Attachment A
Determination of an AEL for n-Propyl Bromide and Its Documentation

1. AEL Derivation

Recommended AEL:	17 ppm (8-hour Time Weighted Average)
Basis and Endpoints:	A decrease in the number of estrous cycles in a 3-week period prior to mating following 7 weeks of exposure
Study:	An Inhalation Two-Generation Reproductive Toxicity Study of 1-Bromopropane in Rats (WIL 2001)
Protocol:	Whole-body inhalation, 6 hours/day, 7 days/week for 70 days prior to mating, during mating, gestation, lactation for two generations
Concentrations:	0, 100, 250, 500, or 750 ppm
BMDL:	162 ppm (mean number of estrous cycles in 3 weeks)
NOAEL:	100 ppm
LOAEL:	250 ppm (estrous cycle, sperm motility and hepatic effects)
BMDL [adj]:	(162 ppm × 6 hours / 8 hours × 7 days / 5 days = 170 ppm)
BMDL [HEC]:	170 ppm
Uncertainty Factors:	10 (composite factor of 3 for animal-to-human extrapolation and 3 for within-human variability to account for differences in individual sensitivity)
Results from BMD Analysis:	Section 3 presents the results from the benchmark dose analyses conducted on estrous cycle data.

1. Discussion of Relevant Literature

ICF has performed a re-evaluation of the literature on n-propyl bromide (1-bromopropane, nPB) for the purpose of assessing potential reproductive toxicity in females. This re-evaluation was prompted by the publication of several peer-reviewed studies (e.g., Sekiguchi 2002; Yamada et al. 2003; Ichihara et al. 2002, 2004 a, b) examining changes in estrous/menstrual cycle parameters in nPB-exposed groups. ICF previously evaluated a broad range of studies for numerous endpoints in order to assess quantitatively the most sensitive level at which adverse health effects occur and to develop an acceptable exposure limit as an 8-hour, time weighted average for the workplace (ICF 2002). Of the health effects examined previously, ICF found liver effects (centrilobular vacuolation), nervous system effects, and reproductive effects to be of concern. In 2002, ICF determined male reproductive effects (sperm motility) to be the most sensitive effect requiring protection at the lowest concentration.

1.1. Ichihara et al.

ICF performed a comprehensive, critical evaluation of all published occupational studies on nPB by Ichihara et al. (1999, 2002, 2004a, 2004b). Workers were examined using neurological, electrophysiological, hematological, biochemical, neurobehavioral and postural sway tests. ICF concluded that these studies were limited because of: small sample sizes; inadequate exposure characterization (e.g., a single time-weighted-average sample from each individual); very short study duration (2-3 days); incomplete information on how interview data were collected and validated; co-occurring exposures to other toxicants, including 2-bromopropane; and failure to adjust for numerous confounding variables. Therefore, ICF found that the Ichihara studies were insufficient for assessing the potential neurological, hematological and reproductive health effects (e.g., amenorrhea, adverse effects on the peripheral nerves and the central nervous system, and anemia) of nPB.

1.2. Sekiguchi

Sekiguchi (2002) observed an increase in estrous cycle length in female rats exposed via inhalation to 1000 ppm nPB for 3 weeks. However, this finding was not statistically significant, possibly because duration of exposure was too short to induce a significant effect.

1.3. Yamada et al.

In another animal study, Yamada et al. (2003), Wistar female rats (N = 9/dose group) were exposed to 0, 200, 400 or 800 ppm nPB for 8 hours/day, 7 days/week for 12 weeks. The Yamada et al. (2003) study found that, relative to controls, treated females exhibited (1) a significant decrease in the number of estrous cycles, due to prolonged diestrous, at 400 and 800 ppm; (2) a change in the subtype distribution of ovarian follicles with dose-dependent statistically significant decreases in antral follicles at 200 and 400 ppm, and in growing follicles at 400 ppm; and (3) an increase in the number of primordial follicles at 400 ppm that was not statistically significant. It should be noted that in this study, rats exposed to 800 ppm became "seriously ill" and were euthanized at week 8. Therefore, ovarian follicle subtype numbers in this group were reported, but excluded from statistical analysis. The decreases in antral follicles at 200 ppm could be a possible indicator of reduced ovulated oocytes and potentially the critical effect in the Yamada et al. (2003) study. However, changes in distribution of follicular subtype as a critical effect are insufficient in the absence of additional data. Statistically significant decreases in both antral follicles and corpora lutea would provide strong evidence of a significant toxicological effect. However, data on corpora lutea were not provided by Yamada et al. (2003). In the absence of data on corpora lutea, it is not possible to interpret the antral follicle findings; therefore, the use of estrous cycle changes instead of decreases in follicles as the critical effect in the Yamada study is scientifically justifiable and defensible. Estrous cycle irregularities in the Yamada et al. (2003)

study were first observed at approximately 2-3 weeks following initiation of exposure in the 800 ppm group and at around 7-9 weeks in the 400 ppm group.

1.4. WIL

Subsequently, ICF reviewed the WIL (2001) multi-generation reproductive/developmental toxicity study and noted findings on female reproductive toxicity similar to those of Yamada et al. (2003). In this study, male and female Sprague-Dawley F_0 and F_1 rats (N = 25/dose group) were exposed via inhalation to 0, 100, 250, 500 and 750 ppm for 70 days prior to mating and during mating, gestation, and lactation. Primordial follicles and corpora lutea were only counted in the controls and high-dose groups of the F_0 and F_1 females, and therefore, no dose-response could be established for these endpoints. However, significant increases in primordial follicles and decreases in corpora lutea were noted in both generations. An increase in estrous cycle length was also observed in each exposure group, relative to controls, prior to mating.

WIL (2001), however, did not statistically evaluate the estrous cycle data. In addition, the estrous cycle data reported by WIL (2001) excluded females that showed no evidence of cycling (1 and 2 animals for the 500- and 750-ppm F_0 groups, respectively). Therefore, ICF re-evaluated the individual data by (1) counting the number of estrous cycles within the three-week period prior to mating; (2) including females who did not complete an estrous cycle; and (3) conducting a statistical analysis of dose-response using an under-dispersed Poisson regression model. Using the new data, ICF performed statistical analyses on two estrous cycle measures: (1) mean cycle length; and (2) mean number of estrous cycles occurring during the 3-week period prior to mating, both following 7 weeks of nPB exposure.

For evaluation of differences in the mean number of estrous cycles occurring within the specified time period, data were assessed using an under-dispersed Poisson regression model. Since the number of cycles is a whole number, a Poisson model, appropriate for count data, was chosen to model the observed counts. The simple Poisson model with a variance equal to the mean fit the data poorly, and so a better fitting under-dispersed Possion model was chosen instead; this model has a variance equal to the mean multiplied by a scale factor. By comparison, the regression model assumes that the logarithm of the mean is a linear function of the dose.

