Materials Submitted to the National Research Council
Part 2: Chemical-Specific Examples

U.S. Environmental Protection Agency

Integrated Risk Information System Program

January 30, 2013

Chemical-Specific Examples Demonstrating Implementation of NRC's 2011 Recommendations

The following are intended to provide the NRC panel with examples of how the IRIS Program is implementing the NRC recommendations included in the 2011 Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. The examples are not to be construed as final Agency conclusions and are provided for the sole purpose of demonstrating the IRIS implementation of the NRC recommendations.

EXAMPLE 1 – Literature Search and Screening

This example demonstrates the implementation of an improved literature search strategy as described in the "Identifying and Selecting Pertinent Studies" section of the draft Handbook for IRIS Assessment Development. The literature search strategy used to identify the studies to be included in the draft assessment, as well as the presentation of the literature search documentation, is shown below.

Literature search for Ethyl tert-butyl ether (ETBE)

1. Initial chemical-specific search conducted in online scientific databases
    a. Pubmed database (http://www.ncbi.nlm.nih.gov/pubmed/) searched (1/8/13) for all articles on Ethyl tert-butyl ether using the following search string:

    "ETBE" OR "Ethyl tert-butyl ether" OR "2-Ethoxy-2-methyl-propane" OR "ethyl tertiary butyl ether" OR "ethyl tert-butyl oxide" OR "tert-butyl ethyl ether" OR "ethyl t-butyl ether" OR "637-92-3"

    Search returned: 188 articles

    b. Toxline and DART searched (1/8/13) using the ToxNet database (http://toxnet.nlm.nih.gov/) using the following search string excluding PubMed records:
    "ETBE" OR "Ethyl tert-butyl ether" OR "2-Ethoxy-2-methyl-propane" OR "ethyl tertiary butyl ether" OR "ethyl tert-butyl oxide" OR "tert-butyl ethyl ether" OR "ethyl t-butyl ether" OR "637-92-3"

    Search returned: 110 articles (110 from Toxline; 0 from DART)

    c. TSCATS 2 (http://yosemite.epa.gov/oppts/epatscat8.nsf/ReportSearch?openform) was searched using the CAS number 637-92-3 for the EPA receipt dates of 1/01/2004-01/01/2013 since Toxline searches TSCATS through 2003.
    Search returned: 1 article

    d. Web of Science database (http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?highlighted_tab=WOS&product=WOS&last_prod=WOS&SID=1Dg72P6B9iG5G14Nd7L&search_mode=GeneralSearch) was searched (1/8/13) using the following search string with lemmatization "on":
    "ETBE" OR "Ethyl tert-butyl ether" OR "2-Ethoxy-2-methyl-propane" OR "ethyl tertiary butyl ether" OR "ethyl tert-butyl oxide" OR "tert-butyl ethyl ether" OR "ethyl t-butyl ether" OR "637-92-3"

    Search returned: 490 articles

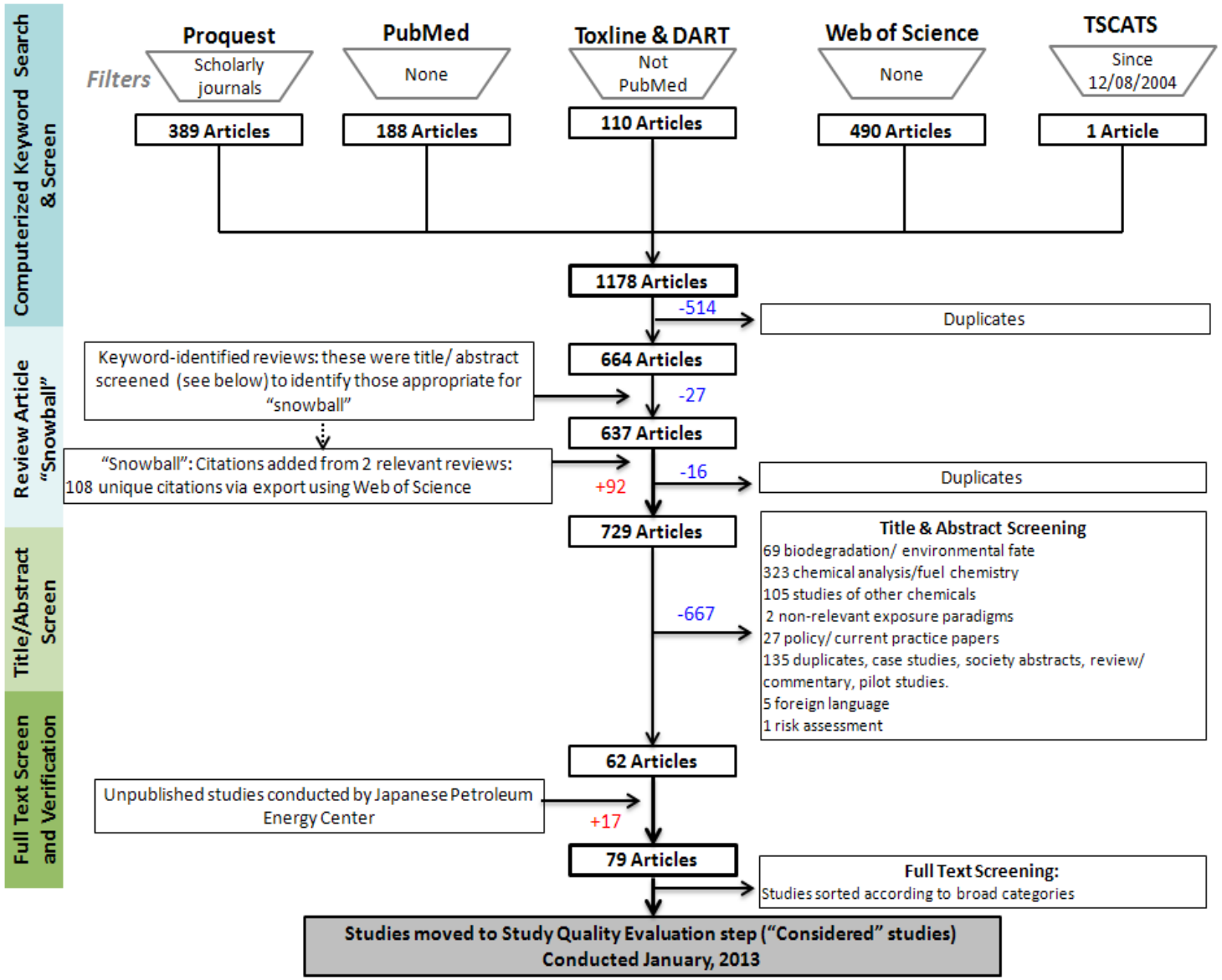
    e. Proquest database (http://search.proquest.com/environmentalscience/index?accountid=102841) was searched (1/8/13) using the following search string including only scholarly journals:

"ETBE" OR "Ethyl tert-butyl ether" OR "2-Ethoxy-2-methyl-propane" OR "ethyl tertiary butyl ether" OR "ethyl tert-butyl oxide" OR "tert-butyl ethyl ether" OR "ethyl t-butyl ether" OR "637-92-3"

Search returned: 389 articles

2. Total articles found: 1178 articles
    a. 514 were duplicates and removed by EPA's HERO search.
3. 664 unique articles imported into an EndNote library from the HERO web site.
    a. 27 references identified as reviews by EndNote query for "review"
    b. Review references removed from list and manually screened.
    c. 2 review references chosen for "snowball" search
        i. McGregor, D. (2007). "Ethyl tertiary-butyl ether: a toxicological review." Critical Reviews in Toxicology **37**(4): 287-312.
        ii. de Peyster, A. (2010). "Ethyl t-butyl ether: Review of reproductive and developmental toxicity." Birth Defects Research, Part B: Developmental and Reproductive Toxicology **89**(3): 239-263.
4. 108 cited references from 2 reviews were identified using Web of Science
    a. 16 references were duplicates and removed by EndNote upon import
    b. 92 unique references were imported into the EndNote library
5. Title and abstract screened manually within the EndNote library for relevance and excluded from further consideration for development of the hazard identification in the Toxicological Review based on the following criteria:
    a. Biodegradation/environmental fate (69)
    b. Chemical analysis/fuel chemistry (323)
    c. Study on non ETBE chemical (105)
    d. Non-relevant exposure (2)
    e. Policy papers (27)
    f. Duplicate, society abstracts, reviews/commentary, case studies, miscellaneous (135)
    g. Foreign language (5)
    h. Risk assessment (1)
6. 62 articles verified by full text review. No articles removed
7. 17 unpublished studies conducted by the Japanese Petroleum Energy Center were provided via direct correspondence. Studies were identified from a submitted report identified in the full text screen. All studies were screened for relevance and none were removed.
8. 79 articles were grouped into broad categories and were evaluated for study quality in the next step ("considered" studies).

Figure 1-1. Literature search documentation for ETBE



EXAMPLE 2 –Evaluation and Display of Individual Studies

This example demonstrates the tables used to evaluate the pertinent studies (including epidemiology and animal toxicology studies) identified through the literature search and screening step with respect to potential methodological considerations, as described in the "Evaluation and Display of Individual Studies" section of the draft Handbook for IRIS Assessment Development. This section will likely be expanded upon in the assessment, but the tables below serve as examples of the table format used to present the study evaluation results.

Table 2-1. Evaluation of observational epidemiology studies of diethyl phthalate - sexual differentiation effects (gray shading indicates a potential weakness or limitation of the study)

Reference, Setting and Design	Participant Selection, Follow-up	Comparability	Exposure Measure and Range	Outcome Measure	Consideration of Likely Confounding	Analysis and Presentation of Results (Estimate and Variability)	Sample Size; Power	Evaluation of Major Limitations
Anogenital Distance								
Suzuki et al. (2011). Japan. Birth cohort	Recruitment process not described. Enrolled at prenatal visit (mean 29 weeks gestation) 120 of 344 enrollees excluded because did not delivery at study hospital	Internal comparison group	Maternal urine (9 – 40 weeks; mean 29 weeks), MEP, 75th percentile = 32 ng/mL (44 ng/mL with SG correction)	Anogenital distance, measured at birth (1-3 days); blinded to exposure. Protocol described; 23 assessors; reliability measures not reported.	Gestational age, birth order, maternal age, maternal smoking and environmental tobacco smoke exposure (stepwise regression); Used SG-corrected urine concentrations	Described as not associated (details not reported)	n = 111 male infants	Relatively low, narrow exposure range.
Swan, 2008; Swan et al., 2005; 2003. United States (3 sites). Birth cohort (first follow-up)	Standardized recruitment process (Sept 1999 – Aug 2002). 85% of cohort agreed to be recontacted. Eligible if pregnancy ended in live birth, was currently 2-36 months of age, and lived within 50 miles of study center. 72% of eligibles participated in follow-up; 75% of participants had maternal urine sample and complete physical exam data (21 enrollees excluded because AGD exam not considered reliable (child too active; 2 declined)	Internal comparison group	Maternal urine (mean 29 weeks), MEP, 75th percentile = 437 ng/mL	Anogenital distance, measured at ages 2 - 36 months; blinded to exposure. Multiple assessors (3 sites); reliability measures not reported	Adjusted for weight percentile and age. Did not adjust for MBP or MEHP.	Percent change per interquartile increase in metabolite and p-value; also presented as metabolite distribution by 3 categories of anogenital distance.	n=106 boys	Is age-size adjustment adequate (considering potential temporal changes in exposure)?

Table 2-1. Evaluation of observational epidemiology studies of diethyl phthalate - sexual differentiation effects (gray shading indicates a potential weakness or limitation of the study)

Reference, Setting and Design	Participant Selection, Follow-up	Comparability	Exposure Measure and Range	Outcome Measure	Consideration of Likely Confounding	Analysis and Presentation of Results (Estimate and Variability)	Sample Size; Power	Evaluation of Major Limitations
Cryptorchidism or Testicular Position								
Main et al., 2006; Boisen et al., 2004. Denmark and Finland. Nested case-control study within birth cohort	Cases identified through standardized examination; all births at two university hospitals (one per country). 1997-2001 (Denmark); 1997–1999 (Finland)	Cases and controls well-matched by maternal characteristics	Breast milk samples collected 1–3 months of age, MEP, upper range not reported	Cryptorchidism, at birth or 3 months; blinded to exposure. Coordination and training of assessors discussed; borderline cases reviewed by two assessors	Analyzed separately by country and combined; no other variables addressed	SE and exact p-value for difference not given, but p > 0.40	n=62 cases, n=68 controls	Exposure measure may not reflect in utero exposure; breast pump use could increase MEP levels
Swan, 2008; Swan et al., 2005; 2003. United States (3 sites). Birth cohort	See entry above	Internal comparison group	Maternal urine (mean 29 weeks), MEP, 75th percentile = 437 ng/mL	One or both testicles not "normal" or "normal retractile" at clinical exam (ages 0– 36 months); blinded to exposure. Multiple assessors (3 sites); reliability measures not reported	Did not adjust for MBP or MEHP	Described as not associated (details not reported)	n=119 boys	Outcome seen in 10% of the study sample; unclear what this represents from clinical perspective
Infant Hormone Levels								
Lin et al., 2011; Wang et al.,2004. Taiwan. Birth cohort	Pregnant women seen in prenatal clinic (≥18 weeks gestation) and intending to deliver in that hospital invited to participate (singleton births, no medical complications). Dec 2000–Nov 2001.	Internal comparison group	Maternal urine (3rd trimester, 28–36 weeks), MEP, 95th percentile–241 ng/mL (346 ug/g creatinine)	Cord blood hormone levels; blinded to exposure	Gestational age, maternal age, gravidity, smoking, body mass index, ever oral contraceptive use, other phthalate metabolites (stepwise regression); Used creatinine-adjusted concentrations	Beta, but no SE, reported for regression analyses (continuous measures)	n=81 boys, 74 girls	Limited analysis

