

Supplementary Appendices

Supplementary Appendix I SEARCH TERMS

CAS numbers searched included: (94-75-7 or 1929-73-3 or 94-80-4 or 2008-39-1 or 5742-19-8 or 1928-43-4 or 1713-15-1 or 25168-26-7 or 94-11-1); 5742-17-6 or 2702-72-9 or 18584-79-7 or 2569-01-9); (5742-17-6 or 2702-72-9 or 18584-79-7 or 2569-01-9); and chemical names:(2(w)4(w)('d' or dichlorophenoxyacet? or dichlorophenoxy acet? or dichloro phenoxy acet? or dichloro phenoxyacet?))

The following search terms were used to identify endocrine relevant papers:

S (ESTROGEN? OR OESTROGEN? OR ESTRADIOL OR OESTRADIOL OR ESTROGEN RECEPTORS+NT/CT OR OESTROGEN RECEPTORS+NT/CT OR OESTROGENS+NT/CT OR ESTROGENS+NT/CT OR HELA 9903 OR HELA9903)

S AND (ANTIESTROGEN? OR ANTIOESTROGEN?)

S AND (ANDROGEN? OR TESTOSTERONE OR ANTIANDROGEN? OR ANTITESTOSTERONE)

S AND (H295R OR STEROIDOGEN? OR STEROID(1W)(SYNTHESIS OR PATHWAY OR METABOLI?) OR HORMONE(1A)(SYNTHESIS OR PATHWAY OR METAB? OR BIOSYNTH?))

S AND (AROMATASE OR CYTOCHROME(W)(P450 OR P 450) OR CYP19)

S ((UTERINE OR UTERUS OR FEMALE GENITALIA OR GENITALIA, FEMALE+NT/CT OR FEMALE GENITAL SYSTEM+NT/CT) AND (SIZE OR WEIGHT OR HYPERTROPH? OR ENLARG?))

S AND HERSHBERGER

S AND (PROSTATE OR SEMINAL VESICLE OR LEVATOR(1A)BULBOCAVERN? OR COWPERS GLAND OR PENIS)

S AND (TESTIS OR TESTES OR EPIDIDYM? OR MALE GENITALIA OR GENITALIA, MALE+NT/CT) AND (SIZE OR WEIGHT OR HYPERTROPH? OR ENLARG? OR HYPOTROPH? OR SHRINK?)

S MALE REPRODUCTIVE SYSTEM+NT/CT OR COWPERS GLAND OR LEVATOR(1A)BULBOCAVERN? OR PENIS OR GENITALIA, MALE+NT/CT OR MALE GENITALIA SYSTEM+NT/CT

S AND (RAT OR RATTUS) AND (PUBERTY OR PUBESC? OR SEX? MATUR?
OR PRE-PUTIAL OR VAGINA?(1A) OPEN?)
S AND (THYROID OR TSH OR THYROTROPIN OR THYROXIN? OR TRIODOTHYRONINE
OR THYROID HORMONES+NT/CT
S AND (CHOLINESTERASE OR CHOLINE ESTERASE)(1A)(INHIBIT? OR BLOCK? OR ANTAGONI?)
OR 9001-08-5(S)(BLOCK? OR INHIBIT? OR ANTAGONI?)
S (FROG OR AMPHIB? OR ZENOPUS OR RANA OR RANIDAE OR ANURA)
S (TOAD OR TADPOLE OR METAMORPHOSIS OR ANURA+NT/CT OR TOAD+NT/CT OR TADPOLE)
S METAMORPHOSIS
S AND (MINNOW OR PIMEPHALES OR MEDAKA OR ORYZIAS OR
CYPRINIDON OR POECILIA OR GUPPY OR ZEBRAFISH)
S(ZEBRA(W)(FISH OR DANIO) OR (DANIO OR BRACHIDANIO OR CYPRINUS)(W) RERIO)
S PITUIT? OR HYPOTHAL? OR HPG AXIS OR HYPOTHAL? PITUITARY GONADAL AXIS
OR HYPOPHYSIS OR HYPOPHYSEAL OR PITUITARY GLAND+NT/CT
OR HYPOTHALAMUS+NT/CT OR PITUITARY DISEASES+NT/CT
OR HYPOTHALMIC DISEASES+NT/CT OR HYPOTHALAMO-HYPOPHYSEAL SYSTEM+NT/CT

Supplementary Appendix II: *In vitro* Studies with Klimisch Scores of 1 or 2¹

A. Regulatory Toxicology (EDSP Tier 1) *in Vitro* Studies

As discussed previously in the publication, the high concentration tested in the EDSP Tier I *in vitro* assays (except steroidogenesis) was restricted below the maximum concentration specified in the EPA Guidelines, such that the concentration did not exceed the TSRC established through TK in the *in vivo* mammalian EOGRT Study.

LeBaron, et al., 2011a; (published in Coady et al., 2014)

In an estrogen receptor (ER) binding assay following the US EPA Guideline (US EPA, 2009a) with the exception regarding the high concentration noted above, uterine cytosol from Sprague Dawley rats was used as the source of ER to conduct a saturation binding experiment and a competitive binding experiment. The saturation binding experiment was conducted to demonstrate that the ER was present in reasonable numbers and was functioning with appropriate affinity for the radiolabeled ligand (³H-17β-estradiol). Subsequently, the competitive binding experiment was conducted in three independent assay runs to measure the binding of a single concentration of [³H]-17β-estradiol (1 nM) in the presence of increasing concentrations of 2,4-D (98.5% a.i) ranging from 10⁻¹¹ M to 10⁻⁴ M. Ethanol was used as a vehicle at a final concentration of <3% (v/v). The adequacy of the experimental conditions for the detection of ER binding was confirmed through the testing of reference chemicals: 17β-estradiol (strong positive control, radioinert), 19-norethindrone (weak positive control), and octyltriethoxysilane (negative control). In addition, control tubes were treated with the solvent used to dissolve the test material (*i.e.*, ethanol) for determination of full binding capacity and calculation of relative binding activity.

In the saturation binding experiment, the maximum binding capacity (B_{max}) was 59.28 fmol/100 μg protein and the dissociation constant (K_d) was 0.1032 nM. These results were within the acceptable range from the validation studies. All other saturation binding performance criteria and results from the competitive binding experiment indicated acceptable performance of the assay. The Scatchard plot indicated a linear response across the concentrations of ligand added. Nonspecific binding as a percent of total binding was 1.7%-8.6% across the entire concentration range in the saturation binding experiment.

In the competitive binding experiment, there were no appreciable alterations in 17β-estradiol ER binding activity at 2,4-D concentrations ranging from 10⁻¹¹ M to 10⁻⁴ M in three independent runs of the assay. Summary results for the ER binding assay of 2,4-D are shown in **Table S1**. No estimated log IC₅₀ and relative binding affinity (RBA) were calculated for 2,4-D because of the absence of binding activity. The logIC₅₀ values for 17β-estradiol and 19-norethindrone were -9.0 and -5.5, respectively. Compared to 17β-estradiol, the RBA for 19-norethindrone was 0.034%. The reference chemicals (17β-estradiol, 19-

¹ See Publication **Tables 3** and **4** for Klimisch scores for the regulatory EDSP and published studies, respectively.

norethindrone, and octyltriethoxysilane) met the QC performance criteria established in the test guideline in all instances, except for some slight deviations for 19-norethindrone in Run 3. However, because the acceptance ranges for the weak positive control are based on use of norethynodrel, these slight deviations from the suggested ranges are considered minor.

Based on the results of this assay, 2,4-D was classified as negative for ER binding.

LeBaron and Kan, 2011; Coady et al., 2014

In an estrogen transcriptional activation assay complying with the EPA Guideline (US EPA, 2009b) with the exception regarding the high concentration noted above, human ER α -HeLa-9903 cells cultured *in vitro* were exposed to 2,4-D (98.5% a.i.) at concentrations of 10⁻¹⁰, 10⁻⁹, 10⁻⁸, 10⁻⁷, 10⁻⁶, 10⁻⁵ and 10⁻⁴ M in DMSO (0.1% (v/v)) for 24 \pm 2 hours. The experiments were performed using 96-well plates and each 2,4-D concentration was tested in triplicate. Cells were exposed to the test agent for 24 \pm 2 hours to induce reporter (luciferase) gene products. Luciferase expression in response to activation of the estrogen receptor by 2,4-D was measured upon addition of a luciferase substrate and detection with a luminometer with acceptable sensitivity.

2,4-D was tested in four independent runs of the ER transcriptional activation assay. Cytotoxicity and precipitation were not observed at any of the test concentrations in these four runs.

The adequacy of the experimental conditions for the detection of ER agonism was confirmed using the following reference chemicals: 17 β -estradiol (strong agonist), 17 α -estradiol (weak agonist), 17 α -methyltestosterone (very weak agonist), and corticosterone (negative control). In addition, vehicle control cultures were treated with the solvent used to dissolve the test material (*i.e.*, DMSO) for determination of basal transcriptional activity. Results of the four concurrent reference chemicals included in each experiment generally fell within the acceptable ranges. There was slight variability between assays, but the overall robustness of the responses for 17 β -estradiol, 17 α -estradiol, 17 α -methyltestosterone, and corticosterone indicated that each assay run included in this assessment performed as expected. In most cases where values were outside of defined ranges, the values indicated increased sensitivity of the assay system (*i.e.*, decreased log PC₅₀ and log PC₁₀ for 17 α -methyltestosterone and 17 α -estradiol).

The RPC_{max} values for 2,4-D (maximum level of response induced by the test chemical expressed as a percentage of the response induced by 1 nM 17 β -estradiol on the same plate) were 0%, 8.8%, 5.3%, and 7.1% in Runs 1, 2, 3 and 4, respectively. The mean RPC_{max} of the four assays was 5.3%. At concentrations of 2,4-D ranging from 10⁻¹⁰ M to 10⁻⁴ M, there were no appreciable increases in estrogen receptor-mediated transcriptional activity. Based on these results, 2,4-D was judged to be negative for estrogen receptor-mediated agonism.

LeBaron et al., 2011b; (published in Coady et al., 2014)

In an androgen receptor (AR) binding assay following EPA Guidelines (US EPA, 2009c) with the exception regarding the high concentration noted above, ventral prostate cytosol from Sprague-Dawley rats was

used as the source of AR to conduct a saturation binding experiment and a competitive binding experiment. Saturation binding experiments were conducted to demonstrate that the AR was present in reasonable numbers and was functioning with appropriate affinity for the radiolabeled ligand (³H-R1881). Subsequently, the competitive binding experiment was conducted in three independent assay runs to measure the binding of a single concentration of ³H-1881 (1 nM) in the presence of increasing concentrations of 2,4-D (98.5% a.i.) ranging from 10⁻¹¹ M to 10⁻⁴ M. Ethanol was used as a vehicle at a final concentration of <3% (v/v). The adequacy of the experimental conditions for the detection of AR binding was confirmed through the testing of reference chemicals: R1881 (strong positive control, radioinert) and dexamethasone (weak positive control), both of which met established quality control criteria. In addition, control tubes were treated with the solvent used to dissolve the test material (*i.e.*, ethanol) for determination of full binding capacity and calculation of relative binding activity.

In the saturation binding experiment, the maximum binding capacity (B_{max}) was 3.245 fmol/100 μ g protein and the dissociation constant (K_d) was 0.4641 nM. Although these values were slightly below the range of values from the validation studies, the results were highly reproducible and all other performance criteria indicated acceptable performance of the assay. The Scatchard plot indicated a linear response across the concentrations of ligand added. Nonspecific binding as a percent of total binding was less than 20% across the entire concentration range in the saturation binding assays (range: 6.2-19.8%), with one exception (24.6%) at the high concentration (10 nM) in one assay.

In the competitive binding experiment, there were no appreciable alterations in R1881 AR binding activity at 2,4-D concentrations ranging from 10⁻¹¹ to 10⁻⁴ M in three independent runs of the assay. Summary results for the AR binding assay of 2,4-D are shown in **Table S2**. The estimated log IC₅₀ and relative binding affinity (RBA) for 2,4-D was not calculated due to the absence of activity. The log IC₅₀ values for R-1881 and dexamethasone were -9.0 and -4.4, respectively. Compared to R1881, the RBA for dexamethasone was 0.0027 %. In all instances, R1881 and dexamethasone met the QC performance criteria established in the test guideline.

Based on the results of this assay, 2,4-D was classified as negative for AR binding.

Coady and Sosinski, 2011; (Published in Coady et al., 2014)

In an aromatase assay conducted according to US EPA Guidelines (US EPA, 2009d) with the exception regarding the high concentration noted above, 2,4-D (98.5% a.i) was incubated with human recombinant aromatase and tritiated androstenedione ($1\text{-}\beta\text{[}^3\text{H(N)]-Androst-4-ene-3,17-dione}$; [^3H]ASDN) in ethanol at concentrations of 0, 10^{-10} , 10^{-9} , 10^{-8} , 10^{-7} , 10^{-6} , 10^{-5} , $10^{-4.5}$ or 10^{-4} M for 15 minutes to assess the effect of 2,4-D on aromatase activity. Aromatase activity was determined by measuring the amount of tritiated water produced at the end of 15 minute incubations for each concentration of chemical. Tritiated water was quantified using liquid scintillation counting (LSC). Three independent runs were conducted and each run included a full activity control, a background activity control, a positive control series (10^{-10} - 10^{-5} M) using 4-hydroxyandrostenedione (4-OH ASDN), a known inhibitor of aromatase, and the test chemical series (10^{-10} - 10^{-4} M) with 3 replicates per concentration. Standard curves generated from Runs 1, 2 and 3 with the known inhibitor, 4-OH ASDN, generally met performance criteria as described in the US EPA guideline. In addition, the standard curves generated for each run had a goodness of fit values that were equal to or exceeded 99%. The full activity and background controls in each of the runs were within the recommended ranges for the assay. Thus, Runs 1, 2, and 3 of the aromatase assay with 2,4-D were considered acceptable for determining whether or not the test material has the potential to inhibit aromatase activity *in vitro*.

The average response from three independent runs of the human recombinant aromatase assay with 2,4-D did not fit the four parameter regression model. Additionally, average aromatase activity for 2,4-D was similar to that of the full activity controls at all concentrations tested. Based on these results, 2,4-D was classified as a non-inhibitor of aromatase activity.

LeBaron et al., 2011c; (Published in Coady et al., 2014)

In a steroidogenesis assay following US EPA Guidelines (US EPA, 2009e), H295R cells cultured *in vitro* in 24-well plates were incubated with 2,4-D (98.5% a.i.) at concentrations of 0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, and 100.0 μM in triplicate for 48 hours. The test chemical vehicle was DMSO (0.1% (v/v)). The highest 2,4-D concentration tested is the assay limit concentration of 10^{-4} M. Testosterone and estradiol levels were measured using LC/APPI-MS/MS. Positive control chemicals prochloraz (an inhibitor) and forskolin (an inducer) at two concentration levels were tested with each run of the assay. In addition, control cultures were treated with the solvent vehicle for determination of acceptable basal hormone production levels (*i.e.*, minimally ~ 500 pg/ml testosterone and ~ 20 pg/ml estradiol). The test chemical, reference chemicals, and solvent controls were tested in replicates of 3/plate.

Three independent runs of the steroidogenesis assay were performed with 2,4-D. No cytotoxicity > 20% or precipitation were observed at any of the test concentrations of 2,4-D. The steroidogenesis assay of 2,4-D was conducted in accordance with OPPT 890.1550 test guidelines and EPA GLP regulations. The

assay performance criteria were met for all three runs of the assay, with minor exceptions not considered to affect the interpretation of the study results.

