Appendix B

Benchmark Concentration Analysis of Diesel Data

B-1. INTRODUCTION TO BENCHMARK

The benchmark dose or benchmark concentration approach, hereafter referred to as the BMC approach, is an alternate to the N/LOAEL option for deriving effect levels. The BMC is currently undergoing extensive consideration by the Agency with promulgation of software and guidelines for application of this methodology (U.S. EPA, 2000). The BMC approach involves fitting a dose-response function to dose and effect information from a single study to derive the best fit of those data. This "best fit"is statistically termed the maximum likelihood estimate but is referred to in the benchmark terminology as the BMC curve. The curve defining the corresponding lower 95% confidence limit of this "best fit"estimate is termed the BMCL curve. This BMCL curve is used to predict the dose that will result in a level of response that is defined *a priori* as the benchmark response "x", $BMCL_x$. In the analyses below, for example, the benchmark response for a 10% increase in incidence[1] of chronic inflammation is defined as a $BMCL_{10}$; the corresponding 10% increase as determined from the BMC curve would be termed the BMC_{10}. This $BMCL_{10}$ would be derived by first using the data and the programs to determine the BMC and BMCL curves. The concentration corresponding to a 10% increase in incidence would then be determined directly from the BMCL. The $BMCL_{10}$ then would be used as the representative value for the effect level or point of departure in the dose-response assessment.

The latest version of the Agency Benchmark Dose Software (BMDS Version 1.2; U.S. EPA, 2000) was used to analyze data on chronic inflammation and pulmonary histopathology present in the chronic studies that were amenable to benchmark analysis. At this time, the Agency BMDS offers sixteen different models total that are appropriate for the analysis of dichotomous data (gamma, logistic, probit, Weibull, log-logistic, multistage, log-probit, quantal-linear, quantal-quadratic), continuous data (linear, polynomial, power, Hill) and nested developmental toxicology data (NLogistic, NCTR, Rai & Van Ryzin). Results from all models include a reiteration of the model formula and model run options chosen by the user, goodness-of-fit information, a graphical presentation for visual inspection and the concentration estimate for the response at the designated $BMCL_x$, as well as the corresponding BMC_x. More details on the modeling results are described and presented in the analysis on dichotomous data following.

The U.S. EPA benchmark dose (BMD/C) methods guidance has not been finalized at this time to provide definitive procedures and criteria (U.S. EPA 1995). Therefore, in this document provisional criteria for minimum data to perform a benchmark analysis are designated such that (1) complete quantitative information on the response of interest should be available (e.g.,

[1]For increases in incidence "extra risk" is used which is response incidence (inc) normalized to the background (BG) incidence; response – BG/1-BG.

incidence as number affected / total, means with variability) and that (2) at least two exposure levels with responses that differ from those of the controls are provided, and (3) a benchmark response of 10% is employed such that outcomes are $BMCL_{10}$s. A response of 10% is at or near the limit of sensitivity in most long-term bioassays as determined from both the typical number of animals used in bioassays and a low spontaneous background rate (e.g., 0.1%) for a given effect (Haseman, 1984; Haseman et al., 1989).

B-2. DIESEL DATA FOR BENCHMARK ANALYSIS

Using the criteria set forth in Section B-1 and the information about the critical effects that have been identified (pulmonary inflammation, pulmonary histopathology including indicators of fibrotic changes such as increases in alveolar-capillary wall thickness) the following rat chronic studies identified in Chapter 6 were analyzed for information suitable for BMC analysis: Ishinishi et al. (1986, 1988), Mauderly et al. (1987a,b; 1988); Heinrich et al. (1986, 1995), and Nikula et al. (1995).