Table 2-1. Evaluation of observational epidemiology studies of diethyl phthalate - sexual differentiation effects (gray shading indicates a potential weakness or limitation of the study)

Reference, Setting and Design	Participant Selection, Follow-up	Comparability	Exposure Measure and Range	Outcome Measure	Consideration of Likely Confounding	Analysis and Presentation of Results (Estimate and Variability)	Sample Size; Power	Evaluation of Major Limitations
	275 of 430 women in cohort provided urine sample; 120 of 275 excluded because of missing cord blood or other data; did not differ by age, body mass index, smoking or alcohol use							
Main et al., 2006 Denmark and Birth cohort	See entry above. Cases and controls combined for this analysis	Internal comparison group	Breast milk samples collected 1–3 months of age, MEP, upper range not reported	Serum hormone levels at 3 months; blinded to exposure	Analyzed separately by country and combined; no other variables addressed	Spearman correlation coefficients and p-values. Did not adjust for MBP	n=130 boys	Exposure measure may not reflect in utero exposure; breast pump use could increase MEP levels
Gender-Related Play								
Swan et al. 2010; Swan et al., 2005; 2003. United States (4 sites – Iowa added 2002-2005). Second follow-up of birth cohort.	See entry above. 128 of 334 eligible not found; 56 of found did not participate. Higher percentage of mothers of participating families were white (88% compared with 78%) and completed college (73% compared with 68%)	Internal comparison group	Maternal urine (mean 29 weeks), MEP, 75th percentile = 437 ng/mL (based on earlier publications)	Pre-school Activities Inventory (24 items, completed by parents; instrument used in previous studies of direct and indirect measures of testosterone); blinded to exposure	Covariates considered: creatinine concentration, child's sex, maternal age, parental education, number of same and opposite sex siblings, clinic location, parental attitude regarding sex-atypical play; kept in model if >10% change in effect estimate (retained: maternal age, boy's age, parental education, parental attitude, and education-attitude interaction)	Described as not associated (details not reported)	n=74 boys, 71 girls	

1

Table 2-2. Evaluation of observational epidemiology studies of diethyl phthalate - neurobehavioral effects (gray shading indicates a potential weakness or limitation of the study)

Reference, Setting and Design	Participant Selection, Follow-up	Comparability	Exposure Measure and Range	Outcome Measure	Consideration of Likely Confounding	Analysis and Presentation of Results (Estimate and Variability)	Sample Size; Power	Evaluation of Major Limitations
Engel et al., 2009; Wolff et al., 2008. United States (Mt Sinai, New York). Birth cohort.	Seen for prenatal care at Mt Sinai hospital or two private practices and delivered at Mt Sinai. Singleton, primiparous pregnancies, delivered May 1998 – July 2001. 475 initially recruited; 404 of these eligible (28 left area; 19 refused; 28 miscellaneous other reasons). Outcome not measured in 93 of 404 enrollees (excluded if in NICU, only in hospital on weekend, parent refused, baby not testable, or study personnel unavailable). Of the 311 with outcome data, 295 also had urine sample. Models were restricted to observations with values >20 mg/dL	Internal comparison group	Maternal urine, MEP (25–40 weeks, mean 32), 75th percentile 1, 025 ng/mL	Brazelton Neonatal Behavioral Assessment Scale (7 domains; 28 behavioral items and 18 primitive reflexes); 4 trained examiners (no information on agreement); blinded to exposure	Covariates considered included maternal age, race, marital status, education, cesarean delivery, delivery anesthesia, infant age, infant sex, infant jaundice, maternal smoking, alcohol, caffeine, and illicit drug use, urinary creatinine, examiner, and maternal urinary organophosphate levels. Dropped from model if <10% change in Beta coefficient compared with full model. Also examined interaction by sex of child.	Beta and 95% CI for summation of low molecular weight metabolites (MEP, MBP, MiBP and MMP)	n=295	Data presented only for summation of low molecular weight metabolites
Engel et al., 2010; Wolff et al., 2008. United States (Mt Sinai, New York). Follow-up(s)	See Engel et al. (2009) for cohort description. BASC F scores > 3 excluded because of questionable validity (n=2); 25 scores of 2	Small differences in education level and age in non-participants at follow-up compared with	Maternal urine, MEP (25–40 weeks, mean 32), MEP (distribution not given but assumed similar to other studies from	Behavior Rating Inventory of Executive Function *86 items, 8 subscales); Behavior	Covariates considered based on relation with phthalates metabolites and outcomes. Also examined interaction by sex of child. Adjusted for race, sex, education and	Beta and 95% CI for summation of low molecular weight metabolites (continuous, tertiles MEP); Beta and p < 0.05 denoted for MEP	n=177	

Table 2-2. Evaluation of observational epidemiology studies of diethyl phthalate - neurobehavioral effects (gray shading indicates a potential weakness or limitation of the study)

Reference, Setting and Design	Participant Selection, Follow-up	Comparability	Exposure Measure and Range	Outcome Measure	Consideration of Likely Confounding	Analysis and Presentation of Results (Estimate and Variability)	Sample Size; Power	Evaluation of Major Limitations
of birth cohort.	or 3 reviewed and 12 excluded because of concerns about language (n=2), random responses (n=7), or overly negative or unrealistic evaluation (n=3). Internal comparison group	participants, but little difference in MEP between groups. Internal comparison group	this cohort)	Assessment System for Children (BASC, 130 items, 9 scales, parent ratings); used in previous studies of executive functioning and behavior; blinded to exposure	marital status of primary caretaker, and urinary creatinine	(continuous)		
Miodovnik et al., 2011; Wolff et al., 2008. United States (Mt Sinai, New York). Follow-up(s) of birth cohort.	See Engel et al. (2009) for cohort description. 137 of original 404 completed 7–9 year follow-up	Higher proportion of lower education (<high school) in non-participants at follow-up compared with participants, but little difference in MEP between groups. Internal comparison group	Maternal urine, MEP (25-40 weeks, mean 32), MEP, 75th percentile 964 ng/mL	Social Responsiveness Scale (65items, completed by caregiver); subscales for Social Awareness, Social Cognition, Social Communication, Social Motivation, and Autistic Mannerisms; used in previous studies of autism behaviors; blinded to exposure	Covariates considered: maternal age, maternal IQ, marital status at the time of follow-up, maternal education, Child's race/ethnicity. Also included and urinary creatinine in models.	Beta and 95% CI (continuous MEP)	n=137	

1

Table 2-3. Evaluation of animal toxicology studies for chemical X

Reference (Species)	Exposure Quality	Test Subjects	Study Design	Toxicity Endpoints	Data and Statistics	Reporting
Smith et al. (1984) (Monkey)	++	++ Note: N=20	++ Note: 102 wk study	++	Not applicable	++
Jones et al. (1986) (Mouse)	+ co-exposure likely	+ N=5; variable ages at onset of exposure across groups	++ Note: 13 wk study	Potential sampling bias; No observer blinding indicated; protocols incompletely reported	+ data represents pooled sexes	+ Surgical procedures not reported
Gray et al. (2012) (Rat)	Test article and exposure methods not specified	Bacterial infection noted in animal colony; N= 3 litters; males only; overt maternal toxicity	No randomization across litters into treatment groups; testing during exposure expected to confound results; acute exposure	++	Not applicable	Results data not reported

Criteria for the six categories developed based on the chemical and hazard type in question. In this example: gray box = examination of relevant study details identified potential limitations that could influence interpretation of the study's results; '+' = criteria not completely met or potential issues identified, but unlikely to directly affect study interpretation; ++ = criteria determined to be completely met. Text accompanying summary table would explain key study details informing these determinations.

EXAMPLE 3 – Evidence Tables

This example demonstrates the evidence table format to be used for presenting the epidemiological and toxicological evidence available for endpoint-specific hazards, as described in the "Evaluation and Display of Individual Studies" section of the draft Handbook for IRIS Assessment Development.

- **Human Evidence**

Table 3-1. Evidence pertaining to male reproductive effects of diethyl phthalate in humans

Reference and Study Design[a]	Results
Reproductive hormones	
Meeker et al., 2009 (United States; Boston) (Tier 1) 425 male partners seen in subfertility clinic, mean age 36 years, 2000–2004 Serum, steroidal and gonadotropin hormones Urine sample, median (90th percentile) MEP 153 (1376) ng/mL	Beta (95% CI) for ln-MEP in relation to hormone 0.0 = no effect Testosterone 8.87 (–7.18, 24.9) Estradiol 0.71 (–0.97, 2.40 1.0 = no effect Free androgen index 1.04 (0.99, 1.09) FSH 0.98 (0.91, 1.06) LH 0.98 0.91, 1.04) Adjusted for age, body mass index, smoking, season and time of sample collection (and time squared), dilution ranking
Jönsson et al., 2005 (Sweden) (Tier 2) 234 men ages 18–21 years (military service) Serum, steroidal and gonadotropin hormones. Urine samples, median (95th percentile) MEP 240 (4400) ng/mL; 83 (1600 nmol/mmol creatinine)	Mean difference (95% CI), highest compared with lowest quartile of MEP Testosterone (nM) –0.3 (–2.3, 1.8) Free testosterone (T/SHBG) 0.06 (–0.05, 0.2) Estradiol (pM) 1.8 (–4.2, 7.7) FSH (IU/L) 0.5 (–0.5, 0.6) LH (IU/L) 0.7 (0.1, 1.2) (Positive difference indicates lower value in highest exposure quartile) Abstinence time and smoking evaluated as confounders
Sperm parameters	
Hauser et al., 2007, 2006 (United States; Boston) (Tier 1) 463 male partners seen in subfertility clinic, mean age 36 years, 2000–2004 [n=379 for damage measures] Semen analysis, sperm damage measures analyzed Urine sample, median (75th , 95th percentile) MEP 158 (535, 2214) ng/mL (specific-gravidity adjusted)	OR (95% CI), by metabolite quartile of MEP MEP: Concentration (< 20 × 10^6/mL); Motility (< 50% motile); Morphology (< 4% normal) 1 (low): 1.0 (referent); 1.0 (referent); 1.0 (referent) 2: 1.5 (0.7, 3.6); 1.1 (0.6, 1.9); 0.8 (0.4, 1.6) 3: 1.0 (0.4,2.5); 0.8 (0.5, 1.5); 0.7 (0.3, 1.3) 4 (high): 1.2 (0.5, 3.0); 1.0 (0.6, 1.8); 0.5 (0.3, 1.1) (trend p): (0.94); (0.84); (0.07) Logistic regression adjusted for age, abstinence time, and smoking. Damage measures, Beta (95% CI) associated with interquartile range increase MEP: Comet extent (μm) 6.06 (0.941, 12.3); Tail distribution (μm) 2.72 (0.46, 5.00); %DNA tail –0.26 (–2.52, 2.02) linear regression, adjusted for age and smoking

Table 3-1. Evidence pertaining to male reproductive effects of diethyl phthalate in humans

Reference and Study Design[a]	Results
Pant et al., 2008 (India) (Tier 2) 300 men, mean age 29 years (100 fertile, 200 infertile), urban and rural Semen analysis DEP concentration in semen for fertile group, mean (±SE) 0.64 (± 0.24) in rural, 0.74 (± 0.04) µg/mL in urban areas	Pearson correlation coefficient between semen DEP and sperm parameter: r (p-value) Sperm concentration –0.19 (p< 0.05) Sperm motility (%) 0.03 Morphology (% abnormal –0.02 Damage (Chromatin integrity) 0.07 (all other p-values > 0.05; exact value not reported)
Zhang et al., 2006 (China) (Tier 2) 52 men seen in Shanghai Institute of Planned Parenthood Research clinics, mean age 32 years Semen analysis DEP concentration in semen, mean 0.47 mg/L	Spearman correlation coefficient between semen DEHP and sperm parameter: r (p-value) Sperm concentration –0.25 (0.15) Sperm motility (%) –0.13 (0.45) Sperm rate of malformations 0.19 (0.28)
Jönsson et al., 2005 (Sweden) (Tier 2) 234 men ages 18–21 years (military service) Semen analysis Urine samples, median (95th percentile) MEP 240 (4400) ng/mL, or 83 (1600 nmol/mmol creatinine)	Mean difference (95% CI), highest compared with lowest quartile MEP Sperm concentration (× 106/mL) 5.0 (–15. 25) Sperm motility (%) –0.4 (–6.4, 5.6) Sperm damage (chromatin integrity) 0.8 (–2.8, 4.4) (Positive difference indicates lower value in highest exposure quartile) Abstinence time and smoking evaluated as confounders
Liu et al., 2012 (China) (Tie 2) 97 male partners seen in subfertility clinic, mean age 32 years Semen analysis Urine sample, median (66th percentile) MEP 12.6 (21.3) ng/mL	OR (95% CI), by metabolite tertile MEP: Concentration (<20 × 10^6/mL); Motility (<50% motile) 1: 1.0 (referent); 1.0 (referent) 2: 1.4 (0.2, 8.8); 0.7 (0.2, 1.9) 3: 1.5 (0.2, 9.6); 0.4 (0.1, 1.2) (trend p): (0.66); (0.10) adjusted for age, body mass index, abstinence time, smoking, alcohol use
Infertility	
Tranfo et al., 2012 epub (Italy) (Tier 2) Case-control study, 56 couples from assisted reproduction center, n=56 control couples (parents), mean age 39 years in both groups. Case-control comparison Urine samples, median (95th percentile) MEP 52 (651) µg/g creatinine (controls); slightly higher in women than men	Comparison between MEP levels in cases and controls Mann-Whitney test p-value Females < 0.001 Males < 0.001 Additional details of sex-stratified results not provided
Pant et al., 2008 (India) (Tier 2) 300 men, mean age 29 years (100 fertile, 200 infertile), urban and rural DEP concentration by fertility status (based on partners who had conceived within 1 year of attempting pregnancy) DEP concentration in semen for fertile group, mean (±SE) 0.64 (± 0.24) in rural, 0.74 (± 0.04) µg/mL in urban areas	DEP concentration in semen, mean ± SE (p-value for difference between fertile and infertile): Rural: Fertile Infertile 0.64 ± 0.24 0.74 ± 0.04 Urban: Fertile Infertile 1.13 ± 0.11 3.11 ± 0.26 (p < 0.05)