Measured concentrations of testosterone and estradiol in the culture media following 2,4-D exposure in Runs 1-3 are shown in **Table S3**. At concentrations of 2,4-D ranging from 10^{-10} M to 10^{-4} M, testosterone production was not statistically different than that of the solvent controls. The limit concentration of 2,4-D (100 μ M, 10^{-4} M) resulted in a small, statistically significant increase in estradiol production that was reproducible across assays.

Based on these data, 2,4-D was judged to produce slightly increased estradiol production at the highest (limit) concentration evaluated and not at any of the lower concentrations. This slight (1.2-fold) increase in estradiol production, however, did not meet the cut-off criteria for increase (*i.e.*, 1.5-fold) of estradiol established in the EDSP validation program as indicating a positive response for the steroidogenesis assay (Hecker et al., 2008) and is not considered biologically relevant.

B. *In Vitro* Studies Identified in the Published Literature

Kojima et al., 2004

Two hundred pesticides, including 2,4-D (purity between 95-100%), were tested for agonist and antagonist activity at the human ER α , human ER β , and human AR. Chinese hamster ovary cells (CHO K1) were plated in 96-well plates at a density of 8,400 cells per well in DMEM/F-12 media containing 5% FBS pretreated with dextran-coated charcoal. One day later, the cells were transiently transfected with ER α , ER β , or AR. Cells were simultaneously transfected with a luciferase reporter plasmid containing either an estrogen responsive element (ERE) or an androgen responsive element (ARE) upstream of the firefly luciferase gene, and a plasmid containing the *Renilla* luciferase gene (used as an internal control for transfection efficiency). Three hours post-transfection, cells were treated with 10^{-8} – 10^{-5} M of each test compound or 0.1% DMSO (vehicle control). To evaluate antagonist activity at the ER α , ER β , and AR, test compounds were added in the presence of either 10^{-11} M E2, 10^{-10} M E2, or 10^{-10} M DHT, respectively. After 24 hours incubation, the cells were lysed and firefly luciferase activity measured and normalized against expression of *Renilla* luciferase activity. A test compound was determined to have agonist activity if it showed an activity level within the range of concentrations tested that was $\geq 20\%$ that induced by 10^{-10} M E2, 10^{-9} M E2, or 10^{-9} M DHT at the ER α , ER β , or AR, respectively. Similarly, a test compound was determined to have antagonist activity if it inhibited the activity of the agonist by at least 20% within the range of concentrations tested. Mean results of three independent experiments were reported. Within the concentration range of 10^{-8} – 10^{-5} M, 2,4-D did not demonstrate agonist or antagonist activity at the ER α , ER β , or AR.

This study is considered to be of high quality and demonstrated an absence of ER α transactivation by 2,4-D, a lack of antagonist activity at the ER α and the absence of both agonist and antagonist activity at ER β , and an absence of ability to transactivate the AR or antagonize the activity of DHT at the AR.

Lin and Garry, 2000

Sixteen agrochemicals, including 2,4-D and 2,4-D isooctyl ester, were assessed for the ability to induce proliferation of an estrogen-responsive human breast cancer cell line. To examine cytotoxicity, MCF-7 cells were seeded in DMEM/F-12 medium (without phenol red; supplemented with 10% FBS [not charcoal-dextran treated] and 1% gentamicin) at a density of 10^5 cells/ml in 24-well plates. Test chemicals or solvent alone were added within 2-4 hours of incubation. After 72 hours in culture, the cells were evaluated for cell number and viability by flow cytometry. Results of cytotoxicity assays were not shown but concentrations tested reported to be non-cytotoxic. To assess cell proliferation, 10^4 MCF-7 cells/ml were seeded into 24 well plates using medium containing either 10% charcoal-dextran treated FBS or 10% non-charcoal-treated FBS. Forty-eight hours later, cells were treated with 10^{-9} M E2, solvent vehicle, or dilutions of test chemicals (0.1 to 10 $\mu\text{g}/\text{mL}$; duplicate wells per sample) and cultured for an additional seven days. The final volume of solvent was $<0.1\%$ of total volume. At end of culturing period, cells were harvested and examined for cell number and viability by flow cytometry. In all cases, three separate experiments were run with at least duplicate samples per concentration tested. Commercial grade 2,4-D LV4 (66.2% purity 2,4-D isooctyl ester) and 2,4-D amine (46.5% purity 2,4-D dimethylamine) both induced MCF-7 cell proliferation, with maximum proliferation at concentrations of 1 $\mu\text{g}/\text{ml}$. Reagent grade 2,4-D isooctyl ester and reagent grade 2,4-D acid, however, had no effect on MCF-7 cell proliferation, suggesting that the proliferative effect was due to additives or contaminants present in the commercial products.

Sun et al., 2012

Sun *et al.* (2012) developed luciferase reporter gene assays to measure the effects of 2,4-D on estrogen receptor alpha ($\text{ER}\alpha$), androgen receptor (AR) and thyroid hormone receptor (TR) activities. Vero cells, derived from African green monkey kidney epithelium, were used because they do not express any of these receptors endogenously. Cells were transfected with one of the $\text{ER}\alpha$, AR, or TR luciferase reporter plasmids and subsequently treated with 0.1-, 1-, 10- and 100-fold concentrations of 2,4-D maximum contamination level (MCL) of 0.03 mg/L in Chinese drinking water, i.e., cells were exposed to 0.003, 0.03, 0.3, and 3.0 mg/L 2,4-D. Agonist activity was measured as 20% relative effective concentration, which was defined as the concentration at which 2,4-D showed 20% of the maximum activity of E2, testosterone, or T3. Similarly, antagonist activity was measured as 20% relative inhibitory concentration when co-treated with 1×10^{-6} mg/L E2, 5×10^{-5} testosterone, or 5×10^{-4} mg/L T3.

2,4-D exhibited no agonist or antagonist activity for $\text{ER}\alpha$ even at concentrations 100 times the MCL. There was also no detectable androgenic or anti-androgenic activity at any of the 2,4-D concentrations tested; however, at 100x the MCL, 2,4-D was reported to enhance the effects of testosterone in the AR antagonist assay. The health significance of this finding is uncertain and it was limited only to the highest concentration tested, and in an assay system designed to test anti-androgenicity and not androgenicity. The concentration showing this effect was roughly equivalent to a 2,4-D serum concentration resulting from an *in vivo* dose greater than 5 mg/kg/day (Saghir *et al.*, 2013), the overall

2,4-D NOEL dose used as the point of departure dose to set the chronic RfD for 2,4-D, and thus this finding is not relevant to human health risk assessment.

Supplementary Appendix III

Ecotoxicological Studies with Klimisch Scores of 1 or 2²

A. Regulatory Toxicology (EDSP Tier 1) Studies

Coady et al., 2010; (published in Coady et al., 2013)

The Tier I EDSP Amphibian Metamorphosis Assay (AMA) is an assay in frogs designed to focus on potential thyroid effects. The study design corresponded with Guidelines: OPPTS 890.1100 (US EPA 2009f) and OECD 231. In brief, African clawed frog (*Xenopus laevis*) tadpoles were exposed to 2,4-D (98.6% purity) under continuous flow-through conditions for 21 days. Nominal test concentrations were selected based on prior acute toxicity studies with *Xenopus laevis* embryos and *Rana pipiens* larvae, which showed that the required high concentration equivalent to 1/3 of the LC50 (estimated maximum tolerated concentration (MTC)) was close to the 100 mg/L guideline limit concentration. The negative control was untreated laboratory dilution water. The study design is summarized in **Table S4**.

The concentration of 2,4-D in the exposure solutions was measured from all replicate tanks on a weekly basis during the test. Mean measured concentrations were 0.273, 3.24, 38.0 and 113 mg acid equivalents (a.e.)/L, representing 68.3, 81.0, 95.0 and 113% of nominal concentrations, respectively. Test substance was not detected in the negative control (Limit Of Quantitation (LOQ) = 0.120 mg a.e./L). A decline in measured concentrations was most likely due to biodegradation, which reduced 2,4-D concentrations in all exposure groups except for the high concentration, and proportionately had a greater effect at the two lowest nominal concentrations.

Hind limb length was normalized by snout-vent length to account for the effects of growth. Length and weight data for tadpoles reaching stages greater than Nieuwkoop and Faber (NF) stage 60 were excluded from statistical analyses in accordance with guideline requirements due to the typical drastic morphological changes in these parameters due to metamorphic progression at this stage and above (US EPA 2009f; Nieuwkoop and Faber, 1994).

The only performance criterion that was not met consistently in the AMA with 2,4-D was that a total of six replicate test vessels (4 replicate test vessels in the 0.273 mg./L 2,4-D treatment group and 2 replicate test vessels in the 3.24 mg /L 2,4-D treatment group) had measured concentrations of 2,4-D with coefficients of variation that exceeded 20%. This was likely due to biodegradation of the test material in the test vessels. Since the validity criteria and most of the performance criteria were fulfilled, and test concentrations were characterized, the study is considered valid, of high quality (Klimisch score of 1), and useful for assessing potential thyroid effects in the WoE.

There were no exposure related effects on survival or behavior. There were no statistically significant effects of 2,4-D on wet weight, snout-vent length (SVL), normalized hind limb lengths (HLL) or median NF developmental stage at either day 7 or day 21. There were no exposure-related histopathologic changes

² See Publication **Table 7** for Klimisch scores and comments for regulatory and published ecotoxicological studies

in the thyroid gland in any of the dosed groups, including no evidence of glandular atrophy, hypertrophy or follicular cell hyperplasia. The incidence of tall columnar cells lining the follicles (follicular cell hypertrophy) did not show any exposure-related differences and was interpreted to be within normal limits at all concentrations of 2,4-D. There was no evidence of any inflammatory or degenerative changes in the thyroid glands examined in any treatment group.

All other histopathologic criteria of all 2,4-D exposed tadpoles, including the overall size of the gland, the follicular lumen area, amount and type of colloid, and the follicular cell type and arrangement, were comparable to those of the controls.

In sum, there were no biologically significant exposure-related effects on the thyroid gland or other endpoints of this assay that are potentially subject to perturbations in the Hypothalamus-Pituitary-Thyroid (HPT) axis (*i.e.*, normalized hind limb length developmental stage, occurrence of asynchronous development). Thus, there is no evidence of a potential interaction with the HPT axis in this Tier 1 EDSP AMA tested to the limit concentration of 100 mg 2,4-D/L.

Marino *et al.*, 2010 (published in Coady *et al.*, 2013)

Marino *et al.* (2010) tested 2,4-D in a Fish Short Term Reproduction Assay (FSTR assay) (US EPA 2009g). Details of this study, which is designed to assess potential interactions with the estrogen, androgen or steroidogenesis pathways, are provided in Coady *et al.*, 2013. The study was conducted in compliance with OPPTS 890.1350 and OECD 229. Sexually mature adult fathead minnows (*Pimephales promelas*) were exposed to 2,4-D (98.6% purity) under continuous flow-through conditions for 21 days. Nominal test concentrations were selected based on acute toxicity tests and an early life stage test with fathead minnows (Alexander *et al.*, 1983; Mayes *et al.* 1990). The negative control was untreated laboratory dilution water. The study design is summarized in **Table S5**. Survival, fecundity, fertilization success, and general observations of health were noted daily during both the 14-day pre-exposure period and the 21-day exposure period. At termination on day 21, lengths and weights of surviving fish were measured, as well as secondary sex characteristics (including tubercle scores), gonado-somatic index (GSI), histopathology of gonads, and plasma vitellogenin concentrations.

The concentration of 2,4-D in the exposure solutions was measured from all replicate test vessels at weekly intervals during the test using a validated analytical method. Mean measured concentrations were 0.245, 3.14, 34.0 and 96.5 mg a.e./L, representing 61.3, 78.5, 85.0 and 96.5% of nominal concentrations, respectively. Test substance was not detected in the negative control (LOQ = 0.10 mg a.e./L). A decline in measured concentrations was most likely due to biodegradation, which reduced 2,4-D concentrations in all treatment groups but proportionately had a greater effect at the two lowest nominal concentrations.

The only performance criterion that was not met consistently was that a total of five replicate test vessels had measured concentrations of 2,4-D with coefficients of variation that exceeded 20%, likely due to degradation as described above. Since the validity criteria and most of the performance criteria were fulfilled, the study is considered valid, of high quality (Klimisch score of 1), and useful for assessing potential estrogen or androgen pathway interactions or effects on steroidogenesis in the WoE.

The overall results of the study are summarized in **Table 5** in the publication. The only significant effect compared to the controls was a decrease in fecundity (considered a non-specific finding) among fish exposed to 96.5 mg a.e./L 2,4-D. In the absence of effects upon the other, more specific endocrine-mediated endpoints, the isolated effect on fecundity at 100 mg a.i./L is most likely due to systemic toxicity and a generalized stress response. This concentration is relatively high (approximately 1/3 of the acute LC50 value in fish), is the limit concentration for the FSTR assay, and is a concentration which exceeds the Maximum Acceptable Toxicant Concentration (MATC) for larval fish survival in an early life stage toxicity test with fathead minnows (Mayes *et al.*, 1990).

In conclusion, 2,4-D does not appear to interact with the estrogen, androgen or steroidogenic pathways, or with the HPG axis, in fathead minnows tested up to the limit concentration in this EDSP Tier 1 FSTR assay.

Mitchell *et al.*, 1999

An avian single generation reproductive toxicity study (Mitchell *et al.*, 1999) of 2,4-D shows limited toxicity to quail and a lack of potentially endocrine-related effects.

In this study at Wildlife international, 2,4-D Acid (96.9% pure) was administered to adult Northern Bobwhite quail (*Colinus virginianus*) for 21 weeks via the diet at the following nominal concentrations: 0, 160, 400, and 1000 ppm active ingredient. Analysis of feed during the study indicated that measured concentrations of 2,4-D were close to nominal values (92-96% nominal). All adult birds were observed daily for signs of toxicity and abnormal behavior. Adult body weights were measured at test initiation, weeks 2, 4, 6, 8, and at adult termination. Feed consumption was measured weekly throughout the test. Following egg production, eggs were set weekly for incubation. During incubation, a subset of the eggs was assessed weekly for eggshell thickness. Eggs were additionally examined to detect eggs cracks, abnormal eggs, infertile eggs or embryo mortality. Eggs were hatched on Day 21 of incubation. Hatchlings were assessed (by pen) for group body weight. Weight of surviving offspring was again assessed at 14 days of age.

Study findings are summarized in **Table 6** of the Publication. There were no treatment-related mortalities, overt signs of toxicity or treatment-related effects upon body weight or feed consumption at any of the concentrations tested. Additionally, there were no treatment-related effects upon any of the reproductive parameters measured at the 160, 400 or 1000 ppm a.i. test concentrations, including fertility, egg shell thickness, embryo-mortality or hatching success. The no-observed effect concentration for northern bobwhite exposed to 2,4-D acid in the diet during the study was 1000 ppm a.i., the highest concentration tested. This study is considered valid; it predicts a very low hazard of reproductive toxicity of 2,4-D to birds and a low likelihood of endocrine-related effects on birds.

