoms responses. Inhal Toxicol 14:745–764.

Adams WC. 2003. Comparison of chamber and face mask 6.6-hour exposure to 0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses. Inhal Toxicol 15:265–281.

Adams WC. 2006a. Comparison of chamber 6.6-h exposures to 0.04-0.08 PPM ozone via square-wave and triangular profiles on pulmonary responses. Inhal Toxicol 18:127–136.

Adams WC. 2006b. Human pulmonary responses with 30-minute time intervals of exercise and rest when exposed for 8 hours to 0.12 ppm ozone via square-wave and acute triangular profiles. Inhal Toxicol 18:413–422.

Bennett WD, Hazucha MJ, Folinsbee LJ, Bromberg PA, Kissling GE, London SJ. 2007. Acute pulmonary function response to ozone in young adults as a function of body mass index. Inhal Toxicol 19:1147–1154.

Bush ML, Asplund PT, Miles KA, Ben-Jebria A, Ultman JS. 1996. Longitudinal distribution of O3 absorption in the lung: gender differences and intersubject variability. J Appl Physiol 81:1651–1657.

Chadha TS, Birch S, Sackner MA. 1987. Oronasal distribution of ventilation during exercise in normal subjects and patients with asthma and rhinitis. Chest 92:1037–1041.

Colucci AV. (1983). Pulmonary dose/effect relationships in ozone exposure. In: Lee SD, Mustafa MG, Mehlman MA, ed. International Symposium on the Biomedical Effects of Ozone and Related Photochemical Oxidants. Princeton, NJ: Princeton Scientific, vol. V, 21–44.

Drechsler-Parks DM, Bedi JF, Horvath SM. 1987. Pulmonary function responses of older men and women to ozone exposure. Exp Gerontol 22:91–101.

Folinsbee LJ, Drinkwater BL, Bedi JF, Horvath SM. (1978). The influence of exercise on the pulmonary function changes due to low concentrations of ozone. In: Folinsbee LJ, Wagner J A, Borgia JF, Drinkwater BL, Gliner JA, Bedi JF, eds. Environmental Stress. New York: Academic Press, 125–145.

Hazucha MJ. 1987. Relationship between ozone exposure and pulmonary function changes. J Appl Physiol 62:1671–1680.

Hazucha MJ, Folinsbee LJ, Bromberg PA. 2003. Distribution and reproducibility of spirometric response to ozone by gender and age. J Appl Physiol 95:1917–1925.

Horstman DH, Folinsbee LJ, Ives PJ, Abdul-Salaam S, McDonnell WF. 1990. Ozone concentration and pulmonary response relationships for 6.6-hour exposures with five hours of moderate exercise to 0.08, 0.10, and 0.12 ppm. Am Rev Respir Dis 142:1158–1163.

Kim CS, Alexis NE, Rappold AG, Kehrl H, Hazucha MJ, Lay JC, Schmitt MT, Case M, Devlin RB, Peden DB, Diaz-Sanchez D. 2011. Lung function and inflammatory responses in healthy young adults exposed to 0.06 ppm ozone for 6.6 hours. Am J Respir Crit Care Med 183:1215–1221.

Kriebel D, Smith TJ. 1990. A nonlinear pharmacologic model of the acute effects of ozone on the human lungs. Environ Res 51:120–146.

Larsen RI, McDonnell WF, Horstman DH, Folinsbee LJ. 1991. An air quality data analysis system for interrelating effects, standards, and needed source reductions: Part 11. A lognormal model relating human lung function decrease to O3 exposure. J Air Waste Manage Assoc 41:455–459.

McDonnell WF, Stewart PW, Smith MV. 2010. Prediction of ozone-induced lung function responses in humans. Inhal Toxicol 22:160–168.

McDonnell WF, Stewart PW, Smith MV. 2007. The temporal dynamics of ozone-induced FEV1 changes in humans: an exposure-response model. Inhal Toxicol 19:483–494.

McDonnell WF, Smith MV. 1994. Description of acute ozone response as a function of exposure rate and total inhaled dose. J Appl Physiol 76:2776–2784.

McDonnell WF, Kehrl HR, Abdul-Salaam S, Ives PJ, Folinsbee LJ, Devlin RB, O'Neil JJ, Horstman DH. 1991. Respiratory response of humans exposed to low levels of ozone for 6.6 hours. Arch Environ Health 46:145–150.

McDonnell WF, Horstman DH, Abdul-Salaam S, Raggio LJ, Green JA. 1987. The respiratory responses of subjects with allergic rhinitis

to ozone exposure and their relationship to nonspecific airway reactivity. Toxicol Ind Health 3:507–517.

McGregor M, Adam W, Sekelj P. 1961. Influence of posture on cardiac output and minute ventilation during exercise. Circ Res 9:1089–1092.

Messineo TD, Adams WC. 1990. Ozone inhalation effects in females varying widely in lung size: comparison with males. J Appl Physiol 69:96–103.

Schelegle ES, Morales CA, Walby WF, Marion S, Allen RP. 2009. 6.6-hour inhalation of ozone concentrations from 60 to 87 parts per billion in healthy humans. Am J Respir Crit Care Med 180:265–272.

Schelegle ES, Adams WC, Walby WF, Marion MS. 2012. Modelling of individual subject ozone exposure response kinetics. Inhal Toxicol 24:401–415.

Seal E Jr, McDonnell WF, House DE, Salaam SA, Dewitt PJ, Butler SO, Green J, Raggio L. 1993. The pulmonary response of white and black adults to six concentrations of ozone. Am Rev Respir Dis 147:804–810.

Silverman F, Folinsbee LJ, Barnard J, Shephard RJ. 1976. Pulmonary function changes in ozone-interaction of concentration and ventilation. J Appl Physiol 41:859–864.

Smith MV, McDonnell WF. 1999. A statistical model for FEV1 response to arbitrary dynamic ozone exposure conditions. J Air Waste Manag Assoc 49:289–298.

Weinmann GG, Weidenbach-Gerbase M, Foster WM, Zacur H, Frank R. 1995. Evidence for ozone-induced small-airway dysfunction: lack of menstrual-cycle and gender effects. Am J Respir Crit Care Med 152:988–996.