[a] "Tier" reflects evaluation of confidence in study results based on evaluation of risk of specific types of bias. *(In the assessment, details of evaluation will be shown in Supplemental Information tables and text).*

1

- **Animal evidence**

Table 3-2. Evidence pertaining to female reproductive effects of diethyl phthalate in animals

Reference and Study Design	Results				
Fertility and birth outcomes					
Fujii et al. (2005) Rat (Sprague Dawley); 21–24/sex/group 0, 600, 3,000, 15,000 ppm (0, 40, 197, 1016 mg/kg-day in F0 males; 0, 51, 255, 1297 mg/kg-day in F0 females; 0, 46, 222, 1150 mg/kg-day in F1 males; 0, 56, 267, 1375 mg/kg-day in F1 females) Diet 105 days for F0 and F1 parental males 119 days for F0 and F1 parental females (exposure through 10 weeks premating + 3 weeks mating + weaning)	No. of implantations *(percent change compared to control)*				
		0	51/56	255/267	1297/1375
	F0 parental females	-	2%	1%	1%
	F1 parental females	-	0%	4%	3%
	Fertility Index *(percent change compared to control)*				
		0	51/56	255/267	1297/1375
	F0 parental females	-	0%	4%	0%
	F1 parental females	-	0%	0%	0%
	Gestation length (days) *(percent change compared to control)*				
		0	51/56	255/267	1297/1375
	F0 parental females	-	0%	0%	-1%
	F1 parental females	-	0%	0%	-1%*
	No. of pups delivered *(percent change compared to control)*				
		0	51/56	255/267	1297/1375
	F0 parental females	-	-1%	1%	1%
	F1 parental females	-	4%	7%	2%
Hardin et al. (1987) Mouse (Swiss); 50 females/group 0, 4500 mg/kg-day Gavage GD6–GD13	*(percent change compared to control)*				
		0		4500	
	No. of live pups/litter	-		0%	
	Percent survival	-		-4%	
	Birth weight	-		-6%	
Howdeshell et al. (2008) Rat (Sprague Dawley); 3–5 female (dams)/group and 9 control dams 0, 100, 300, 600, 900 mg/kg-day Gavage GD8–GD18	*(percent change compared to control)*					
		0	100	300	600	900
	No. of implantations	-	5%	3%	4%	13%
	No. of live fetuses	-	7%	5%	-6%	16%
	Total resorptions	-	0%	0%	325%*	0%
	Fetal mortality (%)	-	0%	0%	283%*	0%

Table 3-2. Evidence pertaining to female reproductive effects of diethyl phthalate in animals

Reference and Study Design	Results				
NTP (1984) Mouse (Swiss); 20/sex/group 0, 0.25, 1.25, 2.5% Diet 7 days premating + 98 days cohabitation + 21 days segregation (126 days total) (F0 males and females), and 0, 2.5% (0, 3640 mg/kg-day) in utero + lactation, and then in the diet through a 7-day mating period at 74±10 days old (F1 females were allowed to deliver litters)	*(percent change compared to control)*				
	F0 females	0	0.25	1.25	2.5
	No. of live pups/litter	-	23%*	14%	3%
	Live pup weight	-	-2%	-2%	1%
	F1 females	0			2.5
	No. of live pups/litter	-			-14%*
	Fertility index (%)	-			0%
	Live pup weight	-			-3%
NTP (1988) Rat (Sprague Dawley); 31–32 females (dams)/group 0, 0.25, 2.5, 5% (0, 198, 1909, 3214 mg/kg-day) Diet GD6–GD 15	*(percent change compared to control)*				
		0	198	1909	3214
	Corpora lutea per dam	-	4%	2%	1%
	Implantation sites per litter	-	4%	1%	2%
	Resorptions per litter	-	5%	13%	-11%
	Percent resorptions per litter	-	2%	7%	-18%
	Live fetuses per litter	-	4%	-2%	3%
Singh et al. (1972) Rat (Sprague Dawley); 5 time-mated females/group 0, 0.506, 1.012, 1.686 mL/kg Intraperitoneal injections on GD5, 10, and 15 (termination on GD 20)		Untreated	0.506	1.012	1.686
	No. of corpora lutea	60	65	59	57
	No. of resorptions	0	28	0	2
	No. of live fetuses	59	35	57	54

Table 3-2. Evidence pertaining to female reproductive effects of diethyl phthalate in animals

Reference and Study Design	Results				
Anogenital distance					
Fujii et al. (2005) Rat (Sprague Dawley), 21–24/sex/group 0, 600, 3,000, 15,000 ppm (0, 40, 197, 1016 mg/kg-day in F0 males; 0, 51, 255, 1297 mg/kg-day in F0 females; 0, 46, 222, 1150 mg/kg-day in F1 males; 0, 56, 267, 1375 mg/kg-day in F1 females) Diet 105 days for F0 and F1 parental males, 119 days for F0 and F1 parental females (exposure through 10 weeks premating + 3 weeks mating + weaning)	*(percent change compared to control)*				
	Females	0	40–56	197–267	1016–1375
	F1 pups at PND 0	-	-5%	-5%	1%
	F1 pups at PND 4	-	-3%	-2%	-1%
	F2 pups at PND 0	-	-2%	0%	-1%
	F2 pups at PND 4	-	-1%	-1%	-2%
Reproductive organ weights					
Fujii et al. (2005) Rat (Sprague Dawley), 21–24/sex/group 0, 600, 3,000, 15,000 ppm (0, 40, 197, 1016 mg/kg-day in F0 males; 0, 51, 255, 1297 mg/kg-day in F0 females; 0, 46, 222, 1150 mg/kg-day in F1 males; 0, 56, 267, 1375 mg/kg-day in F1 females) Diet 105 days for F0 and F1 parental males, 119 days for F0 and F1 parental females (exposure through 10 weeks premating + 3 weeks mating + weaning)	Absolute ovary weight *(percent change compared to control)*				
		0	51/56	255/267	1297/1375
	F0	-	-4%	-10%	-6%
	F1	-	1%	2%	4%
	F1 pup	-	4%	-8%	-4%
	F2 pup	-	0%	0%	-4%
	Relative ovary weight *(percent change compared to control)*				
		0	51/56	255/267	1297/1375
	F0	-	-5%	-8%	-5%
	F1	-	0%	0%	2%
	F1 pup	-	7%	-3%	17%
	F2 pup	-	-3%	-3%	0%
	Absolute uterus weight *(percent change compared to control)*				
		0	51/56	255/267	1297/1375
	F0	-	2%	4%	-4%
	F1	-	4%	7%	-1%
	F1 pup	-	3%	7%	-22%*
	F2 pup	-	-11%	-17%	-27%*
	Relative uterus weight *(percent change compared to control)*				
		0	51/56	255/267	1297/1375
	F0	-	0%	6%	-3%
	F1	-	4%	4%	0%
	F1 pup	-	5%	9%	-5%
	F2 pup	-	-12%	-17%	-20%*

Table 3-2. Evidence pertaining to female reproductive effects of diethyl phthalate in animals

Reference and Study Design	Results			
Pereira et al. (2007c) Rat (Wistar); 6/sex/group 0, 50 ppm (F0) (0, 2.85 mg/kg-day) 0, 25 ppm (F1) (0, 1.425 mg/kg-day) Diet 150 days/generation	Relative ovary weight *(percent compared to control)*			
	F0 parental females		F1 adult females	
	0	2.85	0	1.425
	-	40%*	-	23%*
NTP (1984) Mouse (Swiss); 20/sex/group 0, 0.25, 1.25, 2.5% Diet 7 days premating + 98 days cohabitation + 21 days segregation (126 days total) (F0 males and females), and 0, 2.5% (0, 3640 mg/kg-day) in utero + lactation, and then in the diet through a 7 day mating period at 74±10 days old (F1 females were allowed to deliver litters)	Ovary weight *(percent change compared to control)*			
		0		3640
	Absolute	-		-3%
	Relative	-		3%
	Uterus weight *(percent change compared to control)*			
		0		3640
	Absolute	-		-4%
	Relative	-		-4%

*Statistically significant ($p < 0.05$) based on analysis of data by study authors.
Percent change compared to control = $\frac{\text{treated value – control value}}{\text{control value}} \times 100$

1

EXAMPLE 4 – Evidence Integration

The example below demonstrates the integration of evidence from epidemiology studies in order to draw conclusions about the hazards associated with chemical exposure to humans, as described in the "Evaluating the Overall Evidence of Each Effect" section of the draft Handbook for IRIS Assessment Development.

Example of Synthesis of Epidemiology Studies Evaluating Associations with Lymphohematopoietic Cancers in Formaldehyde-Exposed Populations

In subsequent sections, the evidence of an association for each cancer-subtype in relation to formaldehyde exposure was evaluated using a weight-of-evidence approach as outlined in the U.S. EPA's *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 2005a), and described in general terms in the IRIS preamble. Causal considerations follow from the Bradford-Hill (1965) aspects of causality and include consistency, strength of association, specificity, temporality, evidence of an exposure-response relationship, and biological plausibility. Potential sources of bias were also considered, including selection bias, information bias, and confounding.

This following example, currently under development, includes a draft evaluation of evidence for one of the cancer-subtypes under consideration in the draft IRIS assessment of Formaldehyde.

1.3.1.1.1. Hodgkin Lymphoma

Hodgkin lymphoma is a specific type of lymphohematopoietic cancer originating from white blood cells. Historically, the diagnosis of Hodgkin lymphoma (previously called Hodgkin's disease) used in epidemiologic studies has been ascertained from death certificates according to the version of the International Classification of Diseases (ICD) in effect at the time of study subjects' deaths [i.e., ICD-8 and ICD-9: Code 201 (WHO, 1967; 1977)].

Epidemiologic evidence

Evidence describing the association between formaldehyde exposure and the specific risk of Hodgkin lymphoma was available from 13 epidemiologic studies – one case-control study (Gerin et al., 1989) and 12 cohort studies (Beane Freeman et al., 2009; Pinkerton et al., 2004; Coggon et al., 2003; Andjelkovich et al., 1995; Hansen and Olsen, 1995; Hall et al., 1991; Hayes et al., 1990; Matanoski, 1989; Robinson et al., 1987; Stroup et al., 1986; Walrath and Fraumeni, 1984; 1983b). Study details are provided in the evidence table for Hodgkin lymphoma (Table 4-1).

Causal Evaluation

The evidence of an association was evaluated using a weight-of-evidence approach as outlined in the U.S. EPA's *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 2005a). The epidemiologic data on Hodgkin lymphoma provided the strongest evidence regarding causation

with respect to two particular considerations: 1) the strong evidence of an exposure-response relationship observed in the single largest cohort study; and 2) the inconsistent pattern of risks across studies – many of which had fewer than 5 exposed cases.

Conclusion

Conclusion not available until draft is completed.

Consistency of the observed association

The results of the 13 studies were not consistent. The study of the largest cohort of formaldehyde-exposed workers (Beane Freeman et al., 2009) reported an elevated risk of dying from Hodgkin lymphoma for the cohort as a whole (SMR=1.42; 95% CI: 0.96-2.1) and a pronounced increase in risk among those workers with the highest peak formaldehyde exposures (RR=3.96; 95% CI: 1.31-12.02). However, the results of the other 12 studies were more consistent, with the absence of an effect of formaldehyde exposure on the risk of developing and dying from Hodgkin lymphoma.

Compared with other lymphohematopoietic cancers, the survival rate for Hodgkin lymphoma is relatively high and mortality is rare. This rarity results in very low statistical power and may have contributed to the apparently discordant results. Aside from the Beane Freeman et al. (2009) study which reported 25 deaths from Hodgkin lymphoma, only two other cohort studies observed more than five deaths from Hodgkin lymphoma, Coggon et al. (2003), which reported 6 observed deaths against 8.5 expected deaths, and Hansen and Olsen (1995), which reported 12 observed deaths against 12.2 expected deaths. The case-control study (Gerin et al., 1989) observed only 8 cases of Hodgkin lymphoma and did not report an elevated risk associated with working in formaldehyde-exposed jobs.

The study results presented in Table 4-1 (by publication date) detail all of the reported associations between exposures to formaldehyde and the risks of developing and dying from lymphatic leukemia. Results are plotted in Figure 4-1.

Strength of the observed association

Summary effect estimates for the association between formaldehyde exposure and Hodgkin lymphoma were highly variable and the risk of developing or dying from Hodgkin lymphoma were predominantly less than one (unity) and ranged from zero to 3.33.

While the summary effect estimate from the study by Beane Freeman et al. (2009) was RR=1.42 (95% CI: 0.96, 2.10), the strength of the association was substantially stronger among those workers exposed to the highest peak levels (RR=3.96). Beane Freeman et al. (2009) further showed plots presenting the RR from the internal analyses for each endpoint and for each year of follow-up. The association of Hodgkin lymphoma with formaldehyde exposure is not only seen for the complete 2004 follow-up when the average length of follow-up was 42 years, but throughout

the cohort experience (see Beane Freeman et al., 2009; Figure 1). These plots show that during the 1970's and 1980's, the RR≈8 and remained elevated at about RR = 4 through the end of follow-up in 2004.

