DRAFT - DO NOT CITE OR QUOTE **EPA/635/R-10/002A**
www.epa.gov/iris

TOXICOLOGICAL REVIEW OF FORMALDEHYDE - INHALATION ASSESSMENT

(CAS No. 50-00-0)

In Support of Summary Information on the Integrated Risk Information System (IRIS)

VOLUME IV of IV

Appendices

June 2, 2010

NOTICE

This document is an ***External Review draft***. This information is distributed solely for the purpose of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by EPA. It does not represent and should not be construed to represent any Agency determination or policy. It is being circulated for review of its technical accuracy and science policy implications.

U.S. Environmental Protection Agency
Washington, DC

DISCLAIMER

This document is a preliminary draft for review purposes only. This information is distributed solely for the purpose of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by EPA. It does not represent and should not be construed to represent any Agency determination or policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

CONTENTS IN BRIEF

CONTENTS—TOXICOLOGICAL REVIEW OF FORMALDEHYDE
(CAS No. 50-00-0)

VOLUME IV

CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

Appendix A

1

APPENDIX A

SUMMARY OF EXTERNAL PEER REVIEW
AND PUBLIC COMMENTS AND DISPOSITIONS

[NOTE: This is a placeholder for Appendix A which will be drafted following External Peer review and receipt of public comments.]

Appendix B

1

APPENDIX B

SIMULATIONS OF INTERINDIVIDUAL AND ADULT-TO-CHILD VARIABILITY IN REACTIVE GAS UPTAKE IN A SMALL SAMPLE OF PEOPLE (GARCIA ET AL., 2009)

Garcia et al. (2009) used computational fluid dynamics to study human variability in the nasal dosimetry of model reactive, water-soluble gases in 5 adults and 2 children, aged 7 and 8 years old. They considered two model categories of gases, corresponding to maximal and moderate absorption at the nasal lining. This Appendix was developed in response to EPA reviewers' suggestions that results from the Garcia et al. (2009) work should be used to inform the uncertainty factor considered for interhuman variability in this document. Furthermore the tumor incidence in F344 rats have been used to extrapolate the risk of cancer in the human respiratory tract. This extrapolation was based on internal dose metrics derived using a CFD model constructed from the nasal passages of a single individual (Subramaniam et al. 1998). The adults considered in the Garcia et al. study included that individual.

Garcia et al. (2009) mapped out the nasal airway (including the nasopharynx) geometries of these individuals using magnetic resonance imaging or computed tomography scans. The scans chosen for the analysis were from individuals who had normal nasal anatomies with no pathology (as per a review carried out by a ear-nose-throat surgeon). The minute volumes of these individuals were ranged from 6.8 to 9.0 L/min (adults) and 5.5 to 5.8 L/min (children). The sample size in this study is too small to consider the results representative of the population as a whole (as also recognized by the authors). Nonetheless, various comparisons with the characteristics of other study populations add to the strength of this study; we therefore evaluated this study further in this document partly with the goal of impacting on research directions and future interpretations for specific gases. The range of adult minute volumes in this study is reported by the authors to be in good agreement with that obtained in many other studies in the literature. Minute volumes for the children in the study were found to be similar to the average minute volume of 6.1 ± 1.7 L/min obtained by Bennett and Zeman (2004) in a study of 36 children aged 6 to 13 years; the range of nasal surface area values for the adults agreed well with that obtained by Guilmette et al. (1997) for 45 adults; and the range of values for the surface area to volume ratio is in good agreement with that obtained for 40 adult Caucasians studied by Yokley (2006). The surface area to volume ratio is useful for comparing the rate of diffusional transport of a gas out of different cavities; however in the case of the highly nonhomogeneously shaped nasal lumen, this should only be considered a gross indicator.

We focus here only on the "maximum uptake" simulations in Garcia et al. (2009). In this case, the model gas was considered so highly reactive and soluble that it was reasonable to assume an infinitely fast reaction of the absorbed gas with compounds in the airway lining. Although such a gas could be reasonably considered a proxy for formaldehyde, these results cannot be utilized to inform quantitative estimates of formaldehyde dosimetry (and that does not appear to have been the intent of the authors either). This is because the same boundary condition corresponding to maximal uptake was applied on the vestibular section as well as on the respiratory and transitional epithelial lining of the nasal cavity. This is not appropriate for formaldehyde as the lining on the nasal vestibule is made of keratinized epithelium which is considerably less absorbing than the transitional or respiratory epithelium (Kimbell et al., 2001a).

Table B-1 provides results obtained by Garcia et al. (2009) for gas nasal uptake in five adults and two children for the maximal uptake scenario. Although the nasal cavities of the children were smaller in surface area, volume and length, the surface-area-to-volume ratios were similar in the two age groups. Overall uptake efficiency, average flux (rate of gas absorbed per unit surface area of the nasal lining) and maximum flux levels over the entire nasal lining did not vary substantially between adults (1.6-fold difference in average flux and much less in maximum flux), and the mean values of these quantities were comparable between adults and children. The comparisons between adults and children are in agreement with conclusions reached by Ginsberg et al. (2005) that overall extrathoracic absorption of highly and moderately reactive and soluble gases (corresponding to category 1 and 2 reactive gases as per the scheme in EPA [1994]) is similar in adults and children. However, the interindividual variations in each of these three quantities alone are limited in their ability to characterize variability in the interaction of the gas with the nasal lining. For a very reactive and soluble gas, regional absorption of the gas is highly nonhomogeneously distributed over the nasal lining; interindividual variations due to differing spatial patterns of this distribution between individuals could potentially be diluted when flux is averaged over the whole nose. Estimates of maximum gas flux, on the other hand, correspond to extremely small localized regions of hot spots (see Chapter 3), and thus interindividual differences in this quantity provide limited perspective on interindividual variability in flux distribution patterns over the whole nose. Furthermore, numerical error in the calculation (such as mass balance and irregularly shaped elements of the finite-element mesh) is likely to be most pronounced when estimates are considered over extremely small regions. We do not know to what extent these errors impact upon the accuracy in calculations of maximum flux.

Table B-1. Variations in overall nasal uptake, whole nose flux, and key parameters

	% nasal uptake	MV (L)	SA/V (1/mm)	Avg flux 10^{-8} kg/(s.m^2)		Maximum flux 10^{-8} kg/(s.m^2)	
				left cavity	right cavity	left cavity	right cavity
adult1	93.5	9	1.12	1.8	1.5	10.8	10.0
adult1	92.4	6.8	1.09	1.5	1.5	10.8	10.4
adult1	93.1	9	0.88	1.6	1.3	11	10.6
adult1	89.2	7.1	0.87	1.2	1.2	10.6	10.2
adult1	91.5	6.9	0.95	1.4	1.5	10.8	10.0
child1	92	5.5	1.13	1.9	1.5	11.8	11.0
child2	88.2	5.8	0.95	1.6	1.5	12.3	11.6

MV = minute volume, SA = nasal surface area, V = nasal volume.
Source: Garcia et al. (2009).

On the other hand, Figure 6A of Garcia et al. (2009), reproduced here as Figure B-1, provides a different perspective on interhuman variability in flux values at specific points on the nasal walls. In this figure gas flux across the nasal lining is plotted as function of distance from the nostril along the septal axis of the nose, normalized by the total nasal length along the septal axis of each subject. The local flux of formaldehyde varies among individuals by a factor of 3 to 5 at various normalized distances along the septal axis of the nose. However, interpretation of the values in this plot is problematic for reasons explained in their paper:[1]

> The greater variability among individuals seen for wall fluxes at specific sites of the nasal passages (Figure 6) in comparison to the minimal variability in total uptake (Table 2) and whole-nose dose (Tables 3 and Tables 4) indicates that fluxes of equal magnitude do not exactly overlay the same anatomical regions of the nasal cavity in each individual. This implies that specific anatomical regions subtended by maximum flux could be offset from one individual to another.

Notwithstanding this difficulty in interpretation, we believe the extents of vertical bars on each point plotted in Figure B-1 provide a better perspective of the interindividual (adult) variability in local flux than the variation in whole nose average or in maximum flux presented in Table B-1.

[1] The figures and tables in the cited text refer those in Garcia et al. (2009).

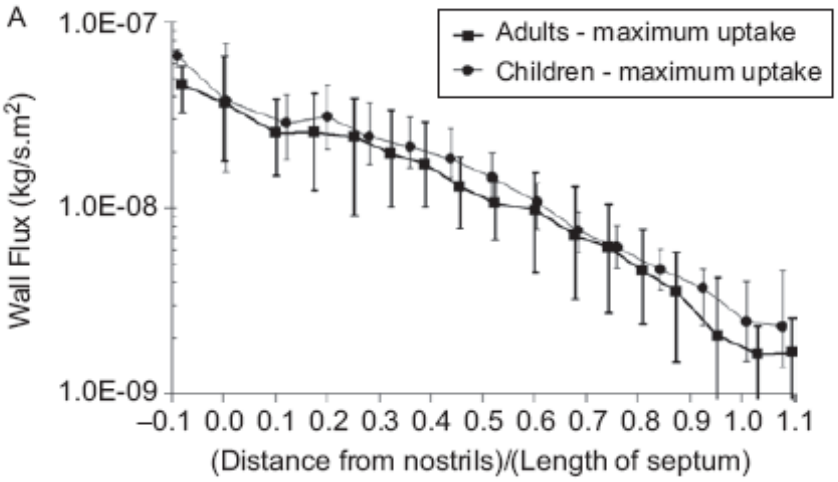

Figure B-1. Gas flux across the nasal lining for the case of a "maximum uptake" gas in Garcia et al. (2009) as a function of axial distance from the nostril. The vertical bars show range of variation. See the paper for further details. Figure is reproduced from Garcia et al. (2009).

Clearly, multiple measures of variability in dose can be developed depending on the adverse response. The advantage of models such as that developed by Garcia et al. is that they make it possible to explicitly carry out these calculations. For example, if deficit in pulmonary function is the adverse response, and the mechanism of action was a function of total dose to the lung, then interindividual variation in mean whole nose flux or overall nasal uptake efficiency would be most useful. It is possible to conceive of allergic or irritation responses being triggered by some threshold value of local flux. In such a case it may be preferable to calculate the variability associated with the net surface area receiving flux values greater than that threshold. On the other hand, the probability of developing a tumor at a nasal site may be nonlinearly related to the flux at that site and linearly related to the number of cells at that site. In this case, the appropriate metric may be the nasal surface area associated with some intermediate levels of local flux (see appendix in Subramaniam et al., 2008).

Various caveats presented by the authors as limitations of their study should be noted: Possible nonuniform distribution of epithelial types, enzymes, glands and other cellular metabolic or clearance machinery were not considered in the model; only effects pertaining to resting breathing were considered; the study sample size was small; children younger than 7 years old were not studied; and, the model assumed a rigid nasal geometry.

Garcia et al. (2009) conclude their paper as follows:

> "…, our simulations predicted no differences in the nasal dosimetry of reactive, water-soluble gases between children and adults, suggesting that the risk factor of 10 typically used to accommodate interhuman variability is adequate."

In addition to the caveats already recognized by the authors, the above conclusion needs further qualification:

1. While the uncertainty factor of 10 that is typically applied for interhuman variability is generally considered to be protective of children, it is not based on variations between children and adults. (If there is reasonable evidence that children are more sensitive than adults, the 10-fold factor may be considered inadequate.)
2. Assuming that the adverse response under consideration is one for which the localized nature of reactive gas flux across the nasal lining is important, the calculations such as those shown in Figure B-1 for the model gas are very relevant to the discussion of interindividual variability. The 3 to 5-fold variation in the local gas flux between adults (and also between the children) in the small sample size in this simulation may be compared with the value of 3.3 used for the pharmacokinetic component of the uncertainty factor for interhuman variability in susceptibility. (EPA practice is often to split this 10-fold uncertainty factor into pharmacokinetic and pharmacodynamic components of 3.3 each.)

Appendix C

1

APPENDIX C

LIFETABLE ANALYSIS

A spreadsheet illustrating the extra risk calculation for the derivation of the lower 95% bound on the effective concentration associated with a 0.05% extra risk (LEC_{0005}) for nasopharyngeal carcinoma (NPC) incidence is presented in Table C-1.

Table C-1. Extra risk calculation[a] for environmental exposure to 0.0461 ppm formaldehyde (the LEC0005 for NPC incidence)[b] using a log-linear exposure-response model based on the cumulative exposure trend results of Hauptmann et al. (2004), as described in Section 5.2.2

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Interval number (i)	**Age interval**	**All cause mortality ($\times 10^5$/yr)**	**NPC incidence ($\times 10^5$/yr)**	**All cause hazard rate (h*)**	**Prob of surviving interval (q)**	**Prob of surviving up to interval (S)**	**NPC cancer hazard rate (h)**	**Cond prob of NPC incidence in interval (Ro)**	**Exp duration mid interval (xtime)**	**Cum exp mid interval (xdose)**	**Exposed NPC hazard rate (hx)**	**Exposed all cause hazard rate (h*x)**	**Exposed prob of surviving interval (qx)**	**Exposed prob of surviving up to interval (Sx)**	**Exposed cond prob of NPC in interval (Rx)**
1	<1	728.7	0	0.0073	0.9927	1.0000	0.00000	0.000000	0	0.0000	0.0000	0.0073	0.9927	1.0000	0.00000
2	1−4	32.9	0.05	0.0013	0.9987	0.9927	0.00000	0.000002	0	0.0000	0.0000	0.0013	0.9987	0.9927	0.00000
3	5−9	16.4	0.03	0.0008	0.9992	0.9914	0.00000	0.000001	0	0.0000	0.0000	0.0008	0.9992	0.9914	0.00000
4	10−14	20.9	0.09	0.0010	0.9990	0.9906	0.00000	0.000004	0	0.0000	0.0000	0.0010	0.9990	0.9906	0.00000
5	15−19	68.2	0.12	0.0034	0.9966	0.9896	0.00001	0.000006	2.5	0.3506	0.0000	0.0034	0.9966	0.9896	0.00001
6	20−24	96	0.16	0.0048	0.9952	0.9862	0.00001	0.000008	7.5	1.0517	0.0000	0.0048	0.9952	0.9862	0.00001
7	25−29	99	0.23	0.0050	0.9951	0.9815	0.00001	0.000011	12.5	1.7528	0.0000	0.0050	0.9951	0.9815	0.00001
8	30−34	116.3	0.48	0.0058	0.9942	0.9766	0.00002	0.000023	17.5	2.4539	0.0000	0.0058	0.9942	0.9766	0.00003
9	35−39	162.2	0.55	0.0081	0.9919	0.9710	0.00003	0.000027	22.5	3.1550	0.0000	0.0081	0.9919	0.9710	0.00003
10	40−44	237.3	1.14	0.0119	0.9882	0.9631	0.00006	0.000055	27.5	3.8561	0.0001	0.0119	0.9882	0.9631	0.00008
11	45−49	356	1.3	0.0178	0.9824	0.9518	0.00007	0.000061	32.5	4.5572	0.0001	0.0178	0.9823	0.9517	0.00009
12	50−54	518.6	1.72	0.0259	0.9744	0.9350	0.00009	0.000079	37.5	5.2583	0.0001	0.0260	0.9744	0.9349	0.00012
13	55−59	801.8	1.69	0.0401	0.9607	0.9111	0.00008	0.000075	42.5	5.9594	0.0001	0.0401	0.9607	0.9110	0.00012
14	60−64	1257.9	1.9	0.0629	0.9390	0.8753	0.00010	0.000081	47.5	6.6605	0.0002	0.0630	0.9390	0.8751	0.00014
15	65−69	1928.2	2.87	0.0964	0.9081	0.8219	0.00014	0.000112	52.5	7.3616	0.0003	0.0965	0.9080	0.8217	0.00021
16	70−74	2968.1	2.1	0.1484	0.8621	0.7464	0.00011	0.000073	57.5	8.0627	0.0002	0.1485	0.8620	0.7461	0.00014
17	75−59	4556.6	2.19	0.2278	0.7963	0.6434	0.00011	0.000063	62.5	8.7638	0.0002	0.2279	0.7962	0.6431	0.00013
18	80−84	7399.6	1.98	0.3700	0.6907	0.5123	0.00010	0.000042	67.5	9.4649	0.0002	0.3701	0.6907	0.5120	0.00009
							Ro =	0.000725						**Rx =**	0.001225
Extra Risk = (Rx−Ro)/(1−Ro) = 0.0005															

Table C-1. Extra risk calculation[a] for environmental exposure to 0.0461 ppm formaldehyde (the LEC0005 for NPC incidence)[b] using a log-linear exposure-response model based on the cumulative exposure trend results of Hauptmann et al. (2004), as described in Section 5.2.2 (continued)

Column B: 5-year age interval (except <1 and 1−4) up to age 85.
Column C: all-cause mortality rate for interval i (× 10^5/year) (2000 data from NCHS).
Column D: NPC incidence rate for interval i (× 10^5/year) (1996−2000 SEER data).
Column E: all-cause hazard rate for interval i (h^*_i) (= all-cause mortality rate × number of years in age interval).[c]
Column F: probability of surviving interval i without being diagnosed with NPC (q_i) (= $\exp(-h^*_i)$).
Column G: probability of surviving up to interval i without having been diagnosed with NPC (S_i) ($S_1 = 1$; $S_i = S_{i-1} \times q_{i-1}$, for $i > 1$).
Column H: NPC incidence hazard rate for interval i (h_i) (= NPC incidence rate × number of years in interval).
Column I: conditional probability of being diagnosed with NPC in interval i (= $(h_i/h^*_i) \times S_i \times (1-q_i)$), i.e., conditional upon surviving up to interval i without having been diagnosed with NPC [Ro, the background lifetime probability of being diagnosed with NPC = the sum of the conditional probabilities across the intervals].
Column J: exposure duration (in years) at mid-interval (xtime).
Column K: cumulative exposure mid-interval (xdose) (= exposure level (i.e., 0.0461 ppm) × 365/240 × 20/10 × xtime) [365/240 × 20/10 converts continuous environmental exposures to corresponding occupational exposures].
Column L: NPC incidence hazard rate in exposed people for interval i (hx_i) (= $h_i \times (1 + \beta \times \text{xdose})$, where $\beta = 0.05183 + (1.645 \times 0.01915) = 0.08333$) [0.05183 per ppm × year is the regression coefficient obtained, along with its SE of 0.01915, from Dr. Hauptmann (see Section 5.2.2.1). To estimate the LEC_{0005}, i.e., the 95% lower bound on the continuous exposure giving an extra risk of 0.05%, the 95% upper bound on the regression coefficient is used, i.e., MLE + 1.645 × SE].
Column M: all-cause hazard rate in exposed people for interval i (h^*x_i) (= $h^*_i + (hx_i - h_i)$).
Column N: probability of surviving interval i without being diagnosed with NPC for exposed people (qx_i) (= $\exp(-h^*x_i)$).
Column O: probability of surviving up to interval i without having been diagnosed with NPC for exposed people (Sx_i) ($Sx_1 = 1$; $Sx_i = Sx_{i-1} \times qx_{i-1}$, for $i > 1$).
Column P: conditional probability of being diagnosed with NPC in interval i for exposed people (= $(hx_i/h^*x_i) \times Sx_i \times (1-qx_i)$) [Rx, the lifetime probability of being diagnosed with NPC for exposed people = the sum of the conditional probabilities across the intervals].

[a] Using the methodology of BEIR IV (1988).
[b] The estimated 95% lower bound on the continuous exposure level of TCE that gives a 0.05% extra lifetime risk of NPC.
[c] For the cancer incidence calculation, the all-cause hazard rate for interval i should technically be the rate of either dying of any cause or being diagnosed with the specific cancer during the interval, i.e., (the all-cause mortality rate for the interval + the cancer-specific incidence rate for the interval—the cancer-specific mortality rate for the interval [so that a cancer case isn't counted twice, i.e., upon diagnosis and upon death]) × number of years in interval. This adjustment was ignored here because the NPC incidence rates are small compared with the all-cause mortality rates.

MLE = maximum likelihood estimate, SE = standard error

Appendix D

1

APPENDIX D

MODEL STRUCTURE & CALIBRATION IN CONOLLY ET AL. (2003, 2004)

The various studies indicated in Section 5.4.1 were followed by the development of a biologically motivated dose-response model for formaldehyde-induced cancer in the respiratory tract. These efforts are represented in a series of papers and in a health assessment report (CIIT model) (Conolly et al., 2004, 2003, 2000; Conolly, 2002; Kimbell et al., 2001a, b; Overton et al., 2001; CIIT, 1999). The CIIT modeling and available data, and alternatives based on their original model were evaluated extensively for the purpose of this assessment and utilized in calculating the cancer potency. EPA's cancer guidelines (U.S. EPA, 2005a) suggest using a BBDR model for extrapolation when data permits since it facilitates the incorporation of MOA in risk assessment

In Conolly et al. (2003), tumor incidence data in the above long-term bioassays were modeled by using an approximation of the two-stage clonal growth model (Moolgavkar et al., 1988) and allowing formaldehyde to have directly mutagenic action. Conolly et al. (2003) combined these data with historical control data on 7,684 animals obtained from National Toxicology Program (NTP) bioassays. These models are based on the Moolgavkar, Venzon, and Knudson (MVK) stochastic two-stage model of cancer (Moolgavkar et al., 1988; Moolgavkar and Knudson, 1981; Moolgavkar and Venzon, 1979), which accounts for growth of a pool of normal cells, mutation of normal cells to initiated cells, clonal expansion and death of initiated cells, and mutation of initiated cells to fully malignant cells.

The MVK model for formaldehyde accounted for two MOAs that may be relevant to formaldehyde carcinogenicity:

- An indirect MOA in which the regenerative cell proliferation in response to formaldehyde cytotoxicity increased the probability of errors in DNA replication. This MOA was modeled by using labeling data on normal cells in nasal mucosa of rats exposed to formaldehyde.
- A possible direct mutagenic MOA, based on information indicating that formaldehyde is mutagenic (Speit and Merk, 2002; Heck et al., 1990; Grafstrom et al., 1985), was modeled by using rat and rhesus monkey data on formaldehyde production of DPXs.

The human model for formaldehyde carcinogenicity (Conolly et al., 2004) is conceptually very similar to the rat model. The model uses, as input, results from a dosimetry model for an anatomically realistic representation of the human upper airways and an idealized

representation of the lower airways. However, the model does not incorporate any data on human responses to formaldehyde exposure. The rat and human formaldehyde models are detailed further below.

The following notations are used in the rest of this appendix:

N cell, normal cell

I cell, initiated cell

LI, labeling index (number of labeled cells/(number labeled + unlabeled cells)

ULLI, unit length labeling index (number labeled cells/length of basement membrane)

N, number of normal cells that are eligible for progression to malignancy

α_N, division rate of normal cells ($hours^{-1}$)

μ_N, rate at which an initiated cell is formed by mutation of a normal cell (per cell division of normal cells)

$\alpha_{I,}$ division rate of an initiated cell ($hours^{-1}$)

$\beta_{I,}$ death rate of an initiated cell ($hours^{-1}$)

$\mu_{I,}$ rate at which a malignant cell is formed by mutation of an initiated cell (per cell division of initiated cells)

A novel contribution of the CIIT model is that cell replication rates and DPX concentrations are driven by local dose, which is formaldehyde flux to each region of nasal tissue expressed as $pmol/mm^2$-hour. This dosimetry is predicted by computational fluid dynamics (CFD) modeling using anatomically accurate representations of the nasal passages (see Chapter 3). Such a feature is important to incorporating site-specific toxicity in the case of a highly reactive gas like formaldehyde, for which uptake patterns are spatially localized and significantly different across species (see Chapter 3). In the CIIT model, each of these parameters is characterized by local flux. The inputs to the two-stage cancer modeling consisted of results from other model predictions as well as empirical data as follows:

- Regional uptake of formaldehyde in the respiratory tract predicted by using CFD modeling in the F344 rat and human (Kimbell et al., 2001a, b; Overton et al., 2001; Subramaniam et al., 1998)
- Concentrations of DPXs predicted by a PBPK model (Conolly et al., 2000) calibrated to fit the DPX data in F344 rat and rhesus monkey (Casanova et al., 1994, 1991) and subsequently scaled up to humans

- α_N inferred from LI data on rats exposed to formaldehyde (Monticello et al., 1996, 1991, 1990)

D.1. DPX AND MUTATIONAL ACTION

Formaldehyde interacts with DNA to form DPXs. These cross-links are considered to induce mutagenic as well as clastogenic effects. Casanova et al. (1994, 1989) carried out two studies of DPX measurements in F344 rats. In the first study, rats were exposed to concentrations of 0.3, 0.7, 2, 6, and 10 ppm for 6 hours and DPX measurements were made over the whole respiratory mucosa of the rat, while, in the second study, the exposure was to 0.7, 2, 6, or 15 ppm formaldehyde for 3 hours and measurements were made at "high" and "low" tumor sites. Overall, these studies showed statistically significantly elevated levels of DPXs at concentrations ≥2 ppm, with the trend also indicating elevated DPXs at 0.7 ppm. In Conolly et al. (2003), DPX formation is considered proportional to the intracellular dose that induces mutations. Conolly et al. (2000) used data from the second study to develop a PBPK model that predicted the time course of DPX concentrations as a function of regional formaldehyde flux (estimated in the CFD modeling and expressed as pmol/mm^2-hour). In Conolly et al. (2003), this PBPK model was then used to predict regional DPX concentrations (that is, as a function of regional formaldehyde flux) (see Figure 5-11, Chapter 5). These data were incorporated into the two-stage clonal expansion model by defining the mutation rate of normal and initiated cells as the same linear function of DPX concentration as follows:

$$\mu_N = \mu_I = \mu_{Nbasal} + KMU \times DPX \qquad \text{(D-1)}$$

The unknown constants μ_{Nbasal} and KMU were estimated by fitting model predictions to the tumor bioassay data.