Females who did not cycle (acyclic) during this period were considered to be in prolonged diestrous, and therefore were included in the analysis. Statistically significant decreases in the mean number of estrous cycles were observed at ≥ 250 ppm in F_0 females and at 500 ppm in F_1 females. (It should be noted that F_0 females exposed to 750 ppm produced no live litters and therefore, there was no F_1 generation at this exposure level.) Female reproductive toxicity findings are presented in Table A.1 and Table A.2.

Table A.1: Female Reproductive Endpoints[a]

Endpoint	0 ppm	100 ppm	250 ppm	500 ppm	750 ppm
F_0 Final Body Wt (g)	331±20.7[b]	330±22.3	327±24.8	332±38.3	319±25.5
F_1 Final Body Wt (g)	321±27.3	325±28.1	318±26.7	309±29.5	-
F_0 Ovaries (g)	0.1227±0.0259	0.1265±0.0240	0.1152±0.02360	0.1119±0.01514	0.09575±0.02798**
F_0 Relative Ovaries (g/100 g)	0.037±0.0078	0.038±0.0068	0.035±0.0072	0.034±0.0056	0.031±0.0079**
F_1 Ovaries (g)	0.1131±0.0155	0.1077±0.0317	0.1056±0.02791	0.1062±0.02302	-
F_1 Relative Ovaries (g/100 g)	0.035±0.0027	0.022±0.0032	0.022±0.0042	0.021±0.0045	-
F_0 Fertility index (%) N (Number of animals at exposure level)	92.0 N =25	100 N =25	88.0 N =25	52.0** N =25	0.0** N =25
F_1 Fertility index (%) N	100.0 N =25	84.0 N =25	80.0 N =25	100.0 N =25	-
F_0 Mating index (%) N	96.0 N =25	100 N =25	100 N =25	84.0 N =25	68.0* N =25
F_1 Mating index (%) N	88.0 N =25	68.0 N =25	64.0 N =25	72.0 N =25	-
F_0 Evidence of mating w/out delivery (no.)	1	0	3	10	17
F_1 Evidence of mating w/out delivery (no.)	3	4	4	8	-
Number of F_0 Implantation sites N	15.3±2.53 N = 23	14.3±3.09 N = 25	13.8±4.23 N =22	9.0±4.54** N =11	NA
Number of F_1 Implantation sites N	15.5±2.11 N =22	15.8±3.29 N =17	13.5±4.12 N =16	9.8±4.93** N =17	
F_0 Number of born N	15.0±2.42 N =23	13.6±3.09 N =25	12.5±4.27 N =22	8.5±4.41** N =11	NA
F_1 Number of born	14.9±1.97 N =22	15.1±3.35 N =17	13.1±4.12 N =16	8.6±4.51** N =17	-

Endpoint	0 ppm	100 ppm	250 ppm	500 ppm	750 ppm
N					
F_0 Number of Unaccounted Implantation Sites N	0.3±0.57 N =23	0.7±0.95 N =25	1.3±1.36** N =22	0.5±0.69 N =11	NA
F_1 Number of Unaccounted Implantation Sites N	0.5±0.86 N =22	0.6±1.22 N =17	0.4±0.63 N =16	1.2±1.09 N =17	

[a] Data were provided on pp. 123-124, 207, 272-275, and 356 of the study report
[b] Mean ± standard deviation

Table A.2: Estrous Cycle Data

Endpoint	0 ppm	100 ppm	250 ppm	500 ppm	750 ppm
F_0 Estrous cycle length (days)[b] N	4.2±0.49[a] N =25	4.5±1.05 N =25	4.7±0.90 N =25	5.5±2.17* N =23	5.6±1.79* N =22
F_1 Estrous cycle length (days)[b] N	4.5±1.25 N =24	4.5±0.91 N =24	4.9±1.43 N =22	5.1±1.68 N =21	-
F_0 Mean no. of estrous cycles within 3 weeks[c] N	3.96±0.54 N =25	3.84±0.62 N =25	3.52±0.65* N =25	2.88±1.17** N =25	2.56±1.26** N =25
F_1 Mean no. of estrous cycles within 3 weeks[c] N	3.64±1.15 N =25	3.68±1.11 N =25	2.88±1.36 N =25	2.68±1.35* N =25	-
F_0 Mean no. of estrous cycles within 3 weeks excluding acyclic females[d] N	3.96±0.54 N =25	3.84±0.62 N =25	3.52±0.65** N =25	3.00±1.02*** N =24	2.78±1.04*** N =23
F_1 Mean no. of estrous cycles within 3 weeks excluding acyclic females[d] N	3.64±1.15 N =25	3.68±1.11 N =25	3.13±1.10 N =23	2.91±1.12* N =23	-

[a] Mean ± standard deviation
[b] Statistical analysis calculated by ICF
[c] Calculated by ICF using an under-dispersed Poisson log-linear regression model
[d] Calculated by ICF excluding females that did not have a full estrous cycle assuming an under-dispersed Poisson log-linear regression model.
* Significantly different from control, $p<0.05$.
** Significantly different from control, $p<0.01$.
*** Significantly different from control, $p<0.001$.

To gain support for the selection of the estrous cycle endpoint, an expert panel was formed to consider the occupational exposure limit for nPB and the toxicological significance of a statistically significant decrease in the number of estrous cycles within a given time period. Dr. George Daston, Dr. Ulrike Luderer, and Dr. Jodi Flaws were asked to serve on the panel because of their expertise on this subject and the experience of the first two scientists as panel members

for the Center for the Evaluation of Risks to Human Reproduction's 2001 review of the reproductive toxicity of nPB (CERHR 2001). The peer reviewers offered constructive comments that were, in general, supportive of the critical endpoint selected for derivation of the industrial AEL and of the modeling approach used by ICF (2004).

2. AEL Determination

The WIL (2001) study was selected as the principal study for the AEL determination because it (1) was a well-conducted experiment performed in accordance with standard test guidelines for multi-generation reproductive toxicity; (2) was documented using Good Laboratory Practice procedures; (3) underwent an independent audit; (4) had sufficiently large sample sizes; and (5) provided raw data for review and analysis. As identified earlier, the Yamada et al. (2003) study was not used in the AEL determination because (1) limited data were provided, precluding audit and independent analysis; (2) the study sample sizes were small; (3) the high dose greatly exceeded the maximum tolerated dose, and animals in this group had to be euthanized at week 8 because of severe illness; and (4) fewer reproductive parameters were measured than in the WIL (2001) study.