Results from this analysis yielded only a few data sets from a single study, that of Nikula et al. (1995), that could be used for BMC analysis. The basis for not including data from the other studies varied. Information on pulmonary histopathology in the studies of Ishinishi et al. (1986, 1988), for example, was supplied only in narrative form with no quantitative information given. A similar situation was found for those reports of the ITRI study; Wolff et al. (1987) reports on clearance alterations due to DPM exposure; Henderson et al. (1988) does give information on hydroxyproline but only in graphical form; the 1988 study of Mauderly et al. deals with pulmonary function as a function of DPM lung loading; the 1987a reference of Mauderly et al. discusses tumor prevalence only and the Mauderly 1987b reference reports on diesel exhaust in developing lung to a single exposure concentration of DPM with no dose-response information available. Those reports on the General Motor study contain extensive information relating not to the critical effects, but mostly to precursors of inflammation such as levels of polymorphonuclear neutrophils and lymphocytes in bronchoalveolar lavage from DPM exposed rats (Strom, 1984) and guinea pigs (Barnhart et al., 1981) as well as information on collagen biosynthesis (Misiorowski et al., 1980) all of which is presented in graphical rather than tabular form amenable for benchmark analysis. The information on noncancer histopathology reported by Heinrich et al. (1995) is in text form only and this author's 1986 study deals primarily with clearance and mortality. Nikula et al. (1995), however, do present extensive quantitative dose-response information (incidence / dichotomous data) on several measures of the critical effect including chronic inflamation (presence of focal aggregates of neutrophils), focal fibrosis with epithelial hyperplasia (nodular fibrosis rimmed by hyperplasia), and septal fibrosis (interstitial fibrosis within

alveolar septa) although the study had but 2 exposure concentrations both of which are different from the controls, a minimal number on which benchmark analysis should be performed.

B-3. BENCHMARK ANALYSIS OF DIESEL DATA

These data from Nikula et al. (1995) were extracted, HEC concentrations calculated using the model of Yu et al. (1991; Appendix A), and analyzed using all 9 applicable models for dichotomous data. Because the benchmark models were ran with the HEC, general from the model of Yu et al. (1991), the $BMCL_{10s}$ are also HECs. The results and data are presented in Table B-1. Results were evaluated based on the nature of the data set, visual inspection of the graphical output, and on the goodness-of-fit parameters, including p values and the AIC. When p values were generated for model fits, values for p that were less than 0.1 were considered to reflect a minimal fit to the data and were disqualified from further consideration. However, the small set of only 3 data points was often matched by the number of parameters fitted in several of the models such that the outcome of the model exactly fit the data and thus no p value is generated; these model fits are often referred to as being overparameterized, and are indicated as "NA" in Table B-1. Values for p that were less than 0.1 were considered to reflect a minimal fit to the data. The AIC (Akaike Information Coefficient; Akaike, 1973; Stone, 1998) is a parameter generated for the models in U.S. EPA (2000) that allows for a general comparison among models run on the same data set. The AIC is defined as -2 log L + 2 p where log L is the log likelihood of the fitted model, and p is the number of parameters estimated; smaller values indicate better fits.

The overall results of this mathematical analysis is reasonable in a biologically mechanistic sense in that chronic inflammation is more prevalent and apparently occurs at lower concentrations (i.e., has lower $BMCL_{10}$ values) than does focal fibrosis. The information on septal fibrosis were not interpretable as the data were not amenable (no or zero background and then total incidence) to any meaningful benchmark or other dose-response analysis. The most sensitive endpoint, chronic inflammation, is therefore the most sensitive benchmark concentration followed by focal fibrosis.

The choice for the most appropriate $BMCL_{10}$ from among the various modeled values for chronic inflammation requires analysis of both the statistical and graphical outputs of the data. The shape of the dose-response curve from information given in Chapter 6 (Table 6-2) gives evidence of considerable "S" character, e.g., several low HECs without any reported effects up to about 0.2 mg/m^3. The shape of the dose-response curves generated by several of the models, including gamma-hit, Weibull, multistage, and quantal linear were all a uniformly upward sloping arc from the origin (graphs not shown) with minimal evidence of any "S" character, a shape not

concordant with the data array in Table 6.2. Models that did generate curves with "S" character included log-logistic, logistic, probit, quantal-quadratic, and log-probit. Because of their concordance with this independent data array on dose-response, the latter outputs are further analyzes.