D.2. CALIBRATION OF MODEL

The rat model in Conolly et al. (2003) involved six unknown statistical parameters that were estimated by fitting the model to the rat formaldehyde bioassay data shown in Table 5-24 in Chapter 5 (Monticello et al., 1996; Kerns et al., 1983) plus historical data from several thousand control animals from **all** the rat bioassays conducted by the NTP. These NTP bioassays were conducted from 1976 through 1999 and included 7,684 animals with an incidence of 13 SCCs (i.e., 0.17% incidence). The resulting model predicts the probability of a nasal SCC in the F344 rat as a function of age and exposure to formaldehyde. The fit of the Conolly et al. (2003) model to the tumor incidence data is shown in Figure 5-12, Chapter 5.

D.3. FLUX BINS

The spatial distribution of formaldehyde over the nasal lining was characterized by partitioning the nasal surface by formaldehyde flux to the tissue (rate of gas absorbed per unit surface area of the nasal lining), resulting in 20 "flux bins" (see Figure 5-13, Chapter 5). Each bin is comprised of elements (not necessarily contiguous) of the nasal surface that receive a particular interval of formaldehyde flux per ppm of exposure concentration (Kimbell et al., 2001a). The spatial coordinates of elements comprising a particular flux bin are fixed for all exposure concentrations, with formaldehyde flux in a bin scaling linearly with exposure concentration (ppm). The number of cells at risk varies across the bins, as shown in Figure 5-14, Chapter 5.

D.4. USE OF LABELING DATA

Cell replication rates in Conolly et al. (2003) were obtained by pooling labeling data from two phases of a labeling study in which male F344 rats were exposed to formaldehyde gas at similar concentrations (0, 0.7, 2.0, 6.0, 10.0, or 15.0 ppm). The first phase employed injection labeling with a 2-hour pulse labeling time, and animals were exposed to formaldehyde for early exposure periods of 1, 4, and 9 days and 6 weeks (Monticello et al., 1991). The second phase used osmotic minipumps for labeling with a 120-hour labeling time to quantify labeling in animals exposed for 13, 26, 52, and 78 weeks (Monticello et al., 1996). The combined pulse and continuous labeling data were expressed as one exposure TWA over all sites for each exposure concentration. α_N was calculated from these labeling data by using an approximation from Moolgavkar and Luebeck (1992). A dose-response curve for normal cell replication rates (i.e., α_N as a function of formaldehyde flux) was then calculated as shown in Figure D-1. These steps are carefully detailed and evaluated in Subramaniam et al. (2008), and discussion of the data will continue in Appendix E in the section on uncertainties in characterizing cell replication rates.

D.5. UPWARD EXTRAPOLATION OF NORMAL CELL DIVISION RATE

The extensive labeling data collected by Monticello et al. (1996, 1991) present an opportunity to use precursor data in assessing cancer risk. However, these empirical data could be used to determine α_N(flux) only for the lower flux range, 0–9,340 pmol/mm^2-hour (see Subramaniam et al. [2008] for the reasons), as shown by the solid line in Figure D-1, whereas the highest computed flux at 15.0 ppm exposure was 39,300 pmol/mm^2-hour. Therefore Conolly et al. (2003) introduced an adjustable parameter, α_{max}, that represented the value of α_N(flux) at the maximum flux of 39,300 pmol/mm^2-hour. α_{max} was estimated by maximizing the likelihood of

the two-stage model fit to the tumor incidence data. For 9,340 < flux ≤ 39,300 pmol/mm^2-hour, α_N(flux) was determined by linear interpolation from α_N(9,340) to α_{max}, as shown by the dashed line in Figure D-1.

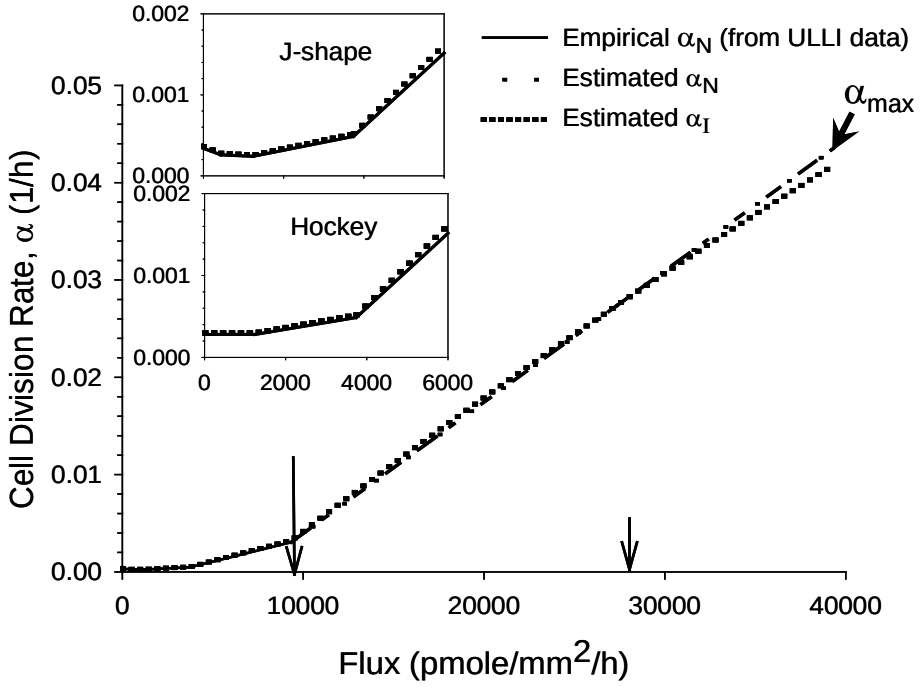

Figure D-1. Dose response of normal (α_N) and initiated (α_I) cell division rate in Conolly et al. (2003).

Note: Empirically derived values of α_N (TWA over six sites) from Table 1 in Conolly et al. (2003) and optimized parameter values from their Table 4 were used. The main panel is for the J-shape dose response. Insets show J-shape and hockey-stick shape representations at the low end of the flux range. The long arrow denotes the upper end of the flux range for which the empirical unit-length labeling data are available for use in the clonal growth model. α_{max} is the value of α_N at the maximum formaldehyde flux delivered at 15 ppm exposure and estimated by optimizing against the tumor incidence data. $\alpha_I < \alpha_N$ for flux greater than the value indicated by the small vertical arrow. Conolly et al. (2004, 2003) assumed $\beta_I = \alpha_N$ at all flux values.

Source: Subramaniam et al. (2008).

D.6. INITIATED CELL DIVISION AND DEATH RATES

The pool of cells used for obtaining the LI data in Monticello et al. (1996, 1991) consists of largely normal cells with perhaps increasing numbers of initiated cells at higher exposure concentrations. Since the division rates of initiated cells in the nasal epithelium, either background or formaldehyde exposed, could not be inferred from the available empirical data, Conolly et al. (2003) made what they perceived to be a biologically reasonable assumption for α_I, assuming α_I to be linked to α_N via a two-parameter function:

$$\alpha_I = \alpha_N \times \{\text{multb} - \text{multc} \times \max[\alpha_N - \alpha_{N(\text{basal})}, 0]\} \quad \text{(D-2)}$$

where $\alpha_N \equiv \alpha_N(\text{flux})$, $\alpha_{N(\text{basal})}$ is the estimated average cell division rate in unexposed normal cells, and multb and multc are unknown parameters estimated by likelihood optimization against the tumor data.[2] The value of $\alpha_{N(\text{basal})}$ was equal to 3.39×10^{-4} hours^{-1} as determined by Conolly et al. (2003) from the raw averaged unit length labeling index data. The ratio α_I/α_N is plotted against flux in Figure D-2, and $\alpha_I(\text{flux})$ is shown by the dotted line in Figure D-1.

Death rates of Initiated cells (β_I) are assumed to equal the division rates of Normal cells (α_N) for all formaldehyde flux values, that is

$$\beta_I(\text{flux}) = \alpha_N(\text{flux}) \quad \text{(D-3)}$$

Conolly et al. (2003) stated that this formulation for α_I and β_I provided the best fit of the model to the tumor data.

D.7. STRUCTURE OF THE CIIT HUMAN MODEL

Subsequent to the BBDR model for modeling rat cancer, Conolly et al. (2004) developed a corresponding model for humans for the purpose of extrapolating the risk estimated by the rat model to humans. Also, rather than considering only nasal tumors (as in the rat model), the human model was used to predict the risk of all human respiratory tumors. The human model for formaldehyde carcinogenicity (Conolly et al., 2004) is conceptually very similar to the rat model and follows the schematic in Figure 5-11, Chapter 5. The following points need to be noted:

[2] multb and multc were equal to 1.072 and 2.583, respectively (J-shaped α_N), and 1.070 and 2.515, respectively (hockey-stick shaped α_N).

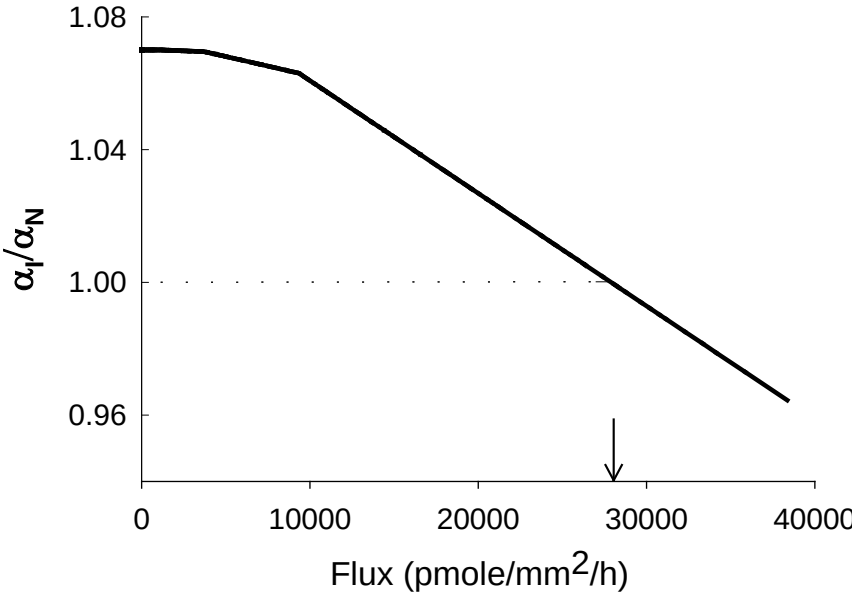

Figure D-2. Flux dependence of ratio of initiated and normal cell replication rates (α_I/α_N) in CIIT model.
Note: Cell replication rate of initiated cells is less than normal cell replication rate at flux exceeding the value denoted by the arrow. By assumption, the Y-axis also represents (α_I/β_I).

Source: Subramaniam et al. (2008).

- The model does not incorporate any data on human responses to formaldehyde exposure.
- The model is based on an anatomically realistic representation of the human nasal passages in a single individual and an idealized representation of the LRT. Local formaldehyde flux to the tissue is estimated by a CFD model for humans (Subramaniam et al., 1998; Kimbell et al., 2001a; Overton et al., 2001).
- Rates of cell division and cell death are, with a minor modification, assumed to be the same in humans as in rats.
- The concentration of formaldehyde-induced DPXs in humans is estimated by scaling up from values obtained from experiments in the F344 rat and rhesus monkey. This scaling up was discussed in chapter 3.
- The statistical parameters for the human model are either estimated by fitting the model to the human background data, assumed to have the same value as obtained in the rat model, or, in one case, fixed at a value suggested by the epidemiologic literature. The delay, D, is fixed at 3.5 years, based on a fit to the incidence of lung cancer in a cohort of British doctors (Doll and Peto, 1978). The two other parameters in the rat model that affect the background rate of cancer (multb and μ_{basal}) are estimated by fitting to U.S. cancer incidence or mortality data. These parameters affect the baseline values for the

human $\alpha_{I,}$ μ_N, and $\mu_{I.}$ Since α_{max}, multfc, and KMU do not affect the background cancer rate, they cannot be estimated from the (baseline) U.S. cancer incidence rates. Therefore, in Conolly et al. (2004, 2003), α_{max} and multfc are assumed to have the same values in humans as in rats, and the human value for KMU is obtained by assuming that the ratio KMU/μ_{basal} is invariant across species. Thus,

$$KMU_{(human)} = KMU_{(rat)} \times \frac{\mu_{Nbasal(human)}}{\mu_{Nbasal(rat)}} \quad \text{(D-4)}$$

This page intentionally left blank.

Appendix E

1

APPENDIX E

EVALUATION OF BBDR MODELING OF NASAL CANCER IN THE F344 RAT: CONOLLY ET AL. (2003) AND ALTERNATIVE IMPLEMENTATIONS

A biologically based dose-response model for formaldehyde-induced cancer was developed in a series of papers and in a health assessment report (CIIT model) (Conolly et al., 2004, 2003, 2000; Conolly, 2002; Kimbell et al., 2001a, b; Overton et al., 2001; CIIT, 1999). The model structure, notations, and calibration have been described in Appendix D. In Chapter 5, an evaluation of the uncertainties of this model and alternative approaches based on its conceptual framework was presented in a summary form. This Appendix provides the relevant details of that evaluation and presents a range of dose-response curves for tumor risk in the rat. It is divided into the following major sections. First, an overview of all the issues that were evaluated is provided in tabular form. The rest of the Appendix then presents only those issues which have a significant impact on model predictions: the use of history controls, the uncertainty and variability in the dose-response for normal cell-replication rates, and sensitivity of model results to uncertainty in the kinetics of initiated cells. These issues have significant impact on inferences regarding mode of action, and this is discussed in some detail in this Appendix. Assumptions and uncertainties related to the human formaldehyde model are discussed in Appendix F.

E.1. TABULATION OF ALL ISSUES EVALUATED IN THE RAT MODELS

Table E-1 summarizes model uncertainties and their impact as evaluated by EPA. The key uncertainties are discussed in considerably more detail in additional sections in this Appendix and in published manuscripts (Klein et al., 2010; Crump et al., 2008; Subramaniam et al., 2008, 2007). The results in Subramaniam et al. (2007) and Crump et al. (2008) have been debated further in the literature (Conolly et al., 2009; Crump et al., 2009). Other alternatives to the CIIT biological modeling (but based on that original model) are also further explored and evaluated in this appendix.

Table E-1. Evaluation of assumptions and uncertainties in the CIIT model for nasal tumors in the F344 rat

Assumptions, approach, and characterization of input data in model	Rationale for assumption/ approach	EPA evaluation	Further elaboration of evaluation[a]
Hoogenveen et al. (1999) solution method, which is valid only for time-independent parameters, is accurate enough.	Errors due to this assumption thought to be significant only at high concentration and not at human exposures.	EPA implemented a solution method valid for time-dependent parameters. Results did not differ significantly from those obtained assuming Hoogenveen et al.(1999) solutions. However, impact was not evaluated for the case where cell replication rates vary in time.	Crump et al. (2005); Subramaniam et al. (2007)
All observed SCC tumors are rapidly fatal; none are incidental tumors.	Death is expected to occur typically within 1–2 weeks of observed tumor (personal communication with R. Conolly).	1) Overall, assumption does not impact model calibration or prediction. 2) However, since 57 animals were observed to have tumors at interim sacrifice times, EPA implementation distinguished between incidental and fatal tumors. Time lag between observable tumor and time of death was significant compared to time lag between first malignant cell and observable tumor.	Subramaniam et al. (2007)
Historical controls from entire NTP database were lumped with concurrent controls in studies.	Large number of control animals (7,684). Intercurrent mortality was not expected to be substantial.	1) Tumor incidence in "all NTP" 10-fold higher than in "all inhalation NTP" controls. Including all NTP controls is considered inappropriate. 2) Low-dose response curve is very sensitive to use of historical controls. 3) Model inference regarding relevance of formaldehyde's mutagenic potential to its carcinogenicity varies from "insignificant" to "highly significant," depending on controls used. (See Appendix F for impact on human risk.)	Table E-2; Subramaniam et al. (2007); Sec E.3.1
LI was derived from experimentally measured ULLI.	Derived from correlating ULLI to LI measured in same experiment.	Significant variation in number of cells per unit length of basement membrane. Spread in ULLI/LI ~25%. Impact on risk not evaluated.	Subramaniam et al. (2008);

Table E-1. Evaluation of assumptions and uncertainties in the CIIT model for nasal tumors in the F344 rat (continued)

Assumptions, approach, and characterization of input data in model	Rationale for assumption/approach	EPA evaluation	Further elaboration of evaluation[a]
Pulse and continuous labeling data were combined in deriving α_N from LI.	All continuous LI values were normalized by mean ratio of pulse to continuous LI for controls.	Formula used for deriving α_N from LI is not applicable for pulse labeling data. Pulse labeling is measure of number of cells in S-phase, not of their recruitment rate into S-phase; not enough information to derive α_N from pulse data. Impact on risk predictions could not be evaluated.	Subramaniam et al. (2008); Section E.3.2.2
To construct dose response for α_N, labeling data were weighted by exposure time (t) and averaged over all nasal sites (TWA). At an exposure concentration, flux was averaged over all nasal sites.	Site-to-site variation in LI was large and did not vary consistently with flux. No reasonable approach was available for extrapolating observed time variation in labeling in rats to humans.	1) TWA assigns low weight to early time LI values, but α_N for early time (t) is very important to the cancer process. Since pulse ULLI was used for $t < 13$ weeks, impact of these ULLIs on risk could not be evaluated. 2) Time dependence in α_N derived from continuous ULLI does not significantly impact model predictions. 3) Site-to-site variation of α_N is at least 10-fold and has major impact on model calibration. Variation in tumor incidence data across sites is 10-fold. 4) Large differences in number of cells across nasal sites (see Table E-3), so averaging over sites is problematic. 5) TWA is also problematic because histologic changes, thickening of epithelium and metaplasia occur at later times for the higher dose and would affect replication rate.	Figures E-1, E-2, E-3**;** Subramaniam et al. (2008); Section E.3.2.3
Steady-state flux estimates are not affected by airway and tissue reconfiguration due to long-term dosing.	Histopathologic changes not likely to be rate-limiting factors in dosimetry.	1) Thickening of epithelium and squamous metaplasia occurring at later times for the higher dose (Kimbell et al. 1997b) will reduce tissue flux. Not incorporated in model. 2) These effects will push regions of higher flux to more posterior regions of respiratory tract. Likely to affect calibration of rat model. Uncertainty not evaluated quantitatively. 3) Calibration of PBPK model for DPXs was seen to be highly sensitive to tissue thickness.	Subramaniam et al. (2008); Cohen-Hubal et al. (1997); Klein et al. (2010).

Table E-1. Evaluation of assumptions and uncertainties in the CIIT model for nasal tumors in the F344 rat (continued)

Assumptions, approach, and characterization of input data in model	Rationale for assumption/approach	EPA evaluation	Further elaboration of evaluation[a]
TWA α_N(flux) rises above baseline levels only at cytolethal dose. Above such dose, α_N(flux) rises sharply due to regenerative proliferation.	Variability in α_N(flux) is partly represented by also considering hockey-stick (threshold in dose) when TWA indicates J-shape (inhibition of cell division) description of α_N(flux).	1) Uncertainty and variability in α_N were quantitatively evaluated to be large. In addition, there are several qualitative uncertainties in characterization of α_N(flux) from LI. 2) Several dose-response shapes, including a monotonic increasing curve without a threshold, were considered in order to adequately describe highly dispersed cell replication data. This has substantial impact on low dose risk.	Figures E-1, E-2, E-3, E-4, E-5; Subramaniam et al. (2008); Section E.3.2
Dose response for α_I was obtained from α_N, assuming ratio (α_I /α_N) to be a two-parameter function of flux (see Figures 5-7, 5-9). Parameters were estimated by optimizing model predictions against tumor incidence data.	(α_I /α_N) was >1.0 in line with the notion of I cells possessing a growth advantage over N cells. Satisfies Occam's razor principle (Conolly et al., 2009).	1) α_I /α_N in CIIT modeling is <1.0 (growth disadvantage) for higher flux values and is >1.0 only at lower end of flux range in model (see Figure 5-9). 2) Since there are no data to inform α_I, sensitivity of risk estimates to various functional forms was evaluated. Risk estimates for the rat were extremely sensitive to alternate biologically plausible assumptions for α_I(flux) and varied by many orders of magnitude at ≤1 ppm, including values lower than baseline risk. All these models described tumor incidence data and cell replication and DPX data equally well.	Figures D-2, E-5, E-6; Subramaniam et al. (2008); Crump et al. (2009, 2008); Section E.3.3
Death rate of I cells is assumed equal to division rate of N cells i.e. β_I(flux) = α_N(flux).	Based on homeostasis (α_N = β_N) and assumption that formaldehyde is equally cytotoxic to N cells and I cells. Satisfies Occam's Razor principle (Conolly et al., 2009).	1) In general, data indicate I cells are more resistant to cytolethality and that ADH3 clearance capacity is greater in transformed cells. Therefore, plausibility of model assumption, that β_I = α_N, is tenuous. 2) Alternate assumption, β_I proportional to α_I, was examined. Risk estimates were extremely sensitive to assumptions on β_I (see Figure 5-12).	Subramaniam et al. (2008); Crump et al. (2009, 2008); Section E.3.3

Table E-1. Evaluation of assumptions and uncertainties in the CIIT model for nasal tumors in the F344 rat (continued)

Assumptions, approach, and characterization of input data in model	Rationale for assumption/approach	EPA evaluation	Further elaboration of evaluation[a]
DPX is dose surrogate for formaldehyde's mutagenic potential. DPX clearance is rapid and complete in 18 hours.	Casanova et al. (1994).	Half-life for DPX clearance in in vitro experiments on transformed cell lines was 7-times longer than estimated by Conolly et al. (2004, 2003) and perhaps 14-times longer with normal (nontransformed) human cells. Some DPX accumulation is therefore likely. However, model calibration and dose response in rat was insensitive to this uncertainty. See section E.3 for effect on scale-up of model to humans.	Quievryn and Zhitkovich, (2000); Subramaniam et al. (2007); Section 3.6.6.3
Formaldehyde's mutagenic action takes place only while DPX's are in place.		DNA lesions may remain after DPX repair and incomplete repair of DPX can lead to mutations (Barker et al. 2005). There is some potential for formaldehyde-induced mutation after DPX clearance. Thus, it is possible that formaldehyde mutagenicity may be underrepresented in model. Could not quantitatively evaluate uncertainty (no data on clearance of secondary lesions).	Subramaniam et al. (2008); Section 4.3.3.3

[a]References stated here are in addition to Conolly et al. (2004, 2003).

Note: Risk estimates discussed in this table are for the F344 rat.

E.2. STATISTICAL METHODS USED IN EVALUATION

Parameters of the alternate models shown here were estimated by maximizing the likelihood function defined by the data (Cox and Hinkley, 1974). Such estimates are referred to as maximum likelihood estimates (MLEs). Statistical confidence bounds were computed by using the profile likelihood method (Crump, 2002; Cox and Oakes, 1984; Cox and Hinkley, 1974). In this approach, an asymptotic $100(1 - \alpha)\%$ upper (lower) statistical confidence bound for a parameter, β, in the animal cancer model is calculated as the largest (smallest) value of β that satisfies

$$2[L_{max} - L^*(\beta)] = x_{1-2\alpha} \quad \text{(E-1)}$$

where L indicates the likelihood of the rat bioassay data, L_{max} is its maximum value, $L^*(\beta)$ is, for a fixed value of β, the maximum value of the log-likelihood with respect to all of the remaining parameters, and $x_{1-2\alpha}$ is the $100(1-2\alpha)$ percentage point of the chi-square distribution with one degree of freedom. The required bound for a parameter, β, was determined via a numerical search for a value of β that satisfies this equation.

The additional risk is defined as the probability of an animal dying from an SCC by the age of 790 days, in the absence of other competing risks of death, while exposed throughout life to a prescribed constant air concentration of formaldehyde, minus the corresponding probability in an animal not exposed to formaldehyde. The MLE of additional risk is the additional risk computed using MLEs of the model parameters.

The method described above for computing profile likelihood confidence bounds cannot be used with additional risk because additional risk is not a parameter in the cancer model. Instead, an asymptotic $100(1 - \alpha)\%$ upper (lower) statistical confidence bound for additional risk was computed by finding the parameter values that presented the largest (smallest) value of additional risk, subject to the inequality

$$2[L_{max} - L] \leq x_{1-2\alpha} \quad \text{(E-2)}$$

being satisfied, with the resulting value of additional risk being the required bound. This procedure was implemented through use of penalty functions (Smith and Coit, 1995). For example, the profile upper bound on additional risk was computed by maximizing the "penalized added risk," defined as (*additional risk – penalty*), where

$$penalty = W \times \{[(L_{max} - L) - x_{1-2\alpha}/2]^+\}^2 \quad \text{(E-3)}$$

and $[]^+$ equals the quantity in the brackets whenever it is positive and zero otherwise. The multiplicative weight, *W*, was selected by trial and error so that the final solution satisfied the following equation sufficiently well.

$$2(L_{max} - L) = x_{1-2\alpha} \quad \text{(E-4)}$$

The computer code was written in Microsoft Excel® 2002 SP3 Visual Basic. Either the regular Excel Solver or the Frontline Systems Premium Solver® was used to make the required function optimizations. Computation of confidence bounds was highly computationally intensive, and, consequently, confidence bounds were computed only for selected parameters in selected runs. For select cases, the bootstrap method was also used to calculate confidence bounds in order to confirm their accuracy. Values so calculated were found to be in agreement with those calculated by using the likelihood method.