Regulatory guidance for interpreting the biological significance of estrous cycle irregularities in the absence of additional reproductive effects is not conclusive. Although EPA's Guidelines for Reproductive Toxicity Risk Assessment (USEPA 1996) notes that these effects are to be considered indicative of potential reproductive toxicity, NRC (2001) concludes that changes in the distribution of estrous cycle length alone are not a reliable predictor of reproductive toxicity. In the WIL (2001) study, statistically significant changes in the number of estrous cycles in a three-week period prior to mating, due to an increase in the cycle length and especially the diestrous phase, were an early precursor to a functional reproductive effect that occurred with increasing dose, as noted in Table A.1 and Table A.2.

Further, the relevance of this endpoint was discussed with Drs. Sally Darney and Ralph Cooper, reproductive toxicity experts at EPA's NHERL (National Health and Environmental Effects Research Lab at Research Triangle Park) (Birgfeld 2004). Both scientists agreed that estrous cycle length is a relevant endpoint to use to determine an AEL and that performing a BMD analysis on estrous cycle length might be possible. Dr. Darney further suggested that the data be transformed into number of estrous cycles within a three-week period to avoid problems in the data generated by acyclic rats. Therefore, estrous cycle length is considered to be the critical effect because it is the effect occurring at the lowest concentration along a continuum of adverse reproductive outcomes that increase in frequency and severity at higher doses. The number of estrous cycles in a 3-week period prior to mating was used instead of the estrous cycle length in order to allow inclusion of the data from acyclic females, for which an estrous cycle length is not defined.

Although nPB inhalation exposure produces reproductive toxicity in both male and female rats, a weight-of-evidence comparison of the reproductive hazard findings suggested that females in the F_0 generation were slightly more sensitive than the F_1 females (with regard to estrous cycle length), the F_0, or the F_1 males (see Table A.2), ensuring that this endpoint will be sufficiently protective. After considering the appropriate uncertainty factors for each endpoint, the reproductive endpoints for both males and females result in lower AELs than those for neurotoxicity or liver effects (ICF 2002). Further, as seen in Table A.1 and Table A.2, the F_0 females show a much clearer dose-response for many of the reproductive endpoints than do the F_1 females. The pattern of exposure in the F_0 generation resembles that occurring in occupational exposure scenarios and is relevant to occupational exposure in humans. Therefore, the BMDL for

F_0 female reproductive toxicity was used as the point of departure for development of the AEL.

3. Benchmark Dose Methods and Analysis

3.1. Background

The Benchmark Dose (BMD) analysis utilized in the determination of the AEL for nPB was based on the mean number of estrous cycles for the F_0 generation within the 3-week period during which estrous cycle measurements were made, prior to mating[1] (Table A.3). Based on a weight-of-evidence hazard characterization and biological relevance, this endpoint in the F_0 generation was considered to be the most sensitive reproductive endpoint for the study. Acyclic females were included in the analysis.

Table A.3: F_0 Mean Number of Estrous Cycles in Female Sprague-Dawley Rats Administered n-Propyl Bromide via Inhalation for 70 Days[a]

Dose (ppm)	F_0 Animals (N)	Estrous Cycles (mean)
0	25	3.96
100	25	3.84
250	25	3.52
500	25	2.88
750	25	2.56

[a] Estrous cycles measured for three-week period prior to mating
N, number of animals per group

The data sets considered for dose-response modeling are discrete and categorical, since the number of estrous cycles is a whole number. The EPA's Benchmark Dose Software (BMDS) (Version 1.3.2, EPA 2000a) was used to accomplish all of the model fitting and BMD estimation. This most recent version of BMDS is designed for either dichotomous data, with only two possible responses, or for continuous data with infinitely many responses. The BMDS model required that the data be treated as both approximately continuous and approximately normally distributed. The continuous endpoints of interest with respect to toxicity were quantitatively summarized by group means and measures of variability (standard errors or standard deviations). The models used to represent the dose-response behavior of those continuous endpoints are those implemented in EPA's BMDS. These models were the power models, the Hill models, and the polynomial models, including the linear model. The BMDS methods and models applied to continuous endpoints are presented in Section 4.

3.2. Definition of the Benchmark Response (BMR) and Corresponding BMD and BMDL

BMDs were implicitly defined as follows:

$$\frac{\mu(BMD) - \mu(0)}{\mu(0)} = 0.1 \qquad \text{(Eq. 1)}$$

In other words, the BMR was defined as a 10 percent change in mean. BMDLs were defined as the 95 percent lower bound on the corresponding BMD. Confidence intervals were calculated

[1] For the rest of this document, this endpoint is simply referred to as "estrous cycles," without the clarifier that they were measured within a 3-week period.

using a profile likelihood method.

Dr. George Daston, a member of the expert panel, recommended that the 10 percent difference in mean for this endpoint not be used as a BMR because of its difference from a typical 10 percent change in the probability of a response for a quantal variable (ICF 2004). It was stated that the 10 percent change in the mean number of estrous cycles might still be within the range of normal values for this endpoint in female rats. Dr. Daston further suggested that, ideally, ICF should find a scientific consensus on the normal range for this value, if such agreement exists. However, no scientific consensus exists with regard to this value at present. Solicitation was not pursued given that such a discussion among reproductive experts would likely not yield a definitive answer. It would, of necessity, rely on archives of historical data and would therefore be a major science policy undertaking. Further, research areas for a few laboratories indicate that measuring estrous cycles in various animal species is ongoing (IZW 2005). These points illustrate that data are not readily available and would require a significant research effort to further investigate. Finally, the cycle ranges of the control rats in the nPB studies cited are more relevant for comparison than what is considered to be the normal range for all rats. As Dr. Luderer discussed, estrous cycle lengths at the 250, 500, and 750 ppm doses were outside the normal range found in the WIL (2001) study (see discussion below). In the absence of an agreed normal range, Dr. Daston listed several other options for determining a biologically significant change, such as using the mean response equal to 0.5 control standard deviations from the control mean (ICF 2004). However, these options were not consistent with EPA guidance, which recommends the use of one standard deviation, and were thus not used.

Alternate approaches, based on data variability, were considered and discarded. There are several reasons why these alternate approaches were not used and why ICF is not basing the AEL on a BMDL value based on one standard deviation of variability. First, the number of estrous cycles is not truly a continuous variable, but instead is a categorical variable with whole number values. It does not neatly fit into the two types of variables addressed by EPA's Benchmark Dose Technical Guidelines: quantal and continuous. Secondly, the control data failed tests for normality of distribution (Shapiro-Wilk, Anderson-Darling, Cramer–von Mises, and Kolmogorov-Smirnov; $p<0.05$). This suggested that the standard formula used to calculate the standard deviation, which is based upon assumptions of normality, would not be a good (i.e., statistically efficient) estimate of the true standard deviation of the data. Rather than attempt to calculate a more precise standard deviation estimate based upon some alternate distribution function for the data, which also would have been inconsistent with the normal approximation used for the dose-response and BMD modeling, the choice was made to base the BMR upon a 10 percent relative change.