Specificity of the observed association

Specificity refers to an increased inference of causality if a single cause is associated with a single effect or disease (Hill, 1965). An example of specificity is seen with respect to a specific infectious disease caused by a specific virus. Based on an understanding that many agents cause cancer at multiple sites (e.g., tobacco), specificity is generally not considered to be a necessary condition for making causal inferences regarding cancer.

Nonetheless, the specificity of the diagnoses of cancer is important – especially for lymphohematopoietic cancers, which are heterogeneous in nature and arise from different cell lines. This point concerning specificity was not discussed in Hill's paper on causality (1965). In an epidemiology study, increasing the specificity of a diagnosis is likely to increase the precision of an observed association because the exposure, if it is causally associated, is relevant to the cases under study (e.g., cases are not diluted with diagnoses that are not relevant to the exposure). In this section, only the specific diagnosis of Hodgkin lymphoma was considered. The most specific level of Hodgkin lymphoma diagnosis that is commonly reported across the epidemiologic literature has been based on the first three digits of the Eighth or Ninth Revision of the ICD code (i.e., Hodgkin's disease ICD-8/9: 201).

Temporal relationship of the observed association

Only one study (Beane Freeman et al., 2009) reported on analyses of the temporal relationship showing that risks were highest 15–25 years since first formaldehyde exposure. Such a pattern is consistent with the expected time-course of disease and mortality following exposure to formaldehyde; however, this finding with respect to formaldehyde is without corroboration for Hodgkin lymphoma.

Exposure-response relationship

An exposure-response relationship showing increasing effects associated with greater exposure strongly suggests cause and effect, especially when such relationships are also observed for duration of exposure (USEPA, 2005a: p. 2-14). None of the studies evaluated the effect of duration of formaldehyde exposure on the mortality risk of Hodgkin lymphoma. There were only two studies that evaluated any form of exposure-response for increasing measures of formaldehyde exposure (Coggon et al., 2003; Beane Freeman et al., 2009). Coggon et al. (2003) reported a lower risk of dying from Hodgkin lymphoma among 'highly' exposed workers based on a single death.

Beane Freeman et al. (2009) reported a clear exposure-response relationship between increasing levels of peak formaldehyde and increased risk of dying from Hodgkin lymphoma among

exposed workers (p=0.01). Compared to exposed workers in the lowest exposure category of peak exposure, those in the middle category were at more than threefold higher risk (RR=3.30; 95% CI: 1.04, 10.50) while those workers in the highest category were at fourfold higher risk (RR=3.96; 95% CI: 1.31, 12.02). Beane Freeman et al. (2009) also reported an exposure-response relationship between increasing levels of average formaldehyde intensity and increased risk of dying from Hodgkin lymphoma among exposed workers (p=0.05).

Biologic plausibility

The reader is referred to the section on mode of action for lymphohematopoietic cancers.

Potential impact of selection bias, information bias, confounding bias, and chance

Selection bias is an unlikely bias in the epidemiologic studies of Hodgkin lymphoma as the case-control study evaluated exposure status without regard to outcome status and had a participation level of 83% and each of the cohort studies included at least 72% of eligible participants and lost fewer than 9% of participants over the course of mortality follow-up.

The healthy-worker effect and the healthy-worker survivor effect could obscure a truly larger effect of formaldehyde exposure in analyses based on "external" comparisons with mortality in the general population (Walrath and Fraumeni 1983b; 1984; Stroup et al. 1986; Matanoski 1989; Hayes et al., 1990; Hall et al., 1991; Robinson et al., 1987; Andjelkovich et al., 1995; Hansen and Olsen, 1995; Coggon et al., 2003; Pinkerton et al., 2004; Beane Freeman et al., 2009), but would not influence analyses using "internal" or matched comparison groups (Gerin et al., 1989; Beane Freeman et al., 2009).

Information bias is unlikely to have resulted in bias away from the null; however, random measurement error or non-differential misclassification is almost certain to have resulted in some bias toward the null among these studies of Hodgkin lymphoma.

Chemical exposures that have not been independently associated with Hodgkin lymphoma are not expected to confound results. The main support for a suggestive association of formaldehyde exposure with increased risk of Hodgkin lymphoma is from the results for peak exposures reported by Beane Freeman et al. (2009) who specifically examined the potential for confounding from 11 substances including benzene and found that controlling for these exposures did not meaningfully change the results. This provides evidence against potential confounding by these co-exposures. There does not appear to be any evidence of confounding that would provide an alternative explanation for the observed association of formaldehyde exposure with increased risk of Hodgkin lymphoma reported by Beane Freeman et al. (2009).

The reported results for the risk of Hodgkin lymphoma associated with exposure to formaldehyde were inconsistent. There were 12 small studies, each with 12 or fewer exposed cases and only 44 exposed cases among them, showing a consistent pattern of risks across studies indicating a lack of an association. However, the single largest study in terms of study population

and number of formaldehyde exposed cases (n=25) showed increased risks of Hodgkin lymphoma mortality as a cohort compared to the general population (SMR=1.42; 95% CI: 0.96, 2.10) and statistically significant increased risks with increasing levels of peak exposure (p-trend among exposed workers =0.01). The evidence of an association with peak exposures reported by Beane Freeman et al. (2009) suggests an association whose risk increases with greater exposure. However, there was only one statistically robust observation of an exposure-response relationship showing increased risks with peak exposures and this finding is tempered by the lack of corroborative epidemiologic evidence.

Conclusion

Conclusion not available until draft is completed.

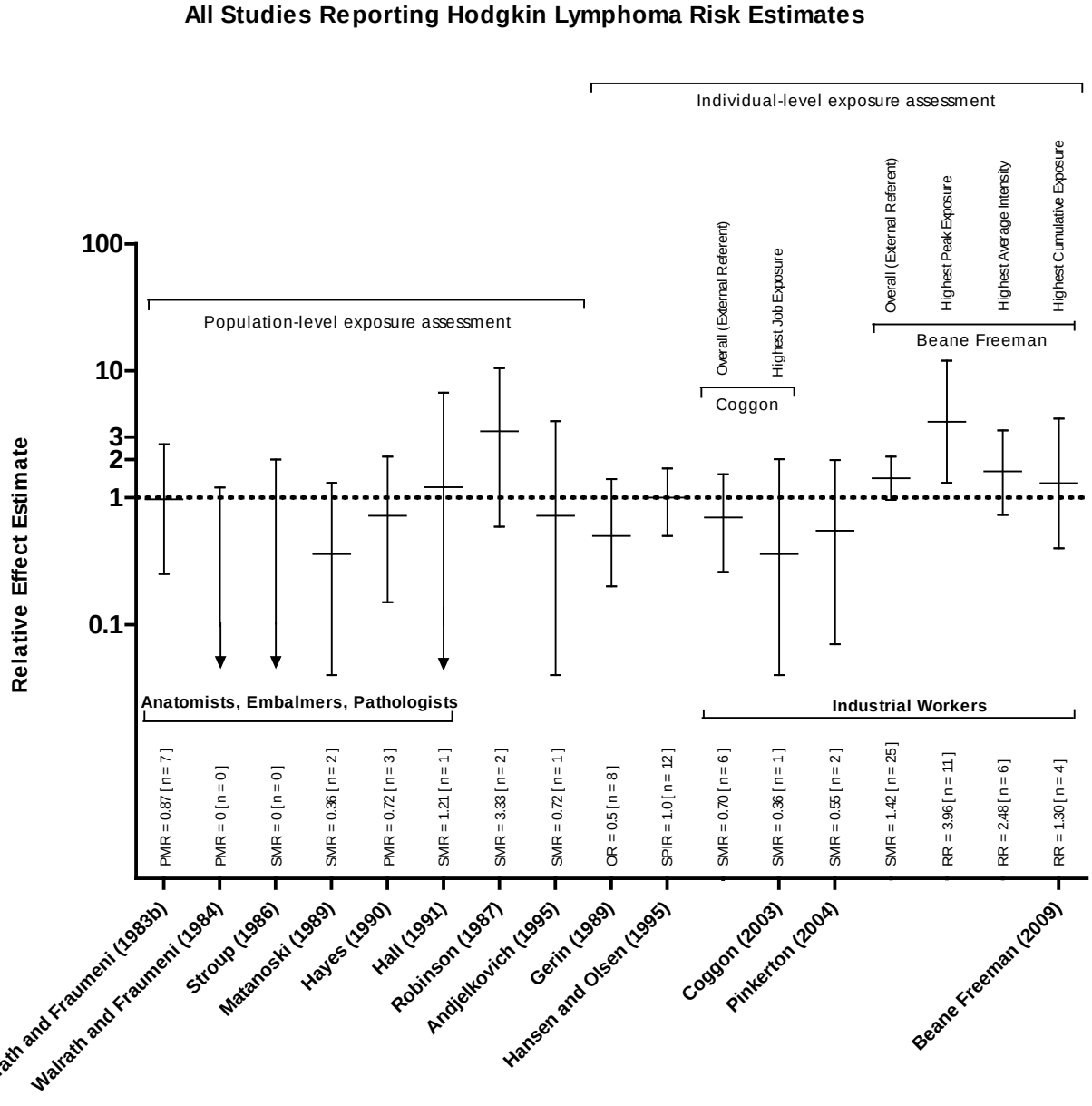


Figure 4-1. Epidemiologic studies reporting multiple Hodgkin lymphoma estimates. SMR: standardized mortality ratio. PMR: proportionate mortality ratio. RR: relative risk. OR: odds ratio. For each measure of association, the number of exposed cases is provided in brackets (i.e., [n=7]). For studies reporting results on multiple metrics of exposure, each metric is included; however, only the highest category of each exposure metric is presented in the figure.

Table 4-1. Epidemiologic studies of formaldehyde exposure and risk of Hodgkin lymphoma

Study	Exposures	Results: Effect estimate (95% CI) [# of cases]

Reference: Beane Freeman et al. (2009) with supplemental online tables.
 	
Population: 25,619 workers employed at 10 formaldehyde using or formaldehyde producing plants in the U.S. followed from either the plant start-up or first employment through 2004. Deaths were identified from the National Death Index with remainder assumed to be living. Vital status was 97.4% complete and only 2.6% lost to follow-up.
 	
Outcome definition: Death certificates used to determine underlying cause of death from Hodgkin disease (ICD-8: 201).
 	
Design: Prospective cohort mortality study with external and internal comparison groups.
 	
Analysis: RRs estimated using Poisson regression stratified by calendar year, age, sex, and race; adjusted for pay category compared to workers in lowest exposed category. Lagged exposures were evaluated to account for cancer latency.
 	
SMRs calculated using sex, age, race, and calendar-year-specific U.S. mortality rates.
 	
Related studies:
Blair et al. (1986)
Hauptmann et al. (2003)
	Exposure assessment: Individual-level exposure estimates based on job titles, tasks, visits to plants by study industrial hygienists, and monitoring data from 1966 through 1980.
	
	Median time weighted average (over 8 hours) =0.3ppm (range 0.01–4.3).
	
	Median cumulative exposure=0.6 ppm-years (range 0–107.4).
	
	Multiple exposure metrics including peak, average, and cumulative exposures were evaluated using categorical and continuous data.
	
	Duration and timing: Exposure period from <1946–1980. Median length of follow-up: 42 years. Duration and timing since first exposure were evaluated.
	
	Variation in exposure:
	For all variations in exposure:
	Level 1 (unexposed)
	
	Peak exposure:
	Level 2 (>0 to <2.0 ppm)
	Level 3 (2.0 to <4.0 ppm)
	Level 4 (≥4.0 ppm)
	Average intensity:
	Level 2 (>0 to <0.5ppm)
	Level 3 (0.5 to <1.0 ppm)
	Level 4 (≥1.0 ppm)
	Cumulative exposure:
	Level 2 (>0 to <1.5 ppm-yrs)
	Level 3 (1.5 to <5.5 ppm-yrs)
	Level 4 (≥5.5 ppm-yrs)
	
	Co-exposures: Exposures to 11 other compounds were identified and evaluated as potential confounders.
		Internal comparisons:
		Peak exposure
		1994 Follow-up:
		Highest peak RR=3.30 (0.98–11.10) (p-trend=0.04)
		2004 Follow-up:
		Peak exposure
		Level 1 RR=0.67 (0.12–3.6) [2]
		Level 2 RR=1.00 (Ref. value) [6]
		Level 3 RR=3.30 (1.04–10.50) [8]
		Level 4 RR=3.96 (1.31–12.02) [11]
		p-trend (exposed) = 0.01;
		p-trend (all) = 0.004
		
		Average intensity
		Level 1 RR=0.53 (0.11–2.66) [2]
		Level 2 RR=1.00 (Ref. value) [10]
		Level 3 RR=2.48 (0.84–7.32) [9]
		Level 4 RR=1.61 (0.73–3.39) [6]
		p-trend (exposed) = 0.05;
		p-trend (all) = 0.03
		
		Cumulative exposure
		Level 1 RR=0.42 (0.09–2.05) [2]
		Level 2 RR=1.00 (Ref. value) [14]
		Level 3 RR=1.71 (0.66–4.38) [7]
		Level 4 RR=1.30 (0.40–4.19) [4]
		p-trend (exposed) = 0.08;
		p-trend (all) = 0.06
		
		Duration of exposure
		No evidence of association (data not shown).
		