E.3. PRIMARY UNCERTAINTIES IN BBDR MODELING OF THE F344 RAT DATA

In their evaluation, Subramaniam et al. (2007) first attempted to reproduce the Conolly et al. (2003) results under similar conditions and assumptions as employed in their paper, which included the assumption that tumors were rapidly fatal. Figure 5-12 in Chapter 5 shows the results for this case. The predicted probabilities shown in this figure were obtained by Subramaniam et al. (2007) by using the source code made available by Dr. Conolly. These are compared with the best-fitting model and plotted against the Kaplan-Meier (KM) probabilities. Although the results are largely similar, there are some residual differences, and these are detailed in Subramaniam et al. (2007).

Given the scope of issues to examine for the uncertainty analyses, the evaluation proceeded in stages. First, the Hoogenveen et al. (1999) solution was replaced by one that is valid for a model with time varying parameters (Crump et al., 2005; first entry in Table E-1), and tumors found at scheduled sacrifices were assumed to be incidental rather than fatal (second entry in Table E-1). Second, weekly averaged solutions for DPX concentration levels were used instead of hourly varying solutions (predicted by a PBPK model). The log-likelihood values and tumor probabilities remained essentially unchanged. Upon quantitative evaluation, these factors, although important from a methodological point of view, were not found to be major determinants of either calibration or prediction of the model for the F344 rat data (Subramaniam et al., 2007).

Following Georgieva et al. (2003), Subramaniam et al. (2007) used the DPX clearance rate constant obtained from in vitro data instead of the assumption in Conolly et al. (2003) that

all DPXs cleared within 18 hours (Subramaniam et al., 2007). With this revision, weekly average DPX concentrations were larger than those in Conolly et al. (2003) by essentially a constant ratio equal to 4.21 (range of 4.12–4.36) when averaged over flux bin and exposure concentrations. Accordingly, cancer model fits to the rat tumor incidence data using the two sets of DPX concentrations (everything else remaining the same) provided very similar parameter estimates, except that the parameter KMU_{rat} in eq D-1 (and eq D-4) (Appendix D) was 4.23 times larger with the Conolly et al. (2003) DPX concentrations. In other words, the product KMU × DPX remained substantially unchanged. However, it is important to note that the different clearance rate does significantly impact the scale-up of the two-stage clonal growth model to the human since the parameter KMU_{human} is not estimated separately but related to KMU_{rat} (see eq D-4).

After making the above modifications, the impact of the other uncertainties in Table E-1 were examined. Of the issues in Table E-1, the following uncertainties had large impacts on the modeling of the F344 rat data, and will be discussed in considerably more detail:

1. use of historical controls,
2. uncertainty and variability in characterizing cell replication rates from the labeling data, and
3. uncertainty in model specification of initiated cell kinetics.

E.3.1. Sensitivity to Use of Historical Controls

E.3.1.1. *Use of Historical Controls*

Conolly et al. (2003) combined the historical controls arising from the entire NTP database of bioassays. Tumor and survival rates in control groups from different NTP studies are known to vary due to genetic drift in animals over time and differences in laboratory procedures, such as diet, housing, and pathological procedures (Haseman, 1995; Rao et al., 1987). In order to minimize extra variability when historical control data are used, the current NTP practice is to limit the historical control data, as far as possible, to studies involving the same route of exposure and to use historical control data from the most recent studies (Peddada et al., 2007).

Bickis and Krewski (1989) analyzed 49 NTP long-term rodent cancer bioassays and found a large difference in determinations of carcinogenicity, depending on the use of historical controls with concurrent control animals. The historical controls used in the CIIT modeling controls came from different rat colonies and from experiments conducted in different laboratories over a wide span of years, so it is clearly problematic to assume that background

rates in these historical control animals are the same as those in the concurrent control group. There are considerable differences among the background tumor rates of SCCs in all NTP controls (13/7,684 = 0.0017), NTP inhalation controls (1/4,551 = 0.0002), and concurrent controls (0/341 = 0.0). The rate in all NTP controls is significantly higher than that in NTP inhalation controls (p = 0.01, Fisher's exact test). Given these differences, the inclusion of any type of historical controls is problematic and is thought to have limited value if these factors are not controlled for (Haseman, 1995).

E.3.1.2. *Influence of Historical Controls on Model Calibration and on Human Model*

To investigate the effect of including historical controls in the CIIT model, the analyses in Subramaniam et al. (2007) were conducted by using the following sets of data for controls (the fraction of animals with SCCs is denoted in parentheses):

a) only concurrent controls (0/341),

b) concurrent controls plus all the NTP historical control data used by Conolly et al. (2003) (13/8,031),

c) concurrent controls plus data from historical controls obtained from NTP inhalation studies (1/4,949) (NTP, 2005).[3]

The results of the evaluation are shown in Table E-2. For these analyses, the same normal cell replication rates and the same relationship (see eq D-2 in Appendix D) between initiated cell and normal cell replication rates as used in Conolly et al. (2003) were used. In all cases, weekly averaged values of DPX concentrations were used. Model fits to the tumor incidence data were similar in all cases to that shown in Figure 5-12 (see Subramaniam et al. [2007] for a more complete discussion). The biggest influence of the control data was seen to be on the estimated basal mutation rate in rats, $\mu_{Nbasal(rat)}$, which, in turn, influences the estimated mutation effect in humans through eq D-4 (Appendix D). α_{max} was also seen to be a sensitive parameter and is discussed later. See Subramaniam et al. (2007) for other parameters in the calibration.

[3] Three animals in the inhalation historical controls were diagnosed with nasal SCC. Of these, two of the tumors were determined to have originated in tissues other than the nasal cavity upon further review (Dr. Kevin Morgan and Ms. Betsy Gross Bermudez, personal communication). These two tumors were therefore not included on the advice of Dr. Morgan. See Subramaniam et al. (2007) for more details.

Table E-2. Influence of control data in modeling formaldehyde-induced cancer in the F344 rat

Case	A	D	B	E	C	F
Control animals (combined with concurrent controls)	**All NTP historical[a]**	**All NTP historical[a]**	**NTP inhalation historical[a]**	**NTP inhalation historical[a]**	**Concurrent only[a]**	**Concurrent only[a]**
Cell replication dose response	**J-shaped**	**Hockey stick**	**J-shaped**	**Hockey stick**	**J-shaped**	**Hockey stick**
Log-likelihood	−1692.65	−1693.68	−1,493.21	−1,493.35	−1,474.29	−1,474.29
μ_{Nbasal}	1.87×10^{-6}	2.12×10^{-6}	7.32×10^{-7}	9.32×10^{-7}	0.0	0.0
KMU	1.12×10^{-7}	0.0	6.84×10^{-7}	6.18×10^{-7}	1.20×10^{-6}	1.20×10^{-6}
KMU/μ_{Nbasal}	0.06 (0.0, 0.40)	0.0 (0.0, 0.25)	0.94 (0.26, 6.20)	0.66 (0.2, 5.20)	∞ (0.42, ∞)	∞ (0.41, ∞)
α_{max}	0.045 (0.029, 0.045)	0.045 (0.029, 0.045)	0.045 (0.026, 0.045)	0.045 (0.027, 0.045)	0.045 (0.027, 0.045)	0.045 (0.027, 0.045)

[a]Values in parentheses denote lower and upper 90% confidence bounds.

Source: Adapted from Subramaniam et al. (2007).

The ratio KMU/μ_{Nbasal} is of particular interest because extrapolation to human in Conolly et al. (2004) assumed its invariance as given by eq D-4 (Appendix D). Now, μ_{Nbasal} in the human is estimated independently by fitting a scaled-up version of the two-stage model to human baseline rates of tumor incidence. Thus, a decrease in the value of μ_{Nbasal} estimated in the rat modeling increases the formaldehyde-induced mutational effect in the human.

The MLE of $KMU_{rat}/\mu_{Nbasal(rat)}$ is zero in Conolly et al. (2003). However, in the various cases examined in Subramaniam et al. (2007) it takes a range of values from 0 to 0.9 mm^3/pmol and undefined (or infinite, when $\mu_{Nbasal} = 0$). The 95% upper confidence bound on this ratio ranges from 0.25–6.2 (these values would be four times larger had the Conolly et al. [2003] DPX concentrations been used) to infinite. Thus, the extrapolation to human risk by using the approach in Conolly et al. (2004) becomes particularly problematic when only concurrent controls are used, because then the mutational contribution to formaldehyde-induced risk in humans becomes unbounded. This issue will be discussed again toward the end of the discussion on historical controls.

It may be noted, however, that absence of tumors in the limited number of concurrent animals does not imply that the calculation will necessarily predict a zero background probability of tumor (i.e., a parameter estimate of $\mu_{Nbasal} = 0$). Subramaniam et al. (2007) observed such a counterexample estimate for μ_{Nbasal} in simulations involving the alternate dose-response curves for α_N and α_I that are discussed in Section E.3.4. Nonetheless, when $\mu_{Nbasal} = 0$, an upper bound for μ_{Nbasal} using the concurrent controls could be inferred. Accordingly, the 90% statistical lower confidence bound on the ratio KMU/μ_{Nbasal} is also reported in Table E-2. Such a value would of course provide a lower bound on risk by using this model and would therefore not be conservative.

Conolly et al. (2003) estimated KMU to be zero for both their hockey-stick and J-shape dose response models for cell replication. However, the estimate for the coefficient KMU (obtained using the solution of Crump et al. [2005]) is zero only for the case of the model with the hockey-stick curve for cell replication and with control data as used by Conolly et al. (2003). It is positive in all other cases and statistically significantly so in all cases in which either NTP inhalation control data or concurrent controls were used. With concurrent controls only and the J-shape cell replication model, the MLE estimate for KMU (1.2×10^{-6}) is larger than the statistical upper bound obtained by Conolly et al. (2003) (8.2×10^{-7}). It should also be kept in mind that the estimate would be about 4.2 times larger still had the Conolly et al. (2003) DPX model been used.

E.3.1.3. *Influence of Historical Controls on Dose-Response Curve*

Subramaniam et al. (2007) showed that inclusion of historical controls had a strong impact on the tumor probability curve below the range of exposures over which tumors were observed in the formaldehyde bioassays. As shown there, the MLE probabilities for occurrence of a fatal tumor at exposure concentrations below 6 ppm were roughly an order of magnitude higher when all the NTP historical controls were used, compared with MLE probabilities predicted when historical controls were drawn only from inhalation bioassays, and many orders of magnitude higher than MLE probabilities predicted when only concurrent controls were used in the analysis. (Note that this comparison should not be inferred to apply to upper bound risk estimates since there were many fewer concurrent than historical controls, so error bounds could be much larger in the case where concurrent controls were used.)

However, as shown by these authors, model fits to the tumor data in the 6–15 ppm exposure concentration range were qualitatively indifferent to which of these control data sets was used. This observation emphasizes the statistical aspect of the CIIT modeling—that significant interplay among the various adjustable parameters allows the model to achieve a good fit to the tumor incidence data independent of the control data used. On the other hand, the results in Subramaniam et al. (2007) show that changes in the control data affect parameter KMU, resulting in significantly different tumor predictions at lower exposure concentrations. Therefore, the strong influence of using all the NTP historical controls on the low-dose region of the time-to-tumor curves presented in Subramaniam et al. (2007) suggests that large uncertainties may arise in extrapolating to both human and rat (in the low-dose region) from such considerations alone.

E.3.1.4. *Problem Including 1976 Study for Inhalation Historical Control*

A crucial point needs to be noted with regard to the use of inhalation NTP historical controls (i.e., cases B and E) in the two-stage clonal growth modeling. The single relevant tumor in the NTP inhalation studies came from the very first NTP inhalation study, dated 1976, and the animals in this study were from Hazelton Laboratories, whereas the concurrent animals were all from Charles River Laboratories. Similar problems arise with inclusion of several other NTP inhalation studies. As mentioned before, genetic and other time-related variation can lead to different tumor and survival rates, and in general it is recommended that use of historical controls be restricted to the same kind of bioassays and to studies within a 5–7 year span of the concurrent animals (Peddada et al., 2007). Thus, it is problematic to assume that the tumor in the 1976 NTP study is representative of the risk of SCCs in the formaldehyde bioassays. Even if it were appropriate to consider the 1976 study, this leads to the unstable situation in which,

despite all of the "upstream" mechanistic information used to construct the BBDR model, the only piece of data that might keep the model predictions of human risk bounded is a single tumor found among several thousand rats from NTP bioassays (Crump et al., 2008). In summary, although it can be argued that the rate of SCCs among the controls in the rat bioassay is probably not zero, it is also problematic to assume that this rate can be adequately represented by the background rate in NTP historical controls or even in NTP inhalation historical controls.

E.3.1.5. *Effect of Control Data on MOA Inferences*

Subramaniam et al. (2007) also examined the contribution of the DPX component (which represents the directly mutagenic potential of formaldehyde in the model) to the calculated tumor probability, choosing for their case study the optimized models that use the NTP inhalation control data. In the range of exposures where tumors were observed (6.0–15.0 ppm), the DPX term was found to be responsible for 58–74% of the added tumor probability. Below 6.0 ppm the estimated DPX contribution was extremely sensitive to whether the hockey-stick shape or J-shape was used to characterize the dose response for cell replication, and varied between 2% and 80%.

The CIIT BBDR cancer modeling has contributed to the weight-of-evidence process in various formaldehyde risk assessment efforts and papers by lending weight to the argument that the direct mutations induced by formaldehyde are relatively irrelevant compared to the importance of cytotoxicity-induced cell proliferation in explaining the observed tumorigenicity in rodent bioassays and in projecting those observations to human exposures (Conolly et al., 2004, 2003; Slikker et al., 2004; Bogdanffy et al., 2001, 1999; Conolly, 1995). The reanalyses in Subramaniam et al. (2007) (in particular, the results in the above paragraph) indicate that, if the CIIT mathematical modeling were utilized to inform this debate, it would in fact indicate the contrary—that a large contribution from formaldehyde's mutagenic potential may be needed to explain formaldehyde carcinogenicity. This discussion is resumed in the context of uncertainties in model specification for initiated cells.

E.3.2. Characterization of Uncertainty-Variability in Cell Replication Rates

E.3.2.1. *Dose-Response for α_N as Used in the CIIT Clonal Growth Modeling*

Monticello et al. (1996, 1991) used unit length labeling index (ULLI) to quantify cell replication within the respiratory epithelium. ULLI is a ratio between a count of labeled cells and the corresponding length (in millimeters) of basal membrane examined, whereas the per-cell labeling index (LI) is the ratio of labeled cells to all epithelial cells, in this case, along some length of basal membrane and its associated layer of epithelial cells. Monticello et al. (1996,

1991) published ULLI values averaged over replicate animals for each combination of exposure concentration, exposure time, and nasal site. These values are plotted in Figure E-1.

In order to utilize the ULLI data in clonal growth modeling, ULLI needed to be related to LI, and thereby to cell replication rate (α_N) of normal cells. Conolly et al. (2003) adopted the following procedure in using these values (Subramaniam et al., 2008):

1. The injection labeled ULLI data were first normalized by the ratio of the average minipump ULLI for controls to the average injection labeled ULLI for controls.
2. Next, these ULLI average values were weighted by the exposure times in Monticello et al. (1996, 1991) and averaged over the nasal sites. Thus, the data were combined into one TWA for each exposure concentration.
3. LI was linearly related to the measured ULLI by using data from a different experiment (Monticello et al., 1990) where both quantities had been measured for two sites in the nose.
4. Cell replication rates of normal cells (α_N) were then calculated as $\alpha_N = (-0.5/t)\log(1 - LI)$ (Moolgavkar and Luebeck, 1992), where LI is the labeling index and t is the period of labeling.
5. This was repeated for each exposure concentration of formaldehyde, resulting in one value of α_N for each exposure concentration.
6. Correspondingly, for a given exposure concentration, the steady-state formaldehyde flux into tissue, computed by CFD modeling, was averaged over all nasal sites. Thus, the α_N(flux) constructed by Conolly et al. (2003) consisted of a single α_N and a single average flux for each of six exposures.

This yielded a J-shaped dose-response curve for cell replication (when viewed on a nontransformed scale for α_N), as shown in Figure D-1 (Appendix D) for the full range of flux values used in their modeling. The authors also considered a hockey-stick threshold representation of their J-shaped curve for α_N in order to make a health-protective choice, and the differences between the two can be seen from the insets in Figure D-1. In these curves, the cell replication rate is less than or the same as the baseline cell replication rate at low formaldehyde flux values. The shape of the dose-response curve for cell replication as characterized in Conolly et al. (2003) is seen as representing regenerative cell proliferation secondary to the cytotoxicity of formaldehyde (Conolly, 2002). Considerable uncertainty and variability, both quantitative and qualitative, exist in the use and interpretation of these labeling data for characterizing a dose response for cell replication rates. The primary issues are discussed here. Unlike the preceding sections, these have largely not been published elsewhere, so more details are provided.

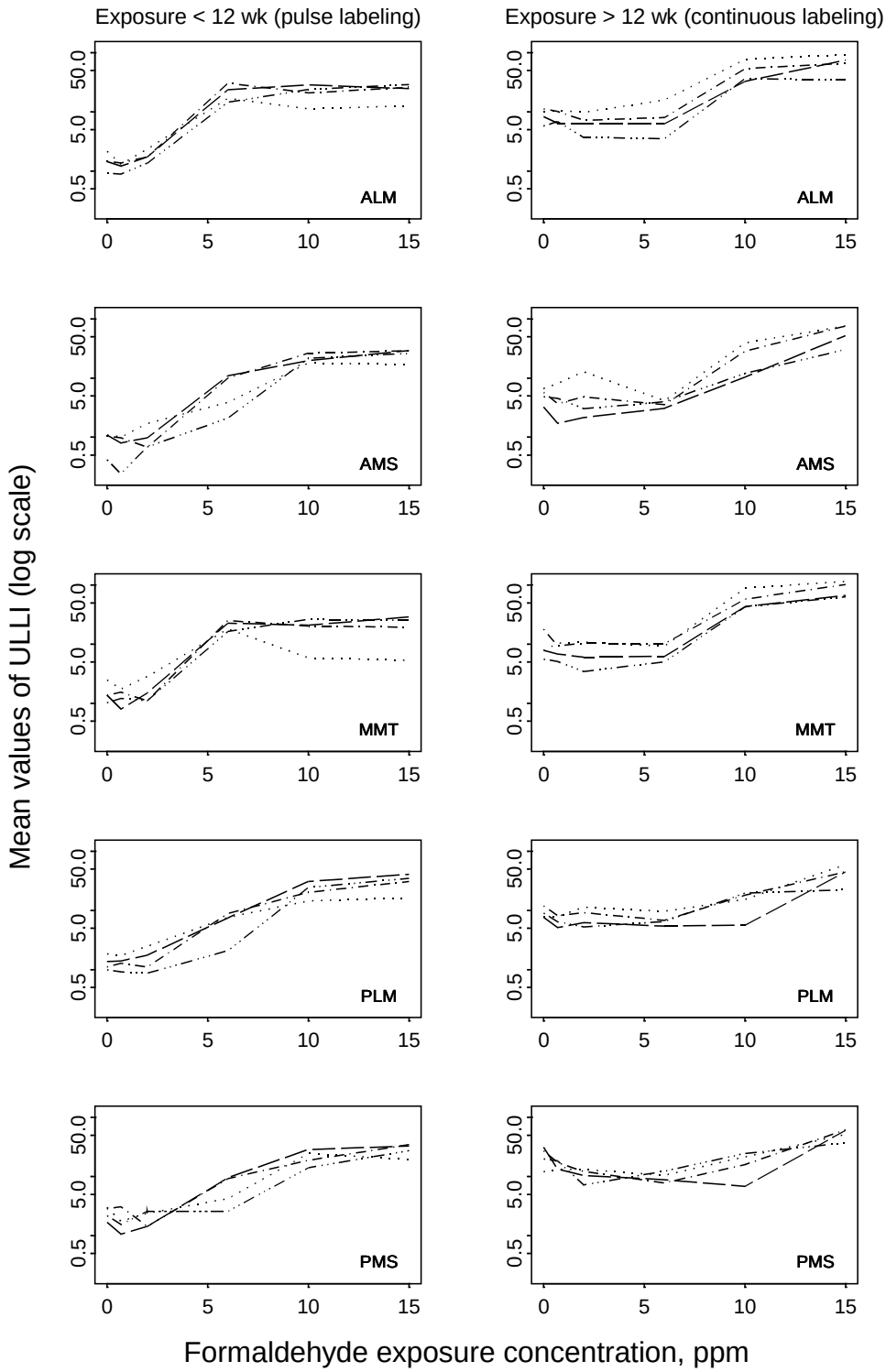

Figure E-1. ULLI data for pulse and continuous labeling studies.

Note: Data are from pulse labeling study, left-hand side, at 1–42 days of exposure and from the continuous-labeling study, right-hand side, at 13–78 weeks of exposure for five nasal sites ALM, AMS, MMT, PLM, and posterior mid septum [PMS]). Within each graph, lines with more breaks correspond to shorter exposure times. Data source: Monticello et al. (1996, 1991).

E.3.2.2. *Time Variability in Labeling Data*

E.3.2.2.1. ***Short-time exposure effects on cell replication.***

Figure E-1 shows the site and time variation in the raw unit-length labeling index (ULLI) data for 1 day to 78 weeks of exposure duration. The temporal variation in ULLI is quite different between the "early time" (left panel) and "later time" (right panel) and these early-time effects may be quite important to the cancer modeling. At the earliest times in the left panel, the data show an increased trend in labeling at 2 ppm for the sites anterior lateral meatus (ALM), anterior medial septum (AMS), posterior lateral meatus (PLM), and medial maxilloturbinate (MMT) relative to control. Such an increase is generally indicated for low flux values also for the 13-week exposure time. This can be seen in the dose-response plotted as a function of flux in Figure E-4.

The early times would be important if, say, repeated episodic exposures were considered, where adequate time has not elapsed for adaptive effects to take place. Such an exposure scenario may be the norm in the human context. In the CIIT cancer modeling, the LI was weighted by exposure time. As a consequence, the contribution of the early-time labeling data is minimized in their modeling.

E.3.2.2.2. ***Uncertainty due to combining pulse and continuous labeled data.***

The formula used for obtaining α_N from LI in Conolly et al. (2003) was due to Moolgavkar and Luebeck (1992) who derived this formula for continuous LI, cautioning that it is not applicable for pulse labeled data. However, Conolly et al. (2003) applied this formula to the injection (pulse) labeled data also. Such an application is problematic because 2-hour pulse labeled data represent the pool of cells in S-phase rather than the rate at which cells are recruited to the pool, and because the baseline values of α_N obtained in this manner from both data sets differ considerably. As such, we are not aware of any reasonable manner to derive cell replication rates from these pulse data without acquisition of data at additional time points. Because of these problems in incorporating the pulse-labeled data, further quantitative analysis of cell replication rates is restricted in this document to the continuous labeled data (Monticello et al., 1996), which do not include measurements made before 13 weeks of exposure. It is unfortunate that the continuous labeled data do not include any early measurements.

E.3.2.3. *Site and Time Variability in Derived Cell Replication Rate*

In the remainder of this section, the factors that are considered in order to represent the uncertainty and variability in the cell replication data when developing alternate dose-response curves for α_N(flux) will be elaborated.

The ULLI data for individual animals were provided by CIIT, which were transformed to LI values using the linear relationship from step 3 in Section E.3.2.1. For these replicate data, cell replication rates of normal cells (α_N) were then calculated as $\alpha_N = (-0.5/t)\log(1 - LI)$ as in Step 4. Figure E-2 (adapted from Subramaniam et al., 2008) shows the variability in α_N due to replicated animals, exposure times, and nasal sites in the continuous labeled data obtained by Monticello et al. (1996). In this figure, log α_N versus site-specific flux are plotted for six sites and four exposure times for four to six replicate animals in each case. (The mean ULLI over these replicates were shown in Figure E-1 for each site and time as a function of exposure concentration.) It needs to be noted that these nasal sites differ considerably in the number of cells estimated at these locations as shown in Table E-3. Each point in Figure E-2 represents data from a single site for a single animal at a given time. For comparison, the α_N(flux) in Conolly et al. (2003) is also plotted in this figure at their averaged flux values (filled circles). For flux >9,340 pmol/mm^2-hour, Conolly et al. (2003) extrapolated this empirically derived α_N(flux) by using a scheme discussed in Appendix D (see Section D.5) on the upward extrapolation of cell replication rate. The curves shown connecting the filled circles in the figure represent their linear interpolation (long dashes) between the six points. Their linear extrapolation for flux value >9,340 pmol/mm^2-hour is also shown (short dashes). Note that the linear interpolation and extrapolation are shown transformed to a logarithmic scale in this plot.