The results for both chronic inflammation and focal fibrosis for those models with outputs having appreciable "S" character suggest that females may be more sensitive than males for these endpoints as the incidences are higher and the $BMCL_{10}$ values are generally lower for females than for males. However, the model fits of the $BMCL_{10}$s to the chronic inflammation data segregated by sex were generally inadequate as judged from the p values (most being far less than 0.1) or from visual inspection of the fits to the data, several of which (e.g., log-logistic and log-probit) were lacking any appreciable "S" character. However, combining female and male data improved data fitting as judged by the increased p values to where nearly all were >0.1 and to where the visual fits were concordant with the independent information on dose-response. Too, most of the combined $BMCL_{10}$s were either intermediate between the female and male values or somewhat closer to the female values such that the combined $BMCL_{10}$ values were not much different from the females $BMCL_{10}$s.

From among the combined male and female model outputs in Table B-1, the logistic, probit, and quantal quadratic results were all excluded based on the high AIC value relative to the log-logistic and log-probit results. The log-logistic results were excluded based on the shape of the lower portion of the dose-response curve which was upward sloping near the origin (graph not shown) and not as concordant with the independent dose-response information in Table 6-2 as was the fit of the log-probit model (Figure B-1). This leaves the fit of the log-probit model as being most reflective of the information in Table 6-2. The $BMCL_{10}$ of the log-probit curve at 0.37 mg/m^3 remains and, by elimination, appears to be the most defensible choice from among the $BMCL_{10}$s arrayed in Table B-1. Figure B-1 shows the graphical representation of the log-probit model fit to the data and the origin of the $BMCL_{10}$. This graph also shows the relationship of the $BMCL_{10}$ of 0.37 mg/m^3 to the variability that exists around the control value and that the value of 0.37 mg/m^3 is not far removed from the outer range of this variability. The log-probit $BMCL_{10}$ for focal fibrosis (combined) of 1.3 mg/m^3 noted as being representative of this lesion from the BMC analysis in Table B-1.

Characterization of this benchmark value indicates that it may not be a suitable candidate for use as a point of departure for development of a dose-response assessment such as the RfC. An attribute of the benchmark method is that the response (such as the 10% as used here) is near the range of the actual experimental values, such that extrapolation is not far below the observed experimental range. However, due to the paucity of data points overall and lack of any values below an HEC of nearly 2 mg/m^3 in the Nikula et al. (1995) study, the extrapolation of this BMC

to the 10% response level is considerable, the $BMLC_{10}$ of 0.37 mg/m^3 being > 5-fold below the nearest observed value of 1.95 mg/m^3. Also, the high experimental exposures used in this study are in the range of those resulting in pulmonary overload conditions in rats and therefore in the range of the model assumptions of Yu et al. (1991) about this phenomenon in humans for calculation of the HECs (Chapter 3). The $BMCL_{10}$ of 0.37 mg/m^3 is considerably greater than other NOAELs in the DPM data base of 0.144 mg/m^3 and 0.128 mg/m^3 (Table 6-2 in Chapter 6), possibly indicating that these NOAELs represent actual incidence levels that are considerably less than10%; from the same log-probit model the corresponding $BMCL_{05}$ was 0.21 mg/m^3 (near the range of these NOAELs) and the corresponding $BMCL_{01}$ was 0.07 mg/m^3 (below the range of these NOAELs). These limitations on this $BMCL_{10}$ make it a less than optimal candidate for consideration as a point of departure in the development of dose-response assessments.