		Time since first exposure
		>0–15 yrs RR=1.00 (Ref. value)
		>15–25 yrs RR=1.54 (0.42–5.62)
		>25–35 yrs RR<1.54
		>35 yrs RR<1.54
		
		External comparisons:
		$SMR_{Unexposed}$ = 0.70 (0.17–2.80) [2]
		$SMR_{Exposed}$ = 1.42 (0.96–2.10) [25]

Table 4-1. Epidemiologic studies of formaldehyde exposure and risk of Hodgkin lymphoma

Study	Exposures	Results: Effect estimate (95% CI) [# of cases]
Reference: Pinkerton et al. (2004) **Population:** 11,039 workers in 3 U.S. garment plants exposed for at least 3 months. Women comprised 81.7% of the cohort. Vital status was followed through 1998 with 98.3% completion and only 1.7% lost to follow-up. **Outcome definition:** Death certificates used to determine both the underlying cause of death (UCOD) as well as all contributing multiple causes of death (MCOD) from Hodgkin's disease (ICD: 201). **Design:** Prospective cohort mortality study with external comparison group. **Analysis:** SMRs calculated using sex, age, race, and calendar-year-specific U.S. mortality rates. Results presented here are UCOD unless otherwise noted. **Related studies:** Stayner et al. (1985) Stayner et al. (1988)	**Exposure assessment:** Individual-level exposure estimates for 549 randomly selected workers during 1981 and 1984. Geometric 8-hr time-weighted average exposures ranged from 0.09–0.20 ppm. Overall geometric mean concentration of formaldehyde was 0.15 ppm, (GSD 1.90 ppm). Area measures showed constant levels without peaks. Historically earlier exposures may have been substantially higher. **Duration and timing:** Exposure period from 1955–1983. Median duration of exposure was 3.3 years. More than 40% exposures <1963. Median time since first exposure was 31.7 years. Duration and timing since first exposure were evaluated. **Variation in exposure:** Not evaluated **Co-exposures:** Study population specifically selected because industrial hygiene surveys at the plants did not identify any chemical exposures other than formaldehyde that were likely to influence findings.	**External comparisons:** SMR=0.55 (0.07–1.98) [2]
Reference: Coggon et al. (2003) **Population:** 14,014 British men employed in 6 chemical industry factories which produced formaldehyde. Cohort mortality followed from 1941 through 2000. Vital status was 98.9% complete and only 1.1% lost to follow-up. **Outcome definition:** Death certificates used to determine cause of deaths from Hodgkin's disease (ICD-9: 201). **Design:** Cohort mortality study with external comparison group. **Analysis:** SMRs based on English and Welsh age- and calendar-year-specific mortality rates. **Related studies:** Acheson et al. (1984) Gardner et al. (1993)	**Exposure assessment:** Exposure assessment based on data abstracted from company records. Jobs categorized as background, low, moderate, high, or unknown levels. **Duration and timing:** Occupational exposure during 1941–1982. Duration and timing since first exposure were not evaluated. **Variation in exposure:** Time weighted average exposure Level 1 (low) Level 2 (moderate) Level 3 (high) **Co-exposures:** Not evaluated. Potential low-level exposure to styrene, ethylene oxide, epichlorhydrin, solvents, asbestos, chromium salts, and cadmium.	**External comparisons:** SMR=0.70 (0.26–1.53) [6] **Within-study external comparisons:** Worked in 'High' exposure jobs SMR=0.36 (0.01–2.01) [1]

Table 4-1. Epidemiologic studies of formaldehyde exposure and risk of Hodgkin lymphoma

Study	Exposures	Results: Effect estimate (95% CI) [# of cases]
Reference: Andjelkovich et al. (1995) Cohort mortality study of 3,929 automotive industry iron foundry workers exposed from 1960−1987 and followed through 1989. SMRs calculated using sex-, age-, race-, and calendar-year-specific U.S. mortality rates.	Exposure assessment based on review of work histories by an industrial hygienist.	**External comparisons:** $SMR_{Unexposed}$ = 0.70 (0.01–3.88) [1] $SMR_{Exposed}$ = 0.72 (0.01–4.00) [1]
Reference: Hansen and Olsen (1995) **Population:** 2,041 men with cancer who were diagnosed during 1970–1984 and whose longest work experience occurred at least 10 years before cancer diagnosis. Identified from the Danish Cancer Registry and matched with the Danish Supplementary Pension Fund. Ascertainment considered complete. Pension record available for 72% of cancer cases. **Outcome definition:** Hodgkin's disease (ICD-7: 201) listed on Danish Cancer Registry file. **Design:** Proportionate incidence study with external comparison group. **Analysis:** Standardized proportionate incidence ratio calculated as the proportion of cases for a given cancer in formaldehyde-associated companies relative to the proportion of cases for the same cancer among all employees in Denmark. Adjusted for age and calendar time.	**Exposure assessment:** Individual occupational histories including industry and job title established through company tax records to the national Danish Product Register. Subject were considered to be exposed to formaldehyde if: 1) they had worked in an industry known to use more than 1 kg formaldehyde per employee per year; and 2) subjects longest single work experience (job) in that industry since 1964 was ≥10 years prior to cancer diagnosis All subjects were stratified based on job title as either low exposure (white-collar worker), above background exposure (blue-collar worker), or unknown (job title unavailable). **Duration and timing:** Exposure period not stated. Based on date of diagnosis during 1970–1984, and the requirement of exposure more than 10 years prior to diagnosis, the approximate period was 1960–1974. **Variation in exposure:** Not evaluated. **Co-exposures:** Not evaluated.	**External comparisons:** Overall (exposure to formaldehyde ≥10 years prior to cancer diagnosis) SPIR=1.0 (0.5–1.7) [12]
Reference: Hall et al. (1991) Cohort mortality study of 4,512 pathologists from the Royal College of Pathologists and the Pathological Society of Great Britain from 1974–1987. Vital status obtained from the census, a national health registry, and other sources. SMRs developed from the English and Welsh populations. **Related studies:** Harrington and Shannon (1975) Harrington and Oaks (1984)	Presumed exposure to formaldehyde tissue fixative.	**External comparisons:** SMR= 1.21 (0.03–6.71) [1]

Table 4-1. Epidemiologic studies of formaldehyde exposure and risk of Hodgkin lymphoma

Study	Exposures	Results: Effect estimate (95% CI) [# of cases]
Reference: Matanoski (1989) **Population:** 3,644 deceased U.S. male pathologists, derived from membership rolls of the American Association of Pathologists and Bacteriologists (1900-), the American Society for Experimental Pathology (1913-), and the American Medical Association (1912−1950). Mortality was followed through 1978. Death certificates obtained for 94% of potential study subjects (n=3,425), 3% from obituary notices (n=101) and 3% presumed dead (n=118). **Outcome definition:** Death certificates and obituary notices used to determine cause of death from Hodgkin's disease (ICD-8: 201). **Design:** Prospective mortality cohort study with two external comparison groups. The first comparison group was the U.S. male population. The second comparison group was comprised of members of a professional society of psychiatrists. **Analysis:** SMRs calculated using sex, race, age, and calendar-year-expected deaths from the U.S. population and psychiatrists.	**Exposure assessment:** Presumed exposure to formaldehyde tissue fixative. **Duration and timing:** Occupational exposure preceding death during 1900–1978. Duration and timing since first exposure were not evaluated. **Variation in exposure:** Not evaluated. **Co-exposures:** Not evaluated.	**External comparisons:** <u>Compared to the U.S. male population</u> SMR=0.36 (0.04–1.31) [2] <u>Compared to the psychiatrists</u> SMR=0.34 (0.06–1.12)† [2] †**Note:** EPA derived CIs using the Mid-P Method (See Rothman and Boice, 1979)
Reference: Hayes et al. (1990) **Population:** 4,046 deceased U.S. male embalmers and funeral directors, derived from licensing boards and funeral director associations in 32 states and the District of Columbia who died during 1975–1985. Death certificates obtained for 79% of potential study subjects (n=6,651) with vital status unknown for 21%. **Outcome definition:** Death certificates and licensing boards used to determine cause of death from Hodgkin's disease (ICD-8: 201). **Design:** Proportionate mortality cohort study with external comparison group. **Analysis:** PMRs calculated using sex, race, age, and calendar-year-expected deaths from the U.S. population.	**Exposure assessment:** Presumed exposure to formaldehyde tissue fixative. Exposure based on occupation which was confirmed on death certificate. Authors subsequently measured personal embalming exposures ranging from 0.98 ppm (high ventilation) to 3.99 ppm (low ventilation) with peaks up to 20 ppm. Authors state that major exposures are to formaldehyde and possibly gluteraldehyde and phenol. **Duration and timing:** Occupational exposure preceding death during 1975–1985. Of 115 deaths from lymphohematopoietic cancer, 66 (57%) were aged 60–74 years. Duration and timing since first exposure were not evaluated. **Variation in exposure:** Not evaluated. **Co-exposures:** Not evaluated.	**External comparisons:** PMR=0.72 (0.15–2.10) [3]

Table 4-1. Epidemiologic studies of formaldehyde exposure and risk of Hodgkin lymphoma

Study	Exposures	Results: Effect estimate (95% CI) [# of cases]
Reference: Gerin et al. (1989) **Population:** Male residents of Montreal, Canada aged 35–70 years. 4,510 eligible incident cancer cases were identified during 1979–1985 from 19 major area hospitals which report to the Quebec Tumor Registry over 97% of all cancer diagnoses from the Montreal area. Interviews and questionnaires completed for 3,726 subjects (83% of eligible cases). 18% of interviews were completed by next-of-kin. **Outcome definition:** Histologically confirmed diagnosis of Hodgkin's lymphoma (ICD: 201) **Design:** Population-based case-control study of 53 formaldehyde exposed men with Hodgkin lymphoma. Cases were compared with two groups; first, against other cancer cases excluding those diagnosed with lung cancer (n=2,599), and second against 533 male population controls selected from electoral list in the Montreal area. **Analysis:** ORs calculated by levels of a composite exposure index using logistic regression controlling for age, ethnic group, socio-economic status, smoking, and dirtiness of jobs held (white vs. blue collar). **Related studies:** Siemiatycki et al. (1987)	**Exposure assessment:** Individual-level exposure estimates developed based on a complete and detailed occupational history ascertained by interviewers using a standardized questionnaire. A team of chemists and hygienists translated each job into a list of potential formaldehyde exposures based on their confidence level, the frequency of exposure, and the duration of exposure. **Duration and timing:** Exposure period based on occupational histories prior to cancer diagnosis. Duration of exposure was evaluated. **Variation in exposure:** For cancer sites with fewer than 30 cases exposed to formaldehyde, results for the exposure subgroups were not shown. **Co-exposures:** Additional occupational and non-occupational potential confounders were included when the estimated exposure-disease OR changed by more than 10%.	**External comparisons:** <u>Compared to other cancers</u> OR=0.5 (0.2–1.2) [8] <u>Compared to population controls</u> OR=0.5 (0.2–1.4) [8]

Table 4-1. Epidemiologic studies of formaldehyde exposure and risk of Hodgkin lymphoma

Study	Exposures	Results: Effect estimate (95% CI) [# of cases]
Reference: Robinson et al. (1987) **Population:** 2,283 plywood mill workers employed at least one year during 1945–1955 followed for mortality until 1977 with vital status for 98% and death certificates for 97% of deceased. **Outcome definition:** Death certificates used to determine underlying cause of death from Hodgkin's disease as coded by trained nosologist using ICD-7:201. **Design:** Prospective cohort mortality study with external comparison group. A subcohort of 818 men co-exposed to formaldehyde and pentachlorophenol were also evaluated. **Analysis:** SMRs calculated using sex, age, race, and calendar-year-specific U.S. mortality rates.	**Exposure assessment:** Presumed exposure to formaldehyde-based glues used to manufacture and patch plywood. Sub-cohort of 818 men co-exposed to formaldehyde and pentachlorophenol worked for one year or more in the relevant exposure categories of veneer pressing and drying, glue mixing, veneer and panel gluing and patching. **Duration and timing:** Exposures during 1945–1955. Duration and timing since first exposure were not evaluated. **Variation in exposure:** Duration of exposure Latency (time since first exposure) **Co-exposures:** Pentachlorophenol	**External comparisons:** <u>Whole cohort of mill workers (n=2,283)</u> SMR=1.11(0.20–3.50) [2] <u>Sub-cohort of highly exposed workers (n=818)</u> SMR=3.33(0.59–10.49) [2]
Reference: Stroup et al. (1986) **Population:** 2,239 white male members of the American Association of Anatomists from 1888–1969 who died during 1925–1979. Death certificates obtained for 91% with 9% lost to follow-up. **Outcome definition:** Hodgkin's disease (ICD-8: 201) listed as cause of death on death certificates. **Design:** Cohort mortality study with external comparison group. **Analysis:** SMRs calculated using sex, race, age, and calendar-year-expected number of deaths from the U.S. population.	**Exposure assessment:** Presumed exposure to formaldehyde tissue fixative. **Duration and timing:** Occupational exposure preceding death during 1925–1979. Median birth year was 1912. By 1979, 33% of anatomists had died. Duration and timing since first exposure were not evaluated. **Variation in exposure:** Not evaluated. **Co-exposures:** Not evaluated.	**External comparisons:** SMR= 0 (0–2.0) [0]

Table 4-1. Epidemiologic studies of formaldehyde exposure and risk of Hodgkin lymphoma