The 10 percent response is scientifically justified for several reasons. The dose-response curve in the WIL study indicates that the number of estrous cycles decreases in F_0 females with increasing dose (Table A.2). The differences from controls are statistically significant at 250, 500 and 750 ppm. The 500-ppm dose included one acyclic female. At 750 ppm the rats were completely incapable of producing an F_1 generation. Therefore, the dose-response curve is very steep from animals with normal estrous cycle lengths at the low dose to ones that cannot produce any offspring at all at the high dose.

These data indicate that the number of estrous cycles is a sensitive early indicator of reproductive success. The decrease in the mean number of estrous cycles at 250 ppm was 11 percent, which is slightly higher than the 10 percent change modeled using BMD according to Eq. 1 above. The next higher dose, 500 ppm, began to produce acyclic females. Thus, a dose somewhere between 100 and 250 ppm is the maximum one that will cause a decrease in the number of estrous cycles without disrupting the ovarian cycle of any of the animals in the study. Therefore, a 10 percent

difference in the mean number of estrous cycles in a 3-week period is a protective endpoint to calculate, because all of the female animals would still exhibit a normal range of cyclicity (e.g., comparable to controls) and reproduction should not be impaired.

Therefore, without additional information on the exact dose at which this decrease in estrous cycles prevents estrous cycling altogether and thus prevents reproduction for at least one animal, the 10 percent level was chosen as appropriate. It is appropriate because of its consistency with BMD technical guidance and because the 10 percent level is a value consistent with the BMDL levels chosen for other adverse outcomes measured in the animals from the WIL study (e.g., sperm motility in males and liver effects in males and females). Consistent with the BMD guidance (EPA 2000), the data have also been modeled as continuous, using one standard deviation from the mean as an appropriate level of change, and presented in the graph and BMDL using the linear model (the model with the best fit) at the end of Section 3.3.1. The BMDL for this model was 208 ppm, which is greater than that obtained with the models using a 10 percent change in the mean (see analysis below).

Additional support for selection of this endpoint and the measure of response came from Dr. Ulrike Luderer, of the peer review panel. Dr. Luderer indicated in her comments that there were four reasons why the measured number of estrous cycles over a period of time was a valid endpoint (ICF 2004):

- There is a clear dose-response relationship of decreasing estrous cycles with increasing dose;
- The alterations in estrous cycle length fall outside the historic range of estrous cycle duration for the laboratory that conducted one of the studies (Stump 2001 [sic] refers to WIL 2001 report);
- Alterations in estrous cycle length are not isolated findings, but occur in the context of other effects on the female reproductive system; and,
- Effects on estrous cycling were observed in two independent studies of nPB exposure (Stump 2001; Yamada et al. 2003).

These arguments from Dr. Luderer provide significant support for ICF's approach.

3.3. Benchmark Dose-Response Modeling Results and Choice of BMDL

The results of the benchmark modeling are presented in Table A.4. Background material regarding the models and software used, as well as an extensive explanation of the decision tree (below) and the criteria for deciding which models provided the best fit, are provided in Section 4. The following guidelines were used in the selection of BMDLs for each data set:

1. Models with an unacceptable fit (including consideration of local fit in the low-dose region) were excluded. Visual fit, particularly in the low-dose region, was assessed for models that had acceptable global goodness of fit.
2. If the BMDL values for the remaining models for a given endpoint were within a factor of three, no model dependence was assumed, and the models were considered indistinguishable in the context of the precision of the methods. The models were then ranked according to the AIC, and the model with the lowest AIC was chosen as the basis for the BMDL.
3. If the BMDL values were not within a factor of three, some model dependence was assumed, and the lowest BMDL was selected as a reasonable conservative estimate, unless it was an outlier compared to the results from all of the other models. Note that when outliers are removed, the remaining BMDLs may then be within a factor of three, and so the criteria given in item two would be applied.

These criteria were applied to the BMDLs reported in Table A.4 for each endpoint.

Table A.4: Comparison of BMD Modeling Results and Final Decision

Model-Variance Model	AIC	P-Value	BMD	BMDL	Sufficient Data?	Good Visual Fit in the Low-Dose Portion of the DR?	BMDLs Within a Factor of 3?	Lowest AIC?
Linear-Homogeneous	98	0.88	201	168	Y	Y	Y	N
Linear-Heterogeneous[1]	**75**	**0.42**	**200**	**162**	**Y**	**Y**	**Y**	**Y**
Polynomial (2)-Homogeneous	100	0.75	178	109	Y	Y	Y	N
Polynomial (2)-Heterogeneous	77	0.29	176	122	Y	Y	Y	N
Polynomial (3)-Homogeneous	102	1.00	232	102	Y	Y	Y	N
Polynomial (3)-Heterogeneous	76	0.64	234	146	Y	Y	Y	N
Power-Homogeneous	102	0.43	201	168	Y	Y	Y	N
Power-Heterogeneous	77	0.25	200	162	Y	Y	Y	N
Hill-Homogeneous	104	0.84	233	103	Y	Y	Y	N
Hill-Heterogeneous	77	0.41	249	143	Y	Y	Y	N

[1] Model that best fits the data; these results were used to derive the AEL
Reference: WIL Study 2000

3.3.1. Decision Tree Based Upon EPA Benchmark Dose Technical Guidance Document[2]

- Assess goodness-of-fit, using a value of $\alpha = 0.1$ to determine a critical value using the Chi Square test.
  All models pass.
- Further reject models that apparently do not adequately describe the relevant low-dose portion of the dose-response, examining residuals and graphs of model and data.
  All models pass. Graphs of the model fits with heterogeneous variance are included in Figure A.1. As dose increases the number of estrous cycles goes down and therefore the dose response curve is downward sloping.

As the models remaining have met the default statistical criteria for adequacy and visually fit the data, any of them theoretically could be used for determining the BMDL. The remaining criteria for selecting the BMDL are adopted as defaults.

- If the BMDL estimates from the remaining models are within a factor of three of each other, then they are considered to show no appreciable model dependence and will be considered indistinguishable in the context of the precision of the methods.
  All BMDL estimates within a factor of three.
- Models are ranked based on the values of their Akaike Information Criterion (AIC), a measure of the deviance of the model fit adjusted for the degrees of freedom, and the model with the lowest AIC is used to calculate the BMDL. If this is not unique, the simple average or geometric mean of the BMDLs with the lowest AIC is used.
  Linear model with heterogeneous variance has lowest AIC (75). BMDL from this model (162 ppm) is used.