B-4. SUMMARY

The recently developed EPA Benchmark dose software (U.S. EPA, 2000) and preliminary guidance was utilized to analyze diesel data by the benchmark approach. Data from only one of the array of principal studies identified elsewhere (Chapter 6) was found to contain data amenable to benchmark analysis. The data from this study, that of Nikula et al. (1995) on pulmonary inflammation and histopathology, was extracted and analyzed as dichotomous data using all available models and designating a 10% response level such that $BMCL_{10}$s were calculated; as the models were ran with HECs, the $BMCL_{10s}$ were also HECs.

The analysis resulted in an array of $BMCL_{10}$s from 3 different effects in two sexes (both separate and combined) with 9 different models. These $BMCL_{10}$s were each considered from a perspective of biological relevance, known dose-response character, and from the individual fit to the data by the models from statistical parameters and visual judgments. The $BMCL_{10}$ that emerged after the above considerations was 0.37 mg/m^3 for the combined male plus female incidence of chronic active pulmonary inflammation. A $BMCL_{10}$ of 1.3 mg/m^3 for pulmonary focal fibrosis was also noted in this analysis. Characterization of these benchmark values indicates that neither may be a suitable candidate for use as a point of departure in development of a dose-response assessment such as the RfC but that they are concordant with other quantitative dose-response aspects of the DPM database.

Table B-1. BMC analysis of pathology incidence data in male and female F344 rats from the study of Nikula et al. (1995) using the different models available from U. S. EPA benchmark dose project (U.S. EPA, 2000) for dichotomous data based on 10% extra risk (i.e., a 10% increase relative to a total that has been adjusted for background) and no threshold term. The concentrations used in the analysis are human continuous equivalent concentrations (HECs) obtained from the interspecies extrapolation model of Yu et al. (1991). The table listings include the $BMCL_{10}$ (the benchmark response level of 10% obtained from the lower 95% limit of the benchmark curve in mg/m^3), the BMC_{10} (the corresponding estimate at 10% response from the best fit benchmark curve, also in mg/m^3), P = goodness-of-fit values. NA indicates a G-O-F value was not available, usually due to the lack of degrees of freedom. AIC = Akaike Information Coefficient (see U.S. EPA, 2000 and below) which may be used for model comparison on the same data set.