B. Other Ecotoxicological Studies Identified in the Published Literature

Reptiles

Crain et al., 1997

American alligator (*Alligator mississippiensis*) eggs collected from Lake Woodruff, Florida (relatively pristine environment) were topically dosed with 17 β -estradiol (positive control) or the herbicides atrazine or 2,4-dichloro-phenoxyacetic acid (2,4-D) (Chem Service, West Chester, PA; 97.6% purity) prior to the time of sexual differentiation. Atrazine will not be discussed further. Each egg received a single dose of 0.014, 0.14, 1.4, or 14 ppm estradiol or 0.14, 1.4, or 14 ppm 2,4-D applied topically to the eggshell in 50 ml of 95% ethanol. (This is a technique frequently used to transport compounds inside reptilian eggshells.) An ethanol solvent control and an untreated control were also included in the study design. Upon pipping, chorio-allantoic fluid (CAF) was collected and analyzed for estradiol and testosterone using radioimmunoassay. Hatchlings were then housed individually for 10 days. Just prior to euthanasia, blood was collected for estradiol and testosterone analyses. Following euthanasia, the right gonadal-adrenal mesonephros (GAM) complex was removed for an aromatase bioassay. The left GAM was removed and analyzed histopathologically to determine sex.

Results indicated that *in ovo* exposure of American alligators to 2,4-D did not alter sexual differentiation and had no effect on plasma steroids, or on gonadal aromatase activity. This study is included in the WoE because hormone levels were directly measured post 2,4-D exposure, the method of application has been used previously and penetration of test material demonstrated, and a positive control was tested.

Crain et al., 1999

This study extends from the study by Crain et al., (1997) above. The same test organisms, methods, and test compounds were used; however, hepatic aromatase activity was measured in hatchling alligators and a more complete histopathological analysis of the gonad was completed. Results indicated that 2,4-D exposure *in ovo* prior to the time of sexual differentiation had no effect on hepatic aromatase activity or testicular histopathology. This study is included in the WoE because aromatase was directly measured post 2,4-D exposure.

Spiteri et al., 1999

This study extends from the study by Crain et al., (1997) and Crain et al., (1999) above. The same test organisms, methods, and test compounds were used. Five eggs from each dose-treatment group were incubated at either a temperature that produces 100% males (33°C) or a temperature that produces 100% females (30°C). At ten days post-hatch, gonadal histology and hepatic steroidogenic activity were assessed in the hatchlings. Hatchlings in all the estradiol treatment groups were 100% female regardless of their incubation temperature. Hatchlings in the 2,4-D treatment group developed as males or females, as was expected based on their incubation temperature. Thus, there were no signs of sex reversal due to herbicide exposure. Compared to controls, 14 ppm estradiol caused a significant increase in the degeneration of the ovarian medulla and a significant increase in the height of Müllerian

duct epithelial cells. (The high dose of estradiol necessary to elicit these responses indicates that these histopathological endpoints are not as sensitive as gender reversal.) No significant histopathological differences were noted among the gonads of hatchlings exposed to 2,4-D. The publication also reiterated the results from hepatic aromatase assays of the hatchling alligators summarized in Crain *et al.*, 1999 above; it is not clear that any new data were developed on this parameter. The design and methods employed in the present study indicate that the study is scientifically valid. The results of this study suggest that egg exposure to concentrations up to 14 ppm 2,4-D do not cause significant alterations in gonadal structure or hepatic steroidogenic enzyme activity of hatchling American alligators. In contrast, all doses of estradiol (even as low as 0.014 ppm) caused feminization of prospective males. These data have relevance for the weight of evidence for endocrine activity since steroidogenesis and gonadal histopathology are relevant endpoints; also the test system was demonstrated to respond to a positive control. Under the conditions of exposure, 2,4-D, did not affect steroidogenesis or ovarian activity.

Supplementary Appendix IV

Mammalian Toxicological Studies with Klimisch Scores of 1 or 2³

A. Regulatory Toxicology Studies

Reproductive Toxicity Evaluations

Marty *et al.*, 2010; Marty *et al.*, 2013

The EOGRT study of 2,4-D (Marty *et al.*, 2010, published in Marty *et al.*, 2013) was specifically designed to provide sufficient information to assess whether endocrine targets are, in fact, altered with *in vivo* exposure, and to provide the basis for robust risk assessment of 2,4-D, including risk assessment protective for any potential endocrine effects. The study design was based on a modified reproductive toxicity study protocol that evaluated multiple endpoints across life stages, including estrogen, androgen and thyroid effects (Cooper *et al.*, 2006). This study design provides a reliable basis for establishing the potential of 2,4-D to interact with the estrogen, androgen or thyroid pathways and is considered a Tier 2 EDSP-equivalent assessment.

As noted, this study incorporated the extensive toxicokinetic information on 2,4-D in dose setting, to inform dose selection. The high dose dietary exposure concentrations, which differed for males and females, were predicted to just exceed the TSRC. This study also included additional measures of internal dosimetry. Based on these latter data, the high dose in males (800 ppm or approximately 40 mg/kg/day) adequately approached the TSRC, but the high dose in females (600 ppm or approximately 25-30 mg/kg/day) clearly exceeded the TSRC.

This study is reported in depth in a publication by Marty *et al.*, 2013. A summary of the study design and endocrine-related parameters for the EOGRT study follows:

CD®rats (27/sex/dose) were fed diets containing 0, 100, 300, and 600 (females) or 800 (males) ppm 2,4-D (98.75% purity), supplying approximately 0, 7, 21, or 40 mg/kg/day 2,4-D for adult females and 0, 6, 17, or 45 mg/kg/day 2,4-D for adult males for four weeks prior to breeding and continuing through breeding (up to two weeks), gestation (three weeks), and lactation (three weeks). Exposure of P (parental generation) males continued to ensure dosing covered a full spermatogenic cycle. P females were exposed until LD 22 (the end of the lactation period). P males and females were evaluated for systemic toxicity, as well as functional and structural evaluations of the reproductive systems. In addition, a satellite group of P females (12/dose) was exposed during pre-breeding, breeding, and GD 0-17, when they were euthanized to assess gross pathology, clinical pathology, thyroid hormones, TK and selected reproductive parameters during gestation.

³ Klimisch scores and comments for regulatory developmental, subchronic and chronic toxicity studies are in publication **Table 10**, and for published mammalian toxicological studies in publication **Table 11**.

Table S6 summarizes the initial group assignments:

Dietary concentrations of 2,4-D were adjusted during lactation (LD 7-21) by a factor of 2x or 3x to account for the large increase in feed consumption (2-3-fold) typical for rat dams in mid- to late lactation- LD 7-14 (Hanley *et al.*, 1985). Diets to weanlings were also adjusted until PND 35. These dietary concentration adjustments were intended to maintain more consistent doses of 2,4-D (mg/kg/day basis) across study phases and were based on TK information from the TK/rangefinding study of 2,4-D (Saghir *et al.*, 2008a). Dose adjustments are shown in **Table S7**.

Selected F1 offspring were maintained on the test diet until PND 60 (10/sex/dose), ~PND 70 (Sets 1a; 10/sex/dose and 2a; 10/sex/dose) or ~PND 90 (Set 2b; 10/sex/dose) and ~PND 139 (Set 3; 20/sex/dose). F1 Offspring were evaluated for effects on systemic toxicity (Set 1a), the nervous system (Set 1b), immune system (Set 2a and 2b), and the reproductive system and thyroid function (Set 3). 2,4-D TK was assessed in the Set 3 F1 offspring on PND 63 and 84.

Systemic toxicity: Systemic toxicity was limited to body weight decreases and decreased feed consumption in P females during lactation (evident before diet concentration adjustment), and decreased pup body weights in pups at 600 ppm. Animals during these life stages still received more compound on a mg/kg/day basis (despite dose adjustment) than did animals in other life stages and high-dose females were considered to have exceeded the TSRC. Renal toxicity was observed at the high dose in males and females, consistent with other studies of 2,4-D. Very slight renal effects were seen at the mid-dose; these are considered toxicologically insignificant. There was no indication of developmental neurotoxicity or developmental immunotoxicity in the study.

Reproductive toxicity: 2,4-D had no effects on estrous cyclicity or P reproductive indices, including mating, fertility, time to mating, gestation length, pre- and post-implantation loss and *corpora lutea* number (examined in satellite dams). There were slight non-statistically and non-biologically significant decreases in the fertility indices at 300 and 600 ppm in the P animals; these differences were within the laboratory historical control data (HCD) range. Litter size, pup survival, female reproductive organ histopathology and ovarian follicle counts were unaffected by 2,4-D exposure.

Decreased bilateral testis size was found in one P male in each of the 300 and 800-ppm groups. This finding was not considered exposure-related; the incidence was well within the laboratory HCD range. A similar but unilateral finding was noted in a control F1 male. This finding was not made in 2,4-D-exposed adult F1 offspring with longer 2,4-D exposures, which included higher mg/kg/day exposures during critical windows of development, and there was no overall exposure-related impact on testes weight. Decreases in testis size are a spontaneous occurrence in adult Crl:CD(SD) rats (Pettersen *et al.*, 1996). There were no exposure-related changes in sperm parameters or histopathological changes in male reproductive organs. Overall, there was no indication of reproductive toxicity caused by 2,4-D in this study.

Potential endocrine toxicity

Findings are summarized in publication **Table 8**.

Estrogenicity/Anti-estrogenicity: There were no exposure-related effects on developmental landmarks, including AGD or age and body weight at vaginal opening. There were no effects on female reproductive organ histopathology in either P or F1 offspring, or on quantitative ovarian follicle counts in F1 offspring evaluated for this parameter. There were no significant, exposure-related changes in reproductive organ weights in P or F1 PND 22 females or in F1 adult males and females. As noted above, reproductive indices and litter size and pup survival were not affected by 2,4-D.

When uterine weights were examined at termination in P and F1 females, there was a suggestion of increased uterine weight in the high-dose group. The increases in uterine weights were not statistically significant compared to control, and showed high variability because the stage of the estrous cycle was not uniform at necropsy. There was general correlation with a higher (but not abnormally higher) incidence of 600 mg/kg/day females compared to control in proestrus or estrus at the time of necropsy. Proestrus and estrus are normal stages of the estrous cycle; these show the greatest fluid imbibition in the uterus, and therefore have higher uterine weights and hypertrophy compared to uteri at other stages of the estrous cycle. There were no signs of enlarged uteri noted on gross examination, and no abnormal findings on histopathological evaluation. Examination of HCD showed control uterine weights fell below HCD. Further, there was no suggestion of a compound-related effect on estrous cyclicity (animals were cycling normally, with no signs of persistent estrus, difference from control in cycle length, or of irregular cycles). There was no effect on time to mating (often prolonged if cycles are irregular). Based on these factors, the perceived increase in uterine weight is considered to reflect normal variability and not an exposure-related finding.

Androgenicity/Anti-androgenicity: In P adult males, decreased seminal vesicle and prostate weights were seen at ≥ 300 ppm; prostate weights were not statistically different from control. These findings were not considered exposure-related because both absolute and relative seminal vesicle and prostate weights in the control group exceeded the laboratory HCD ranges. Organ weights for the 800 and 300 ppm-exposed males were within the laboratory HCD range. Additionally, there were no associated histopathological findings. Differences in prostate and seminal vesicle weight seen in parental males were not reproduced in F1 offspring with longer 2,4-D exposures, including exposures during critical life stages. There were no exposure-related alterations in reproductive or accessory sex gland weights or histopathology in adult F1 males. Neither P nor F1 males (PND 139) showed effects on sperm parameters.

Decreased testis weights in 600 ppm PND 22 F1 weanlings, which lacked corresponding histopathological findings, were attributed to decreased body weights. Body weight findings correlated strongly with the testis weight findings, and resulted from artifactual differences in PND 22 male pup body weights introduced during group assignment, possibly exacerbated by 2,4-D-related toxicity and/or palatability issues in the 600 ppm group, which exceeds the TSRC. Note a previous feed restriction study indicated that weanling organ weights, including testes, decrease with alterations in body weight

(Carney *et al.*, 2004). This differs from results in adult male rats, where testes weights are conserved in the presence of body weight decrements (Chapin *et al.*, 1993). No effects on testis weights were seen in adult males.

There was a slight delay (+1.6 days) in F1 preputial separation at 800 ppm, which was attributed to high-dose body weight decrements and decreased growth during lactation and post-weaning. Body weight at the time of puberty onset was similar in 800 ppm males and controls. High-dose males weighed the same as controls 1.6 days later when puberty onset occurred, indicating that 800 ppm 2,4-D affected the rate of growth in peri-pubescent male rats. The magnitude of this delay (1.6 days with 7-8% differences in body weights on PND 28-42) was consistent with a 1.8-day delay in age at preputial separation in a feed restriction study with a 10% body weight decrement (Marty *et al.*, 2003).

Other androgen-sensitive endpoints including anogenital distance (AGD) and nipple retention in male F1 offspring were not altered. These endpoints are considered highly sensitive to anti-androgenicity (Clark, 1999; McIntyre *et al.*, 2001, Wolf *et al.*, 2002, Hotchkiss *et al.*, 2004). No effects were seen on other androgen-sensitive endpoints examined in the F1 generation.

Thyroid Assessment: Thyroid hormones (T3, T4, and TSH), thyroid weights and/or histopathology were evaluated at multiple life stages. There was no consistent pattern of effects on thyroid parameters across life stages. The only thyroid findings considered likely to be exposure-related were in high-dose satellite GD 17 dams given 600 ppm 2,4-D. These females had non-statistically significant decreases in T4 and T3 with a corresponding increase in TSH. Thyroid histopathological alterations were seen in 3 of 12 dams in this group. The histopathological findings were comprised of smaller thyroid follicles with small vacuoles in the colloid that were suggestive of colloid resorption. There were no adverse pathological alterations and thyroid changes were not observed in LD 21 main study dams, indicating that this effect was transient. The thyroid findings are considered adaptive, as decreased colloid is a normal feedback initiated response to slight fluctuations in circulating thyroid hormone levels. The NOAEL for this study for thyroid effects is the highest dose tested.

High-dose alterations in thyroid function during pregnancy are plausible given that high-doses of 2,4-D have been shown to compete with T4 for serum protein binding (Florsheim *et al.*, 1963 and Van den Berg *et al.*, 1991)) and because pregnancy affects all aspects of thyroid hormone economy (Larsen and Ingbar, 1992; Fukuda *et al.*, 1980). This adaptive change in thyroid function in GD 17 dams occurred only at 600 ppm, an exposure level above TSRC (nonlinear TK was particularly marked in GD 17 dams compared to non-pregnant adult females). There were no thyroid effects at lower dose levels in GD 17 dams and no statistically or biologically significant effects on thyroid endpoints at any of the other life stages examined.

Importantly, no developmental neurotoxicity (DNT) effects were observed in F1 males and females. Adverse effects on DNT parameters have been reported for compounds that are thyroid-active. The thyroid-active agent, 6-propyl-2-thiouracil (PTU), has been associated with decreased brain size, altered brain development (i.e., impaired neuronal migration and white matter hypoplasia) (Behnam-Rassoli *et*

al., 1991; Schoonover *et al.*, 2004; Lavado-Autric *et al.*, 2003; Shibutani *et al.*, 2009); effects not seen in the EOGRT study. There were also no changes seen in myelin deposition, which was reported as a concern by Duffard's group in Argentina (Duffard *et al.* 1996). Special staining was performed in the EOGRT study to evaluate potential effects on myelin deposition in developmentally and peri-natally exposed animals; none were observed.

Thyroid perturbations during development could affect motor activity, which would present as incoordination during clinical observations in pups and persistent hyperactivity (increased motor counts) in post-weaning animals during motor activity testing (Goldey *et al.*, 1995); these effects were not seen in the EOGRT study. Goldey *et al.* (1995) also reported that thyroid perturbations in offspring increased startle response amplitude in adults, again suggesting a hyper-reactive response. There were no effects on startle amplitude in the EOGRT.