The uncertainty analysis associated with the selection of the models, as recommended by the

[2] EPA 2000b

EPA's Benchmark Dose Technical Guidance, is represented in Table A.5 below. This guidance addresses two types of variables: quantal and continuous. The chosen endpoint, the number of estrous cycles, does not represent either type as it is not truly a continuous variable, but instead is a categorical variable with whole number values. Therefore, the choice was made to base the BMR upon a 10 percent relative change as discussed above in Section 3.2.

Table A.5: Uncertainty Analysis Recommended for Each Type of Data

Type of Data	Uncertainty Analysis Recommended by Benchmark Does Technical Guidance		
	Percent Change	**One Standard Deviation[c]**	**Percent Change and Compare to Standard Deviation**
Quantal (Dichotomous) Data			
• Sensitive studies (e.g., reproductive, developmental, epidemiology)[a]	✓		
• All other quantal data[b]	✓		
Continuous Data			
• Minimal level of change in significant endpoint			✓
• Level of adverse response known			✓
• Level of adverse response unknown		✓	
Type of Data	**Uncertainty Analysis Recommendations Not Covered by Benchmark Dose Technical Guidance**		
• Categorical variables with whole number values[d]			✓

[a] Reproductive and developmental studies typically use five percent change uncertainty, and epidemiology studies typically use one percent change.

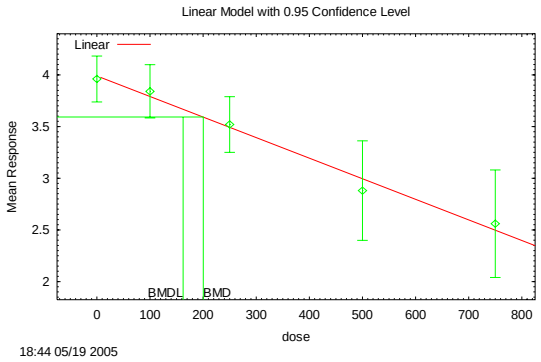
[b] All other quantal data typically uses 10 percent change uncertainty.

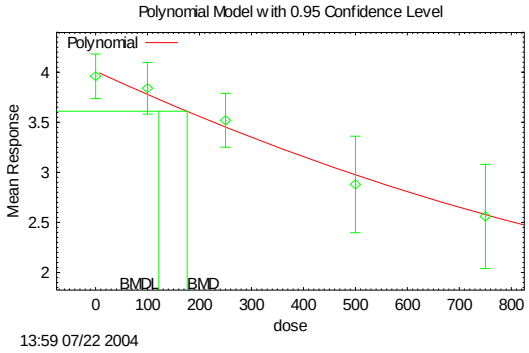
[c] An uncertainty of one standard deviation is generally only applied to normally distributed data. Source: EPA. 2000a. Benchmark Dose Technical Guidance Document. EPA/630/R-00/001. External Review Draft. Risk Assessment Forum, Washington, DC. October 2000.

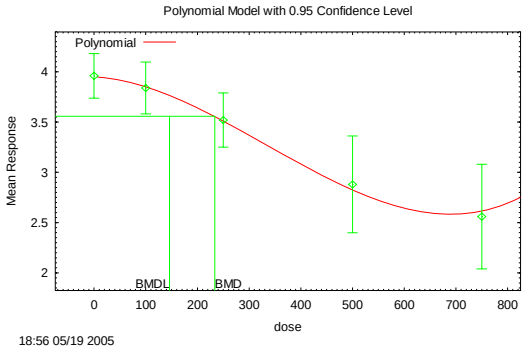
[d] In the absence of EPA guidance on how to assign uncertainty to categorical, whole number data, an expert decision was made to base the BMR upon a 10 percent relative change and compare these results to one standard deviation uncertainty, as discussed above in Section 3.2.

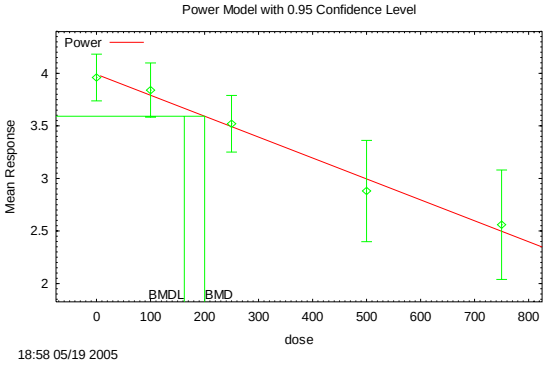
Figure A.1: Graphs of Heterogeneous Model Fits to Data

Dose is measured in parts per million and response is measured as the mean number of estrous cycles in 3 weeks. As dose increases the number of estrous cycles every 3-weeks goes down and therefore the dose response curve is downward sloping.









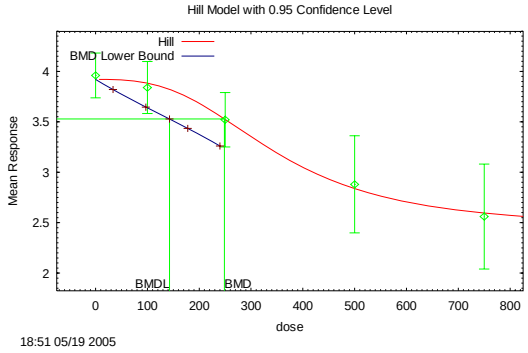
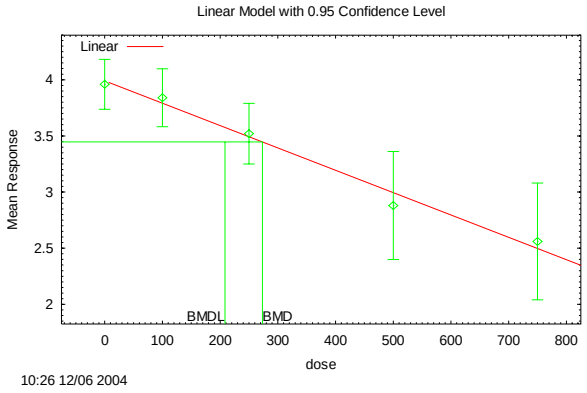


Figure A.2: Graph of Linear Model Fit to Data Using One Standard Deviation
Dose = Parts per million
Response = Mean number of estrous cycles in 3 weeks