Effect (from Table 5 and 6, p 86, Nikula et al., 1995)	Inc @ 0 mg/m^3	Inc @ 1.95 mg/m^3 HEC	Inc @ 5.1 mg/m^3 HEC	$BMCL_{10}$ (BMC_{10}) log-logistic	$BMCL_{10}$ (BMC_{10}) log-probit	$BMCL_{10}$ (BMC_{10}) multi-stage	$BMCL_{10}$ (BMC_{10}) - Weibull	$BMCL_{10}$ (BMC_{10}) - gamma	$BMCL_{10}$ (BMC_{10}) - quantal linear	$BMCL_{10}$ (BMC_{10}) - probit	$BMCL_{10}$ (BMC_{10}) - logistic	$BMCL_{10}$ (BMC_{10}) quantal quadratic
Chronic active inflammation >18 mos, grades 1-3, male + female combined	5/177	59/162	118/174	0.32(0.64) P= NA AIC= 483	0.37(.70) P=NA AIC = 483	0.43(.49) P= 0.982 AIC= 481	0.43(.49) P= 0.982 AIC= 481	0.43(.49) P=0.98 AIC= 480	0.43(.49) P= .982 AIC= 481	1.06(1.19) P= 0.000 AIC= 499	1.12(1.26) P=0.000 AIC= 502	1.34(1.45) P= 0.000 AIC = 505
Chronic active inflammation >18 mos, grades 1-3 in males	1/86	19/81	54/85	0.67(1.16) P= NA AIC= 217	0.74(1.22) P = NA AIC = 217	0.56(.95) undefined AIC= 217	.56(1.04) P= NA AIC= 216	.56(1.09) P= NA AIC= 217	0.50(.61) P= 0.15 AIC= 216	1.31(1.55) P= 0.05 AIC= 219	0.67(1.16) P= NA AIC= 217	1.42(1.57) P= 0.055 AIC = 218
Chronic active inflammation >18 mos, grades 1-3 in females	4/91	40/81	64/89	0.18(0.26) P= NA AIC= 257	.016(.30) P = NA AIC = 257	0.33(.40) P= 0.173 AIC= 257	0.33(.40) P= 0.173 AIC= 257	0.33(.40) P= 0.17 AIC= 257	0.33(.40) P= 0.173 AIC= 257	0.83(.96) P= 0.0001 AIC= 272	0.85(1.0) P= 0.000 AIC= 273	1.21(1.35) P= 0.000 AIC = 279
Focal fibrosis with epithelial hyperplasia, grades 1-4 in males and females combined	0/177	18/162	63/174	1.25(1.8) P= 1.000 AIC= 345	1.3(1.8) P = 1.000 AIC = 345	1.21(1.8) P= 1.000 AIC= 345	1.21(1.8) P= 1.000 AIC= 345	1.21(1.8) P= 1.0 AIC= 345	1.1(1.3) P= 0.363 AIC= 345	2.32(2.61) P= 0.013 AIC= 353	2.50(2.8) P= 0.006 AIC= 356	2.14(2.34) P= 0.091 AIC = 347
Focal fibrosis with epithelial hyperplasia, grades 1-4 in males	0/86	5/81	19/85	1.72(2.7) P= 1.00 AIC= 132	1.6(2.7) P = 1.000 AIC = 132	1.79(2.8) undefined AIC= 134	1.79(2.8) P= 1.00 AIC= 132	1.79(2.75 P= 1.0 AIC= 132	1.7(2.4) P= 0.70 AIC= 131	2.98(3.5) P= 0.199 AIC= 134	3.17(3.69) P= 0.153 AIC= 135	2.68(3.1) P=0.552 AIC = 131
Focal fibrosis with epithelial hyperplasia, grades 1-4 in females	0/91	13/81	44/89	0.80(1.4) P= 1.00 AIC= 199	0.87(1.47) P = 1.000 AIC = 199	0.77 P= 0.99 AIC= 199	0.77(1.4) P=1.0 AIC=199	0.71(1.4) P= 1.00 AIC= 199	0.71(.88) P= 0.445 AIC= 198	1.76 P= 0.037 AIC= 205	1.89(2.2) P= 0.02 AIC= 207	1.7(1.9) P= 0.21 AIC = 200
Septal fibrosis, >18 mos, grades 1-4 in males	1/86	79/81	83/85	.003(.008) P= 0.35 AIC= 53	(failed)	0.07(.08) P= 0.000 AIC= 65	0.07(.08) P= 0.000 AIC= 65	0.07(.08) P= 0.000 AIC= 65	0.07(.08) P= 0.000 AIC= 65	0.29(.37) P= 0.000 AIC= 114	0.32(.44) P= 0.000 AIC= 86	0.42(0.47) P= 0.000 AIC = 100
Septal fibrosis, >18 mos, grades 1- 4 in females	2/91	75/81	87/89	0.009 (.05) P= NA AIC= 87	(failed)	0.08(.10) P= 0.003 AIC= 91	0.08(.10) P= 0.000 AIC= 91	0.08(.10) P= 0.003 AIC= 91	0.08(.10) P= 0.003 AIC= 91	0.32(.40) P= 0.000 AIC= 131	0.34(.45) P= 0.000 AIC= 109	0.46(.51) P= 0.000 AIC = 119