As discussed, the raw labeling data plotted in Figure E-1 indicates considerable temporal variability. In Figures E-3, fitted dose-response curves showing $\log_{10}(\alpha_N)$ versus flux with simultaneous confidence limits separately for each time point for two of the largest sites in Table E-3 (ALM and PLM) are plotted for the continuous labeled data. Note that flux levels are different at each site. Simple polynomial models in flux (as a continuous predictor), with time included as a factor (i.e., a class or indicator variable, τ_i representing the effect of the ith time) were used as follows:

$$\log(\alpha_N) = a + b \times flux + c \times flux^2 + d \times flux^3 + \tau_i \qquad \text{(E-5)}$$

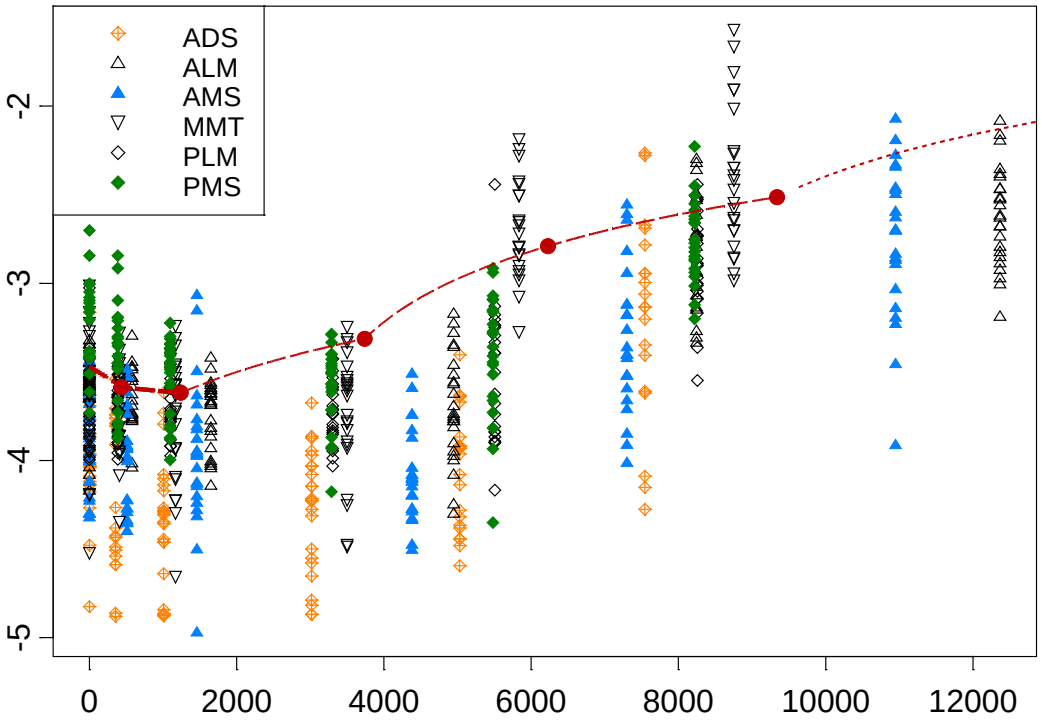

Figure E-2. Logarithm of normal cell replication rate α_N versus formaldehyde flux (in units of pmol/mm^2-hour) for the F344 rat nasal epithelium.

Note: Values were derived from continuous unit length labeled data obtained by Monticello et al. (1996) for four to six individual animals at all six nasal sites (legend, sites as denoted in original paper) and four exposure durations (13, 26, 52, 78 weeks). Each point represents a measurement for one rat, at one nasal site, and at a given exposure time. Filled red circles: α_N(flux) used in Conolly et al. (2003) plotted at their averaged flux values (see text for details). Long dashed lines: their linear interpolation between points. Short dashed line: their linear extrapolation for flux value >9,340 pmol/mm^2-hour (see Figure D-1 for full range of extrapolation). Linear interpolation/extrapolation is shown with Y-axis transformed to logarithmic scale.

Source: Subramaniam et al. (2008).

Table E-3. Variation in number of cells across nasal sites in the F344 rat

Nasal site	No. of cells
Anterior lateral meatus	976,000
Posterior lateral meatus	508,000
Anterior mid septum	184,000
Posterior mid septum	190,000
Anterior dorsal septum	128,000
Anterior medial maxilloturbinate	104,000

Note: Mean number of cells in each side of the nose of control animals.

Source: Monticello et al. (1996).

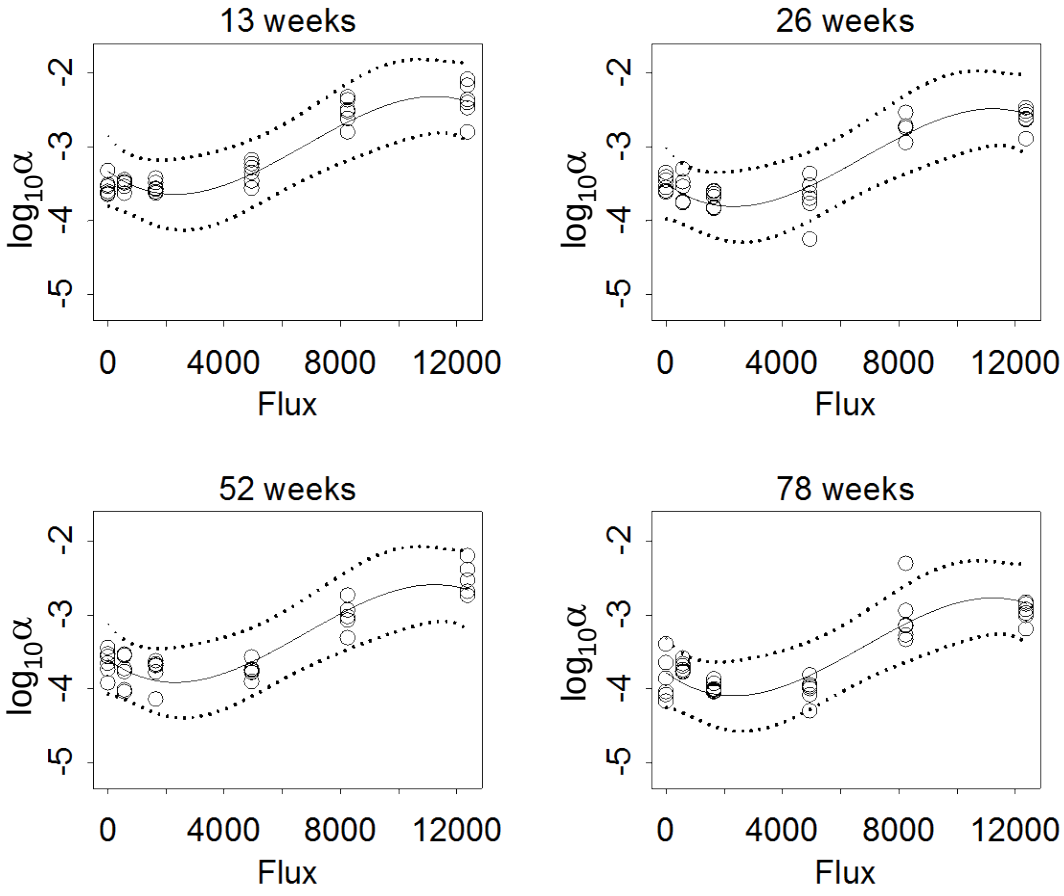

Figure E-3A. Logarithm of normal cell replication rate versus formaldehyde flux with simultaneous confidence limits for the ALM.

Source: Subramaniam et al. (2008).

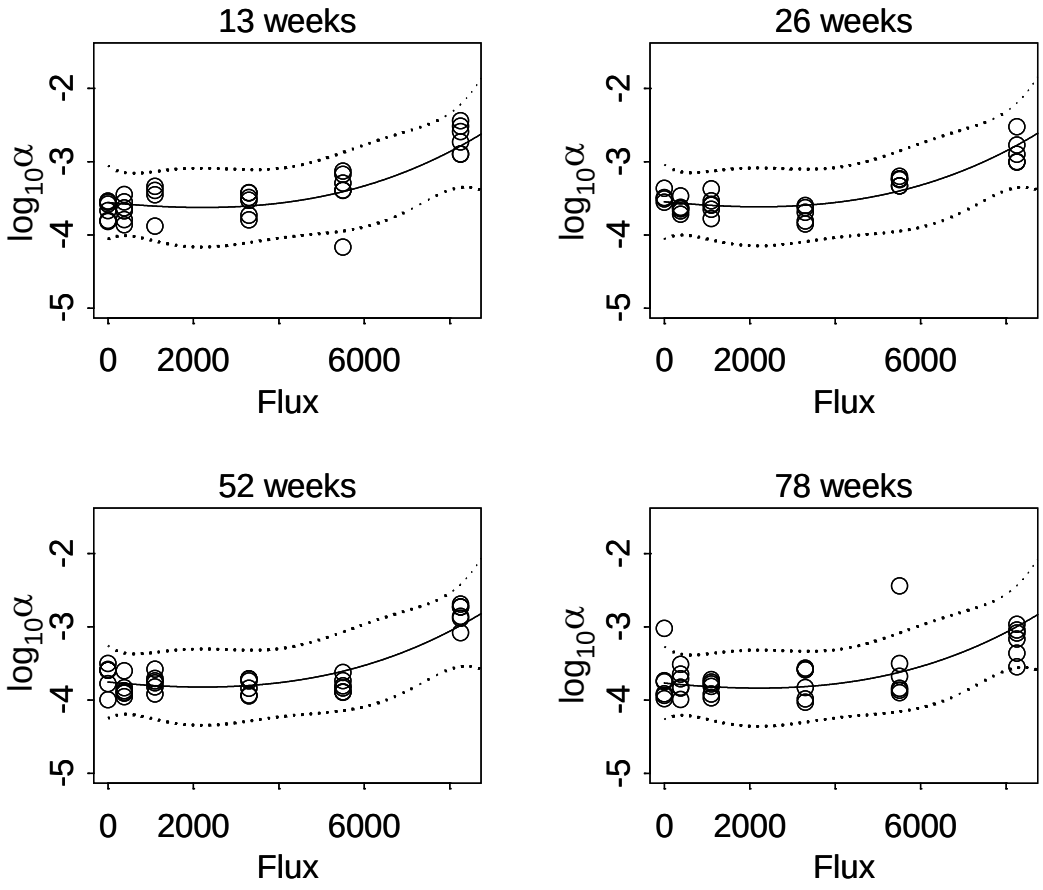

Figure E-3B. Logarithm of normal cell replication rate versus formaldehyde flux with simultaneous confidence limits for the PLM.

Source: Subramaniam et al. (2008).

The variability considered is that among animals and any measurement error as well as any other design-related components of error. Simultaneous 95% confidence limits for $\log(\alpha_N)$ were produced using Scheffe's method (Snedecor and Cochran, 1980). These 95% confidence limits span a range of 0.96 in $\log10(\alpha_N)$, or nearly a 10-fold range in median α_N. There is additional dispersion in these data that does not appear in Figures E-2 and E-3 for α_N, derived using the mean value of ULLI/LI; due to variation in the number of cells per mm basement membrane, the ratio of ULLI/LI had a spread of approximately ±25% (0.45 to 0.71, mean 0.60) among the eight observations considered in Monticello et al. (1990). Thus:

1. As suggested by Table E-3, and Figures E-2 and E-3, the shape of α_N(flux) in Conolly et al. (2003) is therefore likely to be very sensitive to how α_N is weighted and averaged over site and time.
2. Averaging of sites could significantly affect model calibration because of substantial nonlinearity in model dependence on α_N at the 10 and 15 ppm doses associated with high cancer incidence.
3. Monticello et al. (1996) found a high correlation between tumor rate and the ULLI weighted by the number of cells at a site. Therefore, considering these factors while regressing α_N against tissue dose would be important in the context of site differences in tumor response.
4. A further complexity arises because of histologic changes and thickening that occurs in the nasal epithelium over time in the higher dose groups (Morgan, 1997), factors that are likely to affect estimates of local formaldehyde flux, uptake, and replication rates (Subramaniam et al., 2008).

It is clear from Figures E-1 and E-3 that the time dependence in cell replication is significant. It would also be useful to examine if this time dependence affects the results of the time-to-tumor modeling and if early temporal changes in replication rate are important to consider because of the generally cumulative nature of cancer risk. The time window over which formaldehyde-induced cancer risk is most influenced is not known, but the time weighting used by Conolly et al. (2003) assigns a relatively low weight to labeling observed at early times compared with those observed at later time points. Finally, initiated cells are likely to be replicating at higher rates than normal cells as evidenced in several studies on premalignant lesions (Coste et al., 1996; Dragan et al., 1995; Rotstein et al., 1986). Therefore, LI data as an estimator of normal cell replication rate would be most reliable at early times when the mix of cells sampled include fewer preneoplastic or neoplastic cells.

The more relevant question, therefore, is whether the α_N(flux) derived in Conolly et al. (2003) by a TWA over all sites has an effect on low-dose risk estimates. Given the above uncertainties and variability not characterized in CIIT (1999) or in Conolly et al. (2003), it is important to examine whether additional dose-response curves that fit the cell replication data reasonably well have an impact on estimated risk. Such sensitivity analyses are carried out in the sections that follow.

E.3.2.4. *Alternate Dose-Response Curves for Cell Replication*

Clearly, a large number of alternative α_N(flux) can be developed. In conjunction with the other uncertainties, mainly the use of control data and alternative model structures for initiated cell kinetics, the number of plausible clonal growth models to be exercised soon require a

prohibitively large investment of time. Therefore, detailed analyses were restricted to a select set of biologically plausible choices of curves for α_N(flux), which would allow the identification of a range of plausible risk estimates (MLEs and statistical bounds). This discussion is further informed by recently published dose response data for cell replication (Meng et al., 2010), detailed in section F.2.3.

Six alternative equations for α_N were developed by regression analysis of the Monticello et al. (1996) ULLI data. The replicate data corresponding to the summary data presented in this paper were kindly provided to EPA by CIIT for further analyses. In each of these equations, α_N is expressed as a function of formaldehyde flux to nasal tissue (pmol/mm^2-hour) and, in one equation (see eq E-11) that explored time-dependence, the duration of exposure to formaldehyde in weeks. All the graphs use flux/10,000 for the X-axis, and the Y-axis expresses $\log_{10}\alpha_N$.

One source of uncertainty in the cell replication dose response in Conolly et al. (2003) is the large value of α_{max} (the cell replication rate corresponding to the upper end of the flux range at 15 ppm exposure) in the upward extrapolation from the empirically-determined α_N(flux) (see Figure D-1 and surrounding text in Section D.5). The optimal value of α_{max} was found by Conolly et al. (2003) to be 0.0435 hour^{-1}. As noted by the authors, an argument in support of this value is that it corresponds to the inverse of the fastest cell cycle times found in the literature. Since the model treats the induced replication rates as being time invariant, this means that cells in the high-flux region(s) divide at the highest cell turnover rate ever observed throughout most of an animal's life. This does not seem to be biologically plausible (Subramaniam et al., 2008).

Our analysis found that a 20% increase or decrease in the estimated value for α_{max} degraded the fit to the tumor incidence data considerably. Because of the interplay between the parameters estimated by optimization, this sensitivity of the model to α_{max} indicates that it is necessary to examine if other plausible values of α_{max} are also indicated by the data and to what extent low dose estimates of risk are influenced by the uncertainty in its value. The need for such an analysis is also indicated by Figure E-2. The value of α_{max} ($\log_{10}\alpha_{max} = -1.37$) in Conolly et al. (2003) is roughly an order of magnitude greater than the values of α_N(flux) at the highest flux levels in this figure. If the data pooled over all sites and times are to be used for α_N(flux), then, based solely on the trend in α_N(flux) in Figure E-2, it appears unlikely that α_N(flux) could increase up to this value of α_{max}. Visually, these empirically derived data collectively suggest that α_N versus flux could be leveling off rather than increasing 10-fold. Therefore, as an alternative to the approach taken in Conolly et al. (2003) of estimating α_{max} via likelihood optimization against the tumor data, regressions of the empirical cell replication data

in Figure E-2 were used to extrapolate α_N(flux) outside the range of observation (recognizing the uncertainty and model dependence that still results from extrapolating well outside the range of observed data).

In fitting dose-response curves to the cell replication data, a functional form was used that was flexible to allow a variety of monotonic and nonmonotonic shapes, with a parameter that determined the asymptotic behavior of the dose-response function. This allowed the extrapolation of α_N(flux) to higher flux levels by only relying on the empirical cell replication data. Then, there is no need for an adjustable parameter to be estimated by fitting to the tumor data. However, the plausible asymptotes obtained in this manner spanned a large range. In one case below, the asymptote suggested by the fit to the empirical cell replication data was judged to be abnormally high. In this case, the α_N versus flux curve was followed until the biological maximum of α_{max} (as given in Conolly et al. [2003]) was reached.

In three of the six regression models below, the data were restricted to the earliest exposure time (13 weeks) in Monticello et al. (1996) for which the cell proliferation rate (α_N) could be calculated. The interest in using only the 13-week exposure time arises from observations (Monticello et al., 1996, 1991) that at later times there were more frequent and severe histologic changes, which may have altered formaldehyde uptake and cell proliferation response. Consequently, given that the data in Monticello et al. (1991) for times earlier than 13 weeks could not be utilized as explained in Section E.3.2.3, the 13-week responses might better represent proliferation rates for use in a two-stage model of the cancer process than the rest of the Monticello et al.(1996) data.

Second, the LI data showed considerable variation among nasal sites, which may be related to the variation in tumor response among sites. Since the cell replication dose-response curves used in the cancer model represent all of the sites, it was attempted to include this variation by weighting the regression by the relative cell populations at risk at each of the sites. This was carried out for some of the models as stated below.

Finally, in one of the regression models, derived from fitting to all of the Monticello et al. (1996) ULLI data, time-dependence of α_N was considered by using weeks of exposure as a covariate. In this model, time was a regression (continuous) predictor, not a class variable, and its coefficient represents the change in $\log_{10} \alpha_N$ per week of exposure

The following regression models for α_N versus flux, denoted in the equations below as N1–N6 and shown in Figure E-4, as well as the hockey-stick and J-shaped curves used by Conolly et al. (2003), shown in Figure D-1, Appendix D, were next used as inputs to the clonal growth model for cancer:

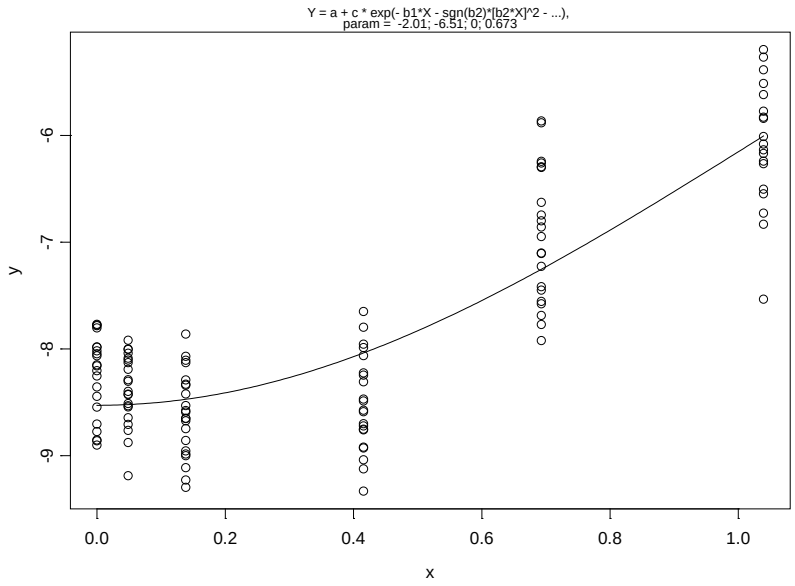

Figure E-4, N1. Various dose-response modeling of normal cell replication rate.

Note: See text for definitions of N1–N6. N1: Quadratic; monotone increasing in flux, derived from fit to all of the Monticello et al. (1996) ULLI data.

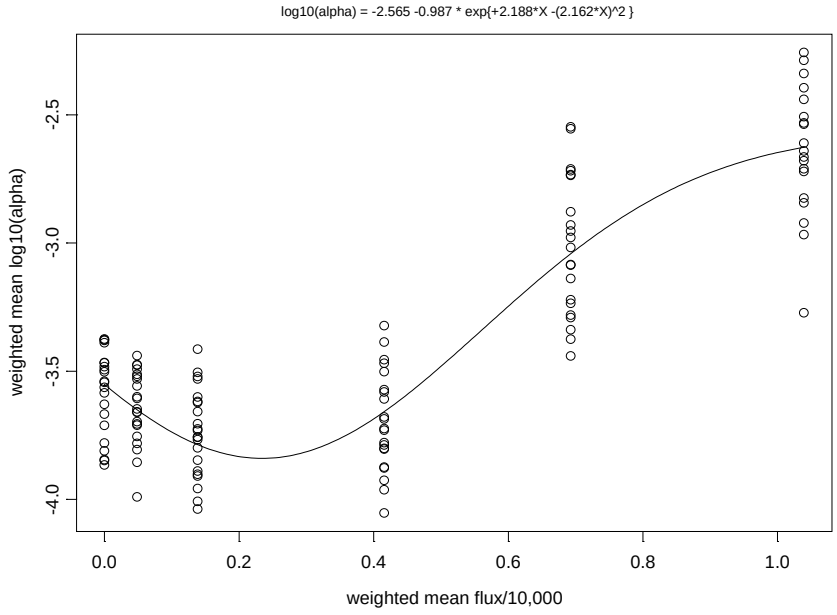

Figure E-4, N2. Various dose-response modeling of normal cell replication rate.

Note: See text for definitions of N1–N6. N2: Linear-quadratic; decreasing in flux for small values of flux, derived from fit to all of the Monticello et al. (1996) ULLI data.

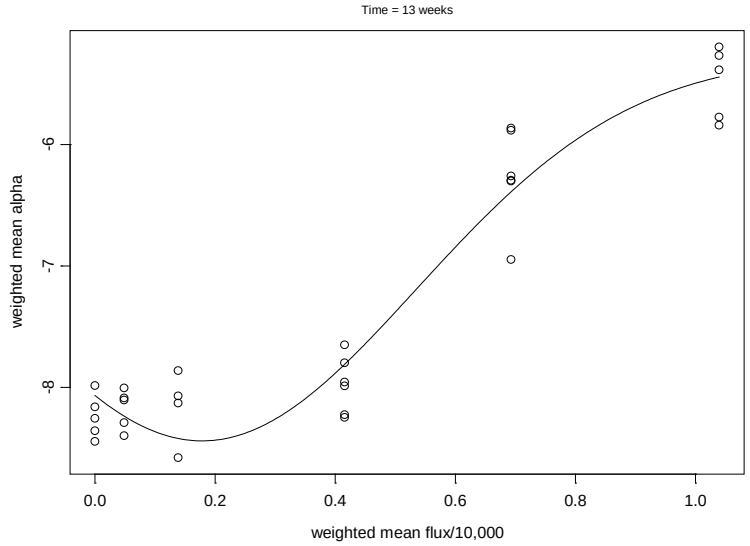

Figure E-4, N3. Various dose-response modeling of normal cell replication rate.
Note: See text for definitions of N1–N6. N3: Linear-quadratic; decreasing in flux for small values of flux, derived from fit to the 13-week Monticello et al. (1996) ULLI data, using average flux over all sites for a given ppm exposure and weighting regression by estimates of the numbers of cells at each of five sites.

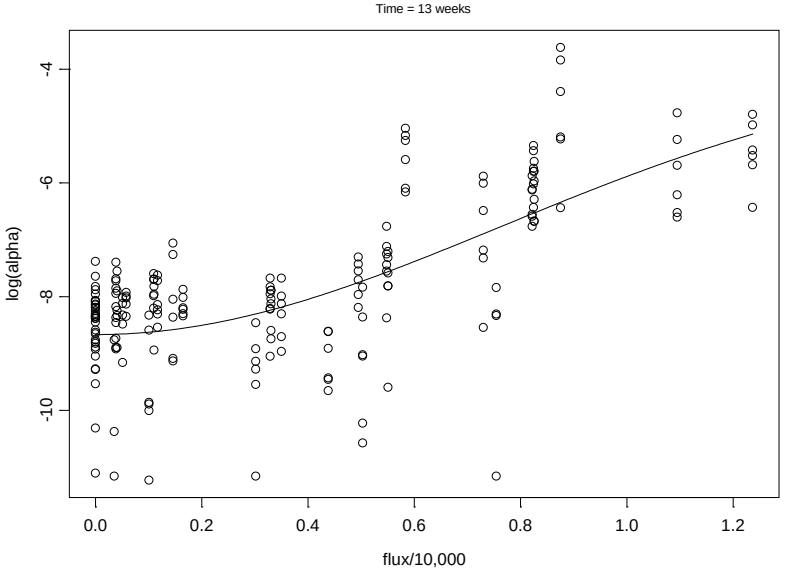

Figure E-4, N4. Various dose-response modeling of normal cell replication rate.
Note: See text for definitions of N1–N6. N4: Quadratic; monotone increasing in flux, derived from unweighted fit to 13-week Monticello et al. (1996) ULLI data.

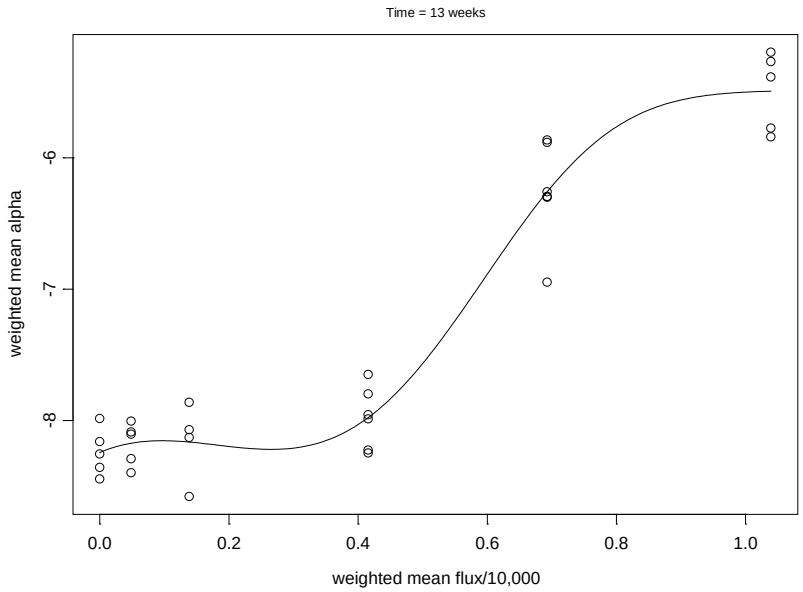

Figure E-4, N5. Various dose-response modeling of normal cell replication rate.