3.3.2. Discussion

From the goodness-of-fit summary statistics and the visual plots, the two best candidates are the linear and quadratic model for the mean with a heterogeneous model for the variance. They have similar AIC values, although the linear model is preferred on the basis of having a lower AIC. In general, if a simpler model gives as good a fit or a better fit to the data than a more complicated model, then the simpler model is preferred. Both models assume the variance is a multiple (alpha) of the mean raised to the power rho, where alpha and rho are estimated from the data. The linear model for the mean assumes that the mean number of estrous cycles is of the form

$$\text{Effect} = \beta_0 + \beta_1 \text{dose}$$

The quadratic model for the mean assumes that the mean number of estrous cycles is of the form

$$\text{Effect} = \beta_0 + \beta_1 \text{dose} + \beta_2 \text{dose}^2$$

There are several statistical tests that can be used to compare the linear and quadratic models, all leading to the same conclusion that the linear model is preferred:

1. The AIC comparison compares the AIC values for the two models, where AIC equals -2L + 2p, L is the log-likelihood, and p is the number of unknown parameters. Models with too many parameters are penalized by the AIC statistic. The better fitting models have the lower AIC values so that the linear model, with AIC = 75, is preferred to the quadratic model, with AIC = 77. Note that the size of the AIC itself does not have a statistical interpretation; only differences between AIC values are meaningful. Note also that there is no general statistical procedure for comparing two AICs to see if the models are statistically significantly different.
2. Because the linear model is a special case of the quadratic model when the quadratic coefficient β_2 equals zero, a standard statistical test is used to examine the quadratic coefficient and test whether it is statistically significantly different from zero. Because the Wald 95 percent confidence interval for β_2 ranges from -2.0E-6 to +3.3E-6, which includes zero, the quadratic coefficient is not statistically significantly different from zero. Thus the more complicated quadratic model does not show a significant improvement over the linear model and so the linear model is preferred.
3. Because the linear model is a special case of the quadratic model when the quadratic coefficient β_2 equals zero, another standard statistical test is a chi-square test based on

twice the difference in log-likelihoods, -2 ΔL, as presented in Section 3.2. Since -2 ΔL = 0 (to one decimal place), which is less than the critical value 3.84, the 95th percentile of a chi-square distribution with one degree of freedom, the chi-square test is not statistically significant. Again, the more complicated quadratic model does not show a significant improvement over the linear model and so the linear model is preferred.

By adding an extra parameter that is not statistically significant, the other parameters are less precisely estimated, so that the predicted values for the quadratic model are less precise than for the linear model. Although the quadratic model has a lower BMD, this predicted dose value is more uncertain than for the linear model. This is reflected in a wider confidence interval for the BMD for the quadratic model.

4. Background on nPB AEL Benchmark Dose Analysis

4.1. BMDS Models

The models used to represent the dose-response behavior of those continuous endpoints are those implemented in EPA's BMDS. These models were the power models, the Hill models, and the polynomial models, including the linear model. These mathematical models fit to the data are defined here. In all cases, μ(d) indicates the mean of the response variable following exposure to "dose" d.

The power model is represented by the equation

$$\mu(d) = \gamma + \beta d^{\alpha}$$

where the parameter $\alpha > 0$.

The Hill model is given by the following equation:

$$\mu(d) = \gamma + (vd^n) / (d^n + k^n)$$

where the parameter k is greater than 0 and n is greater than 1 (USEPA 2000b).The polynomial model is defined as:

$$\mu(d) = \beta_0 + \beta_1 d + ... + \beta_n d^n$$

where the degree of the polynomial, n, was set to less than the number of dose groups in the experiment being analyzed. The linear model is a special case of the polynomial model where n=1.

In the case of continuous endpoints, one must assume something about the distribution of individual observations around the dose-specific mean values defined by the above models. The assumptions imposed by BMDS were used in this analysis, i.e., individual observations were assumed to vary normally around the means with heterogeneous variances given by the following equation:

$$\sigma_i^2 = \alpha[\mu(d_i)]^{\rho}$$

where both α and ρ were parameters estimated by the model. Also fitted were homogeneous variance models where ρ=0. As discussed above, the data used for these analyses were discrete,

integer-valued, and the normal distribution was used as an approximation to their distribution. The validity of this approximation to the joint probability distribution uses the central limit theorem of statistics.

Given the above assumptions about variations around the means, maximum likelihood[3] methods, were applied to estimate all of the parameters, where the log-likelihood to be maximized is (except for an additive constant) given by

$$L = \Sigma [(Ni/2) \ln(\sigma i2) + (Ni - 1)si2/2\sigma i2 + Ni\{mi - \mu(di)\}2/2\sigma i2]$$

where N_i is the number of individuals in group i exposed to dose d_i, and m_i and s_i are the observed mean and standard deviation for that group, respectively. The summation runs over i from 1 to k (the number of dose groups).

4.2. Goodness-of-Fit Analysis

The BMDS software provides three or four different Tests of Fit that the user may use to determine an appropriate model for fitting data. These Tests of Fit are based on asymptotic theories of the likelihood ratio. The likelihood ratio represents the ratio of two likelihood values, many of which are given in the BMDS output. Statistical theory proves that -2*log(likelihood ratio) converges to a Chi-Square random variable as the sample size gets large and the number of dose levels gets large. These values can in turn be used to obtain approximate probabilities to make decisions about model fit.

Each of the ten fitted models has a likelihood value. The BMDS program uses these values to create ratios from two models that form a meaningful test. Suppose the user wishes to test two models for fit, A and B. One assumption that is made for these tests is that the "true" model is in fact B, but it can be simplified in such a way that the simplified model describes the data as well as B. Also suppose A is a much simpler model in that it has much fewer parameter values (the goal is to simplify the model as much as possible without losing information about the data). Assume each model has a maximum likelihood value, call them L(A) and L(B). A ratio can be formulated easily: L(A)/L(B). (Note: The model with a higher number of parameters is always in the denominator of this ratio). Now, using the theory, -2*log{ L(A)/L(B)} approaches a Chi-Square random variable. This can be simplified by using the fact that the log of a ratio is equal to the difference of the logs, or simply put, -2*log{ L(A)/L(B)} = -2*(log{L(A)} - log{L(B)}) = 2*log{L(B)} - 2*log{L(A)}. The likelihood values given by BMDS are in fact the log likelihoods, therefore this becomes a subtraction problem. This value can then in turn be compared to a Chi-Square random variable with a specified number of degrees of freedom.

Each log likelihood value has an associated number of degrees of freedom[4]. The number of degrees of freedom for the Chi-Square test statistic is merely the difference between the two model degrees of freedom. In the mini-example above, suppose L(A) has 5 degrees of freedom, and L(B) has 8. In this case, the Chi-Square value you would compare this to would be a Chi-Square with 8 - 5 = 3 degrees of freedom.

In the A vs B example, what is exactly being tested? In terms of hypotheses, it would be:

[3] Maximum likelihood is an estimate of a population parameter most likely to have produced the sample observations (EPA 2000b).