Pituitary and Adrenal: Absolute and relative pituitary gland weights were significantly decreased in the 800 ppm F1 Set 3 males; however, this finding was not considered exposure related because of the extremely slight nature of the difference. Pituitary weights in P1 or F1 Set1a males and females or F1 Set 3 females were not statistically different. There were no exposure-related histopathological changes in pituitaries at any dose level or life stage.

There was no indication of alterations in adrenal function as there were no effects on absolute or relative adrenal gland weights and no exposure-related histopathological findings in adrenals at any 2,4-D exposure level. Adrenal gland weights and histopathology often are altered in the presence of steroidogenesis inhibitors (USEPA, 2007); therefore, these data do not support effects of 2,4-D on steroidogenesis.

In conclusion, there was no evidence of adversely altered endocrine function in a comprehensive EOGRT study of 2,4-D. Slight adaptive effects were seen on thyroid hormone homeostasis at the high dose in a single life-stage, at a dose exceeding the TSRC and not relevant for human risk assessment.

Rodwell and Brown, 1985

This study was a two-generation OPP 83-4 Guideline reproductive toxicity study in Fischer 344 rats. 2,4-D (97.5% purity) was administered in the diet at nominal dose levels of 0, 5, 20, and 80 mg/kg/day (30/sex/dose) for one full generation and at 0, 5, and 20 mg/kg/day for the second generation. The 80 mg/kg/day group was dropped after the first generation because it exceeded a maximum tolerated dose, based on excessive mortality among the F1b pups following a mis-dosing during gestation and lactation. The mis-dosing resulted in all groups of F1b dams and pups being exposed to greater than nominal doses; high-dose dam exposure was ≥ 100 mg/kg/day. Because of the mis-dosing and several deficiencies this study is scored a Klimisch score of 2.

The study followed standard FIFRA Guideline procedures of the time with the following exceptions:

- Two matings were conducted per generation

- Mating was conducted so one male with one female for a 10-day period; if there was no evidence of mating the female was re-paired with a proven male for an additional five days. The second mating of each generation tested different pairings.
- Pups dying before Day 4 were examined for anomalies externally and internally
- All F1b surviving pups were sacrificed at weaning; these pups, plus a random 10/sex from other dose levels were necropsied and kidney, liver, ovary, uterus, testes, epididymides, accessory sex glands and gross lesions evaluated histopathologically.

There were no treatment-related signs of toxicity or effects on survival of parental animals. At 80 mg/kg/day, decreased weight gain was seen during gestation. No decreased feed consumption was seen, except in dams producing the F1b pups. Fertility indices and mean pre-coital intervals were comparable for control and treated groups in three of the four matings in the study. The fertility index was slightly lower (not statistically significant) compared to control at 80 mg/kg/day in the mating to produce the F1b litters; the interval where the mis-dosing occurred.

The F1a litters showed a slightly increased number of stillbirths and decreased pup body weight; the F1b litters had a significantly increased incidence of stillbirths and marked decreases in pup survival (mortality in 68.3% of all pups born). The severity of the effect on the F1b pups was attributed to excessive maternal toxicity due to the unintentional overdose during gestation and lactation, although pup survival may also have been adversely impacted due to direct toxicity to the pups. There was no clear indication that lactation was adversely affected or any other indication that the pup deaths resulted from endocrine-related toxicity. At ≥ 20 mg/kg/day, pup body weights were slightly decreased (in F1b pups only). No adverse effects on pups were seen at 5 mg/kg/day.

Malformations observed in F1b pups at ≥ 80 mg/kg/day (as noted above likely more than 100 mg/kg/day dose to dams and clearly exceeding the TSRC) included generalized edema, mal-aligned sternbrae, and bent limb bones; variations included 14th ribs, bent ribs and decreased ossification of the vertebral arches (a limited number of F1b control pups were evaluated, therefore the exposure-relationship of these findings cannot be determined). The pattern of these developmental effects, however, is similar to that seen in some rat developmental toxicity studies with 2,4-D acid, esters, and salts at maternally toxic doses (Charles *et al.*, 2001) that also exceeded the TSRC and is not characteristic of the types of anomalies associated with endocrine disruption (e.g., urogenital malformations).

The high-dose effects on pup mortality, survival, and malformations are attributed to direct toxicity of 2,4-D on dams and/or pups; there is no evidence of an endocrine-related pattern. Additionally, no effects were noted on pre-coital length, fertility, or on histopathological evaluation of the accessory sex glands, epididymides, testes, ovaries, or uterus.

The gross and microscopic evaluation of uteri in the PND 28 offspring showed no evidence of imbibed uteri or uterine lining proliferation, even at the excessively toxic dose to the F1b weanlings. F344 pups tend to develop more slowly than do CD rats, and PND 28 pups although close to puberty are unlikely to be cycling, thus reducing uterine weight variability and making any estrogenic effect on the uterus more

readily observable. This evaluation therefore provides additional confidence that potential estrogenic effects do not result from 2,4-D exposure.

One finding suggesting possible endocrine toxicity was that the length of gestation was statistically significantly prolonged (by 1 day) in the production of the F1b pups at ≥ 80 mg/kg/day, compared with controls. Gestation may be prolonged because of difficulties in parturition, hormonal imbalance, or delays in implantation. The first alternative is unlikely because no evidence of dystocia (prolonged or difficult labor) was reported. Another possible explanation is that the females were at a more advanced age at the time of the second littering and may have been more susceptible to the test compound at the high dose which was well above the TSRC. The latter supposition is reinforced by the absence of this finding in the F1a littering (exposed at a lower dose) and is consistent with the absence of this finding in the Saghir *et al.* (2008a) range-finding study, in which a similar dose did not result in prolonged gestation. It should be noted that the F344 rat typical gestation period is 21-22 days; a day longer is likely of marginal biological significance. At most, the finding of prolonged gestation in the F1b litters provides equivocal evidence of a potentially treatment-related hormonal imbalance resulting from 2,4-D exposure at a dose significantly exceeding the RCST (and exceeding a classically defined MTD).

A second finding potentially showing endocrine toxicity was that the 80 mg/kg/day F1a pups showed a statistically significant change in sex ratio (109 males and 71 females), compared with the controls. This finding was not repeated in the F1b pups (at a higher dose) and is considered unlikely to be exposure related because of the lack of consistency. Additionally, there were no parallel exposure related effects on the sex ratio in the F1-extended one generation dietary toxicity study for 2,4-D (Marty *et al.*, 2010), or in the range-finding for the latter study (Saghir *et al.*, 2008a).

The Rodwell and Brown (1985) two-generation study was a complete and relatively well-designed assessment for its time, including adequate numbers of litters in each dose group for evaluation, with two litterings per generation. It was more comprehensive (because of the additional histopathological evaluations of weanlings) than required under the guidelines current at the time (OPP 83-4).

However, there were some study deficiencies to consider when evaluating these data for evidence of endocrine effects:

- Excessive toxicity occurred in the F1b animals at the high-dose level, likely due to an accidental mis-dosing of this group - actual dose exceeded 100 mg/kg/day; therefore only two dose levels were tested for the second generation.
- Additional parameters now recommended by guidelines on reproductive toxicity testing, which are relevant to assessment of potential endocrine disruption, were not included: estrous cyclicity; pituitary weights; developmental landmarks; and evaluations of sperm motility, morphology, or counts.
- F344 rats are not recommended for evaluation in the EDSP Tier 1 pubertal study guidelines (Sprague-Dawley rats are preferred) (US EPA, 2009h and US EPA 2009i).

- No thyroid-related parameters were evaluated as part of this study.

A summary of results for the Rodwell and Brown (1985) study is presented in publication **Table 9**. In the 2,4-D reproductive toxicity study, there were no treatment-related effects on pre-coital length, suggesting that the females had generally normal estrous cycling activity. There was no evidence of adverse treatment-related effects on the testes, which were evaluated microscopically in both adults and pups. The general reproductive success in this study suggests an absence of any potent androgen or estrogen-related toxicity; thyroid related endpoints were not assessed. Although pup weight gain and viability were compromised at ≥ 80 mg/kg/day, the severity of the effect on the F1b pups was attributed to the excessive maternal toxicity at this dose caused by inadvertent overdosing (to at least 100 mg/kg/day). There was no indication that lactation was adversely affected or other indication that the pup deaths resulted from endocrine-related toxicity. As discussed previously, the pattern of malformations seen at the high and maternally toxic dose tested generally paralleled findings in developmental toxicity studies of 2,4-D and related compounds at doses that far exceeded the TSRC, and were not typical for malformations relating to endocrine toxicity.

Developmental Toxicity

Rodwell, 1983

This study was a conventional OPP 83-3 Guideline developmental toxicity evaluation in F344 rats. The study design and results were published by Charles *et al.* (2001) along with the results of evaluations of selected 2,4-D esters and salts, which is summarized subsequently with relevant published data. 2,4-D (97.5% pure) was administered by gavage at doses 0, 8, 25, and 75 mg/kg/day from gestation day (GD) 6-15; rats were sacrificed and caesarian-sectioned on GD 20. No adverse effects were seen on corpora lutea, number of implantations, fetal survival, fetal sex ratio, urogenital malformations, any variations or malformation or fetal body weight after doses to the dams up to 75 mg/kg/day on gestational days (GD) 6-15. The high dose produced relatively slight maternal toxicity, although data from other studies would predict that the high dose by gavage would far exceed the renal clearance saturation threshold. There was no evidence of endocrine disruption (based on absence of uro-genital malformations in this study).

Hoberman, 1990

This study was a conventional OPP 83-3 Guideline developmental toxicity in New Zealand white rabbits. The study design and results were published by Charles *et al.* (2001) along with the results of evaluations of selected 2,4-D esters and salts, which are summarized subsequently with relevant published data. Rabbits were dosed by gavage with 0, 10, 30 or 90 mg/kg/day 2,4-D (97.5% pure) from GD 6-18. No adverse effects were seen on corpora lutea, number of implantations, fetal survival, urogenital malformations, any variations or malformation or fetal body weight after a dose to the dams of 90 mg/kg/day (HDT) on gestational days (GD) 6-18. The high dose produced maternal toxicity, including two abortions. The fetal sex ratio was altered at the high dose; this finding is considered unlikely to reflect endocrine toxicity because genotypic sex determination occurs prior to the start of dosing. Dantzer and Wright (2003) reported OAT-1 transporter in rabbit renal proximal tubules that actively transported p-aminohippurate, the classical renal organic anion transporter substrate. Consistent with

the presence of OAT-1 transporter in kidney, intravenous dosing of pregnant rabbits has demonstrated that renal clearance of 2,4-D was saturated at doses of 10 and 40 mg/kg (Sandberg *et al.*, 1996), suggesting that high-dose oral effects limited to 90 mg/kg/day likely occurred under conditions of non-linear toxicokinetic behaviors in rabbits. There was no evidence of endocrine disruption (based on absence of urogenital malformations in this study).

Subchronic and Chronic Studies

Schulze, 1991a

Male and female F344 rats were fed 2,4-D (96.1% pure) in the diet at 0, 1, 15, 100, or 300 mg/kg/day for approximately 13 weeks (10/sex/dose) in an OPP 82-1 Guideline study. The study design and results were published by Charles *et al.* (2006a) along with the results of evaluations of a selected 2,4-D ester and salt, which are summarized subsequently with relevant published data. Endocrine-related parameters in the Schulze (1991a) study included thyroid hormones (T3 and T4), ovary, vagina, testes with epididymides, thyroid, adrenal and pituitary weights and histopathology, and uterus histopathology.

Females fed 300 mg/kg/day exhibited depressed activity. At 13 weeks, mean body weight and body weight gain of male rats fed 100 or 300 mg/kg/day were decreased (7.5% and 23%, respectively). Mean body weights of female rats fed 300 mg/kg/day at 13 weeks were decreased (28%). The two high doses in this study clearly exceeded the renal clearance saturation threshold, and the high dose exceeded a classically defined MTD. Systemic toxicity was marked at 300 mg/kg/day. Renal toxicity was present at 100 and 300 mg/kg/day based on renal weight changes and/or renal histopathology. High dose effects also included effects on eyes, hearts and lungs.

Findings potentially related to endocrine modulation were limited to high systemically toxic doses above the TSRC:

- T4 was depressed in the males and females at 100 or 300 mg/kg/day at weeks 6 and 13. T3 levels were decreased in females fed 100 and 300 mg/kg/day at 13 weeks. T3 was depressed in the males fed 300 mg/kg/day at 13 weeks. (TSH was not evaluated). Males had higher absolute thyroid/parathyroid weights at 300 mg/kg/day and relative weights at 100 and 300 mg/kg/day. Females at 300 mg/kg/day had higher relative thyroid/parathyroid weights. Increased follicular cell hypertrophy was observed in the thyroid glands of females at 300 mg/kg/day (8/10). No exposure-related findings were noted in the thyroid of males.
- Grossly, the cortex of the adrenal glands was pale in males fed 300 mg/kg/day (5/10). No changes were noted in the medulla. Males fed 100 and 300 mg/kg/day had dose related higher relative adrenal gland weights. Females fed 300 mg/kg/day had lower absolute adrenal gland weights. There was an increased incidence of hypertrophy in the zona glomerulosa of the adrenal gland cortex in males fed 300 mg/kg/day, and females fed 100 and 300 mg/kg/day.

- The ovary weight was increased at 300 mg/kg/day; there was no histopathological correlate.
- The testes at 300 mg/kg/day were soft (8/10) and small (7/10). Epididymides showed similar findings at this dose. Lower absolute and relative testes weights were noted at 300 mg/kg/day. Increased testicular atrophy was seen histopathologically at this dose.
- Females fed 300 mg/kg/day had lower absolute and relative pituitary weights. Males fed 300 mg/kg/day had lower relative pituitary weights. There was no histopathological correlate to this finding in either sex and it is considered unlikely to be exposure-related.
- Histopathological changes were not observed in the following tissues: pituitary, parathyroid glands, epididymides, ovary, uterus, and vagina.

Gorzinski et al., 1981a

Male and female F344 rats were fed 2,4-D (97.3% pure) in the diet at doses approximating 0, 15, 60, 100, or 150 mg/kg/day for approximately 13 weeks (15/sex/dose) in an OPP 82-1 Guideline study. Endocrine relevant parameters included thyroid hormone (T4); ovary, testes with epididymides weights and histopathology; and thyroid, pituitary, adrenal, prostate, seminal vesicle, mammary gland, oviduct and uterus histopathology.

Males fed 150 mg/kg/day showed a 5% decrease in body weight. Females fed 100 and 150 mg/kg/day had a 5-7% and 10% decrease in body weight, respectively. The three high doses exceeded the TSRC for both males and females. Systemic toxicity was apparent at 150 mg/kg/day, including renal toxicity characterized by renal weight changes, slight swelling of the kidneys noted during gross pathology, and/or renal histopathology.

Findings potentially related to endocrine modulation were:

- Serum T4 was reduced in a dose dependent manner in females fed 100 and 150 mg/kg/day. No statistically significant reduction was noted in males. T3 and TSH were not measured. Thyroid weights were not measured. No histopathological changes in the thyroids were noted in either sex.
- The relative testes weight was decreased in male rats fed 100 and 150 mg/kg/day. No correlating histopathological changes were observed.