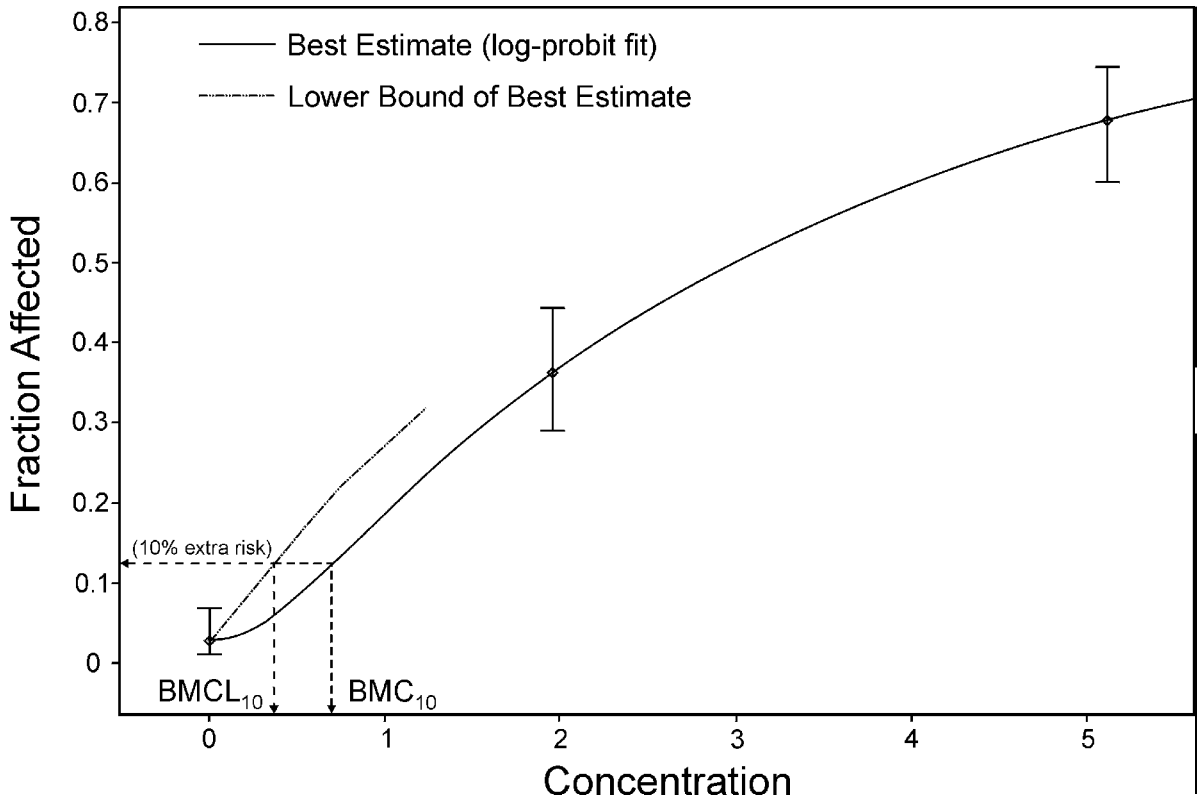

Figure B-1. Benchmark concentration analysis (log-probit) of chronic pulmonary inflammation in rats exposed to DPM from Nikula et al. (1995). $BMCL_{10}$, the lower confidence estimate of the concentration of DPM associated with a 10% incidence (extra risk); BMC_{10}, the corresponding estimate from the best (log-probit) fit. (◇) data with 95% error bounds.

B-5. REFERENCES

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In: Proceedings of the Second International Symposium on Information Theory, B.N. Petrov and F. Csaki, eds. Akademiai Kiado, Budapest, pp. 267-281

Barnhart, MI; Chen, S-T; Salley, SO; et al. (1981) Ultrastructure and morphometry of the alveolar lung of guinea pigs chronically exposed to diesel engine exhaust: six months' experience. J Appl Toxicol 1:88-103.

Haseman, JK. (1984) Statistical issues in the design, analysis, analysis and interpretation of animal carcinogenicity studies. Environ Health Persp 58: 385-392.

Haseman, JK; Huff, JE; Rao, GN and Eustis, SL. (1989) Sources of variability in rodent carcinogenicity studies. Fund Appl Toxicol 12: 793-804.