Note: See text for definitions of N1–N6. N5: Linear-quadratic-cubic; initially increasing slightly with increasing flux, then decreasing slightly, and finally increasing, derived from fit to 13-week Monticello et al. (1996) ULLI data, using average flux over all sites for a given ppm exposure and weighting regression by estimates of the numbers of cells at each of five sites.

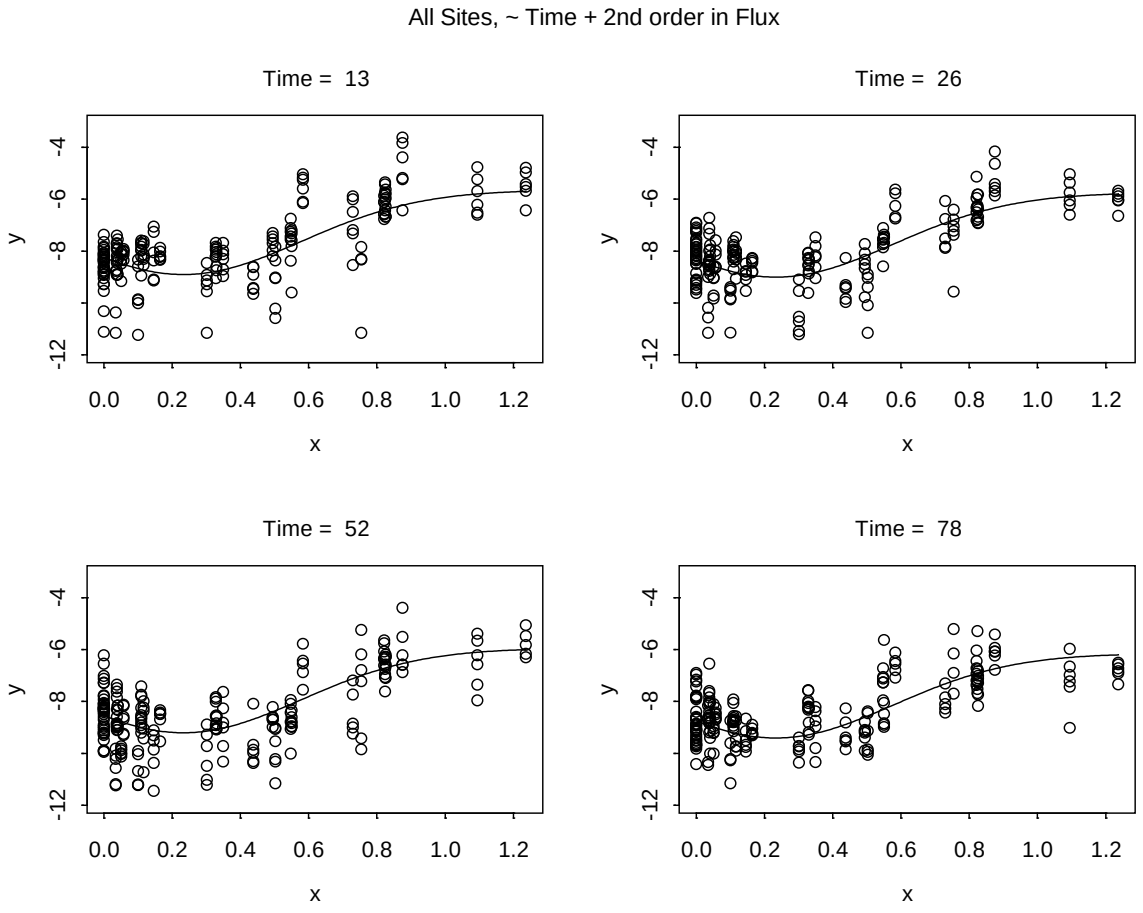

Figure E-4, N6. Various dose-response modeling of normal cell replication rate.

Note: See text for definitions of N1–N6. N6: Linear-quadratic-cubic; initially increasing slightly with increasing flux, then decreasing slightly, and finally increasing, derived from fit to all Monticello et al. (1996) ULLI data, using weeks of exposure as a covariate. In this model, time was a regression (continuous) predictor, not a class variable, and its coefficient represents the decrease in $\log_{10}$ α_N per week of exposure time.

N1: Quadratic; monotone increasing in flux, derived from fit to all of the Monticello et al. (1996) ULLI data.

$$\alpha_N = Exp\{-2.015 - 6.513 \times Exp[-(6.735\times10^{-4} \times flux)^2]\} \quad \text{(E-6)}$$

N2: Linear-quadratic; decreasing in flux for small values of flux, derived from fit to all of the Monticello et al. (1996) ULLI data.

$$\alpha_N = Exp\{-5.906 - 2.272 \times Exp[2.188\times10^{-4} \times flux - (2.162\times10^{-4} \times flux)^2]\} \quad \text{(E-7)}$$

N3: Linear-quadratic; decreasing in flux for small values of flux, derived from fit to the 13-week Monticello et al. (1996) ULLI data, using average flux over all sites for a given ppm exposure and weighting regression by estimates of the numbers of cells at each of five sites.

. $$\alpha_N = Exp\{-5.274 - 2.792 \times Exp[1.407\times10^{-4} \times flux - (1.986\times10^{-4} \times flux)^2]\} \quad \text{(E-8)}$$

N4: Quadratic; monotone increasing in flux, derived from unweighted fit to 13-week Monticello et al. (1996) ULLI data.

$$\alpha_N = Exp\{-3.858 - 4.809 \times Exp[-(9.293\times10^{-5} \times flux)^2]\} \quad \text{(E-9)}$$

N5: Linear-quadratic-cubic; initially increasing slightly with increasing flux, then decreasing slightly, and finally increasing, derived from fit to 13-week Monticello et al. (1996) ULLI data, using average flux over all sites for a given ppm exposure and weighting regression by estimates of the numbers of cells at each of five sites.

$$\alpha_N = Exp\{-5.488 - 2.755 \times Exp[-7.808\times10^{-5} \times flux + (2.349\times10^{-4} \times flux)^2 - (2.166\times10^{-4} \times flux)^3]\} \quad \text{(E-10)}$$

N6: Linear-quadratic-cubic; initially increasing slightly with increasing flux, then decreasing slightly, and finally increasing, derived from fit to all Monticello et al. (1996) ULLI data, using weeks of exposure as a covariate. In this model, time was a regression (continuous) predictor, not a class variable, and its coefficient represents the decrease in $\log_{10} \alpha_N$ per week of exposure time.

$$\alpha_N = Exp\{7.785\times10^{-3} \times (weeks) - 5.722 - 2.501 \times Exp[1.103\times10^{-4} \times flux - (7.223\times10^{-5} \times flux)^2 - (1.575\times10^{-4} \times flux)^3]\} \quad \text{(E-11)}$$

E.3.3. Uncertainty in Model Specification of Initiated Cell Replication and Death

E.3.3.1. *Biological Implications of Assumptions in Conolly et al. (2003)*

The results of a two-stage MVK model are extremely sensitive to the values for initiated cell division (α_I) and death (β_I) rates, particularly in the case of a sharply rising dose-response curve as observed of formaldehyde. The pool of cells used for obtaining the available LI data (Monticello et al., 1996, 1991) consists of largely normal cells with perhaps increasing numbers of initiated cells at higher exposure concentrations. As such there is no way of inferring the division rates of initiated cells in the nasal epithelium, either spontaneous (baseline) or induced by exposure to formaldehyde, from the available empirical data. Conolly et al. (2003) considered α_I(flux) as a function of α_N(flux) as given by eq D-2 in Appendix D. As shown in Figure D-1 (Appendix D), α_I is estimated in Conolly et al. (2003) to be very similar to α_N. That is, with eq D-2 assumed to relate α_I(flux) to α_N(flux), a J- or hockey-shaped dose-response curve for α_N(flux) necessarily results in a J or hockey shape for α_I(flux).

The J shape for the TWA α_N(flux) in Conolly et al. (2003) could plausibly be explained, as suggested by the examples in Conolly and Lutz (2004), by a mathematical superposition of dose-response curves describing the effects of the inhibition of cell replication by the formation of DPXs (Heck and Casanova, 1999) and cytotoxicity-induced regenerative replication (Conolly, 2002). However, as explained earlier, there is considerable uncertainty and variability, both qualitative and quantitative, in the interpretation of the LI data and in the derivation of *normal* cell replication rates from the ULLI data. While the TWA values of ULLI indicate a J-shaped dose response for some sites, as also concluded by Gaylor et al. (2004), this is not consistently the case for all exposure times and sites as discussed earlier. Notwithstanding this uncertainty and variability, and in the absence of data, the following essential questions have a significant impact on risk predictions and need resolution if the model structure in eq D-2 is to be used in a biologically based (or motivated) sense:

- Should mechanisms that might explain a J-shaped dose response for normal cell replication be expected to prevail also for initiated cells? An identical question can be posed for the hockey-stick-shaped curve which indicates a cytotoxicity-driven threshold in dose response.
- Would the formaldehyde flux at which the cell replication dose-response curve rises above its baseline be similar in value for both normal and initiated cells as inferred by the CIIT model in Figure D-1?

The next critical assumption in Conolly et al. (2003) was that made for β_I (the death rate of initiated cells), namely, $\beta_I(\text{flux}) = \alpha_N(\text{flux})$ (see eq D-3). The rationale for this assumption is explained by assuming formaldehyde to be equally cytotoxic to initiated and normal cells since the mechanism is presumed to be via its general chemical reactivity (Subramaniam et al. 2008). In essence, this assumption brings the cytotoxic action of formaldehyde to bear strongly on the parameterization of the CIIT model.

There are no data to evaluate the strength of these assumptions, so Subramaniam et al. (2008) studied the plausibility of various inferences that arise as a result of these assumptions. These inferences are only briefly listed here (see the paper for further discussion).

- For flux <27,975 pmol/mm^2-hour, $\alpha_I > \alpha_N$ (see Figures D-1 and D-2 of Appendix D). Qualitatively, this concept of a growth advantage is in line with data on epithelial and other tissue types with or without exposure to specific chemicals.
- For higher flux levels, however, the model indicates $\alpha_I < \alpha_N$ (see Figure D-2). There are no data to shed further light on this inference.
- At these higher flux levels, initiated cells in the model die at a faster rate than they divide, indicating the extinction of initiated cell clones in regions subject to these flux levels. There are no data indicating formaldehyde to have this effect.

In evaluating these inferences, Subramaniam et al. (2008) point to various data that indicate that initiated cells represent distinctly different cell populations from that of normal cells with regard to proliferation response (Ceder et al., 2007; Bull, 2000; Schulte-Hermann et al., 1997; Coste et al., 1996; Dragan et al., 1995), have excess capacity to clear formaldehyde and, in general, are considerably more resistant to cytotoxicity, and may already have altered cell cycle control. The resistance to toxicity is manifested variably as decreased ability of the toxicant to induce cell death or to inhibit cell proliferation compared to corresponding effects in normal cells. Therefore, the influence of formaldehyde on apoptosis likely differs between normal and initiated cells.

As concluded in Subramaniam et al. (2008), taken together, there is much data to suggest that inferring $\alpha_I < \alpha_N$ at cytotoxic formaldehyde flux levels is problematic and that death rates of initiated cells are likely to be very different from those of normal cells.

In the absence of data to indicate that eq D-2 and eq D-3 (in Appendix D) are biologically reasonable approaches to link the kinetics of initiated cells with those of normal cells, alternate model structures other than those represented by these relationships considered by Conolly et al. (2003) need to be explored, given that the two-stage model is extremely sensitive to α_I and β_I. Such an evaluation needs to primarily explore if the assumptions in eq D-2 and eq

D-3 significantly impact the intended use of the model, namely extrapolation to low-dose human cancer risk and the calculation of an upper bound on human risk. Any such alternate model structure needs to provide a good fit to the time-to-tumor data.

E.3.3.2. *Plausible Alternative Assumptions for α_I and β_I*

Therefore, in the additional sensitivity analysis presented here,

a) Initiated cell kinetics are considered to be independent of normal cells,
b) Initiated cell replication dose-response cannot take a J shape; this is motivated by the consideration that lower-than-baseline turnover rate represents an increased amount of DNA repair taking place, which may not be consistent with impaired DNA repair in initiated cells.

Thus, two alternatives were considered to eq D-2 for α_I(flux):

I1: $$\alpha_I = \gamma_1 \times [1 + exp(\gamma_2 / \gamma_3)] / \{1 + exp[-(flux - \gamma_2) / \gamma_3]\} \quad \text{(E-12)}$$

I2: $$\alpha_I = max[\alpha_I(\text{I1}), \alpha_{NBasal}] \quad \text{(E-13)}$$

Here γ_1, γ_2, and γ_3 are parameters estimated by fitting the cancer model to the rat bioassay data. In eq E-12, α_I increases monotonically with flux from a background level of γ_1 asymptotically up to a maximum value of $\gamma_1 \times [1 + Exp(\gamma_2 / \gamma_3)]$. The choice of this functional form in eq E-12 and eq E-13 was considered in order to be parsimonious while at the same time allowing for a flexible shape to the dose-response curve. The sigmoidal curve allows for the possibility of a slow rise in the curve at low dose and an asymptote.

Equation E-13 is a modification of eq E-12 that restricts the rate of division of initiated cells to be at least as large as the spontaneous division rate of unexposed normal cells. There is evidence to suggest (e.g., in the case of liver foci) that initiated cells have a growth advantage over normal cells, with or without exposure to specific chemicals (Ceder et al., 2007; Grasl-Kraupp et al., 2000; Schulte-Hermann et al., 1999; Coste et al., 1996; Dragan et al., 1995).

In addition, in most runs, an upper bound (α_{high}) is selected for both α_N and α_I. This value is assumed to represent the largest biologically plausible rate of cell division. Following Conolly et al. (2003), in most cases α_{high} is set equal to 0.045 hours^{-1}. If a value of α_I or α_N computed using one of the above formulas exceeded α_{high}, the value of α_{high} was used in the computation rather than the value obtained by using the formula.

As noted above, Conolly et al. (2003) set the rate of death for intermediate cells, β_I, equal to the division rate of normal cells, $\beta_I = \alpha_N$. On the other hand, apoptotic rates and cell

proliferation rates are thought to be coupled (Schulte-Hermann, 1999; Moolgavkar, 1994), so that death rates of initiated cells would rise concomitantly with an increase in their division rates (Grasl-Kraupp et al., 2000; Schulte-Hermann et al., 1999). Therefore, as an alternative to the Conolly et al. (2003) formulation, it is assumed that the death rate of intermediate cells is proportional to the division rate of intermediate cells.

$$\beta_I = \kappa_\beta \times \alpha_I \qquad \text{(E-14)}$$

where the constant of proportionality, κ_β, is an additional parameter to be estimated by optimization against the tumor incidence data. Such an assumption has also been made by other authors (Luebeck et al., 2000, 1995; Moolgavkar et al., 1993).

E.3.4. Results of Sensitivity Analyses on α_N, α_I, and β_I

E.3.4.1. *Further Constraints*

The number of models that might be constructed if all the possibilities listed above for α_N, α_I, *and* β_I are to be tried in a systematic manner clearly become exponential and daunting. (Optimally, it would have been desirable to elucidate the role of a specific modification while keeping others unchanged to determine risk.) Therefore, in order to carry out a viable sensitivity analysis while at the same time examining the plausible range of risks resulting from variations in parameters and model structures, various uncertainties were combined in any given simulation. By using the constraints described above (see eqs E-6 through E-13 and associated text) for α_I , β_I, and α_N, 19 models were obtained that provided similarly good fits to the time-to-tumor data (which in some cases contained only five dose groups).

However, for many of these models, the optimal α_I(flux) displayed a threshold in flux even when the model utilized for α_N(flux) was a monotonic increasing curve without a threshold (i.e., model N4 for α_N in Figure E-4). Indeed, if a thresholded dose-response curve was plausible for α_I based on arguments of cytotoxicity, then a threshold is all the more plausible for α_N, and such models are removed from consideration.

Secondly, the basal value of α_I was required to be at least as large as the basal value of α_N. Another constraint was placed on the baseline initiated cell replication rate. In the absence of formaldehyde exposure, α_I was not allowed to be greater than two or four times α_N, even if such models described the tumor data, including the control data, very well. There are some data that suggest that baseline initiated cells have a small growth advantage over normal cells, so a huge advantage was thought to be biologically less plausible.

Finally, since most of the SCCs in the rat bioassays occurred in rats exposed to the highest formaldehyde concentration (15 ppm), the data from this exposure level have a big impact on the estimated model parameters. In most runs that incorporated the 15 ppm data, the model appeared, based on inspection of the KM plots, to fit the 15 ppm data quite well but to fit the lower exposure data less well. Because of the high level of necrosis occurring at 15 ppm, it is possible that the data at this exposure may not be particularly relevant to modeling the sharp upward rise in the dose response at 6 ppm. Furthermore, the principal interest is in the predictions of the model at lower levels to which human populations may be exposed. Consequently, in order to improve the fit of the model at lower exposures, some of the alternative models were constructed with the 15 ppm data omitted.

E.3.4.2. *Sensitivity of Risk Estimates for the F344 Rat*

Figure E-5 contains plots of the MLE of additional risk computed for the F344 rat at formaldehyde exposures of 0.001, 0.01, 0.1, and 1 ppm for eight models. Two log-log plots are provided. For those models for which the estimates of additional risk are all positive, the additional risks are plotted (panel A), and, for those for which estimates of additional risk are negative, the negatives of additional risks are plotted (panel B). Only five dose groups were considered (i.e., 15 ppm data omitted) for models 8, 5, 15, and 16. Figure E-6 shows the dose-response curves for α_N and α_I for these eight cases (panels A and B corresponding to those in Figure E-5). The specification and estimated values of the parameters for these models are provided in Tables E-4 and E-5. The primary results are as follows:

1. Among the models considered, negative values for additional risk can arise only in models in which the dose response for normal cells is J shaped. Thus, all of the models with negative dose responses for risk have J-shaped dose responses for normal cells. However, the converse is not necessarily true as may be noted from model 8. This model has both a positive dose response for risk and a J-shaped dose response for normal cells. In this case, the strong positive increase in response of initiated cells at low dose was sufficient to counteract the negative response of normal cells.
2. For doses below which no tumors were observed, the risk estimates predicted by the different models span a very large range. This result points to large uncertainties in model specification (how to relate the kinetics of normal and initiated cells) as well as in parameter values. As mentioned above, the analysis does not attempt to separate the influence of the different sources of uncertainty, so this range also incorporates the uncertainty arising from the use of different control data and that due to α_{max}.

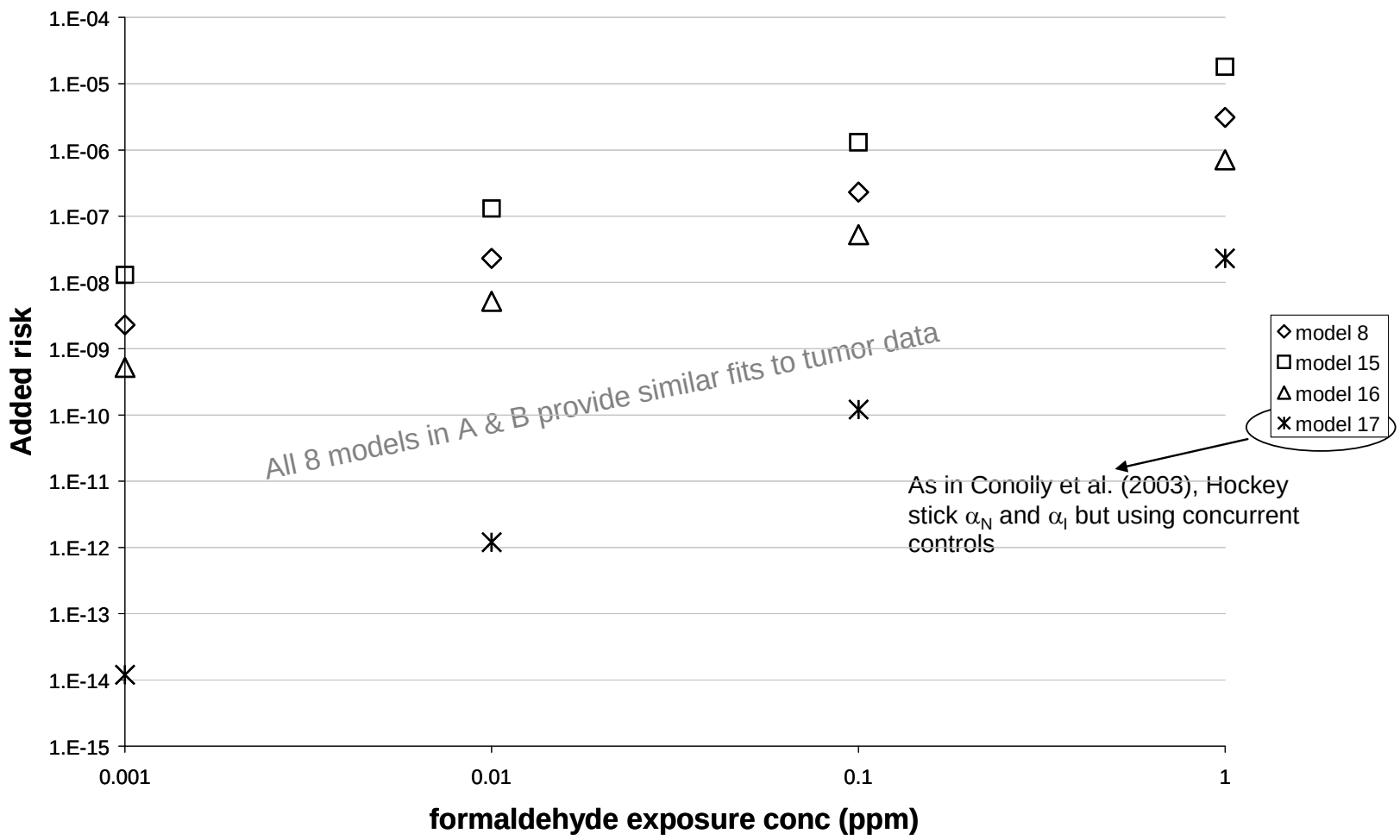

Figure E-5A. BBDR models for the rat—models with positive added risk.

Note: All four models provide "similar" fits to tumor data (see text).

Figure E-5B. BBDR rat models resulting in negative added risk.

Note: All four models provide "similar" fits to tumor data (see text).

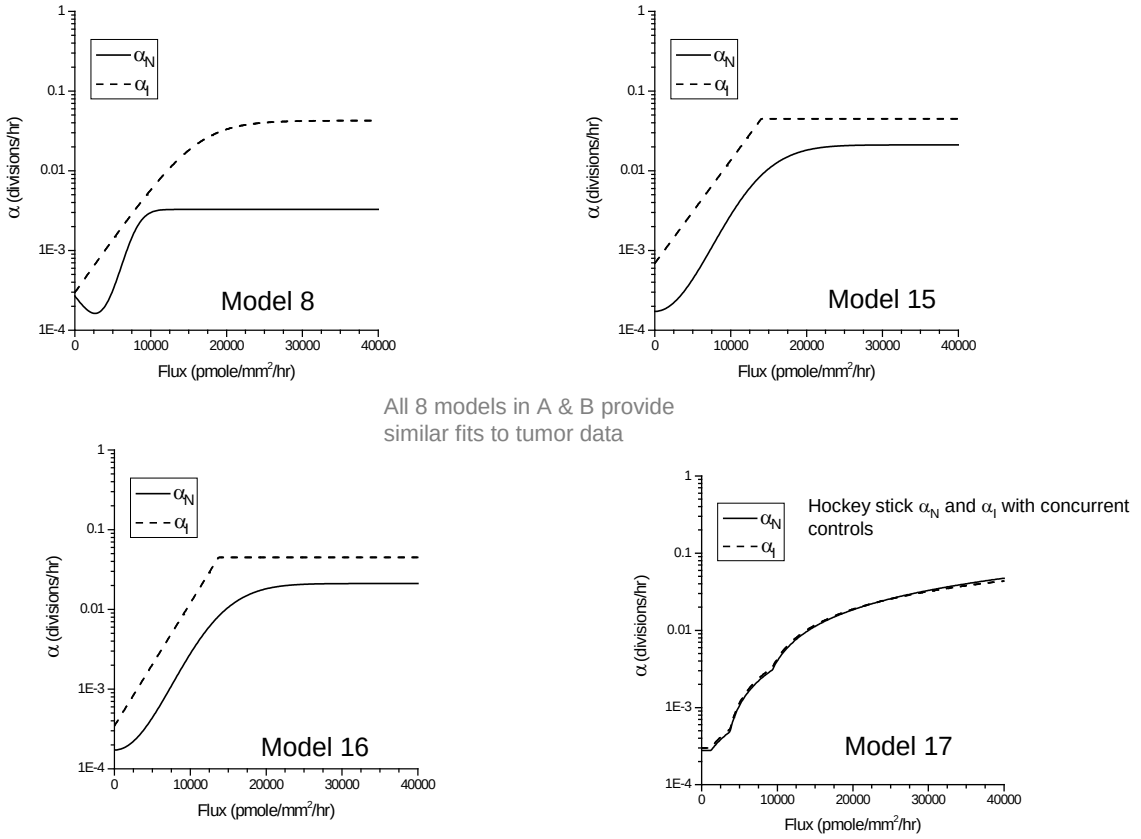

Figure E-6A. Models resulting in positive added rat risk: Dose-response for normal and initiated cell replication.