No histopathological changes were noted in the adrenal glands, ovary, uterus, pituitary glands, epididymides, seminal vesicles, coagulating gland, parathyroid glands, mammary glands or oviducts.

Gorzinski et al., 1981b

Male and female F344 rats were fed 2,4-D (100% pure) in the diet at dietary concentrations equivalent to 0, 15, 60, 100 or 150 mg/kg/day (15/sex/dose) in a non-Guideline sub-chronic toxicity study with purified test material examining endocrine relevant endpoints of thyroid hormone (T4) and testes

weight. Note although this is scored a Klimisch 3 because of very limited evaluations it is summarized here because it provides supplemental information to Gorzinski *et al.* (1996a).

Mean body weight of females was decreased (7.6%) at 150 mg/kg/day and body weight of males was decreased (4.6 and 7.7%) at 100 and 150 mg/kg/day, respectively. The two highest doses exceed the threshold for renal saturation for both males and females. Renal toxicity was noted in males and females fed 100 and 150 mg/kg/day based on renal weight change and swelling noted grossly for kidneys. Renal histopathological changes were seen in males at 150 mg/kg/day.

Findings potentially related to endocrine modulation were limited to decreased T4 in females fed 60, 100, or 150 mg/kg/day. T4 levels in males were variable. There were no exposure-related effects on testes weights.

Note no organ weights or histopathology was performed on endocrine-sensitive organs such as the adrenal glands, thyroid glands, ovary, uterus, pituitary glands, epididymides, seminal vesicles coagulating glands, parathyroid glands, mammary glands or oviducts. Thus, this study is limited in the amount of information relevant to assessing endocrine modulation.

Jeffries *et al.*, 1995

In this rat chronic toxicity/oncogenicity study, male and female F344 rats were administered 2,4-D (96.45% pure) in the diet at 0, 5, 75, or 150 mg/kg/day for approximately 12 months (10/sex/dose) or 24 months (50/sex/dose). Endocrine-relevant endpoints included: thyroid hormone measurement (T4); ovary, testes, thyroid, adrenal weights and histopathology; and pituitary, oviducts, cervix, uterus, vagina, mammary gland, epididymides, seminal vesicle, coagulating gland and prostate histopathology. The only deviation from the guideline at the time the study was performed is that the high dose female dose exceeded an MTD.

Mid-and high doses exceeded the TSRC and the high dose females exceeded a classically defined MTD based on bodyweight gain depression. There was no exposure-related mortality. There was no indication of increased tumor incidence, endocrine or otherwise. Statistically decreased body weight was reported for males at 150 mg/kg/day and females at 75 mg/kg/day (both at 12 months and 24 months). Primary effects were seen in the eyes, liver, and lungs of male and females at 150 mg/kg/day, the heart of males at 150 mg/kg/day, and the lungs of females at 75 mg/kg/day. Secondary effects were considered related to the decreased body weight and feed consumption at 75 and 150 mg/kg/day. At 12 months, a non-statistically significant increased incidence of very slight or slight degeneration of the descending proximal tubule of the kidney was indicated at 75 and 150 mg/kg/day. A high incidence of increased panlobular hepatocytes size often accompanied by altered tinctorial properties was reported for both males and females at 150 mg/kg/day, an increased incidence of very slight and decreased slight bile duct hyperplasia with or without inflammation was reported for males at 150 mg/kg/day. Analyses of severity showed no statistically significant changes.

The following findings represent potentially endocrine related effects. It should be noted these were seen only above the TSRC and in the presence of significant other systemic toxicity.

- At 12 and 24 months, thyroid hormone levels (T4) were statistically decreased at 75 and 150 mg/kg/day in both males and females.
- Statistically increased thyroid weights (absolute and relative) were seen at 12 and 24 months in males at 150 mg/kg/day and females at 75 and 150 mg/kg/day; this increase was attributed to thyroid masses. No clear histopathological correlate was evident for these masses, however, females at 150 mg/kg/day at 24 months had an increased incidence of parafollicular cell nodular hyperplasia compared to controls. Parafollicular cell hyperplasia is not typically observed in response to decreased circulating thyroid hormones, and the exposure-relationship of this finding is questionable. [Note that tissue accountability at 24 months was poor for female thyroids due to a labeling problem during processing. The individual animal identification for 35 female thyroid/parathyroid glands was lost during the preparation of these tissues for histological evaluation during the dehydration process as a result of removal of the identifying ink from the tissue containers. This resulted in 8, 9, 9, and 10 thyroid glands missing from the 0, 5, 75, and 150 mg/kg/day groups respectively. However, because the misidentified tissues were spread across dose groups and adequate numbers of appropriately labeled tissues were available for evaluation, this deviation had no effect on interpretation of the study outcomes.]
- At 12 months, females showed a very slight decreased secretory material in the epithelial cells of the thyroid at 150 mg/kg/day compared to control. This change suggests an adaptive response to decreased circulating T4 and is considered likely to be exposure-related. At 24 months in all modes of death (early termination through 24 months), an increased incidence of focal cystic dilatation was observed in follicles of the thyroid gland of females at 75 and 150 mg/kg/day.
- At 24 months, statistically decreased weight of ovaries (absolute and/or relative) at was found at 75 and 150 mg/kg/day; there was no histopathological correlate and this finding is considered likely not to be biologically significant but related to decreased body weight.
- At 12 and 24 months, statistically decreased weight of testes (absolute) was seen at 150 mg/kg/day. Relative testes weights were also decreased at 24 months. No exposure-related histopathological changes were seen in the testes.
- At 24 months in all modes of death (early termination through 24 months), a decrease in very slight focal or multifocal area of altered cells was found in the cortex of the adrenal glands of females at 150 mg/kg/day. No other histopathology was indicated; decreased incidence of primary benign adenoma in the cortex of the

adrenal glands was reported in females at 150 mg/kg/day. This change may be exposure related, however, cannot be characterized as adverse. There were no gross findings in the adrenals at either sacrifice interval, and no exposure-related histopathological changes in the adrenals at the 12-month sacrifice.

- At 24 months in all modes of death (early termination through 24 months), there was a decreased incidence of pituitary mass/nodules in both males and females (not statistically significant. At 24 months (all modes of death), there was a decreased incidence in benign adenomas of the *pars distalis* in the pituitary, which was relatively slight in males, but marked in females at 150 mg/kg/day. The incidence in males was (19/50 (38%), 15/32 (47%), 12/28 (43%), 9/49 (18%)) and females was (21/50 (42%), 15/39 (39%), 9/21 (43%), 1/50 (2%)) at 0, 5, 75, and 150 mg/kg/day, respectively. The decrease in the incidence of this relatively common likely estrogen responsive tumor in females is considered related to marked weight loss at the high dose exposure.
- Gross pathology at 24 months showed a decreased incidence of mammary left inguinal hyperplasia in female rats (16/50, 24/50, 12/50, and 1/50) at 0, 5, 75, and 150 mg/kg/day, respectively. This was not noted in males. At 24 months in all modes of death (early termination through 24 months), a decreased incidence of mammary gland hyperplasia often accompanied by duct ectasia was reported at 150 mg/kg/day (26, 14, 11, 9 in males and 36, 33, 16, 1 in females) at 0, 5, 75, and 150 mg/kg/day, respectively. This decrease may be attributable to decreased estrogens but is most likely related to the marked weight loss at the high dose exposure. No gross or microscopic findings were made in the mammary glands at the 12 month sacrifice.

Other endocrine related tissues showed no exposure related gross or microscopic findings: parathyroid glands, preputial or clitoral glands, prostate, seminal vesicles, oviducts, uterus and vagina.

Schulze, 1991b

Male and female B6C3F1 mice were administered 2,4-D (96.1% pure) in the diet at 0, 1, 15, 100, or 300 mg/kg/day for approximately 13 weeks (10/sex/dose). Endpoints evaluated relevant to endocrine effects include thyroid hormone (T4), ovary, thyroid, pituitary, adrenal and testes weights and histopathology, and uterus and epididymides histopathology.

There were no significant differences between treated animals from control animals in overall body weight or body weight gains. At 13 weeks, the livers of males and females at 300 mg/kg/day had elevated nuclear hyperchromatism in periportal hepatocytes (8/10 animals/sex). The absolute kidney weights decreased (non-statistically) at 300 mg/kg/day for males. Tubular degeneration was observed in 9/10 male mice at 300 mg/kg/day; this was not observed in

females. Other renal changes in males at 300 mg/kg/day include: karyomegaly, loss of brush border, and decreased size of tubular lining. The absolute and/or relative kidney weight in females was increased at 100 or 300 mg/kg/day. The lack of histopathological correlation in the females makes the biological significance of the slight changes in the kidney weight questionable.

Findings potentially related to endocrine modulation include:

- Thyroxine (T4) in males and females was non-statistically significantly decreased at 100 and 300 mg/kg/day. There was a dose-dependent decrease in the males. There were no changes in thyroid weight or histopathology and the T4 decrease, although likely exposure-related, is not considered an adverse effect.
- The absolute weights of the adrenal glands were slightly but statistically increased at 1, 15, and 100 mg/kg/day (with an apparent increase at 300 mg/kg/day) for females; no changes were reported for males. The increase was similar and very slight at all doses except 300 mg/kg/day. The relative weights of the adrenal glands were statistically elevated compared to controls for 1, 15, and 300 mg/kg/day (with an apparent increase at 100 mg/kg/day). No dose dependence was observed. There were no histopathological findings in the adrenal for either sex at any dose, and the non-dose related adrenal weight changes are considered unlikely to be biologically significant or exposure-related.

There were no significant differences in incidences of gross pathology of treated animals from control animals were found in the following tissues: pituitary, adrenal glands (cortex and medulla), thyroid glands, parathyroid glands, liver, kidney, testis, epididymides, ovary, uterus, or mammary glands. No significant differences in absolute or relative organ weights of treated males or females were identified in the following tissues: ovary, liver, testes, thyroid glands or pituitary glands. No histopathological changes in treated animals were reported for the following tissues: pituitary, adrenal glands (medulla and cortex), thyroid glands, parathyroid glands, testes, epididymides, ovary, or uterus. This study does not demonstrate any adverse endocrine-related effects in mice at doses below 100 mg/kg/day in a 90-day study. Although direct data are lacking, it is likely that high-doses of 2,4-D are above the TSRC of 2,4-D in mice. Both rats and mice express comparable levels of OAT-1 transporter mRNA in the kidney (Buist and Klaassen, 2004).

Stott et al. 1995a

In this mouse oncogenicity study, female B6C3F1 mice were administered 2,4-D (96.4% pure) in the diet at 0, 5, 150, or 300 mg/kg/day for approximately 12 months (10/sex/dose) or 24 months (50/sex/dose). There were no long term changes in body weight or body weight gains of mice exposed to 2,4-D. At 12 months absolute and relative kidney weights increased at 150 and 300 mg/kg/day (13 and 16%). At 24 months the relative kidney weight increased at 150 and

300 mg/kg/day (11 and 20%). At 24 months, histopathological effects on kidneys included hypercellularity of the descending portion of the proximal tubule and degeneration/regeneration of the cortical tubules at 150 and 300 mg/kg/day and mineralization of renal tubules at 300 mg/kg/day.

There were no effects on endocrine responsive tissues (based on histopathological evaluation) in this study at 12 or 24 months, including: adrenals, cervix, liver, mammary glands, ovaries, oviducts, parathyroids, pituitary, thyroids, uterus, and vagina.

Stott et al. 1995b

In this mouse oncogenicity study, male B6C3F1 mice were administered 2,4-D (96.4% pure) in the diet at 0, 62.5 or 125 mg/kg/day for approximately 12 months (10/sex/dose) or 24 months (50/sex/dose). There were no changes in body weight or body weight gains reported. There was no exposure related incidence of neoplasms. At 24 months, kidney weight increased at 62.5 and 125 mg/kg/day and histopathological lesions of the kidney were apparent at both 12- and 24-months. In the liver at 24 months, there was an increase in multifocal areas of aggregates of reticuloendothelial cells frequently adjacent to degenerative or necrotic hepatocytes at 125 mg/kg/day.

There were no exposure-related endocrine related effects in this study. At 12 months absolute and relative testes weights were not different from control. At 24 months, testes relative organ weight was increased at 125 mg/kg/day (6%). However, there were no histopathological correlate to the weight change, and the change was slight in nature, and absolute testes weights were not statistically significantly increased. Therefore this finding is not considered exposure-related.

No changes in gross pathology were reported at 12 or 24 months for the following tissues: kidneys, liver, seminal vesicles (12 months only), testes, adrenals (24 month only), and epididymides (24 months only). No histopathological changes occurred in the following tissues at 12 or 24 months: adrenals, coagulating glands, epididymides, parathyroid, prostate, seminal vesicles, testes, or thyroid gland.

Schulze, 1990

In this sub-chronic study, 2, 4-D (96.1% pure) was administered to dogs via capsule at 0, 0.3, 1, 3 or 10 mg/kg/day for 13 weeks. Endocrine related endpoints evaluated included: thyroid hormones (T3 and T4), ovary, thyroid and testes (with epididymides) weights and histopathology, and adrenal, pituitary and uterus histopathology.

There was severe systemic toxicity at the highest dose tested (10 mg/kg/day) (HDT): including mortality, weight loss, renal pathology and altered BUN and creatinine. At 3 mg/kg/day renal pathology and altered BUN and creatinine were evident. The 10 mg/kg/day dose clearly exceeded a Maximum Tolerated Dose in dogs.

Interestingly, despite the severe toxicity in this study there were no exposure related effects on T3 or T4 levels, and no exposure related effects on thyroid hormones (T3 and T4), weights or histopathology. This stands in contrast to the rodent data, and highlights the susceptibility of the rat to changes in thyroid homeostasis. In addition, although thyroid toxicity was absent dogs, 2,4-D serum concentrations likely were substantially higher in dogs than in rats. Administration of a single oral dose of 5 mg/kg 2,4-D to either dogs or rats resulted in a serum AUC that was 232-fold higher in dogs compared to rats (vanRavenzwaay *et al.*, 2003).

Decreased testes weight and testicular findings were present at the high dose. This finding will be discussed in conjunction with the other dog sub-chronic study below.

Dalgard, 1993a

A second subchronic dog study was conducted at the same laboratory as the Schultze (1990) dog study using 2,4-D (96.7% pure) administered in the diet. The study design and results were published by Charles *et al.* (2001b) along with the results of evaluations of selected 2,4-D esters and salts, which are summarized subsequently with relevant published data. In the Dalgard (1993a) study, concentrations were administered to achieve targeted dose levels of 0, 0.5, 1, 3.75 or 10 mg/kg bw/day; after 8 weeks the 10 mg/kg/day dose was lowered to 7.5 mg/kg bw/day based on excessive toxicity. Endocrine relevant parameters evaluated in this study included adrenal, ovary, pituitary, thyroid and testes (without epididymides) weights and histopathology, and epididymides, prostate, mammary gland, uterus and vagina histopathology.

Body weight gain in this study was decreased at the mid and high dose but not statistically significantly. At 7.5 mg/kg/day, testes weight was decreased and relative but not absolute thyroid weight was increased. No effects were seen on any other endocrine-relevant parameter. The toxicological significance of the thyroid findings is not clear given the absence of any correlating histopathological findings, the lack of absolute weight increase in the thyroid, and the lack of effects on thyroid hormone levels in the prior sub-chronic dog study at a higher dose level. Based on these factors, the increased relative thyroid weight is considered unlikely to be exposure related.