[4] Degrees of Freedom is the difference between the number of data points and the number of parameters in the model (EPA 2000b).

H_0: A models the data as well as B
H_1: B models the data better than A

Keeping these tests in mind, suppose 2*log{L(B)} - 2*log{L(A)} = 4.89 based on 3 degrees of freedom. Also, suppose the rejection criterion is a Chi-Square probability of less than 0.10. Looking on a Chi-Square table, 4.89 has a p-value somewhere between 0.10 and 0.25. In this case, H_0 would not be rejected, and it would seem to be appropriate to model the data using Model A.
The BMDS software provides three or four default tests, depending on the variance model the user has specified (constant variance model, or a non-constant variance model where the variance is a function of the mean, namely,

$$\sigma_i^2=\alpha\mu_i^\rho$$

BMDS assumes the rejection criterion is a Chi-Square probability of less than 0.10 for all of the tests; however p values are presented so that the user is free to use any rejection criteria. Each test in each model will be discussed in some detail below.

Test 1: Tests the hypothesis that response and variance don't differ among dose levels. If this test is not rejected, there may not be a dose-response relationship, although it is possible for some data sets with a slightly significant trend to not reject this test. This model implies no differences in the mean or in the variance at each dose level, and thus, there would be no adverse effect as dosage is increased. If this test is rejected, then modeling the data is appropriate, and the user should consider the tests below.

Test 2: Tests the hypothesis that variances are homogeneous. Recall that the goal is to simplify the model. If this test is not rejected, the simpler constant variance model may be appropriate. If this test is rejected, the user may want to run a non-constant variance model, or if the non-constant variance model was run, then the user should look at the second test 3 below to make further decisions.

Test 3 (Test 4 is a test of the variance model): Tests the hypothesis that the model for the mean fits the data. If this test is not rejected, the user has support for the selected model. If this test is rejected, the user may want to try a different model.

Test 4 (Non-constant variance model): Tests the hypothesis that the variances are adequately modeled. Here, the test is to see whether or not the variance model, $\sigma_i^2=\alpha\mu_i^\rho$, is an appropriate assumption. Again, the purpose is to reduce the parameter space, and by modeling the variances as a function of the mean (which also intuitively makes sense that variance may have some dependence on the mean value) we achieve some reduction. If this test is not rejected, it may be appropriate to conclude that the true variances have the form above. If this test is rejected, BMDS has no further way to model variance.

Visual fit, particularly in the low-dose region, was assessed for models that had an acceptable global goodness of fit. Acceptable global goodness of fit was either a p-value > 0.1, or a perfect fit when there were no degrees of freedom for a statistical test of fit. Local fit was evaluated visually on the graphic output, by comparing the observed and estimated results at each data point.

Goodness-of-fit statistics are not designed to compare different models, particularly if the different models have different numbers of parameters. Within a family of models, adding

parameters generally improves the fit. BMDS reports the Akaike Information Criterion (AIC) to aid in comparing the fit of different models. The AIC is defined as –2L+2p, where L is the log-likelihood at the maximum likelihood estimates for the parameters, and p is the number of model parameters estimated. When comparing the fit of two or more models to a single data set, the model with the lesser AIC was considered to provide a superior fit.

5. Selected Output from BMDS

```
 ================================================================
        Polynomial Model. Revision: 2.2  Date: 9/12/2002
        Input Data File: C:\BMDS\DATA\NPB-EST.(d)
        Gnuplot Plotting File:  C:\BMDS\DATA\NPB-EST.plt
                                          Wed Jun 29 17:09:10 2005
 ================================================================

 BMDS MODEL RUN
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

   The form of the response function is:

   Y[dose] = beta_0 + beta_1*dose + beta_2*dose^2 + ...


   Dependent variable = MEAN
   Independent variable = dose
   Signs of the polynomial coefficients are not restricted
   The variance is to be modeled as Var(i) = alpha*mean(i)^rho

   Total number of dose groups = 5
   Total number of records with missing values = 0
   Maximum number of iterations = 250
   Relative Function Convergence has been set to: 1e-008
   Parameter Convergence has been set to: 1e-008



                  Default Initial Parameter Values
                          alpha =     0.811333
                            rho =            0
                         beta_0 =      3.98617
                         beta_1 =  -0.00198177



                                Parameter Estimates

                                                          95.0% Wald Confidence
Interval
       Variable         Estimate        Std. Err.     Lower Conf. Limit   Upper
Conf. Limit
          alpha          73.6006          69.1595            -61.9495
209.151
            rho         -3.98394         0.778685            -5.51013
-2.45774
         beta_0           3.9907        0.0841431             3.82579
4.15562
         beta_1      -0.00199298      0.000310492         -0.00260154         -
0.00138443


           Asymptotic Correlation Matrix of Parameter Estimates

                  alpha          rho       beta_0       beta_1
     alpha            1        -0.99      0.00051       0.0061
       rho        -0.99            1      -0.0012       -0.005
    beta_0      0.00051      -0.0012            1        -0.65
    beta_1       0.0061       -0.005        -0.65            1
```

Table of Data and Estimated Values of Interest

Dose	N	Obs Mean	Obs Std Dev	Est Mean	Est Std Dev	Chi^2 Res.
0	25	3.96	0.539	3.99	0.545	-0.282
100	25	3.84	0.624	3.79	0.603	0.403
250	25	3.52	0.653	3.49	0.71	0.194
500	25	2.88	1.17	2.99	0.965	-0.592
750	25	2.56	1.26	2.5	1.39	0.231

Model Descriptions for likelihoods calculated

Model A1: Yij = Mu(i) + e(ij)
Var{e(ij)} = Sigma^2

Model A2: Yij = Mu(i) + e(ij)
Var{e(ij)} = Sigma(i)^2

Model A3: Yij = Mu(i) + e(ij)
Var{e(ij)} = alpha*(Mu(i))^rho

Model R: Yi = Mu + e(i)
Var{e(i)} = Sigma^2

Likelihoods of Interest

Model	Log(likelihood)	DF	AIC
A1	-46.881357	6	105.762714
A2	-31.698420	10	83.396841
A3	-32.027970	7	78.055939
fitted	-33.432720	4	74.865440
R	-67.085741	2	138.171481