Study	Exposures	Results: Effect estimate (95% CI) [# of cases]
Reference: Walrath and Fraumeni (1984) **Population:** 1,007 deceased white male embalmers from the California Bureau of Funeral Directing and Embalming who died during 1925–1980. Death certificates obtained for all. **Outcome definition:** Hodgkin's disease (ICD-8: 201) listed as cause of death on death certificates. **Design:** Proportionate mortality cohort study with external comparison group. **Analysis:** PMRs calculated using sex, race, age and calendar-year-expected number of deaths from the U.S. population.	**Exposure assessment:** Presumed exposure to formaldehyde tissue fixative. **Duration and timing:** Occupational exposure preceding death during 1916–1978. Birth year ranged from 1847–1959. Median age of death was 62 years. Most deaths were among embalmers with active licenses. Duration and timing since first exposure were not evaluated. **Variation in exposure:** Not evaluated. **Co-exposures:** Not evaluated.	**External comparisons:** Observed: 0 Hodgkin's disease deaths Expected: 2.5 Hodgkin's disease deaths PMR= 0 (0–1.20)† [0] †**Note:** EPA derived CIs using the Mid-P Method (See Rothman and Boice, 1979)
Reference: Walrath and Fraumeni (1983b) **Population:** 1,132 deceased white male embalmers licensed to practice during 1902–1980 in New York who died during 1925–1980 identified from registration files. Death certificates obtained for 75% of potential study subjects (n=1,678). **Outcome definition:** Hodgkin's disease (ICD-8: 201) listed as cause of death on death certificates. **Design:** Proportionate mortality cohort study with external comparison group. **Analysis:** PMRs calculated using sex, race, age, and calendar-year-expected numbers of deaths from the U.S. population.	**Exposure assessment:** Presumed exposure to formaldehyde tissue fixative. **Duration and timing:** Occupational exposure preceding death during 1902–1980. Median year of birth was 1901. Median year of initial license was 1931. Median age at death was 1968. Expected median duration of exposure was 37 years. Duration and timing since first exposure were not evaluated. **Variation in exposure:** Not evaluated. **Co-exposures:** Not evaluated.	**External comparisons:** Observed: 2 Hodgkin's disease deaths Expected: 2.3 Hodgkin's disease deaths PMR= 0.87 (0.15-2.87)† [7] †**Note:** EPA derived CIs using the Mid-P Method (See Rothman and Boice, 1979)

1

EXAMPLE 5 – Selecting Studies for Derivation of Toxicity Values

The example below demonstrates the selection of studies for derivation of toxicity values from a group of studies identified and evaluated as part of the hazard identification, as described in the "Dose-Response Analysis" section of the draft Handbook for IRIS Assessment Development.

Summary of issues covered:

- Overall description of the most suitable studies, given availability of a broader range of study designs.
- Summary of studies judged less suitable (or unsuitable); dismiss if possible or necessary (e.g., non-developmental acute or short-term studies; studies with only one relatively high dose).

Draft assessment text:

In Section 1.2.1, reproductive toxicities in male and female rodents were identified as hazards and liver and kidney toxicities were identified as potential hazards of dipentyl phthalate (DPP) exposure by the oral route. Studies within each effect category were evaluated using general study quality characteristics (as discussed in Section 6 of the *Preamble*) to help inform the selection of studies from which to derive toxicity values. Rationales for selecting studies and effects to represent each of these hazards are summarized below. The first objective was to derive an overall reference dose (RfD) for DPP. The second objective was to derive organ/system-specific reference values for DPP for each of the effects identified as hazards, to facilitate aggregating effects across phthalates when exposure is to a phthalate mixture.

A number of DPP studies supporting hazard identification were not considered for dose-response assessment, due to study designs that were less relevant for developing reference values for lifetime exposure and/or lacked evaluation of dose-response relationships (e.g., one dose level). These studies mainly comprised non-developmental studies with short-term and acute exposures ($\leq$10 daily doses) and evaluation of effects for $\leq$2 days following the last exposure, and included mechanistic studies. Most were conducted at relatively high doses ($\geq$2,000 mg/kg-day), generally using a single dose level, thus providing little information about dose-response relationships. The remaining DPP studies were reproductive or developmental studies. The reproductive study of Heindel et al. (1989; NTP, 1985), while including three dose levels (Task 2-continuous breeding phase), was not considered for dose-response assessment because a very high response (90% decrease in number of live pups per litter) was observed at the lowest dose tested, thus yielding little information about the shape of the dose response. The rat study of Liu et al. (2005) was not considered for dose-response assessment because it included only one dose level and because of the availability of other rat studies that used multiple, lower-dose levels and assessed a number of reproductive/developmental outcomes, including offspring mortality (Hannas et al., 2011; Howdeshell et al., 2008).

The studies selected for dose-response assessment consisted of two gestational exposure studies evaluating endpoints in rats exposed to DPP via gavage (Hannas et al., 2011; Howdeshell et al., 2008). Male reproductive toxicity was demonstrated in both studies. Effects observed included outcomes consistent with the "phthalate syndrome"—decreased fetal testicular testosterone production, decreased anogenital distance (AGD) in male pups, and retention of nipples/areolae in male pups after gestational exposure (Hannas et al., 2011; Howdeshell et al., 2008). Female reproductive toxicity was also demonstrated following gestational exposure to DPP by increased fetal/neonatal mortality (Hannas et al., 2011; Howdeshell et al., 2008). The Heindel et al. (1989; NTP, 1985) 29-week mouse study, while employing only one dose level (following task 3 [crossover mating phase] during week 19 to terminus of study in F0 male and female mice), represents the only evidence available for informing liver and kidney hazard following oral exposure to DPP. Thus, alterations in liver and kidney weights that were observed in adult mice exposed to DPP for up to 29 weeks were considered for dose-response assessment (Heindel et al., 1989; NTP, 1985).

EXAMPLE 6 – Dose-Response Modeling Output

The example below demonstrates the presentation of dose-response modeling results and output as it would appear in the supplemental information of IRIS assessmenst, as described in the "Dose-Response Analysis" section of the draft Handbook for IRIS Assessment Development.

Benchmark Dose Modeling Summary

This appendix provides technical detail on dose-response evaluation and determination of points of departure (POD) for relevant toxicological endpoints. The endpoints were modeled using the U.S. EPA's Benchmark Dose Software (BMDS, version 2.2). The following sections describe common practices used in evaluating the model fit and selecting the appropriate model for each endpoint, as outlined in the *Benchmark Dose Technical Guidance Document* (U.S. EPA, 2012). In some cases it may be appropriate to use alternative methods, based on statistical judgment; exceptions are noted as necessary in the summary of the modeling results.

Noncancer Endpoints

Evaluation of Model Fit

For each dichotomous endpoint (see Table 6-1), BMDS dichotomous models were fitted to the data using the maximum likelihood method. The following parameter restrictions were applied, unless otherwise noted: for the log-logistic model, restrict slope ≥ 1; for the gamma and Weibull models, restrict power ≥ 1; for the multistage models, restrict beta's ≥ 0. Each model was tested for goodness-of-fit using a chi-square goodness-of-fit test (χ^2 *p*-value < 0.10 indicates lack of fit). Other factors were also used to assess model fit, such as scaled residuals, visual fit, and adequacy of fit in the low-dose region and in the vicinity of the benchmark response (BMR).

For each continuous endpoint, BMDS continuous models were fitted to the data using the maximum likelihood method. The following parameter restrictions were applied, unless otherwise noted: for the polynomial models, restrict the coefficients b1 and higher to be nonnegative or nonpositive if the direction of the adverse effect is upward or downward, respectively; for the Hill, power, and exponential models, restrict power ≥ 1. Model fit was assessed by a series of tests as follows. For each model, first the homogeneity of the variances was tested using a likelihood ratio test (BMDS Test 2). If Test 2 was not rejected (χ^2 *p*-value ≥ 0.10), the model was fitted to the data assuming constant variance. If Test 2 was rejected (χ^2 *p*-value < 0.10), the variance was modeled as a power function of the mean, and the variance model was tested for adequacy of fit using a likelihood ratio test (BMDS Test 3). For fitting models using either constant variance or modeled variance, models for the mean response were tested for adequacy of fit using a likelihood ratio test (BMDS Test 4, with χ^2 *p*-value < 0.10 indicating inadequate fit). Other factors were also used to assess the model fit, such as scaled residuals, visual fit, and adequacy of fit in the low-dose region and in the vicinity of the BMR.

Model Selection

For each endpoint, the BMDL estimate (95% lower confidence limit on the benchmark dose [BMD], as estimated by the profile likelihood method) and Akaike information criterion (AIC) value were used to select a best-fit model from among the models exhibiting adequate fit. If the BMDL estimates were "sufficiently close," that is, differed by at most threefold, the model selected was the one that yielded the lowest AIC value. If the BMDL estimates were not sufficiently close, the lowest BMDL was selected as the POD.

Table 6-1. Noncancer endpoints selected for dose-response modeling for 1,2,4-trimethylbenzene

Species (generation) / Sex Endpoint	Doses and Effect Data				
Korsak (1996)					
Rat (Wistar) / Male	Dose (mg/kg-d)	0	123	492	1230
CNS: Pawlick (seconds)	No. of animals Mean ± SD	9 15.4 ± 5.8	10 18.2 ± 5.7	9 27.6 ± 4.6	10 30.1 ± 6.1
CNS: RotoRod	Incidence / Total	0 / 10	1 / 10	2 / 10	4 / 10

Modeling Results

Below are tables, graphs, and BMDS output summarizing the modeling results for each endpoint modeled.

Table 6-2. Summary of BMD modeling results for CNS: Pawlick in male Wistar rats exposed to 1,2,4-trimethylbenzene by inhalation for 3 months (Korsak, 1996); BMR = 1 SD change from the control mean

Model[a]	Goodness of fit		BMD_{1SD} (mg/m^3)	$BMDL_{1SD}$ (mg/m^3)	Basis for Model Selection
	p-value	AIC			
Exponential (M2)[b] Exponential (M3)	0.01	181.65	646	512	Only exponential model 4 provided an adequate fit, so it was selected.
Exponential (M4)	**0.35**	**173.57**	**150**	**80.8**	
Exponential (M5)	NA[c]	174.68	200	89.7	
Hill	NA[c]	174.68	186	88.6	
Polynomial 1°[d] Polynomial 2° Polynomial 3° Power	0.02	178.58	508	380	

[a]Constant variance models are presented (BMDS Test 2 p-value = 0.84), with the selected model in bold. Scaled residuals for the selected model for doses 0, 123, 492, and 1230 mg/kg-d were 0.36, –0.65, 0.53, and –0.19, respectively.

[b]For exponential model 4, the estimate of d was 1 (boundary). The models in this row reduced to exponential model 2.

[c]χ^2 test had insufficient degrees of freedom.

[d]For the power model, the power parameter estimate was 1 (boundary of parameter space). For the polynomial 2° and 3° models, the b2 and b3 coefficient estimates were 0 (boundary of parameter space). The models in this row reduced to the polynomial 1° model.

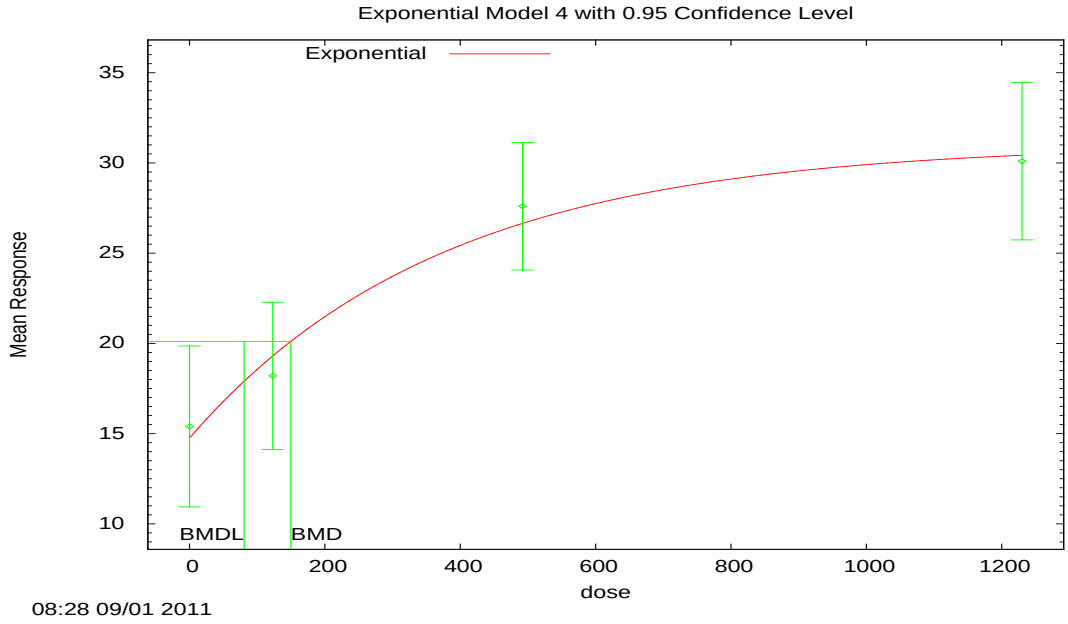


Figure 6-1. Plot of mean response by dose, with the fitted curve for exponential model 4 with constant variance. BMR = 1 SD change from the control mean; dose shown in mg/kg-day.

Exponential Model. (Version: 1.7; Date: 12/10/2009)

The form of the response function is: Model 4: Y[dose] = a * [c-(c-1) * exp{-b * dose}]

A constant variance model is fit.