Regarding the testis weight decreases seen in both subchronic studies, and the dose-related incidence of testicular lesions in the Schulze (1990) study, the evidence suggests that a high number of juvenile dogs in both subchronic studies may have contributed to the observed effects. (The age of the dogs is not defined in the study reports; however a subsequent paper by Charles *et al.* (1991b) indicates the dogs in the Dalgard (1993a) study were 4-6 months of age at study initiation.) A comparison of body weights and testes and thyroid histopathology in both sub-chronic dog studies (see **Table S8**), leads to the conclusion that the majority of the dogs in both the sub-chronic studies were juvenile animals on the low end of the stated age range. (Both studies were conducted at the same laboratory during a similar time frame.)

Comparison of the body weights of the dogs at study initiation supports that the dogs were of a similar age and that they were young (mature male beagle adult weight is 10-12 kg). Testicular giant cells and hypospermatogenesis are common background lesions in beagles (up to 30% per Rehm, 2000). Immature dogs (less than 9 months of age) have been reported to have control incidences as high as 75% of both decreased testes weight and hypospermia (Goedken *et al.*, 2008). In the Goedken *et al.* analyses of data from a large population of control dogs, atrophic/hypoplastic tubules were seen in 26.3% of all dogs with 25%–40% of dogs under twelve months old having this finding, decreasing with increased age to 14%–17% in dogs twelve to thirty-six months old. Additionally, the high incidence of juvenile prostate findings in the second study (prostate was not evaluated in the first study) supports that the dogs were immature. It seems likely that the decreased testes weights in both sub-chronic studies and histopathological findings in the testes of the dogs in the 1990 study are an artifact related to the young age of these animals, although it is possible that the high (and lethal) dose of 10 mg/kg/day contributed to delayed development. Supporting the possibility of artifact, a chronic study in dogs (Dalgard, 1993b, discussed below) showed no exposure-related effects on testes weights or testicular histopathology following a one year exposure to 2,4-D at a high dose level (10 reduced to 7.5 mg/kg/day) generally equivalent to the high dose in the prior sub-chronic studies.

Dalgard, 1993b

In a chronic (1 year) dog study, 2,4-D (96.7% pure) was administered in the diet at target doses of 0, 1, 5 and 10 mg/kg bw/day. After 8 weeks the high dose was dropped to 7.5 mg/kg/day from 10 mg/kg/day. The study design and results were published by Charles *et al.* (1996b) along with the results of evaluations of selected 2,4-D esters and salts sub-chronic toxicity studies. Body weight and feed consumption were measured; compound intake calculated and dietary concentrations confirmed. Potential endocrine related parameters in this study included organ weights and histopathological evaluation of adrenals, testes, ovaries, pituitary and thyroid, and additional histopathological evaluation of epididymides, prostate, uterus, vagina, and mammary gland.

Body weight was markedly decreased in the high dose group, severely in females before lowering the high dose. Renal pathology and altered BUN and creatinine were evident at 10/7.5 and 5 mg/kg/day. The NOAEL was 1 mg/kg/day. There were no effects on endocrine parameters. Testicular and ovarian weights and histopathology were not affected, nor were thyroid weights or histopathology.

The absence of effects on testes weights and histopathology supports that the sub-chronic study findings were primarily related to the immature age of the test animal, as the high dose in the chronic study was similar to that in the sub-chronic studies.

As noted in the publication, the dog is not regarded as a relevant species for human risk assessment given the substantial differences in 2,4-D toxicokinetics in this species relative to rats and humans (Timchalk, 2006; vanRavenzwaay *et al.*, 2003). Thus, 2,4-D toxicity in dog is only relevant to assessment of potential environmental mammalian species impacts. It is also potentially relevant in that it supports the unique susceptibility of rodents (especially rats) to perturbation of thyroid hormone balance.

B. Studies Identified in the Published Literature

Developmental Studies

Charles *et al.*, 2001

Developmental toxicity studies in rats and rabbits, performed with the following chemicals: 2,4-D acid, 2,4-D dimethylamine salt (2,4-D DMA), and 2,4-D 2-ethylhexyl ester (2,4-D EHE); 2,4-D diethanolamine salt (2,4-D DEA); 2,4-D isopropylamine (2,4-D IPA); 2,4-D triisopropanolamine (2,4-D TIPAA); 2-butoxyethyl ester (2,4-D BEE); and 2,4-D isopropyl ester (2,4-D IPE) were summarized in a publication by Charles *et al.* (2001). These were GLP studies designed to comply with USEPA Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) testing guidelines. The study methods and results are reviewed in detail in Charles *et al.* 2001; the studies are briefly summarized below (with the exception of the 2,4-D acid rat (Rodwell, 1983) and rabbit (Hoberman, 1990) data, which are included with the regulatory studies of 2,4-D, based on the data in the study reports).

Rat

The rat developmental toxicity studies comply with OPP 83-3 guidelines and are considered to meet Klimisch criteria 2, because of a lack of sufficient detail in the published studies, although it should be noted current guidelines require a longer exposure period. In the rat developmental toxicity studies, groups of 25 to 35 bred female rats/group were gavaged daily on GD 6-15. Sprague-Dawley rats were used in these studies, except for the study on 2,4-D acid, which used Fischer 344 rats. (The 2,4-D acid developmental study (Rodwell, 1983) is reviewed in the discussion of regulatory toxicity studies; this study is scored a Klimisch 1.) The dose levels for each compound are indicated in **Table 12** in the publication, but range from 8-150 mg/kg/day expressed as 2,4-D acid equivalents (ae). Separate control groups for each compound were administered the appropriate vehicle. The 2,4-D ae values will be used in the following discussion.

Dosing was during the period of major organogenesis. Dams were observed twice daily, and maternal body weights and feed consumption were recorded. Surviving dams were euthanized on GD 20 and examined grossly, and kidney and liver weights recorded. Dams dying or sacrificed moribund before GD 20 had gross necropsies performed. At cesarean section, gravid uterine weight, number of corpora lutea, number and position of implantations, resorptions, and live or dead fetuses were recorded. Uteri with no visible implantations were stained for evidence of early resorptions. Each fetus was individually identified, weighed, sexed, and given a gross examination for external malformations/variations. Approximately one-half of the rat fetuses in each litter were evaluated for visceral malformations/variations, with heads examined for craniofacial defects. The remaining rat fetuses in each litter were processed for evaluation of skeletal alterations.

Developmental toxicity with the various 2,4-D salts and esters was observed only at dose levels causing maternal toxicity and was dose related. In general, maternal body-weight effects in rats began to be apparent at dose levels of 30 mg/kg/day. This dose by gavage has been shown to exceed the renal

clearance saturation threshold. At 30 mg/kg/day, body-weight gain was significantly depressed with 2,4-D EHE, 2,4-D EHE and 2,4-D IPE. At dose levels ≥ 90 mg/kg/day ae, clinical signs of toxicity were reported with 2,4-D DEA, 2,4-D DMA, 2,4-D IPA, and 2,4-D BEE, and maternal mortality was seen with 2,4-D TIPA. The no-observed-adverse-effect level (NOAEL) in dams across the studies ranged from 8 to 17 mg/kg/day.

There were no treatment-related effects on litter size, resorption rates, or fetal sex ratios. Developmental effects generally occurred only at the highest dose levels tested (HDT). Significantly decreased fetal-body weights were seen with 2,4-D DEA, 2,4-D DMA, 2,4-D TIPA, 2,4-D EHE, and 2,4-D IPE at >90 mg/kg/day (the HDT of each compound). Statistically significant treatment-related increases in fetal variations observed with 2,4-D DEA, 2,4-D DMA, 2,4-D TIPA, 2,4-D BEE, and 2,4-D IPE at doses >90 mg/kg/day included slightly delayed skeletal ossification and the presence of extra ribs, either cervical or lumbar. With 2,4-D EHE, a statistically significant increase in the incidence of incomplete or unossified sternbrae was the only effect noted in fetuses at 30 mg/kg/day and above. A statistically significant increase in the incidence of external and visceral malformations and skeletal anomalies was seen only at the high-dose level of 2,4-D TIPA (175 mg/kg/day), consisting of malformations of the eyes and rib variations (wavy, fused). This dose level exceeded an MTD, with maternal toxicity including mortality and severely decreased maternal weight gain. The types of developmental findings were not those typically characteristic of endocrine system disruption.

Rabbit

Groups of 18 to 24 inseminated adult female New Zealand white rabbits were administered the test materials by oral gavage once daily on either GD 6-18 or 7-19. Dose levels are shown in the preceding table but range from 10 to 75 mg/kg/day expressed as 2,4-D ae. Separate control groups were administered the appropriate vehicle.

The studies complied with the OPP 83-3 Test Guidelines. The developmental studies in rabbits (except for the study on 2,4-D acid (Hoberman, 1990) which is scored a 1) are considered Klimisch criteria 2, primarily because of a lack of some detail in the presentation, although it should be noted current guidelines require a longer exposure period. Additionally, the study on 2,4-D DMA was scored a 2 because maternal toxicity (not clearly dose related) limited the number of litters available for evaluation. However, developmental toxicity was considered adequately characterized.

Does were observed twice daily, and maternal body weights recorded at intervals during gestation. Surviving does were killed on GD 28 or 29; gross postmortem examinations were performed on all females (including deaths or moribund sacrifices). At cesarean section, gravid uterine weight, number of corpora lutea, number and position of implantations, resorptions, and live or dead fetuses were recorded. Uteri with no visible implantations were stained with an ammonium sulfide solution to detect early resorptions. Each fetus was individually identified, weighed, sexed, and given a gross examination for external malformations/variations. All fetuses were examined by dissection for evidence of visceral alterations and processed for evaluation of skeletal alterations.

In rabbits, dose-related maternal toxicity was seen with 2,4-D salts and esters at doses at or above 30 mg/kg/day ae. With higher doses (75-90 mg/kg/day), more severe maternal effects were noted. Clinical signs of toxicity accompanied by maternal body-weight losses, and in some cases, significant morbidity and/or mortality (2,4-D DMA, 2,4-D IPA, 2,4-D TIPAA, and 2,4-D BEE) was observed.

In rabbits, embryonic and fetal development was essentially unaffected, even at maternally toxic doses. There were no effects on maternal reproductive measures such as litter size, number of resorptions, or on fetal body weights. Increased 7th cervical ribs were observed in the group exposed to 2,4-D DEA at 40.6 mg/kg/day, a dose level that also produced clinical signs of toxicity and decreased maternal body weight gain. Although statistically significant compared to control, it should be noted this is a finding with a high background incidence. There was no evidence of teratogenicity in rabbits with 2,4-D acid or its salts or esters. Additionally, there was no evidence of urogenital anomalies characteristic of endocrine disruption in these studies.

The types of findings in both the rat and rabbit developmental toxicity studies are characteristic of those associated with maternal toxicity, and do not suggest endocrine-mediated effects. Adverse effects on the developing rat and rabbit fetuses exposed *in utero* to 2,4-D were observed only at dose levels that produced maternal toxicity, with increasing dose levels that exceeded the TSRC (established for rats only, but predicted for rabbits based on similarities in their renal transport capability), causing increasingly more severe maternal effects, with concomitant effects on the developing fetus. Further, the studies do not provide any specific evidence of endocrine disruption. The majority of the 2,4-D forms were not associated with major malformations; the single exception is found in the rat study with 2,4-D TIPAA at a very severely maternally toxic dose. This review of developmental findings for 2,4-D and its forms is included in the WoE because it shows the commonality of findings for this group of related compounds and does not predict endocrine disrupting activity for any of the compounds.

Dinamarca et al., 2007

Eight-week old, ICR/Jcl mice were mated, and were subsequently administered 2,4-D as a “pure compound” (purity unspecified) or as a commercially available formulation available in Chile (unspecified) in drinking water at concentrations providing mg/kg/day doses of 0, 0.01, 0.10 or 100 mg/kg/day from GD 0-9. GD 0 was designated as the day a copulatory plug was observed. There were 10-13 mice/group. Dam body weights were recorded at GD 0, 6 and 9. Water consumption was measured, and mice were monitored for behavior. Feed consumption was stated to have been observed; it was not clear if this was measured. Mice were bled at GD 9 for biochemical evaluations and caesarian-sectioned. Ovaries were evaluated for numbers of corpora lutea and uterine horns were evaluated for number of implantation sites, resorptions and live embryos.

Oxidative stress was assessed by the determination of catalase activity, thiobarbituric reactive species (TBARs) and total antioxidant capacity (TAC). Carbon tetrachloride was administered subcutaneously at 3 ml/kg on GD8 as a positive control.

There were no signs of maternal toxicity nor differences in body weight gain between the dosed groups and the control. Numbers of corpora lutea, implantation sites, resorptions and live embryos were similar between the dose groups and control.

TAC values were significantly decreased only after administration of 100 mg/kg/dose 2,4-D of both “pure” 2,4-D and the formulated test material, suggesting that non-enzymatic antioxidant defenses may be depleted at that dose level which likely exceeds the TSRC in mice. There are some reservations regarding this conclusion because it is not clear how the TAC was assessed; multiple pathways are involved. Catalase activity and TBARS were not changed by exposure.

This study includes a fairly complete description of methods and appropriate analyses of results. One flaw is that details on the purity and source of the “pure” 2,4-D are not provided. The dose spacing in this study was designed to address the low dose hypothesis proposed by Cavieres *et al.* (2002). (The Cavieres *et al.* work is summarized in Supplementary Appendix VII.) The Dinamarca *et al.* study demonstrated that the finding of decreased implantations in mice exposed to 2,4-D in the Cavieres *et al.* (2002) report could not be replicated, even with an exposure period correctly designed to explore this possibility

Male Reproductive Toxicity

Lamb *et al.*, 1981a

In this study, male C57BL/6N mice were dosed via diet for 8 weeks with various concentrations of 2,4-D; 2,4,5-T; and TCDD. Approximate daily exposures were 40 mg/kg/day 2,4-D, 40 mg/kg/day 2,4,5-T, and 2.4 micrograms per kilogram per day ($\mu\text{g}/\text{kg}/\text{day}$) TCDD; 40 mg/kg/day 2,4-D, 40 mg/kg/day 2,4,5-T, and 0.16 $\mu\text{g}/\text{kg}/\text{day}$ TCDD; and 20 mg/kg/day 2,4-D, 20 mg/kg/day 2,4,5-T, and 1.2 $\mu\text{g}/\text{kg}/\text{day}$ TCDD for groups II, III, and IV respectively. Group I (control) received untreated diet. Fertility, sperm number, motility, and morphology were evaluated. Somatic cell sister chromatid exchange frequencies were also evaluated in mice injected intraperitoneally with single doses of similar chemical mixtures. No significant effects were observed in the male mice exposed to 2,4-D in the above mixtures compared to the control groups. Thus, this study provides no evidence of endocrine-disrupting activity of 2,4-D. Its primary weaknesses are that it tests only mixtures, however, this would not be expected to confound the negative results, and the dose and purity of the 2,4-D are defined. The high dose of 2,4-D would be anticipated to approximate the TSRC in mice so exposure to 2,4-D was not limited by the toxicity of the other mixture components.