Explanation of Tests

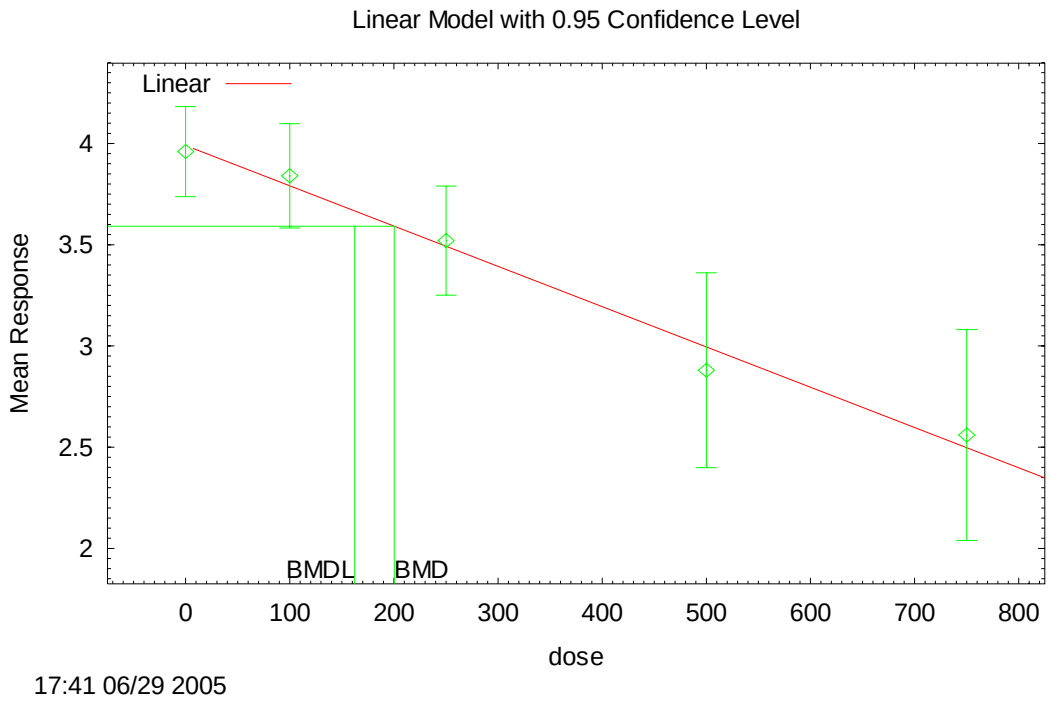
Test 1: Does response and/or variances differ among Dose levels?
(A2 vs. R)
Test 2: Are Variances Homogeneous? (A1 vs A2)
Test 3: Are variances adequately modeled? (A2 vs. A3)
Test 4: Does the Model for the Mean Fit? (A3 vs. fitted)

Tests of Interest

Test	-2*log(Likelihood Ratio)	Test df	p-value
Test 1	70.7746	8	<.0001
Test 2	30.3659	4	<.0001
Test 3	0.659098	3	0.8828
Test 4	2.8095	3	0.4219

The p-value for Test 1 is less than .05. There appears to be a difference between response and/or variances among the dose levels
It seems appropriate to model the data

```
The p-value for Test 2 is less than .05.  A non-homogeneous variance model
appears to be appropriate

The p-value for Test 3 is greater than .05.  The modeled variance appears
 to be appropriate here

The p-value for Test 4 is greater than .05.  The model chosen seems to
adequately describe the data


 Benchmark Dose Computation
Specified effect =           0.1

Risk Type        =     Relative risk

Confidence level =          0.95

              BMD =       200.238


             BMDL =       162.391


BMDL computation failed for one or more point on the BMDL curve.

                    The BMDL curve will not be plotted
```



6. References

Birgfeld E. 2004. Conversation with Sally Darney and Ralph Cooper, May 27, 2004.

CERHR. 2002. NTP-CERHR Expert Panel Report on the Reproductive and Developmental Toxicity of 1-Bromopropane. National Toxicology Program. NTP-CERHR-1-BP-02. March 2002.
EPA. 2001. Help Manual for Benchmark Dose Software. Office of Research and Development, Environmental Protection Agency. Washington, DC. EPA 600/R-00/014F.

EPA. 2000a. Benchmark Dose Software. National Center for Environmental Assessment with technical support from the National Health & Environmental Effects Research Laboratory. www.epa.gov.ncea/bmds.htm

EPA. 2000b. Benchmark Dose Technical Guidance Document. EPA/630/R-00/001. External Review Draft. Risk Assessment Form, Washington, DC. October 2000.

EPA. 1996. Guidelines for Reproductive Toxicity Risk Assessment. U.S. Environmental Protection Agency, Risk Assessment Forum, Washington, DC, 630/R_96/009, 1996.

ICF Consulting. 2004. nPB Expert Panel Response Final Document. Delivered to the U.S. Environmental Protection Agency on December 13, 2004.

ICF Consulting. 2002. Acceptable Industrial Exposure Limit for n-Propyl Bromide. Delivered to the U.S. Environmental Protection Agency on May 3, 2002.

Ichihara et al. 2004a. A Survey on Exposure Level, Health Status, and Biomarkers in Workers Exposed to 1-Bromopropane. Am. J. Ind. Med. 45:63–75. Wiley-Liss, Inc. 2004.

Ichihara et al. 2004b. Neurological Abnormalities in Workers of 1-Bromopropane Factory, Environmental Health Perspectives published by the National Institute of Environmental Health Sciences, National Institutes of Health, U.S. Department of Health and Human Services, June 2004.

Ichihara et al. 2002 Neurological disorders in three workers exposed to 1-bromopropane. Journal of Occupational Health 44:1-7 (2002).

Ichihara G., Jong X., Onizuka J., et al. 1999. Histopathological changes of nervous system and reproductive organ and blood biochemical findings in rats exposed to 1-bromopropane. Abstracts of the 72nd Annual Meeting of Japan Society for Occupational Health. May 1999. Tokyo.

IZW. 2005. Institute for Zoo and Wildlife Research. "Comprehensive Science: Non-invasive Monitoring of Hormones. Accessed on the World Wide Web on May 31, 2005 at http://www.izw-berlin.de/en/research/fg4/index.html?reproduktionsmonitoring.html~rechts.

National Research Council (NRC). 2001 Evaluating Chemical and Other Agent Exposures for Reproductive and Developmental Toxicity, Commission on Life Sciences (CLS), NATIONAL ACADEMY PRESS, Washington, D.C. 2001.

Sekiguchi S, Suda M, Zhai YL, Honma T. 2002. Effects of 1-bromopropane, 2-bromopropane, and 1,2-dichloropropane on the estrous cycle and ovulation in F344 rats. Toxicol. Lett. 126(1):41-9.

Yamada T, Ichihara G, Wang H, et al. 2003. Exposure to 1-bromopropane causes ovarian dysfunction in rats. Toxicol. Sci. 71:96-103.

WIL. 2001. An Inhalation Two-Generation Reproductive Toxicity Study of 1-Bromopropane in Rats. Conducted by Stump, D.G., at WIL Research Laboratories, Inc., Sponsored by Brominated Solvents Consortium. May 24, 2001.