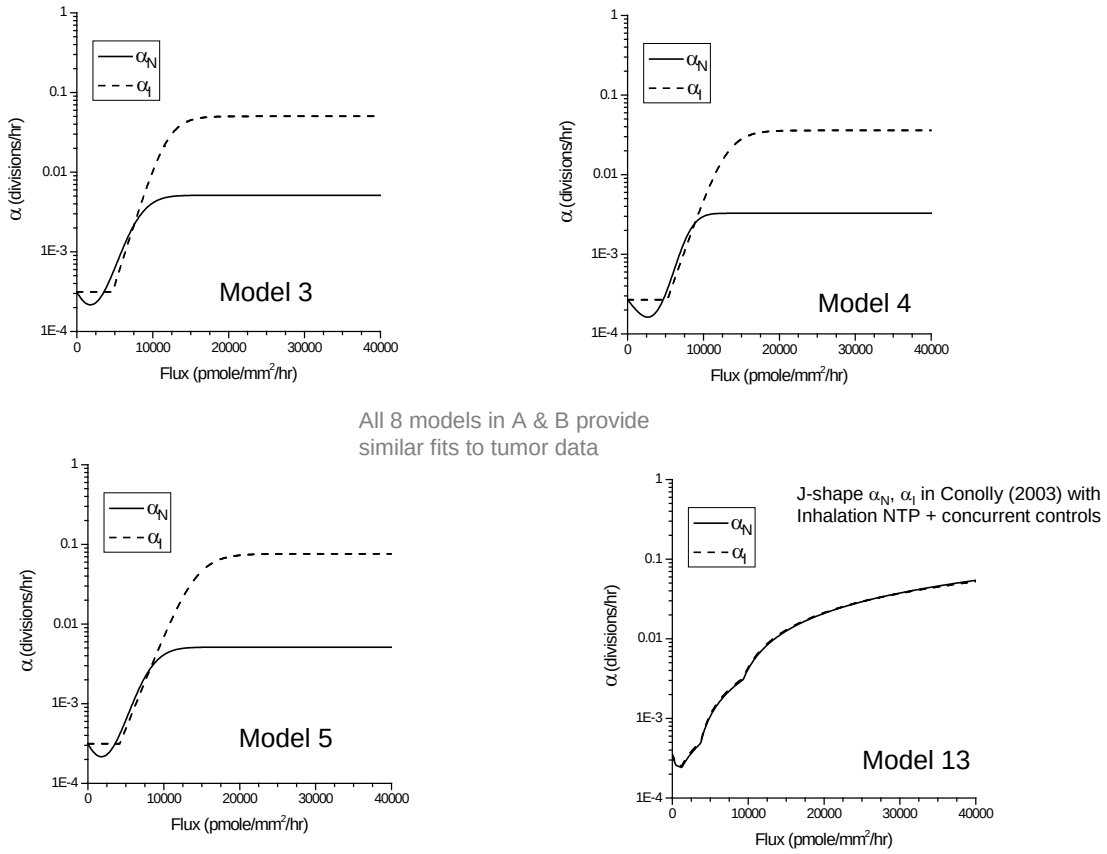

Figure E-6B. Models resulting in negative added rat risk: Dose-response for normal and initiated cell replication.

Table E-4. Parameter specifications and estimates for clonal growth models of nasal SCC in the F344 rat using alternative characterization of cell replication and death rates

Parameters	Model 3	Model 4	Model 5	Model 8	Model 15	Model 16
Historical controls added to concurrent	Inhalation NTP	Inhalation NTP	Inhalation NTP	Inhalation NTP	Inhalation NTP	Inhalation NTP
Number of dose groups	6	6	5	5	5	5
DPX concentration	Subramaniam et al. (2007)	Subramaniam et al. (2007)	Subramaniam et al. (2007)	Subramaniam et al. (2007)	Subramaniam et al. (2007)	Subramaniam et al. (2007)
α_N definition	N3	N6	N3	N6	N4	N4
α_I definition	I2	I2	I2	I1	I1	I1
α_{high}	--	0.045	--	0.045	0.045	0.045
β_I definition	$\beta_I = K_\beta \times \alpha_I$	$\beta_I = K_\beta \times \alpha_I$	$\beta_I = K_\beta \times \alpha_I$	$\beta_I = K_\beta \times \alpha_I$	$\beta_I = K_\beta \times \alpha_I$	$\beta_I = K_\beta \times \alpha_I$
					$\gamma_1 \leq 4\ \alpha_{NBasal}$	$\gamma_1 \leq 2\ \alpha_{NBasal}$
Log-likelihood	-1495.34	-1495.61	-184.02	-184.22	-182.75	-186.37
μ_{NBasal}	7.518E-7	1.664E-6	8.684E-7	9.230E-7	1.037E-6	1.662E-7
KMU	3.884E-7	3.471E-7	0.0	0.0 (0.0, 2.093E-6)	4.582E-6 (1.8E-6,1.86E-5)	0.0
KMX (*KMU* / μ_{NBasal})	0.5166	0.2086	0.0	0.0 (0.0, 4.696)	4.420 (1.53, 17.67)	0.0
D_0[§]	214.3	199.7	261.8	254.2	423.2	245.1
D_{0F}[§]	75.26	79.81	119.7	101.1	100.8	98.83
γ_1	1.164E-5	1.006E-5	3.168E-5	2.967E-4	6.888E-4	3.441E-4
γ_2	1427	1591	1825	3223	4652	2818
γ_3	11944	13017	14207	15989	54334	37896
K_β	0.9893	0.9848	0.9804	0.9504	1.006	0.9660

[§]See Subramaniam et al. (2007) for an explanation of the time delay constants D_0 and D_{0F}.

Table E-5. Parameter specifications and estimates for clonal growth models of nasal SCC in the F344 rat using cell replication and death rates as characterized in Conolly et al. (2003)

Parameters	Model 13	Model 17
Historical controls added to concurrent	All NTP	NO historical controls
Number of dose groups	6	6
DPX concentration	Conolly et al. (2000)	Subramaniam et al. (2007)
α_N definition	J-shape (TWA, Conolly et al. 2003)	Hockey (TWA, Conolly et al., 2003)
α_I definition	eq. D-1	eq. D-1
α_{high}	--	--
β_I definition	$\beta_I = \alpha_N$	$\beta_I = \alpha_N$
Log-likelihood	−1692.68	−1474.29
μ_{NBasal}	1.731E-6	0.0
KMU	0.0	1.203E-6 (1.0E-6,1.427E-6)
KMX (*KMU*/μ_{NBasal})	0.0	Infinite (0.4097,infinite)
D_0[§]	239.5	243.13
D_{0F}[§]	66.31	68.83
multib	1.047	1.078E+0
multic	1.510	3.347
α_{max}	5.153E-2	0.045

[§]See Subramaniam et al. (2007) for an explanation of the time delay constants D_0 and D_{0F}.

3. At the 10 ppb (0.01 ppm) concentration, MLE risks range from -4.0×10^{-6} to $+1.3 \times 10^{-7}$. At this dose, models that gave only positive risks resulted in a five orders of magnitude risk range from 1.2×10^{-12} to 1.3×10^{-7}, while narrowing to a four orders of magnitude risk range from 1.2×10^{-10} to 1.3×10^{-6} at the 0.1 ppm level. This narrowing continues as exposure concentration increases, and the curves coalesce to substantially similar values at 6 ppm and above (not shown). For all these 8 models, the rat added risk at 6.0 ppm ranged from 1.8×10^{-2} to 2.1×10^{-2}.

4. There does not seem to be any systematic effect on additional risk that depends on whether the 15 ppm data are included in the analysis.

5. For all of the models except models 13 and 17 in Figures E-5 and E-6, the additional risk varies substantially linearly with exposure at low exposures between 0.001 and 1.0 ppm (departing only to a small extent from linearity between 0.1 and 1.0 ppm). Models 13 and 17 show a quadratic dependence; these models employ the TWA J-shape and hockey stick dose response curves for α_N used in Conolly et al. (2003) and the same equations used by those authors to relate α_I and β_I to α_N (see eqs D-2 and D-3, Section D-6). However, the control data in Model 17 was different from those used by Conolly et al.; while all NTP controls were added to the concurrent controls in model 13, only concurrent controls were used in model 17.

The various model choices presented in Figure E-5 all provided equally good fits to the time-to-tumor data although within the context of a significant qualification. It was not possible to simply use the maximized log-likelihood values as a means of comparing the goodness-of-fit to the tumor incidence data across all these model choices. This is because many of the model choices differed in the number of doses or in the number of control animals that were used, so the fits were compared across such models only visually.

Wherever results from the BBDR modeling are discussed, values of added risk, as opposed to extra risk, are reported. This is purely for convenience in interpretation. Because of the low background incidence, these values are only negligibly different from the corresponding extra risk estimate. The final risk (or unit risk) estimates provided in this document are based on extra risk estimates.

E.3.4.3. *MOA Inferences Revisited*

The ratio KMU/μ_{Nbasal} represents the added fractional probability of mutation per cell generation ($\mu_N - \mu_{Nbasal}$)/μ_{Nbasal} due to unit concentration of DPXs. As discussed in Sections E.3.1.2 and E.3.1.5 (see Appendix E), this parameter has a critical impact on the extrapolation as well as on inferring whether the mutagenic action of formaldehyde is relevant to explaining the observed tumor incidence or its carcinogenicity at lower concentrations. In that prior discussion, this ratio was found to be extremely sensitive to the choice of historical control data. The analysis indicates that, for a given set of control data that is used, uncertainties associated with α_N and α_I also have a large impact on this ratio.

As discussed in E.3.1.2, this ratio was infinite when concurrent controls were used because the MLE value for μ_{Nbasal} was found to be zero. The use of these concurrent controls, however, does not necessarily imply that μ_{Nbasal} will be determined to be zero. In one of the scenarios examined in the sensitivity analysis, where concurrent controls were used along with the combination of dose-response curves eq D-9 for α_N (see Figure E-4) and eq E-13 for α_I, the

optimal value of the ratio KMU/μ_{Nbasal} was equal to 0.25. For the models in Figure 5-13A, this ratio was 0 for all except model 17 for which it was infinite. For the models in Figure 5-13B with negative added risk, the ratio ranged from 0–4.5. For some of those models where KMU/μ_{Nbasal} was finite, the upper confidence bound on this ratio was found to increase by an order of magnitude from the MLE value.

Thus, we conclude that the modeling does not help resolve the debate as to the relevance of formaldehyde's mutagenic potential to its carcinogenicity.

E.3.4.4. *Confidence Bounds: Model Uncertainty Versus Statistical Uncertainty*

For models 15 and 17 in Figures E-5A and E-6A, 90% CIs for additional risk were calculated by using the profile likelihood method. Table E-6 compares the lower and upper confidence bounds for these models for 0.001 ppm, 0.1 ppm (doses well below the range where tumors were observed), and 6 ppm (the lowest dose where tumors were observed) with the MLE risk estimates at these doses. In both cases, these intervals were quite narrow compared with the differences in risk predicted by different models in Figure E-5. This suggests that model uncertainty is of more consequence in the formaldehyde animal model than is statistical uncertainty. We also estimated confidence bounds using the bootstrap method for select models, and determined that these estimates were in agreement with the bounds calculated using the profile likelihood method. These results are not presented here. We return to the calculation of confidence limits when determining points of departure (PODs).

Table E-6. Comparison of statistical confidence bounds on added risk for two models

Dose (ppm)	Model	Lower bound	MLE	Upper bound
0.001	Model 15	4.4×10^{-9}	1.3×10^{-8}	1.6×10^{-8}
	Model 17	1.2×10^{-14}	1.2×10^{-14}	1.3×10^{-14}
0.1	Model 15	4.5×10^{-7}	1.3×10^{-6}	1.7×10^{-6}
	Model 17	1.2×10^{-10}	1.2×10^{-10}	1.3×10^{-10}
6	Model 15	1.8×10^{-2}	2.1×10^{-2}	2.3×10^{-2}
	Model 17	1.3×10^{-2}	1.8×10^{-2}	3.0×10^{-2}

In conclusion, it is demonstrated that the different formaldehyde clonal growth models can fit the data about equally well and still produce considerable variation in additional risk and biological inferences at low exposures. However, even with these large variations, the highest

MLE added risk for the F344 rat is only of the order of 10^{-6} at 0.1 ppm. Thus, with regard to calculating a reasonable upper bound that includes model and statistical uncertainty, the relevant question is whether the range arising out of uncertainties in the rat model amplifies when extrapolated to the human. Thus, in Appendix F, the human model in Conolly et al. (2004) will be examined.

This page intentionally left blank.

Appendix F

1

APPENDIX F

SENSITIVITY ANALYSIS OF BBDR MODEL FOR FORMALDEHYDE INDUCED RESPIRATORY CANCER IN HUMANS

F.1. MAJOR UNCERTAINTIES IN THE FORMALDEHYDE HUMAN BBDR MODEL

Subsequent to the BBDR model for modeling rat cancer, Conolly et al. (2004) developed a corresponding model for humans for the purpose of extrapolating the risk estimated by the rat model to humans. Also, rather than considering only nasal tumors, it is used to predict the risk of all human respiratory tumors. The human model for formaldehyde carcinogenicity (Conolly et al., 2004) is conceptually very similar to the rat model and follows the schematic in Figure 5-11 in Chapter 5. The model structure, notations, and calibration are described in Appendix D. Unlike the sensitivity analysis of the rat modeling where a number of issues were examined, a much more restricted analysis will be presented here for the sake of brevity. A more extensive analysis was carried out initially that carried forward several of the rat models from Appendix E to the human, and the lessons learned from those exercises are in agreement with the more restricted presentation that follows. Table F-1 lists the major uncertainties and assumptions in the human extrapolation model in Conolly et al. (2004).

Table F-1. Summary of evaluation of major assumptions and results in CIIT human BBDR model

Assumptions[a]	Rationale in Conolly et al. (2003) or CIIT (1999)	EPA evaluation	Further elaboration
Cell division rates derived from rat labeling data were assumed applicable to human (except for assuming different fraction of cells with replicative potential).	There are no equivalent LI data for human or guidance for extrapolating cell division rate across species.	Enzymatic metabolism plays a role in mitosis. Therefore, we expect interspecies difference in cell division rate. Basal cell division rates in humans are expected to be much more variable than in laboratory animals.	Subramaniam et al. (2008)
Parameters for enzymatic metabolism of formaldehyde in human PBPK model for DPX concentrations: K_m varies by order of magnitude between rat and monkey but is same for monkey and human. V_{max}/K_m is similar for rat and monkey but 6-fold lower for human.	See text (Section 3.6.6.2)	See text (Section 3.6.6.2)	Section 3.6.6.2; Conolly et al. (2000); Subramaniam et al. (2008); Klein et al. (2010)
Anatomically realistic representation of nasal passages.	Reduces uncertainty (over default calculation carried out by averaging dose over entire nasal surface).	Computer representation pertains to that of one individual (Caucasian male adult). There is considerable interindividual variability in nasal anatomy. Susceptible individuals are even more variable.	Kimbell et al. (2001a, b); Subramaniam et al. (2008, 1998)
KMU/μ_{Nbasal} is species invariant (used to estimate human).	Human cells are more difficult to transform than rodent, both spontaneously and by exposure to formaldehyde.	μ_{Nbasal} is 0 when concurrent controls or inhalation NTP controls in time frame of concurrent bioassays are used. Leads to infinitely large KMU for human.	Subramaniam et al. (2007); Crump et al. (2009, 2008).
Conservative assumptions were made. Results are conservative in the face of model uncertainties.	1) Hockey-stick dose-response for α_N was included even though TWA indicated J-shape. 2) Overall respiratory tract cancer incidence data for human baseline rates were used. 3) Risk was evaluated at statistical upper bound of the proportionality parameter relating DPXs to the probability of mutation.	Results in Conolly et al. (2004) are not conservative in the face of model uncertainties: (a) Human risk estimates are very sensitive to use of historical controls in the analysis of the animal bioassay. (b) Human risk estimates are unboundedly large when concurrent controls are used in rat model. (c) Minor perturbations in model assumptions regarding division and death rates of initiated cells lead to upper bound risks that were more than 1,000-fold greater than the highest estimates in Conolly et al. (2004).	Conolly et al. (2004); Subramaniam et al. (2007); Crump et al. (2009, 2008).

[a]Assumptions in this table are in addition to those listed for the BBDR model for the F344 rat.

F.2. SENSITIVITY ANALYSIS OF HUMAN BBDR MODELING

Crump et al. (2008) carried out a limited sensitivity analysis of the Conolly et al. (2004) human model. This analysis was limited to evaluating the effect on the human model of the following. These evaluations have been the subject of some debate in the literature and at various conferences (Conolly, 2009; Conolly et al., 2009, 2008; Crump et al., 2009).

1. The use of the alternative sets of control data for the rat bioassay data that were considered in the sensitivity analysis of the rat model in Appendix E.
2. Minor perturbations in model assumptions regarding the effect of formaldehyde on the division and death rates of initiated cells (α_I, β_I).
    - As mentioned in Section D.7 one (of the two) adjustable parameter in the expression for the human α_I in Conolly et al. (2004) was determined from the model fit to the rat tumor incidence data while the second parameter was determined from background rates of cancer incidence in the human. Therefore, variations considered in α_I were constrained to only those that (a) did not meaningfully degrade the fit of the model to the rat tumor incidence data and (b) were in concordance with background rates in the human.
    - Crump et al. (2008) also evaluated these variations with respect to their biological plausibility. The sensitivity analysis on assumed initiated cell kinetics was thought to be particularly important since there were no data to even crudely inform the kinetics of initiated cells for use in the models, even in rats, and the two-stage clonal expansion model is very sensitive to initiated cell kinetics (Gaylor and Zheng, 1996; Crump, 1994a, b).

Crump et al. (2008) note that, since the purpose of their analysis was to carry out a sensitivity analysis, in order to illustrate certain points, only risks to the general U.S. population from constant lifetime exposure to various levels of formaldehyde under the Conolly et al. (2004) environmental scenario (8 hours/day sleeping, 8 hours/day sitting, and 8 hours/day engaged in light activity) are considered. Fits based on the hockey-stick and J-shape models were identical, and, of the three estimated parameters (μbasal, multb, and D), only the estimate of μbasal differed between the two models.

F.2.1. Effect of Background Rates of Nasal Tumors in Rats on Human Risk Estimates

Crump et al. (2008) quantitatively evaluated the impact of different control groups on estimates of additional human risk as follows:

1. Concurrent controls plus all NTP controls:, the same as used by Conolly et al. (2004);
2. Concurrent controls plus controls from NTP inhalation studies;
3. Only concurrent controls;
4. Each set of control data was applied with both the J shape and hockey-stick models in Conolly et al. (2004) for α_N(flux) and α_I(flux) for a total of six analyses;.
5. Uncertainties associated with α_N or α_I are not addressed. Parameters α_{max}, multfc, and KMU were estimated in exactly the same manner as in Conolly et al. (2004).

Crump et al. (2008) present the following dose-response predictions of additional risk in humans from constant lifetime exposure to various levels of formaldehyde arising from exercising the above six cases. Their plots are reproduced in Figure F-1**,** where the corresponding curves based on Conolly et al. (2004) are also shown for comparison.

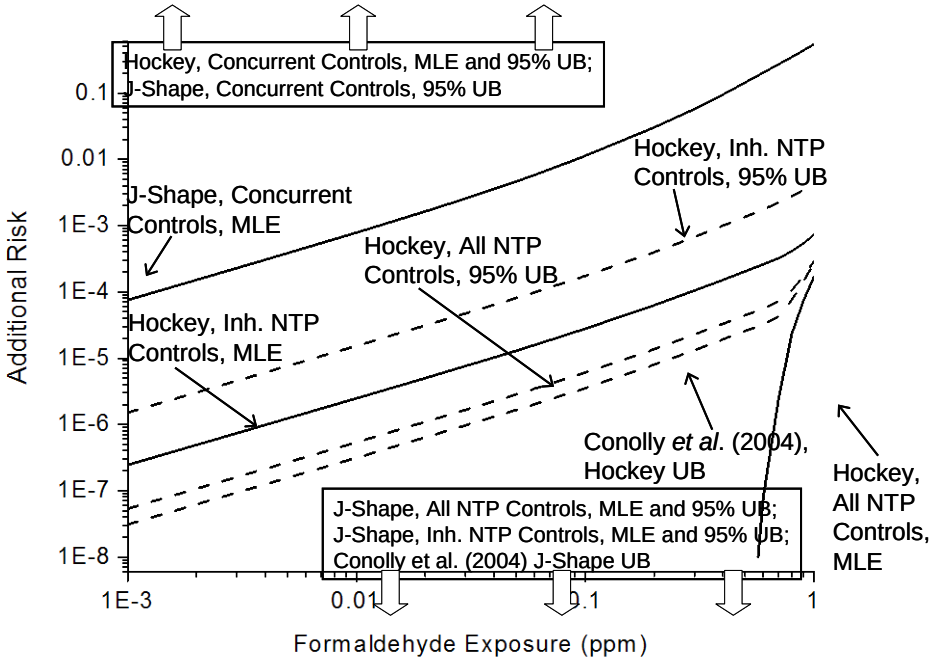

Figure F-1. Effect of choice of NTP bioassays for historical controls on human risk.

Note: Estimates of additional human risk of respiratory cancer by age 80 from lifetime exposure to formaldehyde are obtained by using different control groups of rats.

Source: Crump et al. (2008).

The lowest dotted curve in Figure F-1 represents the highest estimates of human risk developed by Conolly et al. (2004). This resulted from use of the hockey-stick model for cell division rates in conjunction with the statistical upper bound for the parameter KMU. As indicated by the downward block arrows in the figure, their corresponding estimates based on the J-shape model were all negative for exposures below 1 ppm.

Consider next the solid curves in the figure, which show predicted MLE added risks that were positive and less than 0.5. Crump et al. (2008) next examined the added risk obtained when the MLE estimate of (KMU/μ_{basal}) in these cases is replaced by the 95% upper bound of this parameter ratio. The upper bound risk estimates in Conolly et al. (2004) were calculated in a similar manner (but using all NTP historical controls). Except for minor differences, risk estimates corresponding to such an upper bound and using all NTP controls were very similar in the two efforts (Crump et al., 2008; Conolly et al., 2004).

Figure F-1 shows that the choice of controls to include in the rat model can make an enormous difference in estimates of additional human risk. For the J-shaped model for cell replication rate both estimates based on the MLE and those based on the 95% upper bound on KMU/μ_{basal} are negative for formaldehyde exposures below 1 ppm. However, when only concurrent controls are used in the model in Crump et al. (2008), the MLE from the J-shape model is positive and is more than three orders of magnitude higher than the highest estimates obtained by Conolly et al. (2004). Using only concurrent controls, estimates based on the 95% upper bound on KMU/μ_{basal} are unboundedly large (block arrows at the top of the figure). For the hockey-stick shaped model for cell replication rate, when all NTP controls are used, the estimates based on the MLEs are zero for exposures less than about 0.5 ppm. If only inhalation controls are added, the MLEs are about seven times larger than the Conolly et al. (2004) upper bound estimates, and the estimates based on the 95% upper bound on KMU/μ_{basal} are about 50 times larger than the Conolly et al. (2004) estimates. If only concurrent controls are used, both the MLE estimates and those based on the 95% upper bound on KMU/μ_{basal} are unboundedly large.

F.2.2. Alternative Assumptions Regarding the Rate of Replication of Initiated Cells

For the human model, Conolly et al. (2004) made the same assumptions for relating α_I(flux) and β_I(flux) to α_N(flux) as in their rat model (Conolly et al., 2003). That is, these quantities were related by using eqs D-2 and D-3 (see Appendix D). As discussed in the context of the rat modeling, by extending the shape of these curves to humans, the authors' model brings the cytotoxic action of formaldehyde to bear strongly on the parameterization of the human model as well.

In the sensitivity analyses of the rat modeling in Appendix E, it was concluded that other biologically plausible assumptions for α_I and β_I resulted in several orders of magnitude variations in the low dose risk relative to those obtained by models based on the assumptions in Conolly et al. (2003) but that the highest risks were nonetheless of the order of 10^{-6} at the 10 ppb level. This section examines how these uncertainties in the rat model propagate to the human model.

Crump et al. (2008) made minor modifications to the assumed division rates of initiated cells in Conolly et al. (2004), while all other aspects of the model and input data were kept unchanged. Two alternatives were considered for each of the J-shape and hockey-stick models. Figure F-2 shows the hockey-stick model for initiated cells in rats. In the first modification to the hockey-stick model (hockey-stick Mod 1), rather than having a threshold at a flux of 1,240 pmol/m^2-hour, the division rate increases linearly with increasing flux until the graph intersects the original curve at 4,500 pmol/m^2-hour, where it then assumes the same value as in the original curve for larger values of flux. The second modification (hockey-stick Mod 2) is similar, except the modified curve intersects the original curve at a flux of 3,000 pmol/m^2-hour.