Parameter Estimates

Variable	Estimate	Default Initial Parameter Values
lnalpha	3.35713	3.3338
rho	0	0
a	14.756	14.63
b	0.00266447	0.00210148
c	2.10364	2.16029
d	1	1

Table of Data and Estimated Values of Interest

Dose	N	Obs Mean	Est Mean	Obs Std Dev	Est Std Dev	Scaled Resid
0	9	15.4	14.76	5.8	5.358	0.361
123	10	18.2	19.31	5.7	5.358	−0.653
492	9	27.6	26.65	4.6	5.358	0.531
1230	10	30.1	30.43	6.1	5.358	−0.193

Likelihoods of Interest

Model	Log(likelihood)	# Param's	AIC
A1	-82.34222	5	174.6844
A2	-81.92912	8	179.8582
A3	-82.34222	5	174.6844
R	-98.61903	2	201.2381
4	-82.78544	4	173.5709

Tests of Interest

Test	-2*log(Likelihood Ratio)	Test df	p-value
Test 1 (Does response and/or variances differ among Dose levels, A2 vs. R)	33.38	6	<0.0001
Test 2 (Are Variances Homogeneous, A2 vs. A1)	0.8262	3	0.8432
Test 3 (Are variances adequately modeled, A2 vs. A3)	0.8262	3	0.8432
Test 4 (Does the model for the Mean fit, A3 vs. fitted)	0.8864	1	0.3464

Benchmark Dose Computation

BMR = 1 estimated standard deviation from the control mean

BMD = 149.743

BMDL at the 95% confidence level = 80.7575

Table 6-3. Summary of BMD modeling results for incidence of CNS: RotoRod in male Wistar rats exposed to 1,2,4-trimethylbenzene by inhalation for 3 months (Korsak, 1996); BMR = 10% extra risk

Model[a]	Goodness of fit		BMD_{10} (mg/m^3)	$BMDL_{10}$ (mg/m^3)	Basis for Model Selection
	p-value	AIC			
Gamma[b] Multistage 1° Multistage 2° Multistage 3° Weibull	0.93	32.33	229	129	Of the models that provided an adequate fit, the log-logistic model was selected based on lowest BMDL (BMDLs differed by more than threefold).
Log-Logistic	**0.97**	**32.16**	**194**	**93.9**	
Logistic	0.60	35.53	529	342	
Probit	0.63	35.40	490	318	
Log-Probit	0.58	35.40	426	233	

[a]Selected model in bold. Scaled residuals for the selected model for doses 0, 123, 492, and 1230 mg/kg-d were 0, 0.43, –0.15, and –0.09, respectively.

[b]For the gamma and Weibull models, the power parameter estimates were 1 (boundary of parameter space). For the multistage 2° and 3° models, the b2 and b3 coefficient estimates were 0 (boundary of parameter space). The models in this row reduced to the multistage 1° model.

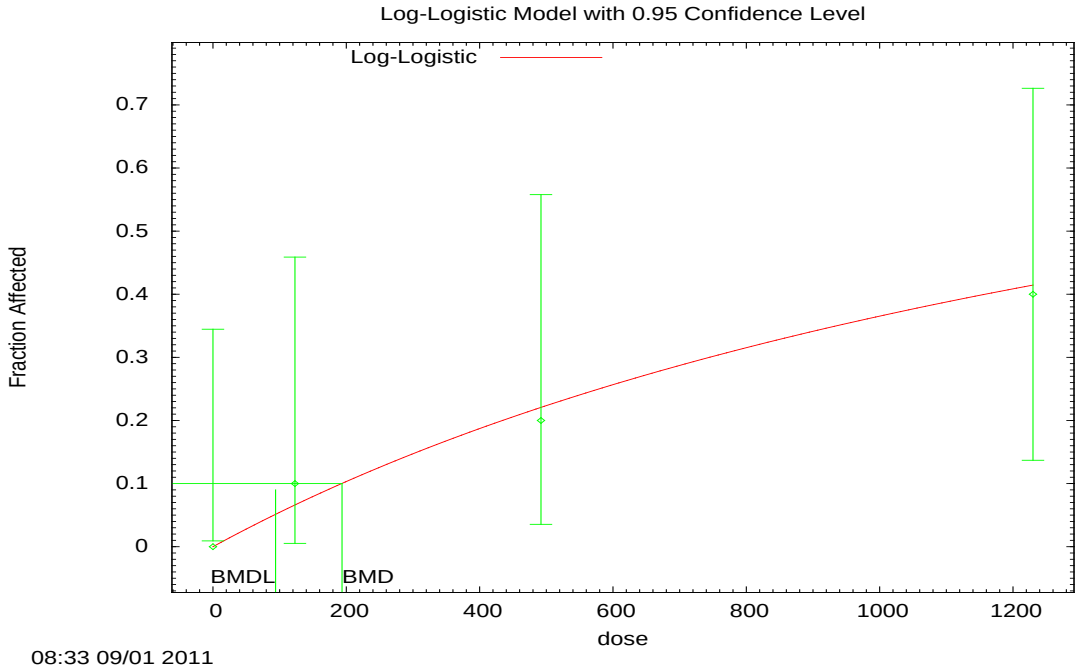


Figure 6-2. Plot of incidence rate by dose, with the fitted curve for the log-logistic model. BMR = 10% extra risk; dose shown in mg/kg-day.

Log-logistic Model (Version: 2.13; Date: 10/28/2009)

The form of the probability function is: P[response] = background+(1-background)/[1+EXP(-intercept-slope*Log(dose))]

Slope parameter is restricted as slope >= 1

Parameter Estimates

Variable	Estimate	(Default) Initial Parameter Values
Background	0	0
Intercept	-7.46289	-7.46166
Slope	1	1

Analysis of Deviance Table

Model	Log(likelihood)	# Param's	Deviance	Test d.f.	p-value
Full model	-14.985	4			
Fitted model	-15.0832	1	0.196433	3	0.9782
Reduced model	-18.5491	1	7.12817	3	0.06792

AIC: 32.1664

Goodness of Fit Table

Dose	Est. Prob.	Expected	Observed	Size	Scaled Resid
0	0.0000	0.000	0	10	0.000
123	0.0659	0.659	1	10	0.434
492	0.2202	2.202	2	10	-0.154
1230	0.4138	4.138	4	10	-0.089

Chi^2 = 0.22 d.f. = 3 P-value = 0.9743

Benchmark Dose Computation

BMR = 10% extra risk

BMD = 193.575

BMDL at the 95% confidence level = 93.947

Cancer Endpoints

For each endpoint (see Table 6-4), multistage cancer models, with coefficients restricted to be non-negative (beta's ≥ 0), were fitted to the data using the maximum likelihood method. Each model was tested for goodness-of-fit using a chi-square goodness-of-fit test (χ^2 p-value < 0.05[1] indicates lack of fit). Other factors were used to assess model fit, such as scaled residuals, visual fit, and adequacy of fit in the low-dose region and in the vicinity of the BMR.

For each endpoint, the BMDL estimate (95% lower confidence limit on the BMD, as estimated by the profile likelihood method) and AIC value were used to select a best-fit model from among the models exhibiting adequate fit. If the BMDL estimates were "sufficiently close," that is, differed by less than threefold, the model selected was the one that yielded the lowest AIC value. If the BMDL estimates were not sufficiently close, the lowest BMDL was selected as the POD.

[1] A significance level of 0.05 is used for selecting cancer models because the model family (multistage) is selected a priori (*Benchmark Dose Technical Guidance Document*, U.S. EPA, 2012).

Table 6-4. Cancer endpoints selected for dose-response modeling for diisononyl phthalate (DINP)

Species / Sex Endpoint	Doses and Effect Data					
Moore (1998b)						
Mice ($B6C3F_1$) / Female	Dose (mg/kg-d)	0	15.89	47.30	127.47	263.72
Hepatocellular adenoma or carcinoma	Incidence / Total	3 / 70	5 / 68	10 / 68	11 / 67	33 / 70

Modeling Results

The modeling results are summarized below.

Table 6-5. Summary of BMD model results for increased incidence of hepatocellular carcinomas and adenomas combined in female $B6C3F_1$ mice exposed to DINP in the diet for 2 years (Moore, 1998b); BMR = 10% extra risk

Model[a]	Goodness of fit		BMD_{10HED}	$BMDL_{10HED}$	Basis of model selection
	p-value	AIC			
Multistage 1°	0.30	281.78	55.6	42.6	All models provided an adequate fit. The multistage-cancer 4° model was selected based on lowest AIC.
Multistage 2°	0.25	282.65	82.0	45.2	
Multistage 3°	0.34	282.04	87.7	47.1	
Multistage 4°	**0.39**	**281.73**	**88.7**	**48.4**	

[a]Selected model in bold. Scaled residuals for the selected model for doses 0, 15.89, 47.30, 127.47, and 263.72 mg/kg-d were –0.48, 0.02, 1.10, –0.66, and 0.06, respectively. The cancer slope factor for the selected model was 0.1 / 48.4 = 0.00206.

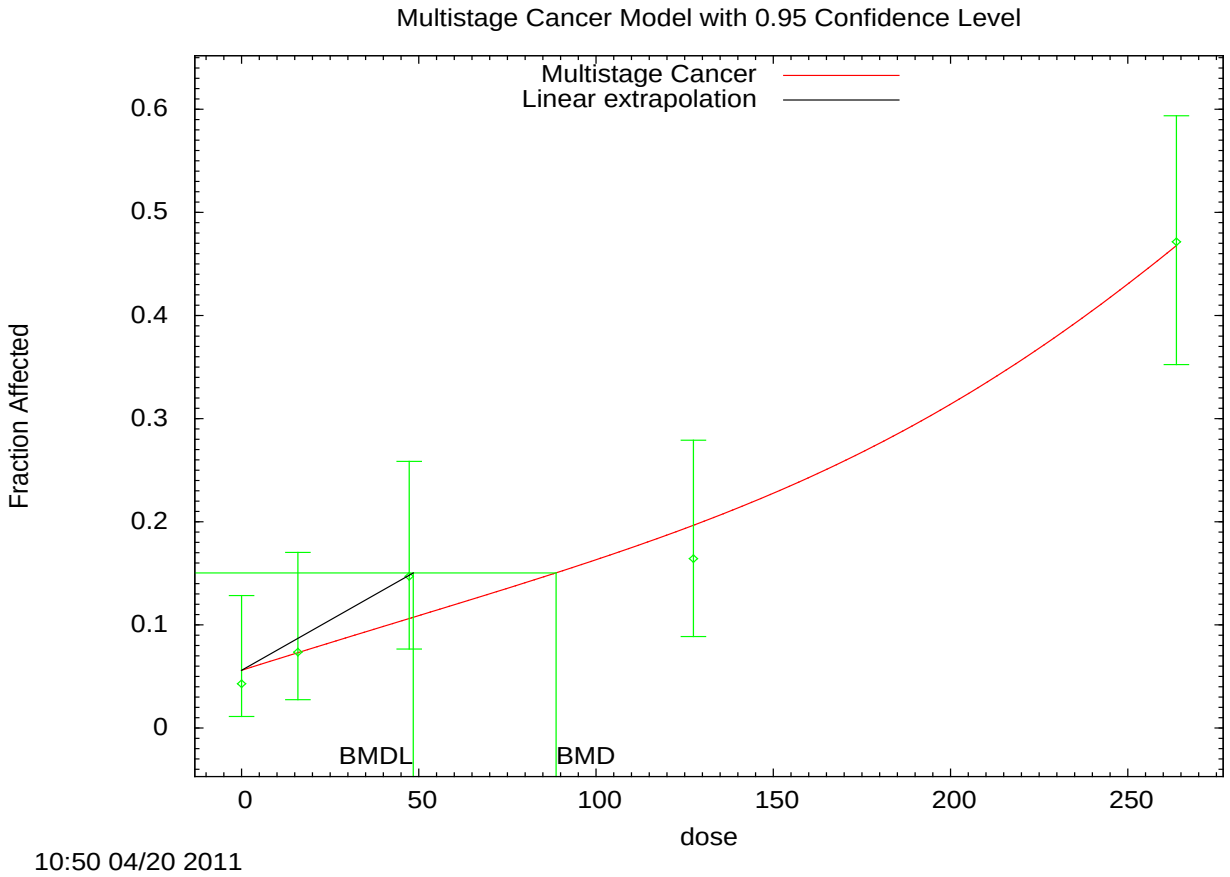


Figure 6-3. Plot of incidence rate by dose, with the fitted curve for the multistage-cancer 1° model. BMR = 10% extra risk; dose shown in mg/kg-day.

Multistage Cancer Model. (Version: 1.9; Date: 05/26/2010)

The form of the probability function is: P[response] = background + (1-background)*[1-EXP(-beta1*dose^1-beta2*dose^2-beta3*dose^3-beta4*dose^4)]

The parameter betas are restricted to be positive

Parameter Estimates

Variable	Estimate	(Default) Initial Parameter Values
Background	0.0559152	0.0659908
Beta(1)	0.00114845	0.000880186
Beta(2)	0	0
Beta(3)	0	0
Beta(4)	5.58108e-011	6.9501e-011

Analysis of Deviance Table

Model	Log(likelihood)	# Param's	Deviance	Test d.f.	P-value
Full model	-136.965	5			
Fitted model	-137.865	3	1.79969	2	0.4066

Reduced model	-162.082	1	50.233	4	<.0001

AIC: 281.73

Goodness of Fit Table

Dose	Est. Prob.	Expected	Observed	Size	Scaled Resid
0	0.0559	3.914	3	70	-0.476
15.89	0.0730	4.963	5	68	0.017
47.30	0.1061	7.214	10	68	1.097
127.47	0.1964	13.160	11	67	-0.664
263.72	0.4676	32.732	33	70	0.064

Chi^2 = 1.88 d.f. = 2 P-value = 0.3915

Benchmark Dose Computation

BMR = 10% extra risk

BMD = 88.7294

BMDL at the 95% confidence level = 48.4306

BMDU at the 95% confidence level = 163.388

Taken together, (48.4306, 163.388) is a 90% two-sided confidence interval for the BMD

Multistage Cancer Slope Factor = 0.00206481

EXAMPLE 7 – Considerations for Selecting Organ/System-Specific or Overall Toxicity Values

The example below demonstrates the derivation or selection of an organ/system-specific toxicity value for each organ or system affected by the agent, as well as an overall toxicity value for the agent to represent lifetime human exposure levels where effects are not anticipated to occur, as described in the "Dose-Response Analysis" section of the draft Handbook for IRIS Assessment Development.