Heinrich, U; Muhle, H; Takenaka, S; et al. (1986) Chronic effects on the respiratory tract of hamsters, mice, and rats after long-term inhalation of high concentrations of filtered and unfiltered diesel engine emissions. J Appl Toxicol 6:383-395.

Heinrich, U; Fuhst, R; Rittinghausen, S; et al. (1995) Chronic inhalation exposure of Wistar rats and two strains of mice to diesel engine exhaust, carbon black, and titanium dioxide. Inhal Toxicol 7:553-556.

Henderson, RF; Pickrell, JA; Jones, RK; et al. (1988) Response of rodents to inhaled diluted diesel exhaust: biochemical and cytological changes in bronchoalveolar lavage fluid and in lung tissue. Fundam Appl Toxicol 11:546-567.

Ishinishi, N; Kuwabara, N; Nagase, S; et al. (1986) Long-term inhalation studies on effects of exhaust from heavy and light duty diesel engines on F344 rats. In: Ishinishi, N; Koizumi, A; McClellan, RO; et al., eds. Carcinogenic and mutagenic effects of diesel engine exhaust: proceedings of the international satellite symposium on toxicological effects of emissions from diesel engines; July; Tsukuba Science City, Japan. (Developments in toxicology and environmental science: v. 13.) Amsterdam: Elsevier Science Publishers BV; pp. 329-348

Ishinishi, N; Kuwabara, N; Takaki, Y; et al. (1988) Long-term inhalation experiments on diesel exhaust. In: Diesel exhaust and health risks: results of the HERP studies. Tsukuba, Ibaraki, Japan: Japan Automobile Research Institute, Inc., Research Committee for HERP Studies; pp. 11-84.

Mauderly, JL; Jones, RK; Griffith, WC; et al. (1987a) Diesel exhaust is a pulmonary carcinogen in rats exposed chronically by inhalation. Fundam Appl Toxicol 9:208-221.

Misiorowski, RL; Strom, KA; Vostal, JJ; et al. (1980) Lung biochemistry of rats chronically exposed to diesel particulates. In: Pepelko, WE; Danner, RM; Clarke, NA, eds. Health effects of diesel engine emissions: proceedings of an international symposium; December 1979. Cincinnati, OH: U.S. Environmental Protection Agency, Health Effects Research Laboratory; pp. 465-480; EPA report no. EPA-600/9-80-057a. Available from: NTIS, Springfield, VA; PB81-173809.

Nikula, KJ; Snipes, MB; Barr, EB; et al. (1995) Comparative pulmonary toxicities and carcinogenicities of chronically inhaled diesel exhaust and carbon black in F344 rats. Fundam Appl Toxicol 25:80-94.

Stone, M. (1998) Akaike's Criteria. In: Encyclopedia of Biostatistics, Armitage, P. and Colton, T., eds. Wiley, New York.

Strom, KA. (1984) Response of pulmonary cellular defenses to the inhalation of high concentrations of diesel exhaust. J Toxicol Environ Health 13:919-944.

U.S. Environmental Protection Agency (U.S. EPA). (1995) The use of the benchmark dose approach in health risk assessment. Washington, DC: Office of Research and Development, Risk Assessment Forum United States Environmental Protection Agency.

U.S. EPA. (2000) Benchmark dose software version 1.2. Washington, DC: National Center for Environmental Assessment, United States Environmental Protection Agency. Available: http://www.epa.gov/ncea/bmds.htm [2000, June 7].

Wolff, RK; Henderson, RF; Snipes, MB; et al. (1987) Alterations in particle accumulation and clearance in lungs of rats chronically exposed to diesel exhaust. Fundam Appl Toxicol 9:154-166.

Yu, CP; Yoon, K.J.; Chen, Y.K. (1991) Retention modeling of diesel exhaust particles in rats and humans. J. Aerosol Res. 4 (2): 79-115.