Lamb *et al.*, 1981b

In this study, male C57BL/6N mice were dosed via diet for 8 weeks with various concentrations of 2,4-D, 2,4,5-T, and TCDD. Approximate daily exposures were 40 mg/kg/day 2,4-D, 40 mg/kg/day 2,4,5-T, and 2.4 $\mu\text{g}/\text{kg}/\text{day}$ TCDD; 40 mg/kg/day 2,4-D, 40 mg/kg/day 2,4,5-T, and 0.16 $\mu\text{g}/\text{kg}/\text{day}$ TCDD; and 20 mg/kg/day 2,4-D, 20 mg/kg/day 2,4,5-T, and 1.2 $\mu\text{g}/\text{kg}/\text{day}$ TCDD for groups II, III, and IV respectively. Group I (control) received untreated diet. Males were mated with untreated female mice. Females

were either cesarean sectioned at gestation day 18 for evaluation of the fetuses, or allowed to deliver and rear their pups until weaning at postnatal day (PND) 21 for evaluation of offspring birth weight and viability. There was no residual impact on reproductive performance of the treated males in this study. Development and survival of fetuses and pups in the offspring of treated groups were similar to that of the control mice, and no effect of these mixtures was apparent. The study provided no evidence of male-mediated reproductive toxicity or of endocrine disrupting activity of 2,4-D for these mixtures of chemicals. Its primary weakness is that it tests only mixtures; however, this would not be expected to confound the negative results. The high dose of 2,4-D would be anticipated to approximate the TSRC in mice so exposure to 2,4-D was not limited by the toxicity of the other mixture components.

Subchronic and Chronic Toxicity

Charles et al., 1996a

This review article presents data from several rat subchronic toxicity studies conducted with 2,4-D acid, 2,4-D dimethylamine salt (DMA), or 2,4-D 2-ethylhexyl ester (2-EHE). These studies were GLP Guideline studies conducted to satisfy US EPA testing requirements. The 2,4-D acid subchronic study (Schultze, 1991a) is discussed with the regulatory studies above, based on the comprehensive data in the study report. In the 2,4-D DMA and 2-EHE studies, Fischer 344 rats (10/sex/dose group) were dosed in the diet with target doses of 0, 1, 15, 100, and 300 mg/kg/day (expressed as acid equivalent doses) for 90 days.

Endocrine endpoints evaluated included: thyroid hormones (T3 and T4); adrenal, ovaries, pituitary, testes, thyroid (and parathyroids) organ weights and adrenal, epididymides, mammary gland, ovaries, pancreas, pituitary, prostate, testes, thyroid (and parathyroids), and uterus histopathological evaluations. Clinical signs, body weight, feed consumption, clinical pathology, and evaluation of standard target organ histopathology were also done in these studies.

Several possible indications of endocrine pathway interactions were noted in these studies. All occurred at high and systemically toxic doses that exceeded the TSRC. These indications included decreased T4 and/or T3 at dose levels of ≥ 100 mg/kg/day, with T4 appearing somewhat more sensitive than T3 and females more sensitive than males. Correlating with these findings were increases in relative thyroid weights (primarily at 300 mg/kg/day); however no correlating histopathological evidence of thyroid follicular cell hypertrophy or hyperplasia was evident. Therefore, these changes are considered slight in severity and non-adverse. Adrenal cortical hypertrophy was also seen at dose levels ≥ 100 mg/kg/day. (Typically adrenal cortical hypertrophy results from increased ACTH release from the pituitary in response to generalized stress; in the case of 2,4-D it reflects the excessive toxicity at the high doses far exceeding the TSRC.) Relative testes weights were decreased at 300 mg/kg/day ae, and testicular atrophy was noted at the same dose level.

Treatment-related deaths were seen for 2-EHE at 300 mg/kg/day. Severe decreases in body weight gain and decreased feed consumption were seen at the high dose level (300 mg/kg/day) for all three compounds. Decreased body weight gain was also seen at 100 mg/kg/day for 2,4-D acid. A trend

towards anemia and/or decreased platelet count was observed at ≥ 100 mg/kg/day. BUN was increased for DMA (females) and 2-EHE at 300 mg/kg/day. Decreased relative liver and increased relative kidney weights were found at ≥ 100 mg/kg/day. Treatment-related histopathological findings included brush border loss and vacuolization of kidney tubular cells, centrilobular hepatocellular hypertrophy, and retinal degeneration and cataract formation (females only). These histopathological changes were primarily evident at 300 mg/kg/day.

Effects on the thyroid were considered relatively slight in severity and non-adverse, and exposure-related findings in the testes were seen only at doses clearly exceeding an MTD and far above the TSRC (and irrelevant for either human or environmental exposures). At those doses it is not clear if the observed findings reflect endocrine-related changes or are secondary to the excessive systemic toxicity. Based on other data, oxidative stress would be anticipated to occur at this very high exposure level. These studies also demonstrate that the salt and ester forms of 2,4-D tested show similar toxicities to 2,4-D when tested on an acid equivalent mg/kg/day basis.

Charles *et al.*, 1996b

This article presents data from several dog subchronic toxicity studies conducted with 2,4-D (acid), 2,4-D DMA, or 2,4-D 2-EHE, and a dog chronic toxicity study on 2,4-D acid. Data from the subchronic and chronic dog studies on 2,4-D acid (Dalgard 1993a and Dalgard 1993b, respectively) are summarized above under regulatory toxicology studies. All of these studies were GLP Guideline studies conducted to satisfy USEPA testing requirements. Beagle dogs (4/sex/dose group) were dosed in the diet with target doses of 0, 1.0, 3.75 and 7.5 mg/kg/day (expressed as acid equivalent doses).

Endocrine endpoints evaluated in these studies included adrenal, ovaries, pituitary, testes, thyroid (and parathyroids) organ weights, and adrenal, epididymides, mammary gland, ovaries, pancreas, pituitary, prostate, testes, thyroid (and parathyroids), and uterus histopathological evaluations. Clinical signs, body weight, feed consumption, clinical pathology, and evaluation of standard target organ histopathology were also performed in these studies. Thyroid hormone analyses were also conducted in the sub-chronic study on 2,4-D acid (Dalgard, 1993a).

There were two findings in these studies that potentially could be related to endocrine modulation, although it is equally possible that they are artefacts due to the young age of the study animals (4-6 months, based on the Charles *et al.* (1996b) publication. As reviewed in the discussion on the regulatory studies, testicular weight decreases and histopathological findings are exceedingly common in young dogs. Testes weights (in relation to body weight) were lower in the mid dose but not high dose of the DMA and EHE subchronic studies, and as noted above were also lower in the high dose of both subchronic studies of the acid. This finding did not occur in the chronic 2,4-D acid study (Dalgard 1993b), however, which used the same high dose as the subchronic 2,4-D acid study (Dalgard 1993a). Additionally, there were no treatment-related histopathological lesions in the testes in any of the subchronic or chronic dog studies the Dalgard 1993a and 1993b studies of the acid, or in the studies of 2,4-D DMA or 2,4-D 2-EHE. Therefore, the biological significance of the decreased testes weight finding

was considered unclear by the authors. It should be noted that an earlier subchronic study of 2,4-D acid in dogs (Schulze *et al.*, 1991b) had shown an increased incidence of testicular lesions albeit at a lethal dose (summarized above for the regulatory toxicity studies on the acid); comparison of age, body weights and the nature of the findings support that this finding was also due to the use of juvenile animals, but may have been exacerbated by systemic toxicity. The second finding in the studies reviewed by Charles *et al.* is that inactive/juvenile prostates were noted in “several high-dose males” in the subchronic studies; this finding is likely to reflect the immature age of the tested animals. Review of the report for the subchronic 2,4-D acid study (Dalgard 1993a) evaluated by Charles *et al.* shows this finding clearly present in all dose groups (see **Table S8**). This finding was also not made in the chronic acid dog toxicity study (Dalgard 1993b). As discussed above in the context of the 2,4-D acid regulatory toxicity studies, a very high incidence (up to 75%) of decreased testes weights and testicular lesions (hypospermia) have been reported in juvenile control dogs up to 9 months of age (Goedken *et al.*, 2008).

These possibly endocrine-related findings were made at doses showing generalized systemic toxicity. In the subchronic studies for all three compounds, the high-dose dogs had decreased body weight gains and decreased feed consumption; the mid-dose 2-EHE group also showed decreased weight gain. Renal toxicity was demonstrated by changes in blood urea nitrogen and creatinine, which was most pronounced at the high-dose level, and liver enzymes differed from control. No correlating histopathological lesions were found in the kidney, and only minimal increases in peri-vascular inflammation were found in the livers of high-dose dogs. Other minor clinical pathological changes were not considered toxicologically significant. None of the dog studies showed adverse effects on the thyroid.

The chronic acid study (Dalgard 1993b) showed decreased weight gain at the mid- and high dose during the first 13 weeks (correlating with the findings in the subchronic studies). No evidence of endocrine-related toxicity, including testicular toxicity, was found in this study. One to two dogs in these groups periodically showed poor weight progression and decreased feed consumption during the study.

As noted previously, dogs are not predictive for human risk assessment due to incomplete excretion of 2,4-D and lack of metabolic capabilities relevant to humans. However, as a susceptible species the dog may predict potential effects on other species deficient in the OAT-1 transporter. Dogs also provide information supporting that rodent species (especially rats) are particularly vulnerable to changes in thyroid hormone economy. Therefore data from the dog studies are included in the weight of the evidence evaluation. The publication by Charles *et al.* (1996b) also supports that findings from the dog studies on 2,4-D acid are generally predictive for those on the salts and esters of 2,4-D (when doses are expressed as acid equivalents).

Other Studies Relevant to Hypothetical Mechanism(s) for Thyroid Effects in Rats

Florsheim and Velcoff, 1962

Florsheim and Velcoff placed male Sprague-Dawley rats on a restricted iodine diet regimen for 26 days with a daily subcutaneous injection of 1 cc iodine solution (0.8 μc I131 and 5 or 10 μg stable iodide). Experimental rats were subcutaneously injected with 80 mg/kg bw/day 2,4-D for the last 7 days of the regimen. After autopsy, a 3 cc serum sample from each animal was evaluated. A significant increase in 24-hour I¹³¹ uptake and significant decreases in serum protein-bound-iodine (PBI) and thyroid : serum radioiodide ratio were observed. However, 2,4-D exposure had no effect on serum or pituitary TSH concentrations, thyroidal cell height, or thyroid histopathology. These findings provide a possible mechanistic explanation for decreases in circulating thyroid concentrations in rats at high dosages of 2,4-D, but do not provide evidence of a biologically significant adverse effect. Although it is not considered a high quality study, this study is cited as part of the WoE because it provides some mechanistic information to possibly explain the high-dose effects of 2,4-D on the rat thyroid.

Florsheim et al., 1963

Florsheim *et al.* (1963) evaluated the distribution of thyroxine after 2,4-D exposure. The dosage and route of administration of the 2,4-D is not specified in this paper. Based on the methodology described in Florsheim and Velcoff, it is likely that a high intraperitoneal dose was used. Rats received 2.0 milliliters (mL) of thyroxine-labeled serum by femoral infusion. No significant differences in thyroxine half-life or thyroxine distribution (30 minutes after tracer thyroxine injection) were observed. Significant differences in the fate of thyroxine 1 hour post-injection of thyroxine- I¹³¹ tracer were observed; however, these results were based upon data from three exposed and three control animals. Brain and liver tissues of 2,4-D treated rats had significantly higher concentrations of thyroxine- I¹³¹ tracer and significantly lower serum thyroxine- I131 tracer concentration compared to controls. Authors concluded that “2,4-D reduces the serum binding capacity for thyroxine...” and “2,4-D lowers serum thyroxine binding by competing with thyroxine for its binding sites, although it appears to be a weak competitor.” The Florsheim *et al.* (1963) and Florsheim and Velcoff (1962) studies support that high doses of 2,4-D in the rat may modulate thyroid hormone levels. However, thyroid histopathology was not affected, even at these high doses, suggesting that the effects on circulating hormone levels had little biological significance. Although it is not considered a high quality study, this study is cited in the WoE primarily because it provides some mechanistic information to possibly explain the high-dose effects of 2,4-D on the rat thyroid.

Van den Berg, et al., 1991

Several chemical classes (including the phenoxy acids such as 2,4-D and 2,4-dichlorophenoxybutyric acid [2,4-DB]) were evaluated for their ability to bind to the T4 binding site of the thyroid hormone carrier, transthyretin. In addition, *in vivo* studies using male rats were conducted to determine the effect of selected chemicals (e.g., 2,4-DB) on plasma thyroid hormone levels. The *in vitro* study indicated that the phenoxy acids, including 2,4-D, had a relatively high affinity for the T4 binding site of transthyretin.

Because the study did not present results for individual phenoxy acids in the *in vitro* study, it cannot be determined which phenoxy compound was most reactive. The *in vivo* study showed that male rats administered 2,4-DB acutely via the intraperitoneal route had lower plasma T4 levels compared to control. 2,4-D itself was not tested in the *in vivo* study; however, the findings (for 2,4-DB) are generally consistent with the results of high-dose exposures to 2,4-D to rats in other *in vivo* studies. The intraperitoneal route of administration is not considered relevant to human risk assessment. The *in vitro* study results support that 2,4-D with its relatively high affinity for the T4 binding site of transthyretin in rats displaces the T4 from the binding protein in rats, thus making the thyroxine more available for liver sequestration or metabolism. This study is included in the WoE discussion only because it contributes a potential mechanistic hypothesis for high-dose 2,4-D effects on the rat thyroid.

Supplementary Appendix V

In vitro studies with Klimisch scores of 3 or 4⁴

General note regarding yeast based assays: In general regulatory authorities (US and EU) have moved away from supporting yeast-based assays for screening for endocrine modulating chemicals in part because of differences in absorption from mammalian systems, and also because of potential lack of stability of transfected yeast cells. Differences in rank-order potency of compounds have been detected, presumably due to differences in uptake and/or extrusion of the test compounds (Gaido *et al.*, 1997). Mammalian-based *in vitro* assays are generally considered superior to yeast, because mammalian assays are thought to be more representative of humans in terms of uptake of xenobiotics, potential biotransformation, and the cellular context of steroid receptors and the various signaling molecules with which the xenobiotics interacts (O'Connor *et al.*, 2002).

Blair *et al.*, 2000

One hundred and eighty-eight chemicals, including 2,4-D, were examined for binding at the rat estrogen receptor (ER). Sprague Dawley rats (245 days of age) were ovariectomized 10 days prior to sacrifice and the uterine tissues excised and homogenized. Chemicals were then examined for the ability to inhibit the binding of 1 nM ³H-17 β -estradiol (E2) to the rat uterine ER using a tiered approach. First, chemicals were examined using two high concentrations spanning three log concentrations. Chemicals that demonstrated ER binding in Tier 1 were then examined in Tier 2 using a wider range of concentrations (typically 10⁻⁹ – 10⁻⁴ M). Radio-inert E2 (10^{-10.5} – 10⁻⁷ M) was also tested as a positive reference chemical. Samples were run in duplicate and each assay was conducted twice. 2,4-D tested at concentrations up to 0.1 mM did not demonstrate an ability to bind to the rat ER, so it was not examined in a wider concentration range. In this study, the uterine tissues were derived from rats that were older than recommended in the EPA testing guidelines for the ER binding assay (EPA, 2009a). This study was conducted using the same test model as prescribed in the EPA guidelines, with the exception noted above. The results demonstrated that 2,4-D does not bind to the ER. This study was well-designed and adequately reported, however only two concentrations of 2,4-D were tested (with the second concentration undefined) and specific results for 2,4-D were not included. Therefore it was considered to score a 3 by the Klimisch criteria.