Draft Assessment Text:

The candidate values presented in the table below are preliminary to the derivation of the organ/system-specific reference values. These candidate values are considered individually in the selection of a representative oral reference value for a specific hazard and subsequent overall RfD for benzo[a]pyrene.

Table 7-1. Effects and corresponding derivation of candidate values

Endpoint and Reference	POD_{HED}[a] (mg/kg-d)	POD type	UF_A	UF_H	UF_L	UF_S	UF_D	Composite UF	Candidate value (mg/kg-d)
Developmental									
Neurodevelopmental alterations in rats **Chen et al. (2012)**	0.09	$BMDL_{1SD}$	10	10	1	1	3	300	3×10^{-4}
Cardiovascular effects in rats **Jules et al. (2012)**	0.15	LOAEL	3	10	10	1	3	1,000	2×10^{-4}
Reproductive									
Decreased ovary weight and ovarian follicles in rats **Xu et al. (2010)**	0.37	$BMDL_{1SD}$	3	10	1	10	3	1,000	4×10^{-4}
Decreased intratesticular testosterone in rats **Zheng et al. (2010)**	0.24	NOAEL	3	10	1	10	3	1,000	2×10^{-4}
Decreased sperm count in mice **Mohamed et al. (2010)**	0.15	LOAEL	3	10	10	10	3	10,000	Not calculated due to UF > 3000[b]
Cervical epithelial hyperplasia in mice **Gao et al. (2011a)**	0.06	$BMDL_{10}$	3	10	1	10	3	1,000	6×10^{-5}
Immunological									
Decreased thymus weight in rats **Kroese et al. (2001)**	1.9	$BMDL_{1SD}$	3	10	1	10	3	1,000	2×10^{-3}
Decreased serum IgM in rats **De Jong et al. (1999)**	1.7	NOAEL	3	10	1	10	3	1,000	2×10^{-3}
Decreased serum IgA in rats **De Jong et al. (1999)**	5.2	NOAEL	3	10	1	10	3	1,000	5×10^{-3}

Endpoint and Reference	POD_{HED}[a] (mg/kg-d)	POD type	UF_A	UF_H	UF_L	UF_S	UF_D	Composite UF	Candidate value (mg/kg-d)
Decreased number of B cells in rats **De Jong et al. (1999)**	5.2	NOAEL	3	10	1	10	3	1,000	5×10^{-3}

[a]HED PODs were calculated using BW3/4 scaling (U.S. EPA,2011) for adult animal studies (Chen et al., 2011; Mohamed et al., 2010; Xu et al., 2010; and De Jong et al., 1999) but not for studies dosing early postnatal animals (Chen et al., 2012).

[b]As recommended in EPA's A Review of the Reference Dose and Reference Concentration Processes (U.S. EPA, 2002), the derivation of a reference value that involves application of the full 10-fold uncertainty factor in four or more areas of extrapolation should be avoided.

UF_A—A value of 3 (100.5 = 3.16, rounded to 3) was applied to account for uncertainty in characterizing toxicodynamic differences between rats and humans when an HED was calculated using BW3/4 scaling as uncertainty in characterizing toxicokinetic differences was accounted for through calculation of an HED using a standard DAF consistent with EPA guidance (U.S. EPA, 2011b). A value of 10 was applied when BW3/4 scaling was not employed to account for uncertainty in extrapolating from laboratory animals to humans because of the absence of information to characterize either the toxicokinetic or toxicodynamic differences between animals and humans following oral exposure to benzo[a]pyrene.

UF_H—A value of 10 was applied to account for potentially susceptible individuals because adequate information is not available to quantitatively estimate variability in human susceptibility. In the case of benzo[a]pyrene, insufficient information is available to quantitatively estimate variability in human susceptibility.

UF_L—A value of 1 was applied when the POD is based on dose-response modeling or a NOAEL; 10 when the POD is a LOAEL. In the case of benzo[a]pyrene, an UF_L of 1 was applied for LOAEL-to-NOAEL extrapolation because a BMR of a 1 SD change from the control mean in neurodevelopmental impairments was selected under an assumption that it represents a minimal biologically significant response level.

UF_S—A value of 1 was applied when dosing occurred during gestation or the early postnatal period that is relevant to developmental effects (U.S. EPA, 1991a); 10 when the POD is based on a subchronic study (studies in this table, other than the developmental toxicity studies, were 42–90 days in duration) to account for the possibility that longer exposure may induce effects at a lower dose.

UF_D—A value of 3 was applied to account for database deficiencies including the lack of a standard multigenerational study or extended 1-generation study that includes exposure from premating through lactation, considering that benzo[a]pyrene has been shown to affect fertility in adult male and female animals by multiple routes of exposure (see Section 1.1.2). Also, the lack of a study examining functional neurological endpoints following a more comprehensive period of developmental exposure (i.e., gestation through lactation) is a data gap, considering human and animal evidence indicating altered neurological development (see Section 1.1.1).

Figure 7-1 presents graphically the candidate values, UFs, and PODs, with each bar corresponding to one data set described in Table 7-1.

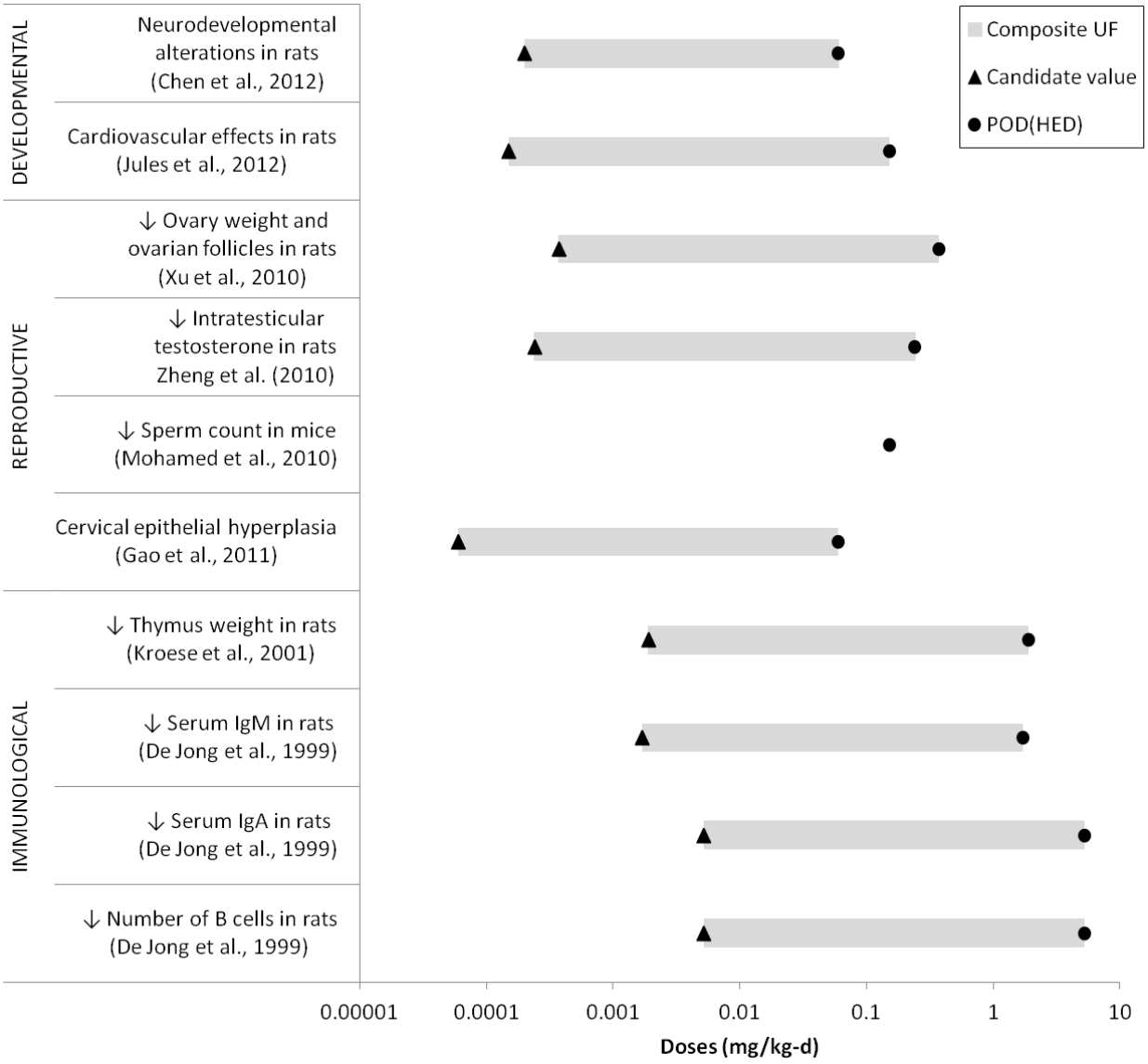


Figure 7-1. Candidate values with corresponding PODs and composite UFs.

Derivation of Organ/System-specific Reference Doses

Table 7-2 distills the candidate reference doses from Table 7-1 into a single value for each organ or system. These organ or system-specific reference values may be useful for subsequent cumulative risk assessments that consider the combined effect of multiple agents acting at a common site.

Table 7-2. Organ/system-specific RfDs and proposed overall RfD for benzo[a]pyrene

Effect	Basis	RfD (mg/kg-d)	Exposure description	Confidence
Developmental	Neurodevelopmental alterations	3×10^{-4}	Critical window of development (postnatal)	MEDIUM
Reproductive	Decreased ovary weight and ovarian follicles	4×10^{-4}	Subchronic	MEDIUM
Immunological	Decreased thymus weight and serum IgM	2×10^{-3}	Subchronic	LOW
Proposed Overall RfD	**Developmental toxicity**	**3×10^{-4}**	Critical window of development (postnatal)	**MEDIUM**

Developmental Toxicity

The candidate value based on neurodevelopmental impairment in rats (Chen et al., 2012) was selected as the organ/system-specific RfD representing developmental toxicity. This candidate RfD was selected because it is associated with the application of the smaller composite UF and because similar effects were replicated across other studies.

Reproductive Toxicity

The candidate RfD based on decreased ovary weight and ovarian follicle numbers in rats from the Xu et al. (2010) study was selected as the organ/system-specific RfD representing reproductive toxicity. The ovarian effects are supported by a large database of animal studies and human studies of exposure to benzo[a]pyrene and PAH mixtures. The data supporting cervical effects associated with oral benzo[a]pyrene exposure are limited to a single study; however, the finding is supported by corollary findings after i.p. exposure and by studies in humans.

Immunotoxicity

The candidate RfDs based on decreased thymus weight (Kroese et al., 2001) and serum IgM levels in rats (De Jong et al., 1999) were selected as the organ/system-specific RfD representing immunotoxicity. The observed decreases in thymus weight, IgM and IgA levels, and number of B cells associated with exposure to benzo[a]pyrene were determined to be representative of immunotoxicity. In combination, these effects provide more robust evidence of immunotoxicity. The candidate RfDs for decreased thymus weight (Kroese et al., 2001) and serum IgM levels in rats (De Jong et al., 1999) were equal and provided the most sensitive candidate RfDs; thus, these candidate RfDs were selected as the organ/system-specific RfDs representing immunotoxicity.

Selection of the Proposed Overall Reference Dose

For benzo[a]pyrene, multiple organ/system-specific reference doses were derived for effects identified as potential hazards from benzo[a]pyrene including developmental toxicity, reproductive toxicity, and immunotoxicity. To estimate an exposure level below which effects from benzo[a]pyrene exposure are not expected to occur, the lowest organ/system-specific RfD (3×10^{-4} mg/kg-day) is proposed as the overall reference dose for benzo[a]pyrene. This value, based on induction of neurodevelopmental alterations in rats exposed to benzo[a]pyrene during a susceptible lifestage is supported by several animal and human studies (see Section 1.1.1).

The overall reference dose is derived to be protective of all types of effects for a given duration of exposure and is intended to protect the population as a whole including potentially susceptible subgroups (U.S. EPA, 2002). Decisions concerning averaging exposures over time for comparison with the RfD should consider the types of toxicological effects and specific lifestages of concern. Fluctuations in exposure levels that result in elevated exposures during these lifestages could potentially lead to an appreciable risk, even if average levels over the full exposure duration were less than or equal to the RfD.

Furthermore, certain exposure scenarios may require particular attention to the risk-assessment population of interest in order to determine whether a reference value based on toxicity following developmental exposure is warranted. For example, the use of an RfD based on developmental effects may not be appropriate for a risk assessment in which the population of interest is post-reproductive age adults.

Confidence Statement

A confidence level of high, medium, or low is assigned to the study used to derive the RfD, the overall database, and the RfD itself, as described in Section 4.3.9.2 of EPA's *Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry* (U.S. EPA, 1994).

Confidence in the principal study (Chen et al., 2012) is medium-to-high. The study design included randomized experimental testing, blinded observations, culling of pups to account for nutritional availability, treatment-randomization, and controls for litter and nursing bias. Some informative experimental details were, however, omitted including the sensitivity of some assays at the indicated developmental ages and lack of reporting gender-specific data for all outcomes. Notably, these study limitations do not apply to the endpoint chosen to derive the RfD, and the overall methods and reporting are considered sufficient. Confidence in the database is medium, primarily due to the lack of a multigenerational reproductive toxicity study given the sensitivity to benzo[a]pyrene during development. Reflecting medium-to-high confidence in the principal study and medium confidence in the database, confidence in the RfD is medium.