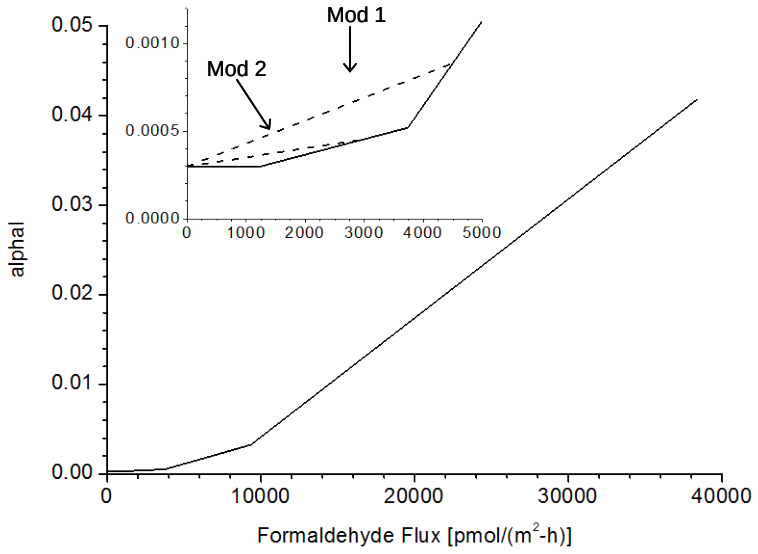

Figure F-2. Conolly et al. (2003) hockey-stick model for division rates of initiated cells in rats and two modified models.

Source: Crump et al. (2008).

Figure F-3 shows the rat J-shape model for initiated cells. In the first modification to this dose response (J-shape Mod 1), rather than having a J shape, the division rate of initiated cells remains constant at the basal value until the original curve rises above the basal value and has the same value as the original curve for larger values of flux. In the second modification (J-shape Mod 2), the J shape is retained but somewhat mitigated. In this modification, the division rate initially decreases in a linear manner similar to that of the original model but with a less negative slope until it intersects the original curve at a flux of 1,240 μm/m^2-hour, where it then follows the original curve for higher values of flux.

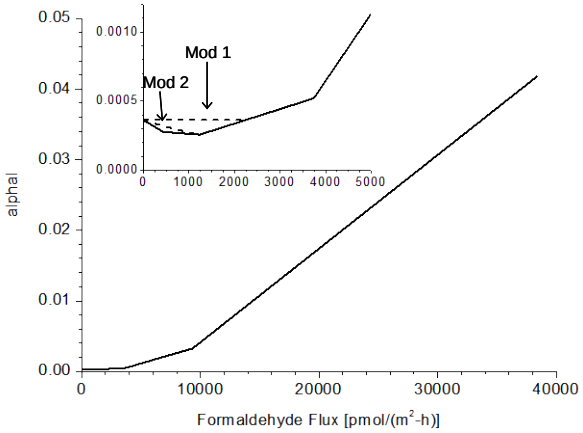

Figure F-3. Conolly et al. (2003) J-shape model for division rates of initiated cells in rats and two modified models.

Source: Crump et al. (2008).

Since the first constraint on the variation in α_I was in concordance with the rat time-to-tumor incidence data, Crump et al. (2008) applied each of the modified models in Figures F-2 and F-3 to the version of the formaldehyde models in Subramaniam et al. (2007) that employed all NTP controls and the hockey-stick curve for α_N. These authors restricted their analysis to this case since their stated purpose was only a sensitivity analysis as opposed to developing alternate credible risk estimates. Figure F-4 reproduces (from Crump et al. [2008]) curves of the cumulative probability of a rat dying from a nasal SCC by a given age for bioassay exposure groups of 6, 10, and 15 ppm. For comparison purposes, the corresponding KM (nonparametric) estimates of the probability of death from a nasal tumor are also shown. Three sets of probabilities are graphed: the original unmodified one and the ones obtained by using hockey-

stick Mod 1 and Mod 2. Crump et al. (2008) state that the changes in the tumor probability resulting from these modifications are so slight that the three models cannot be readily distinguished in this graph.[4] Thus, the modifications considered to the models for the division rates of initiated cells caused an inconsequential change in the fit of the model-predicted tumor incidence to the animal tumor data.

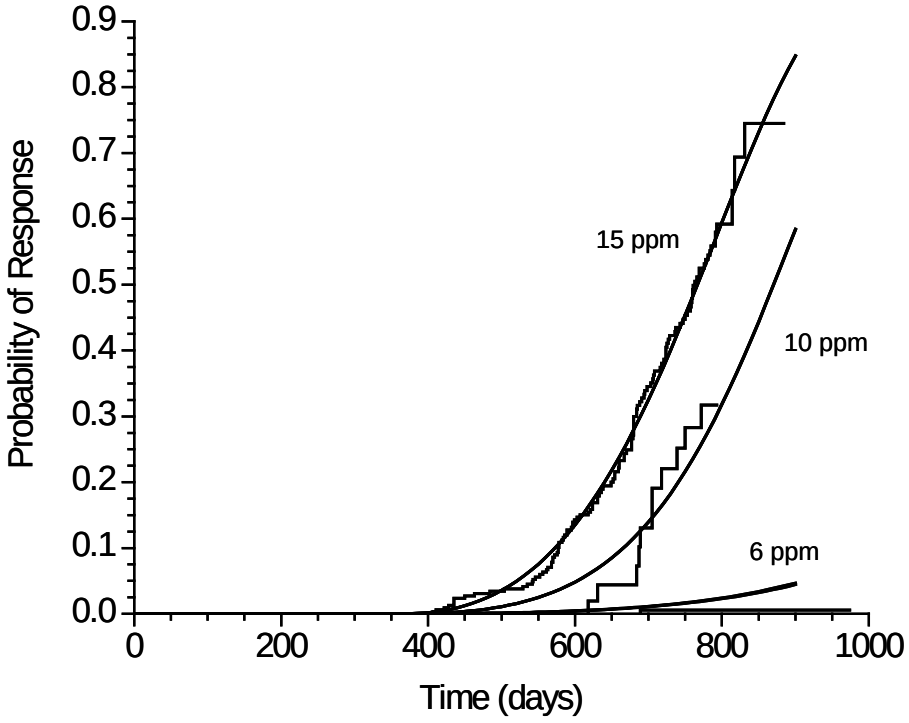

Figure F-4. Very similar model estimates of probability of fatal tumor in rats for three models in Figure F-2.

Note: The differences are visually indistinguishable. Models were derived from the implementation of Conolly et al. (2003) with the hockey-stick curves for α_I(flux) and α_N(flux) and variants derived from modifications (Mod 1 and Mod 2, Figure F-2) to α_I(flux). Model probabilities are compared to KM estimates. The three sets of model estimates are so similar that they cannot be distinguished on this graph.

Source: Crump et al. (2008).

[4] The largest change in the tumor probability resulting from this modification for any dose group and any age up through 900 days was found to be less than 0.002, a change so small that it would be impossible to detect, even in the largest bioassays ever conducted. The changes in tumor probability resulting from the other modifications described earlier were found to be even smaller. These comparisons were made in Crump et al. (2008) without re-optimizing the likelihood. The authors note that re-optimization of the model subsequent to the variations would have made the fit of modified models even better.

The above modifications did not affect the basal rate of cell division in the model and likewise had no effect on the fit to the human background data (Crump et al., 2008).

Crump et al. (2008) noted that, although the threshold model for initiated cells in Conolly et al. (2003) was replaced with a model that had a small positive slope at the origin, the resulting curves, hockey-stick Mod 1 and hockey-stick Mod 2, could have been shifted slightly to the right along the flux axis in order to introduce a threshold for α_I without materially affecting the risk estimates resulting from these modified curves. Thus, "the assumption of a linear no-threshold response is not an essential feature of the modifications to the hockey-stick model; clearly threshold models exist that would produce essentially the same effect" (Crump et al. 2008).

F.2.3. Biological Plausibility of Alternate Assumptions

These very small variations made to the α_I in Conolly et al. (2003) are seen to be

- consistent with the tumor-incidence data (see Figure F-4);
- small compared with the variability and uncertainty in the cell replication rates characterized from the available empirical data (at the formaldehyde flux where α_I was varied);
- supported (qualitatively) by limited data, suggesting increased cell proliferation at doses below cytotoxic;
- perturbations that one should expect on any dose response derived from laboratory animal data because of human population variability in cell replication;
- and biologically plausible because cell cycle control in initiated cells is likely to be disrupted.

The averaged cell replication rate constants as tabulated in Table 1 of Conolly et al. (2003) and shown by the red curve in Figure E-2 of Appendix E (for various exposure concentrations and corresponding average formaldehyde flux values in the F344 rat nose) demonstrate an increase over baseline values only at exposure concentrations of 6 ppm and higher. Increased cell proliferation at these concentrations of formaldehyde, whether transient or sustained, have been associated in the literature with epithelial response to the cytotoxic properties of formaldehyde (Conolly, 2002; Monticello and Morgan, 1997; Monticello et al., 1996, 1991). The labeling data are considered to show a lack of cytotoxicity and regenerative cell proliferation in the F344 rat at exposures of 2 ppm and below (Conolly, 2002). In the Conolly et al. (2003) modeling, it is further assumed that the formaldehyde flux levels at which cell replication exceeds baseline rates remain essentially unchanged when extrapolated to the human and for initiated cells for the rat as well as the human. These assumptions need to be first

viewed in the context of the uncertainty and variability in the data on normal cells discussed in Appendix E.

Arguments for a hockey-stick or J shape over the background have been made in the literature for sustained and chronic cell replication rates. However, the analyses of the cell replication data show that the data are not consistently (over each site and time) indicative of a hockey-stick or J shape as the best representation of the data (see Appendix E). This uncertainty is particularly prominent when examining the cell replication data at the 13-week exposure time and the pooled data from the PLM nasal site from Monticello et al. (1996) (see Figures E-1 [dotted curve], E-3B, and E-4 of Appendix E). The earliest exposure time in this experiment was at 13 weeks, and the 13-week cell replication data appear to be more representative of a monotonic increasing dose response without a threshold; it is possible that early times are of more relevance to the carcinogenesis as well as for considering typical (frequent short duration) human exposures.

Recently, Meng et al. (2010) measured cell replication in the anterior lateral meatus of the F344 rat using continuous labeling on rats exposed to all the concentration levels in the Monticello et al. (1996) experiment. Labeling index (i.e., LI, as opposed to ULLI in the Monticello experiment) was measured as the percentage of BrdU-labeled cells among the total number of cells counted at the nasal site. Their data are reproduced below in Figure F-5, where the asterick denotes the observation of a statistically significant difference from the control group (Dunnett's test, $p < 0.01$). These data appear to be consistent with a monotonically increasing dose-response shape for cell replication. Linear regression provided good fits to all of the data ($R^2 = 0.97$) as well as to the subset of the data obtained by deleting the higher dose data at 10 ppm and 15 ppm exposures ($R^2 = 0.84$). We cite these data in support of considering the modifications carried out in Figure F-2.

For initiated cells, there are no data on which to evaluate the modifications made in Section F.2.2 to these rates. However, some perspective can be gained by comparing them to the variability in the division rates obtained from the data on normal cells used to construct the formaldehyde model. As shown in Figure E-2 and discussed further in Subramaniam et al. (2008), these data show roughly an order of magnitude variation in the cell replication rate at a given flux. As part of a statistical evaluation of these data, a standard deviation of 0.32 was calculated for the log-transforms of individual measurements of division rates of normal cells (Crump et al., 2008). By comparison, the maximum change in the log-transform division rate of initiated cells resulting from hockey-stick Mod 2 was only 0.20, and the average change would be considerably smaller. Thus, although there are no data for initiated cells, it can be said that

the modifications introduced in Crump et al. (2008) for initiated cells are extremely small in comparison to the dispersion in the data for normal cells.

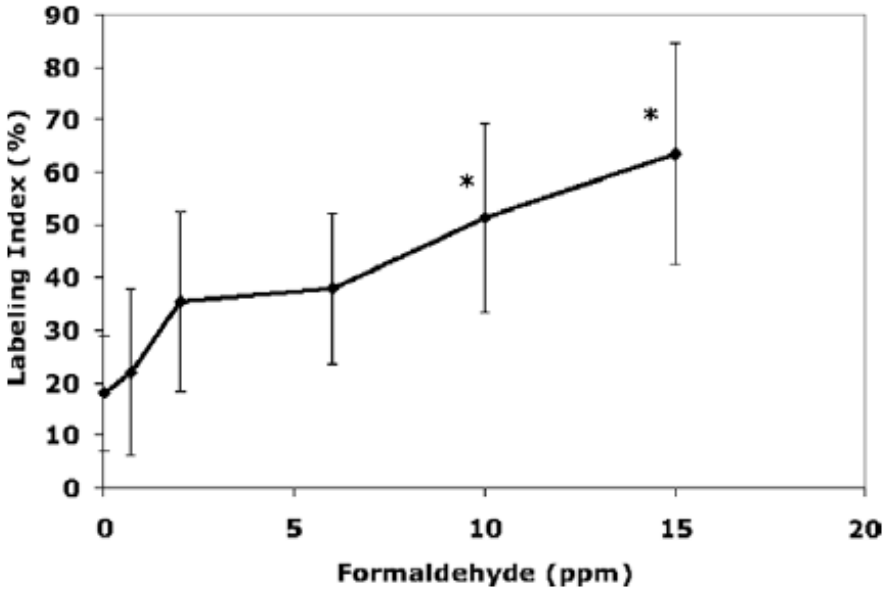

Figure F-5. Cell proliferation data from Meng et al. (2010). The Y-axis shows the percentage of BrdU-labeled cells among the total number of cells counted in the ALM section of the rat nose.

Reproduced with permission from Meng et al. (2010).

Subramaniam et al. (2008) also point to some additional, albeit limited, data, suggesting that exposure to formaldehyde could result in increased cell replication at doses far below those that are considered to be cytotoxic. Tyihak et al. (2001) treated different human cell lines in culture to various doses (0.1–10 mM) of formaldehyde and found that the mitotic index increased at the lowest dose of 0.1 mM. These findings considered along with human population variability and susceptibility (for example, polymorphisms in ADH3 [Hedberg et al., 2001]) indicate that it is necessary to consider the possibility of small increases in the human α_I over baseline levels at exposures well below those at which cytotoxicity-driven proliferative response is thought to occur.

Heck and Casanova (1999) have provided arguments to explain that the formation of DPXs by formaldehyde leads to inhibition of cell replication (i.e., if this effect alone is considered, normal cell replication rate of the exposed cells would be less than the baseline rate). However, this hypothesis was posed for normal cells. Subramaniam et al. (2008) argue that if an initiated cell is created by a specific mutation that impairs cell cycle control, the effect would be

to mitigate the DPX-induced inhibition in cell replication, either partially or fully, depending on the extent to which the cell cycle control has been disrupted. In the absence of data on initiated cells, the above argument provided biological motivation to the modification applied to the J-shape model for cell division (Crump et al. 2008).

Thus, the previous paragraphs suggest that the changes made in the analysis in Crump et al. (2008) to the assumption by Conolly et al. (2003) regarding the dose response for the division rate of initiated cells are plausible.

F.2.4. Effect of Alternate Assumptions for Initiated Cell Kinetics on Human Risk Estimates

Figure F-6 contains graphs of the additional human risks estimated (in Crump et al. [2008]) by applying these modified models for α_I and using all NTP controls, compared with those obtained by using the original Conolly et al. (2004) model. Each of the four modified models presents a very different picture from that of Conolly et al. (2004). At low exposures, these risks are three to four orders of magnitude larger than the largest estimates obtained by Conolly et al. (2004).

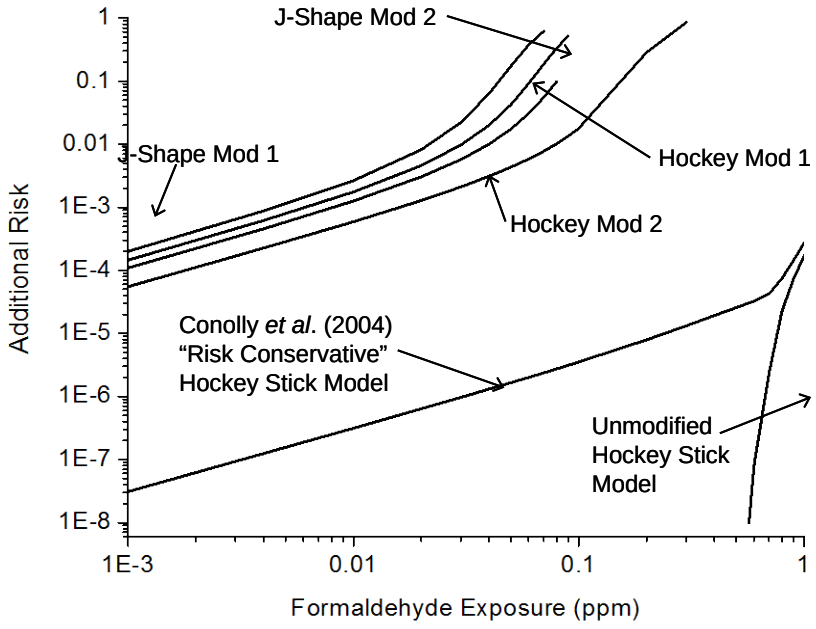

Figure F-6. Graphs of the additional human risks estimated by applying these modified models for α_I, using all NTP controls, compared to those obtained using the original Conolly et al. (2004) model.
Source: Crump et al. (2008).

These results have been criticized by Conolly et al. (2009) as being unrealistically large and above the realm of any epidemiologic estimate for formaldehyde SCC. Thus, they argue that the parameter adjustments made in Crump et al. (2008) are inappropriate. Crump et al. (2009) rebutted these points by arguing that the purpose of their work was not to provide a more reliable or plausible model but to carry out a sensitivity analysis. They argued that the changes made to the model (in their analyses) were reasonable since they did not violate any biological constraints or the available data. Further, they pointed out that "by appropriately mitigating the small modifications [they] made to the division rates of initiated cells, the model [would] provide any desired risk ranging from that estimated by the original model up to risks 1,000-fold larger than the conservative estimate in Conolly et al. (2004)."

Crump et al. (2008) also evaluated the assumption in eq D-3 of the CIIT modeling pertaining to initiated cell death rates (β_I) by making small changes to β_I. They report that they obtained similarly large values for estimates of additional human risk at low exposures. Obtaining reliable data on cell death rates in the nasal epithelium appears to be an unusually difficult proposition (Hester et al., 2003; Monticello and Morgan, 1997), and, even if data are obtained, they are likely to be extremely variable.

Appendix G

1

APPENDIX G

EVALUATION OF THE CANCER DOSE-RESPONSE MODELING OF GENOMIC DATA FOR FORMALDEHYDE RISK ASSESSMENT

G.1. MAJOR CONCLUSIONS IN ANDERSEN ET AL. (2008)

In Chapter 4, the gene microarray data from animal studies on formaldehyde (Andersen et al., 2008; Thomas et al., 2007) were described. The analysis of these animal high throughput data and the conclusions reached in these two groundbreaking papers were closely examined for use in this assessment. Studies on high throughput animal data provide a wealth of information that helps further understanding of the relevant mechanisms. However, such studies have generally not made quantitative bottom-line inferences that inform low dose human risk. The above-mentioned studies are a notable exception due to the breadth of their conclusions on low dose MOAs, their pioneering application of the benchmark dose (BMD) methodology to genomic data, their use of BMD-response analysis that identified dose estimates at which specific cellular processes were significantly altered, and the fact that they were accompanied by recommendation in the literature urging use of these results in setting exposure standards for formaldehyde (Daston, 2008).

We focus here on the conclusions in these papers with regard to modeling the cancer dose-response for formaldehyde. In addition to supporting our disposition of these analyses for this assessment, this write-up serves the purpose of exemplifying critical issues that need to be considered for the future.

The overall BMD determined in Andersen et al. (2008) for all genes with significant dose-response averaged 6.4 ppm. These analyses indicated a general progression with the lowest BMD values (i.e., the most sensitive epithelial responses) for extracellular and cell membrane components and higher BMD values for intracellular processes. Overall, these authors concluded that

- Genomic changes, including those suggestive of mutagenic effects, did not temporally precede or occur at lower doses than phenotypic changes in the tissue
- Genomic changes were no more sensitive than tissue responses
- Formaldehyde, being an endogenous chemical, is well handled until some threshold is achieved. Above these doses, toxicity rapidly ensues with concomitant genomic and histologic changes.
- Linear extrapolations, or extrapolations that specify similar MOAs at high and low doses would be inappropriate.

These findings were judged to have significant implications on the debated MOA for formaldehyde carcinogenicity, confirming results from earlier bioassays and dose-response modeling that the mutagenicity of formaldehyde was too weak to be of relevance to its carcinogenicity. Daston (2008) judged the method in these efforts to be extremely sensitive and therefore suited to examining whether responses at the molecular level take place at doses below which frank adverse effects occur. Daston (2008) argued that "… if there are pleiotropic effects at lower exposure levels that would elicit a different profile of gene expression, those genes would not go unnoticed" and thus concluded that "the gene expression data confirm that the responses are not linear at low doses."

In the analyses that follow, we point to some significant quantitative factors that impact on these conclusions.

G.2. USE OF MULTIPLE FILTERS ON THE DATA

The analyses in these papers involved the following sequence of data filters.

1. Gene probe sets that differed in expression in response to treatment were identified by one-way analysis of variance. Probability values were adjusted for multiple comparisons by using a false discovery rate of 5%.
2. Next, in addition to the above statistical filter, the output was further screened by selecting only those genes that exhibited a change from the control group that was greater than or equal to 1.5-fold (logarithmic).
3. The gene probe sets that demonstrated significant dose-response behavior were then matched to their corresponding biological process and molecular function gene ontology (GO) categories (considering only those involving more than three genes) and grouped into process categories such as cell division, DNA repair, cellular proliferation, apoptosis, and related molecular function categories.

A large number of genes are expressed in these studies; therefore, clearly some appropriate filter needs to be used for meaningful interpretation of the vast database. Tissue pathology served as a phenotypic anchor for the interpretation of microarray results, and the genomic study confirmed (and improved on) the qualitative and quantitative understanding derived from the histopathology and observation of frank effects. It is possible that the combination of filters used by these authors is adequate for an inquiry into some mechanisms associated with the specific phenotypic effects. However, the studies reached bottom-line conclusions with regard to the low-dose MOA and approach to be considered for quantitative extrapolation. These conclusions necessarily involve questions as to whether there were gene

expression changes at low dose and at early exposure times that may be relevant to initiating carcinogenesis and finally as to whether there is a threshold in dose associated with formaldehyde carcinogenesis. However, collectively, the three filters employed in these studies likely constitute overly stringent criteria, taking away the resolution needed to observe critical gene changes needed to delineate low dose effects. An indication that this may indeed be the case can be seen by examining the correlations in their findings with the observed trend in the data on DPXs formed by formaldehyde. This is detailed in the following section.

G.3. DATA FOR LOW-DOSE CANCER RESPONSE

A significant finding in Thomas et al. (2007) is that BMD estimates for the GO categories applicable to cell proliferation and DNA damage were similar to values obtained for cell labeling indices and DPXs in earlier studies and to BMD estimates obtained for the onset of nasal tumors. The mean BMD for the GO category of "positive regulation of cell proliferation" was 5.7 ppm; in comparison, Schlosser et al. (2003) obtained a 10% BMD of 4.9 ppm for the cell labeling index. The GO category associated with "response to DNA damage stimulus," seen as a genomic correlate to a mutagenic effect, had a mean BMD of 6.31 ppm. Thomas et al. (2007) compare this finding with significant increase at 6 ppm of DPXs following a 3-hour exposure in the study by Casanova et al. (1994). The formation and repair of DPXs have been considered to be one of the potential mechanisms associated with the genotoxic action of formaldehyde (Conolly et al., 2003, 2000). Based on earlier work in the same laboratory (Conolly et al., 2004, 2003; Conolly, 2002), Slikker et al. (2004) concluded that there is a dose threshold (at about 6 ppm) to formaldehyde carcinogenicity and that the putative mutagenic action of formaldehyde is not relevant to its carcinogenicity. Therefore, the finding that a significant genomic response (e.g., induction of DNA repair genes) is not observed at doses lower than those that induce tumors in rodent bioassays is seen by these authors (Andersen et al., 2008; Daston, 2008; Thomas et al., 2007) to further buttress the above conclusions related to the mode of action for formaldehyde-induced respiratory cancer.

However, phenotypic anchoring to the DPX data drawn only from Casanova et al. (1994) misses critical low-dose data that informs mode of action. In an earlier study, Casanova et al. (1989) observed statistically significantly elevated (over controls) levels of DPXs at 2 ppm and a trend towards elevated DPXs at 0.7 ppm. In analysis of low-dose data, the trend in the dose-response is critically important because data inherently lack the power to establish statistical significance. Furthermore, the two studies by Casanova and coworkers are different in some respects. The earlier study was a 6-hour exposure, while the later study was a 3-hour study; thus, on this account alone, it appears more relevant to compare with the older study. Exposures in

the earlier study were additionally at 0.3 and 10 ppm, thus affording a lower exposure concentration. In the earlier study, tissue from the whole nose was analyzed, whereas in the later study tissue from two specific regions was obtained from the "high" tumor (Level II) and "low" tumor regions. Together, these data suggest that DPXs occur at exposure concentrations considerably lower than those that elicited transcriptional changes. One possible explanation is that the increase in DPXs was not sufficient to induce DNA repair genes. Alternatively, these discrepancies may be due to the stringent filters and the low statistical power of the Andersen et al. (2008) study. These disparities between the gene array study and the DPXs question the ability of the studies in Andersen et al. (2008) and Thomas et al. (2007) to inform the presence or absence of a mutational MOA for formaldehyde, and in essence, to inform the low-dose response curve for formaldehyde-induced cancer.

In another instance, Andersen et al. (2008) clearly stated that no genes were significantly altered by exposure to 0.7 ppm, yet they state that there was "a trend toward altered expression at 0.7 ppm" in some genes with U and inverted U shape dose-responses (Figures 4 and 5 of their paper). While these changes may not be statistically significant, they could be biologically significant.

G.4. DIFFICULTIES IN INTERPRETING THE BENCHMARK MODELING

The benchmark analyses are summarized in Thomas et al. (2007) as average BMD estimates for genes in a given GO that were statistically significantly dose related. The benchmark modeling was then used by the authors to identify that the dose below individual cellular processes was judged to be "not altered."

The BMD definition used by these authors is quite stringent: it defines an effect so that only 0.005 of controls will be considered affected and sets the BMR corresponding to this dose at 0.105. The net effect is that the BMD is the air level, such that the increase in the mean response is 1.349 × standard deviation. This is essentially an arbitrary definition. For comparison, if 0.05 of controls are considered affected and the BMR is set at 0.1 (common values that are applied to whole animal data), the BMD is the air level such that the increase in the mean response is 0.608 × standard deviation. Thus, if this definition had been used (as is traditionally the case), the BMD estimates would all be 2.2 times smaller than those obtained by Schlosser et al. (2003). Furthermore, the analysis assumes equal variance in all dose groups. Thus, further consideration of these issues with regard to interpretation of the BMR obtained from these studies is needed before it can be used in regulatory exposure setting. Secondly, lower confidence limits on the BMDs need to be derived for the data in Andersen et al. (2008).

G.5. STATISTICAL SENSITIVITY OF THE DATA FOR DOSE-RESPONSE

Another cautionary note pertains to the qualification of gene array studies as being extremely sensitive. Such a qualification should actually refer to the fact that only tiny amounts of mRNA are needed, that is, the sensitivity of the assay per se for measuring gene expression. However, this should not be confused with the sensitivity needed to identify the very small dose-related changes at low dose. Andersen et al. (2008) reports on results of studies that involve small numbers of animals in each dose group (five or eight). Despite the limited power in such studies, the paper equates the absence of a statistically significant effect with no effect. This limitation is generally true of studies of the dose responses of changes in gene expression conducted to date; they have generally relied on very few animals (≤10 per dose group). Since there will likely always be background amounts of gene expression, quantifying the dose response requires statistically significant changes in gene expression as a function of dose. If the genomic data involve even fewer animals per group than the histopathological data, they have even less power to delineate the dose response; in particular, whether there is a threshold at low exposures. This is illustrated by the example in Figure G-1 of the dose responses for epithelial hyperplasia. The data in this figure are from lesion 2 in Andersen et al. (2008); the linear regressions and confidence limits were determined by EPA These appear equally consistent with both a threshold at around 1 ppm and a linear response down to zero.

G.6. LENGTH OF THE STUDY AND STOCHASTIC EVENTS

Another significant consideration with regard to MOA conclusions that are pertinent to the disease process is the length of the study, 15 days. If formaldehyde-induced tumor formation is a stochastic process (e.g., genotoxicity), then exposure of a small number of animals to low concentrations for 15 days may not be long enough to detect changes that might occur under long-term exposure scenarios.

Relatedly, it has been suggested that gene (and protein) expression is a stochastic process whereby steady state gene expression obeys Poisson statistics (i.e., distribution of rare events), and that events of interest may occur in a single cell or small number of cells in which larger tissue samples can average out such stochastic events and prevent the detection of nonaverage behavior (Quakenbush, 2007). Given the implied difficulty in such an analysis, duration of exposure may be one of the most tenable ways of addressing whether a chemical increases the probability of an adverse response.

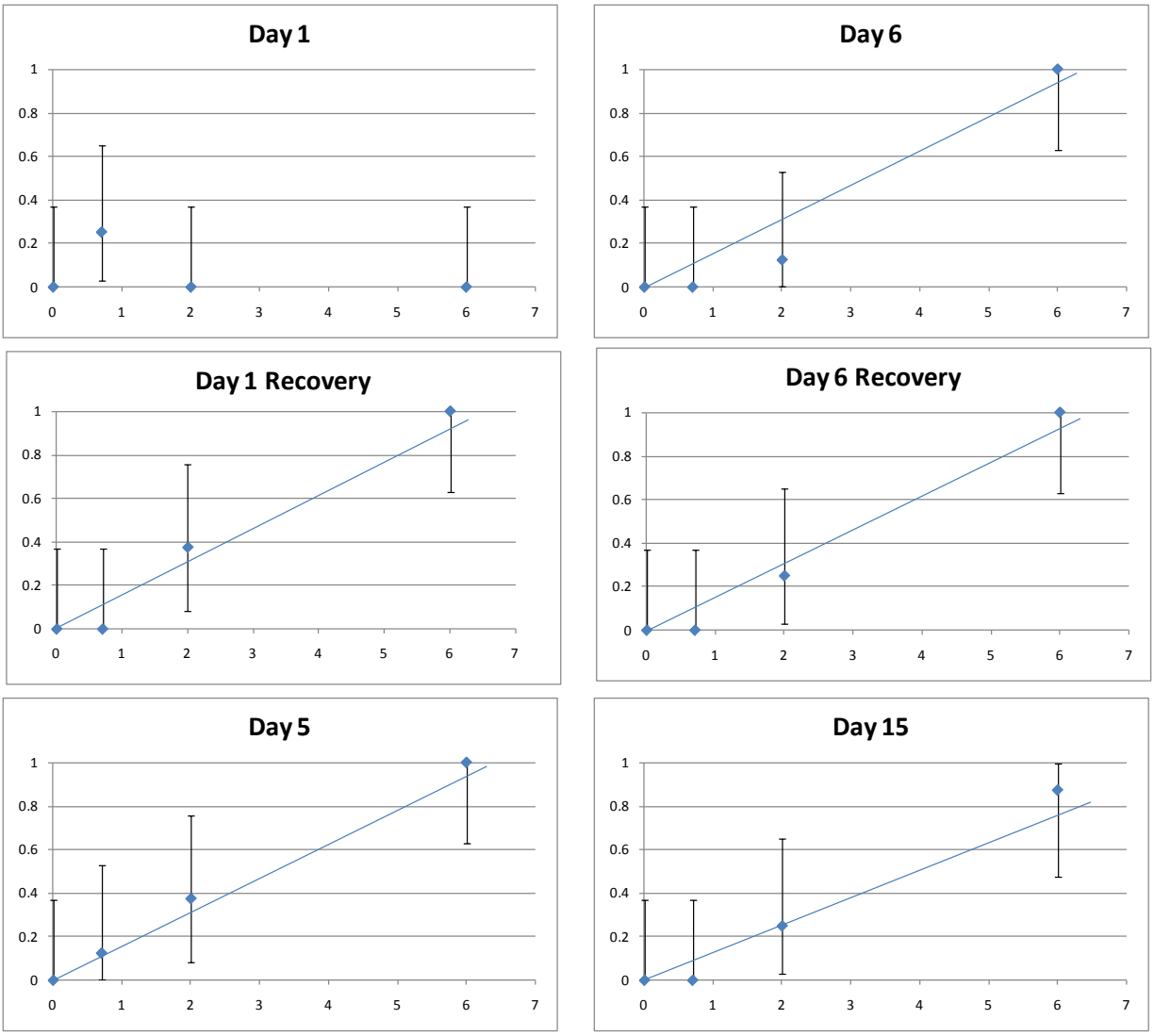

Figure G-1. Graphs of epithelial hyperplasia (Lesion 2) versus formaldehyde concentration (ppm) with 95% confidence intervals (with linear fit by eye).

Source: Fit to data from Andersen et al. (2008).

G.7. OVERALL CONCLUSION

We believe our analyses of the presentations in Andersen et al. (2008) and Daston (2008) are generally useful with regard to future developments in quantitative analyses of genomic data if they are to be of relevance to risk assessment. For risk assessment, rather than focusing on what responses are statistically significant, an analysis should focus on (1) what range of values of critical parameters (e.g., gene expression) are consistent with the data, and (2) what these values imply for whole animal risk. This is of course, an extremely difficult proposition because

we do not know nearly enough about how changes in genes quantitatively affect whole animal risk, or even which genes are important.

Appendix H

1

APPENDIX H

EXPERT PANEL CONSULTATION ON QUANTITATIVE EVALUATION OF ANIMAL TOXICOLOGY DATA FOR ANALYZING CANCER RISK DUE TO INHALED FORMALDEHYDE

The National Center for Environmental Assessment convened an expert panel of scientists for advice on evaluating available approaches for incorporating biological information in analyzing animal tumor data for assessing cancer risk due to inhaled formaldehyde. This Appendix pertains to the major deliberations and results of that meeting and is divided into three sections.

A. Scope and Agenda of Meeting on Quantitative Evaluation of Animal Toxicology Data for Analyzing Cancer Risk due to Inhaled Formaldehyde. October 28 & 29, 2004.

B. Summary of Consultative Meeting on CIIT Formaldehyde Model. October 28 & 29, 2004.

C. Meeting Report from Dr. Rory B. Conolly

A. Scope and Agenda of Meeting on Quantitative Evaluation of Animal Toxicology Data for Analyzing Cancer Risk due to Inhaled Formaldehyde

October 28 & 29, 2004. Washington, DC.

This meeting is to assist EPA in evaluating available approaches for incorporating biological information in analyzing animal tumor data for assessing cancer risk due to inhaled formaldehyde. The CIIT Centers for Health Research (CIIT) has published a novel risk assessment that links site-specific predictions of flux using computational fluid dynamics (CFD) modeling with a two-stage clonal growth model of cancer to analyze nasal tumor incidence in two rodent bioassays. The rodent models are used with corresponding human models for low-dose extrapolation of cancer risk to people.

Key predictions of the CIIT effort are a zero maximum likelihood estimate of the probability of formaldehyde-induced mutation per cell generation in the rat and a *de minimus* additional lifetime risk in nonsmokers due to continuous environmental exposure below 0.2 ppm. The National Center for Environmental Assessment is carrying out sensitivity analyses and examining variations of the CIIT model in order to understand the implications of the model structure and parameters on model predictions. In this meeting, we wish to focus on the strengths and key uncertainties of this model, the extent to which assumptions in the CIIT model are supported by biological data, and examine the impact of uncertainty and variability on the overall quantitative risk characterization.

Broadly, the discussions will focus on the following areas:

- Impact of uncertainties in dosimetry on human risk estimates
- Uncertainties in the use of experimental data on labeling index
- The model structure related to initiated cells and DNA protein cross-links
- Considerations of time-to-tumor in the clonal growth modeling
- Inferences and information on the role of mutation and cytotoxicity in estimating human risk
- Relative merits of benchmark dose modeling vs. the 2-stage clonal growth model

Discussions on Mode of Action are expected to be an integral part of several of the sessions. Therefore a specific time-slot is not set aside for this purpose.

The meeting will have a panel discussion format. There will be no formal presentations unless necessary to elucidate an issue. Various attachments referred to in the Agenda below, as well as the relevant manuscripts will be sent separately.

Specifically, we suggest the following issues upon which to focus the discussion in the above areas, and approximate time frames and discussion leads, although discussants should feel free to bring up other critical issues.

I. Introduction and purpose of discussion

Peter Preuss.

9:00 AM, Oct 28

II. Impact of uncertainties in dosimetry on risk estimates

Lead discussant: Linda Hanna

9:15 - 11 AM Oct 28

Boundary conditions

The CFD modeling specified a mass transfer coefficient as a boundary condition on the nasal lining, adjusting the value of this coefficient on the "absorbing" portion of the lining so as to match simulated overall uptake in the rat nose to the experimentally determined average overall uptake. This value was then used for the corresponding human nasal lining. Are these boundary conditions appropriate surrogates for the underlying pharmacokinetics, including saturation in metabolism and mucociliary clearance, particularly with reference to humans?

Turbulence

Turbulent flow has been seen to occur in experimental models of the human nose at some of the higher flow rates at which the CFD models were used in CIIT's assessment. It is not likely that the CIIT CFD model can reliably identify signatures of transition to turbulent behavior. Turbulent flow can significantly alter regional uptake patterns. Additionally, significant mass balance errors were seen at the higher flow rates in the human flow models. Discuss if these are likely to impact significantly on risk estimates.

Interindividual variability

The CIIT assessment has focused on the nasal anatomy of a single individual. Discuss the implications of interindividual variations in nasal anatomy on the population distribution in risk.

III. Uncertainties in the use of experimental data on labeling index

Lead discussant: George Lucier

11AM – 11:45 AM, 1:00 - 3:15 PM Oct 28

Cell-replication rate and its relationship to flux is a critical determinant of risk. Therefore uncertainties and variability in measurement of the unit length labeling index and its use in the CIIT clonal growth modeling need to be characterized.

1. Discuss the strengths, uncertainties and limitations associated with estimating cell replication rates from the unit length labeling index (ULLI).
    a. For example, a constant ratio of the measured ULLI to the labeling index (LI) that is used in the model is assumed. Is it valid to assume this ratio to be constant across nasal sites, dose and exposure time.
    b. How uncertain is this ratio?

2. Considering the large patterns of variability in the ULLI data, discuss the validity of using ULLI averaged over site and exposure times.
    a. The averaging loses information on the sequential effect of change with time, and on significant differences among sites.
    b. How sensitive is the clonal growth modeling result to these variations in the dose-response function for cell replication rates vs. flux to the tissue? A discussion of this question in this session is intended to serve as input to later deliberations on the issue.

3. Discuss the validity of combining data collected in different experiments using different labeling methods, and the validity of estimating cell replication rates from LI or ULLI measured in a single pulse labeling experiment.

    See attachment C: "ULLI Dose-Response Modeling and Statistical Analysis" for a discussion of these issues, and Moolgavkar and Luebeck (1992).

IV. **Model Structure: Birth and death rates for Initiated cells, Role of DPX**

Lead discussant: Kenny Crump

3:30 - 6:00 PM Oct 28.

Parameters for initiated cells

1. The CIIT analysis of ULLI data allows for a virtual threshold in dose in the replication rate of normal cells. Discuss the validity of ascribing such a behavior to initiated cells considering the sensitivity of 2-stage model results to the initiated cell replication rates.

2. Discuss the treatment of death rate for initiated cells in the model (set equal to birth rate of normal cells in Conolly et al., 2003) and implications for confidence in model predictions.

    Also see Attachment A (memo from Rory Conolly) and Attachment D (EPA discussion of CIIT clonal growth modeling and some sensitivity analyses. . .)

Treatment of DNA protein cross-links (DPX) in clonal expansion model

3. Formaldehyde-induced mutation is modeled as taking place only while DPX are in place with DPX undergoing rapid repair. Discuss the possibility of persistent genetic damage that extends beyond the DPX half-life and enhances mutation. How might this issue be included in the model structure?

V. Considerations of time-to-tumor in the CIIT clonal growth modeling

Lead discussant: Christopher Portier

8:30 – 11:00 AM, Oct 29.

1. A number of issues affect likelihood values and the model fit to the time-to-tumor data. Discuss assumptions in the treatment of time-to-tumor in the CIIT clonal expansion model, and their impact on parameter estimates. For example,
    a. Results in Conolly et al. (2003, 2004) are derived considering all tumors to be fatal. Note in this context that serially sacrificed animals have been combined with those experiencing mortality—the effect of this is visible as irregularities in the time-to-tumor curve.
    b. How is the time variability in ULLI likely to impact on the time-to-tumor predictions?

2. Long delay times are predicted by the model for observation of detectable tumor. Is this compatible with the assumption of rapidly fatal tumors?

3. Discuss the weight to be given to differences in likelihood when comparing with variations on the Conolly et al. (2003) model structure such as in Attachment A or D.

VI. Inferences on the role of formaldehyde-induced mutation and cell proliferation

Lead discussant: Dale Hattis

11:15 – 12:00 PM, 1:00 – 4:00 PM, Oct 29.

1. The model structure in Conolly et al. (2003) predicts a zero maximum likelihood estimate for the constant of proportionality (KMU) linking DPX to the probability of formaldehyde-induced mutation per cell generation. Examine the strength of this conclusion, and the extent to which an insignificant probability of formaldehyde-induced mutation per cell generation is supported by data.

2. Discuss the biological relevance and validity of model-estimated parameters, particularly in the context of low-dose predictions.
    a. Discuss possible avenues to validate CIIT cancer model predictions.

3. Discuss the validity of using cell replication rates determined for the rat to predict human risk in a population.

4. In the face of uncertainties, are the results in Conolly et al. (2003, 2004) conservative in the sense of overpredicting risk?
    a. Discuss the extent to which sensitivity analyses have addressed this issue and the extent to which sensitivity analyses can speak to the strength of the model. *[See Attachments A: Memo from Conolly, and D: EPA discussion of CIIT clonal growth modeling and some sensitivity analyses….]*

VII. Benchmark Dose Modeling

Lead discussant: Kenny Crump
4:15 – 5:30 PM, Oct 29.

Discuss the relative merits of using a benchmark dose approach that incorporates biological modeling (such as estimating flux to tissue or DPX levels) as compared with the CIIT 2-stage model for cancer. (*See attachment E and Schlosser et al., 2003.)*

B. Summary of Consultative Meeting on CIIT Formaldehyde Model October 28 & 29 2004, NCEA, Washington, DC

Date: November 10, 2004

Ravi P. Subramaniam, Ph.D.
Quantitative Risk Methods Group
National Center for Environmental Assessment, ORD, US EPA

This is a broad summary of the most important issues at the formaldehyde meeting.

It was generally felt by consultants that the broad framework of the approach adopted by CIIT, namely the use of a two-stage model for cancer, the linking of localized flux to cell replication rates and DPX concentration, and the expression of formaldehyde-induced mutation as a linear function of DPX, was reasonable.

Potential errors in the dosimetry modeling were seen not to have a significant effect on risk estimates. The boundary conditions used were discussed to be a reasonable representation of the pharmacokinetics for both rats and humans. The discussion on the impact of interindividual variability of nasal anatomy was not particularly conclusive. It was determined that there was likely to be much less variability in reactive gas uptake than that seen in particulates.

Crucial errors were however identified on several fronts in the manner in which the clonal growth model had been implemented in the CIIT effort. Dr. Portier felt that the calculation of probability was seriously flawed on account of lumping serially-sacrificed animals and animals that died of tumor together, while at the same time assuming rapid fatality of all tumors. This was seen to significantly alter the calculation of tumor probability (the shape of the dose-response curve), and his insight was that a correction was likely to allow for a substantially higher value for the probability of formaldehyde-induced mutation at low-dose. The best estimate for this probability is now zero in the model. Drs. Crump, Portier and Hattis argued that replacing this estimate by an upper confidence bound on KMU (the coefficient determining the role of DPX in the probability of mutation per cell generation), keeping other structural problems in the model unexplored, or other parameters fixed, would not be enough. There was a discussion on the need to provide confidence bounds on risk determined by allowing all the parameters to vary. Drs. Crump and Hattis (and Portier?) felt such an estimate would be very different from that calculated based on individual parameters.

Drs. Crump, Hattis and Portier urged us not to be constrained by the optimal likelihood values of a single plausible model, and underscored the need to explore a variety of biologically reasonable model structures as a requisite for utilizing such a model in risk assessment.

Likelihood was seen to be an inadequate expression of what is to be considered an optimal model (okay only for comparing models that were nested, etc.). These models should allow the expression of variability and uncertainty in the data, as well as in underlying assumptions in model specification. Dr. Crump (and Hattis also?) felt that alternate model structures, if explored, could potentially lead to risk estimates, for the range below the observed data, that were higher by several thousands.

Dr. Crump cautioned that extrapolating to human using the hockey or J-shaped cell replication curve used in the rodent carried with it a large uncertainty that had not been characterized in the Conolly modeling.

Dr. Portier expressed concern over the manner in which historical and concurrent controls were lumped together. The thrust of Portier's comments was that such a combination of controls was generally not done. The large number of historical controls was likely to significantly bias the impact of the bioassay data in determining the time-to-tumor fits.

There were various discussions about the pros and cons of constructing a joint likelihood of the cell replication data and the tumor data, and the weights to be assigned to the separate likelihoods. This was considered to be problematic by Dr. Portier.

Dr. Crump's opinion was that the Conolly model, and those explored by EPA, fit the tumor data poorly, and that an improved description of the tumor data was needed before the model could be used for low-dose and interspecies extrapolation.

Drs. Lucier and Hattis placed emphasis on including the early-time cell replication data instead of constructing a time-weighted average. It was felt that the two Monticello experiments could not be combined together as in Conolly et al. Dr. Lucier felt that the early-time data would have a greater impact in the progression of carcinogenesis. In general, the effect of "time" was considered to have significant effects on the time-to-tumor modeling, and they urged us to incorporate time-dependent terms in the modeling. CIIT expressed willingness to provide the original cell replication data to us for further analysis. (Further discussion on this matter did not take place in the open forum.)

Preliminary indications are, particularly based on Dr. Portier's insight, that the currently-held "de-minimus" picture of low-dose risk, as expressed in Conolly et al. (2004), is not likely to be the case if these various suggestions are incorporated in the modeling.

C. Meeting Report from Dr. Rory B. Conolly

Rory B. Conolly, Sc.D., D.A.B.T.
106 Michael's Way
Chapel Hill, NC 27516
Voice: 919.929.2258

July 24, 2005

Dr. Bobette Norse
ORAU Procurement - MS-04
P.O. Box 117
Oak Ridge, TN 37831-0117
Phone: 865-576-3051
Fax: 865-576-9385

Dear Dr. Nourse,

The following is my final written report on the formaldehyde review meeting held at the U.S. EPA in Washington, D.C. on 28-29 October, 2004.

EPA provided no guiding philosophical statement about the criteria being used to evaluate the CIIT assessment. The new Guidelines for Carcinogen Assessment state that the preferred default approach is to use a biologically based model. Since the key components of the CIIT assessment have been published in the peer-reviewed literature and have undergone several peer reviews other than the current NCEA effort, one has to wonder just how high the bar is set for acceptance of biologically based assessments. Given the time and resources expended on the CIIT assessment and the richness of the supporting data base, I find it difficult to imagine what an acceptable biologically-based assessment might look like if in the end the CIIT assessment is deemed not acceptable by NCEA. If this is in fact the outcome it will have major implications for the likelihood that anyone will be willing to commit the significant resources needed to develop of these kinds of risk assessment models.

The documents provided in advance of the October 2004 review meeting were collectively a discussion of uncertainty about the CIIT work. With respect to the clonal growth model, however, no new risk predictions were provided, so there was no way to judge how the uncertainties that NCEA identified might impact predicted risk. Evaluation of the significance of "uncertainties" when the impact of the uncertainties on the predicted risk is not known is itself an uncertain process.

A related concern is that there did not seem to be any consideration of the historical context of the CIIT assessment. EPA developed formaldehyde assessments in 1987 and 1991. The 1987 assessment used ppm as the input and the LMS model for the dose-response prediction. The 1991 assessment used DPX as a dosimeter and the LMS model. BMD assessments have since become available from other sources such as Paul Schlosser's work. The risk predictions of the BMD models are similar to the 1991 LMS assessment. Both the DPX-LMS and BMD assessments predicted somewhat less risk than the 1987 assessment, establishing the trend of less risk with increased incorporation of relevant data. I have always argued

(probably initially at the 1998 Ottawa review) that the historical context is the appropriate context for evaluating the CIIT clonal growth model. For a "level playing field" the uncertainties of the 1987 and 1991 assessments, and of the more recent BMD models, should be analyzed to the same degree as the clonal growth model. Does NCEA think that, because the LMS and BMD approaches used structurally simpler dose-response models and much more limited data inputs, they are less uncertain? The NCEA analysis seemed to be implying that use of more data and of a biologically more realistic model structure actually makes the CIIT approach more uncertain than the LMS and BMD approaches. I encourage NCEA to consider how uncertainties that can be evaluated explicitly in the structurally rich CIIT model compare to hidden uncertainties in the simpler models, where the hidden uncertainties encompass, for example:

1. Missing or incomplete descriptions of the regional dosimetry of formaldehyde.
2. Lack of simultaneous incorporation of the directly mutagenic and cytolethal/regenerative proliferation modes of action.
3. Lack of explicit consideration of the multistage nature of cancer.
4. Lack of consideration of the growth kinetics of initiated cell populations
5. Lack of evaluation of the measured J-shaped dose response for regenerative cellular proliferation.

A careful, balanced comparison of the CIIT assessment with the previous assessments along these lines would be informative with respect to the suitability of the CIIT assessment as the basis for a new IRIS listing for formaldehyde.

A further concern involves the peer-review of the CIIT formaldehyde assessment held in Ottawa in 1998. This review was sponsored by the U.S. EPA and Health Canada and involved what was arguably a world-class review panel. The CIIT assessment was not in its final form at that time, though we did provide a detailed description of the overall approach and the specific methods we were using to generate dose-response predictions. The 1999 CIIT document and the subsequent peer-reviewed publications are responsive to the comments and suggestions raised by the reviewers. My concern is that no information was provided on the role that Ottawa review plays in the ongoing review of the CIIT formaldehyde assessment by NCEA. Should the October 2004 review be viewed as standing on the shoulders of the 1998 review or as being in parallel to it? It was not at all clear to me that the October 2004 review in any way utilized the judgments of the 1998 review. It seems that the 2004 review was more of a parallel effort and that the 1998 review was ignored and was effectively a waste of time and money. I would like to have some clear understanding of how the 2004 review effort should be viewed relative to that of 1998.

In closing, let me reiterate that while the detailed examination of the CIIT formaldehyde assessment is laudable, this examination should be conduced with an eye to the historical context of formaldehyde risk assessment on the one had and, on the other hand, to a concern for encouraging, and not discouraging, development of biologically based risk assessment models.

Sincerely yours,

Rory B. Conolly, Sc.D., D.A.B.T.

- End of Volume IV -