Fang *et al.*, 2003

Two hundred and two chemicals, including 2,4-D, were evaluated for the ability to bind to recombinant rat androgen receptor (AR; purchased commercially). The AR protein (1.84 nM) was incubated with 1 nM of ³H-R1881 and increasing concentrations (4.28 x 10⁻⁹ M to 4.28 x 10⁻⁴ M, in one log-unit concentration intervals) of radio-inert test compounds. Radio-inert R1881 (10⁻⁷ – 10⁻¹¹ M) was run as a reference chemical and a no competitor tube was included to assess total radioligand binding. After 18-20 hours at 4°C, hydroxyapatite slurry was added to separate bound from unbound radioligand.

⁴ See Publication **Table 4** for Klimisch scores and comments for the published *in vitro* studies.

Duplicate tubes were included at each test concentration and all assays were run at least twice. AR binding affinity was reported in terms of relative binding affinity. 2,4-D (purity not reported) failed to compete with ^3H -R1881 for AR binding.

The source of AR used in this study (*Escherichia coli*-expressed recombinant rat AR protein) is different from that indicated in the EPA guidelines for the AR binding assay (rat prostate; EPA, 2009b). As noted above, the purity of 2,4-D was not reported, which is the primary reason for scoring this assay as a Klimisch criteria 3. Nevertheless, the assay described in Fang *et al.* (2003) appears to be robust and to align closely with the general procedures described in the EPA guidelines, including ^3H -R1881 assay concentration, test compound concentrations, incubation times and temperature. It provides support for the finding that 2,4-D does not interact with the AR.

Jung *et al.*, 2004

Fifty chemicals, including 2,4-D, were evaluated for anti-estrogenic activity in a yeast two-hybrid system expressing the rat estrogen receptor ER α and the coactivator TIF2, along with a β -galactosidase reporter gene containing an estrogen-responsive promoter region. The yeast cells were incubated overnight in SD medium free of tryptophan and leucine (required for growth), then treated with 10^{-9} - 10^{-3} M test chemicals in DMSO and 5×10^{-9} M E2 for 4 hours at 30°C. After incubation, the cells were washed, lysed with Zymolyse 20T, mixed with *o*-nitrophenol- β -galactopyranoside, and galactosidase activity determined based on spectrometric absorbance for an aliquot of the lysate. Chemicals found to exhibit anti-estrogenic activity were further evaluated using an estrogen receptor competitive binding assay and a human ER transactivation assay in MCF-7 cells. 2,4-D (form and purity not reported) was negative for anti-estrogenic activity in the yeast two-hybrid system at concentrations tested. The number of replicates run, the number of chemical concentrations tested, and data regarding solubility of the test chemicals and cytotoxicity were not provided. Another limitation of the Jung *et al.* paper involves the use of a yeast-based assay system (see general note above).

Jungbauer and Beck, 2002

Various compounds, including 2,4-D (form and purity not reported), were tested for estrogenic activity in a yeast two-hybrid system expressing the human ER α (the expression of which was under control of a metallothionein promoter) and a reporter plasmid containing the lacZ gene under control of the vitellogenin hormone response element. Human ER expression was induced overnight by the addition of 10^{-5} M CuSO $_4$. Cells were then treated with test chemicals for 4 hours at 30°C, after which the induced expression of β -galactosidase was determined spectrophotometrically. E2 was run as a reference chemical to characterize the system. Potency values within the range of 10^{-4} M were considered not biologically relevant. Exact details regarding assay parameters were limited (e.g., concentrations of test chemicals evaluated, number of replicates and runs of the assay). 2,4-D did not transactivate the estrogen receptor. Another limitation of the Jungbauer and Beck paper involves the use of a yeast-based assay system (see general note above).

Kim et al., 2005

This paper contained the results of several different *in vitro* assays. One concern potentially limiting relevance is that the 10nM concentration of DHT used in most of the assays is potentially a supra-physiological concentration, based on serum and prostate DHT concentrations reported in Belanger et al., 1989.

22Rv1 cell proliferation

This assay evaluated the ability of 2,4-D (98% purity) and its metabolite, 2,4-dichlorophenol (DCP; 98% purity), to alter the response of human prostate cancer cells to dihydroxytestosterone (DHT) treatment. Human 22Rv1 cells, which express the AR, were seeded in 96-well plates at 3×10^3 cells per well in RPMI medium containing 10% fetal bovine serum (FBS) pre-treated with dextran-coated charcoal. Twenty-four hours later, the cells were treated with 10^{-13} – 10^{-5} M 2,4-D, DCP, or DHT or with 10^{-9} – 10^{-5} M 2,4-D or DCP in the presence of 10^{-8} M DHT. After four days, the cells were fixed and stained with sulforhodamine-B to examine cell proliferation spectrophotometrically. Reported results were the mean of three separate experiments. Neither 2,4-D nor DCP alone increased 22Rv1 cell proliferation; however, co-treatment combining either 2,4-D or DCP and DHT was reported to increase cell proliferation substantially above levels induced by DHT treatment alone. No solvent only controls were tested in this assay.

Note in a 2006 abstract by the same group of investigators (not relating to 2,4-D), DHT alone was reported not to have induced proliferation in this cell line, in contrast to what was reported in the 2005 study.

Transactivation in 22Rv1 cells and PC3 cells

Similar results were reported in receptor transactivation experiments. 22Rv1 cells were transiently transfected with a mouse mammary tumor virus-linked luciferase reporter gene, then treated 24 hours later with 10^{-8} M DHT, 10^{-9} – 10^{-5} M 2,4-D or DCP, or 10^{-8} M DHT plus 10^{-12} – 10^{-6} M 2,4-D or DCP. Likewise, PC3 cells, which do not express the AR, were transiently transfected with the luciferase reporter gene as well as a human AR expression vector, then treated 24 hours later with 10^{-8} M DHT, 2,4-D, DCP, or a combination of both DHT plus 2,4-D or DCP. After 24 hours incubation, cells were lysed and extracts assayed for luciferase activity, which was normalized for transfection efficiency using expression of a Tk vector. Neither 2,4-D or DCP alone caused transactivation of the AR in 22Rv1 or PC3 cells; however, both chemicals were reported to substantially increase the AR transactivation induced by treatment with 10^{-8} M DHT. No solvent only controls were tested in this assay.

AR mRNA Expression

Treatment of 22Rv1 cells with 10^{-8} M DHT, 10^{-7} M 2,4-D, 10^{-10} M DCP, or a combination of DHT with 2,4-D or DCP had no significant effect on AR mRNA expression (as assessed by slot blot analysis with normalization against β -actin mRNA expression). Although DHT treatment induced AR protein expression (as assessed by Western blot analysis with normalization against β -actin protein expression)

approximately 50%, treatment with 2,4-D, DCP had no effect. Additionally, co-treatment with DHT and 2,4-D and DCP did not further increase AR protein expression above levels induced by DHT treatment alone. No solvent only controls were tested in this assay and only one concentration of 2,4-D was evaluated

AR Receptor Binding

The ability of 2,4-D and DCP to bind to the AR receptor was evaluated in COS-1 cells (African green monkey kidney cells) transiently transfected with the human AR. Following transfection, cells were cultured in serum-free and phenol-free DMEM for 24 hours, then treated for 2 hours with 5×10^{-9} M ^3H -DHT in the presence or absence of 5×10^{-9} M, 5×10^{-8} M, or 5×10^{-7} M of DHT, 2,4-D or DCP. Both 2,4-D and DCP inhibited binding of ^3H -DHT by approximately 50% at a concentration of 5×10^{-7} M. No solvent only controls were tested in this assay.

AR Nuclear Translocation

Finally, the ability of 2,4-D and DCP to affect AR nuclear translocation was investigated in PC3 cells that were transiently transfected with a green fluorescent protein (GFP)-AR fusion protein expression vector. Cells were treated for 20 min at 37°C to 10^{-8} M DHT, 2,4-D, or DCP or 10^{-8} M DHT plus 10^{-8} M 2,4-D or DCP. After incubation, cells were fixed in 10% formaldehyde and examined for fluorescence using a fluorescence microscope. Treatment with DHT, but not 2,4-D or DCP, induced nuclear translocation of the GFP-AR fusion protein. Co-treatment with DHT plus 2,4-D or DCP was reported to result in nuclear translocation that was increased over that induced by DHT treatment alone. No solvent only controls were tested in this assay.

The authors of this study interpreted the results to suggest that 2,4-D and its environmental metabolite, DCP, do not activate the human AR receptor; but they hypothesized that these compounds may enhance the activity of DHT at the AR by promoting nuclear translocation of the DHT-activated receptor. Kim *et al.* (2005) used a very different, and less robust, model system (COS-1 cells transiently transfected with human AR) than that indicated in EPA guidelines for the assessment of AR binding (2009b). For one, they evaluated binding in whole cells rather than using receptor protein from homogenized tissues or lysed cells. As well, they did not use a reagent such as hydroxyapatite to separate unbound from bound ligand, only evaluated a limited number of test compound concentrations compared to the range recommended in the EPA testing guidelines, and did not report using a solvent/vehicle only control in this assay, or any of the others described in the paper. The findings in Kim *et al.* contradict those of Fang *et al.* (2003), which used recombinant rat AR protein and tested a wider range of test compound concentrations (4.28×10^{-9} to 4.28×10^{-4} M) although the Fang *et al.* assay did not report the purity of 2,4-D tested. The results also contradict the results of the ToxCast™ program screening both in cell-free and cell-based assays. Finally, although 2,4-D and DCP were found to bind the human AR in the study of Kim *et al.* (2005), these compounds alone did not induce proliferation of an androgen-response prostate cancer cell line or activate the human AR, as demonstrated in the receptor transactivation assays. This study is considered limited and the findings are difficult to interpret.

Lee et al., 2006

Lee *et al.* (2006) developed a yeast two-hybrid detection system to examine transactivation via the estrogenic receptor for a variety of compounds. In this experiment, yeast was created that expressed bait, fish and reporter gene. The “bait” were human estrogen receptor α (ER α), either full length or with the transactivation domains missing (truncated), fused to a GAL4 DNA binding domain. This chimeric ER α (full length or truncated) would bind to DNA as a monomer without the need for a ligand. The “fish” were steroid receptor co-activator-1 (SRC-1) or transcriptional intermediate factor-2 (TIF-2) that was fused to the GAL4 transactivation domain. The reporter gene consisted of the GAL4 DNA binding domain controlling a *LacZ* gene. Agonist activation of the DNA bound bait (ER α – Gal4) allows the fish (SRC-1 or TIF-2) to bind to the bait and this induces the reporter gene. This artificial system is quite different than normal mammalian estrogen signaling where agonist activated ER α first binds to another ER α protein forming a ER α homodimer that only then would bind to an ER DNA binding site (ERE) and activate a target gene.

The authors noted increased galactosidase activity at lower concentrations of the test compounds than the original one-hybrid construct. 2,4-D, which was negative with the original one-hybrid construct, had detectable binding with the two-hybrid construct, ranging from 2.09×10^{-4} M to 5.42×10^{-6} M (17 β -estradiol was positive at 10^{-10} M).

There were no negative control compounds included in the Lee *et al.* (2006) study. Another limitation of the Lee *et al.* (2006) paper involves the use of a yeast-based assay system (see note above). Another issue is that agonists and antagonists do not always mimic their *in vivo* agonist or antagonist activity (Pham *et al.*, 1991, 1992). It is possible that some of this discrepancy is due to the use of an artificial system in which the conformational shapes of the receptors and differences in post-translational modifications across species alter their response pattern to exogenous compounds (Mak *et al.*, 1989; Pham *et al.*, 1992; Routledge and Sumpter, 1996; Zysk *et al.*, 1995). Thus, this study is considered limited and the findings are difficult to interpret.

Lemaire et al., 2006

Forty-nine pesticides, including 2,4-D (purity of 95% or greater), were tested for the ability to activate the human ER α and ER β using stably transfected HeLa cells expressing either ER α or ER β and luciferase reporter gene with an upstream estrogen responsive element. Cells were seeded at a density of 30,000 cells per well in phenol red-free DMEM supplemented with 6% dextran-coated charcoal fetal calf serum. After 24 hours, cells were incubated with 10 μ M test chemical for 16 hours, after which luciferase activity was measured using a luminometer. ER α and ER β receptor activity was reported as a percentage of the activity induced by 10 nM 17 β -estradiol and findings are the result of three separate experiments conducted with either duplicate or triplicate wells for each transfected cell line. Chemicals were initially tested at a single concentration of 10 μ M; testing at lower concentrations was conducted only if a positive response was seen at the high initial concentration. The transfected cell lines also expressed a constitutive background level of luciferase activity and the ability of the test chemicals to affect that activity was also evaluated. At a concentration of 10 μ M, 2,4-D exhibited no differences from

a DMSO control for relative activity to both the ER α and ER β . Further, it did not affect the constitutive activity of either receptor protein in the absence of 17 β -estradiol. 2,4-D was not further tested for estrogen receptor binding in this study.

This study evaluates antagonism activity of constitutive ER α activity and the absence of both agonist and antagonist activity at ER β . Although the results were clearly negative, 2,4-D was tested in only a single high concentration; this is the major weakness identified in this study.

Nishihara *et al.*, 2000

In this study, 517 chemicals, including 2,4-D, were tested for agonist activity at the estrogen receptor in a yeast two-hybrid system expressing the estrogen receptor ER α and the coactivator TIF2, along with a β -galactosidase reporter gene containing an estrogen-responsive promoter region. The yeast cells were incubated overnight in SD medium free of tryptophan and leucine (required for growth), then treated with test chemicals in DMSO for 4 hours at 30°C. After incubation, the cells were washed, lysed with Zymolyse 20T, mixed with *o*-nitrophenol- β -galactoside, and galactosidase activity determined based on spectrometric absorbance for an aliquot of the lysate. 2,4-D (form and purity not reported) was negative for estrogenic activity at concentrations up to 1×10^{-4} M. The number of replicates run, the number of chemical concentrations tested, and data regarding solubility of the test chemicals and cytotoxicity were not provided. Another limitation of the Nishihara *et al.* (2000) assay involves the use of a yeast-based assay system (see note above).

Orton *et al.*, 2009

Twelve pesticides, including 2,4-D (>97% purity), were tested for agonist and antagonist activity at both the human estrogen receptor and the human androgen receptor using yeast reporter systems. The DNA sequence of either the human estrogen receptor (yeast estrogen screen; YES) or the human androgen receptor (yeast androgen screen; YAS) was stably integrated into the yeast genome. The yeast cells also contained a lac-Z reporter plasmid, which encodes the β -galactosidase enzyme and could be expressed upon activation of either the estrogen or androgen receptor. For assessing agonist activity, concentrations of 4.9×10^{-7} – 1×10^{-3} M 2,4-D dissolved in ethanol were added to the wells of a 96-well plate and evaporated to dryness. Aliquots (200 μ l) of minimal medium seeded with either the YES or YAS yeast cells and containing the chromogenic substrate chlorophenol red- β -D-galactopyranoside (CPRG) were added to the wells and incubated at 32°C for 3 days. Samples were run in triplicate over three plates and the assay was run two separate times. For the YES assay, 17 β -estradiol was run as a positive control; for the YAS assay, testosterone was run as a positive control. Both solvent and media-only controls were also run. Conversion of the yellow CPRG substrate to a red product was then assessed by measuring absorbance at 540 nm. Cell growth (based on culture turbidity measured at 620 nm) was also measured to assess cytotoxicity. To evaluate antagonist activity, the assays were run similarly, except the pesticide was added in combination with either 0.25 nM 17 β -estradiol (for the YES system) or 2.5 nM testosterone (for the YAS system). Positive controls were either 0.01-25 μ M 4-hydroxytamoxifen (to assess anti-estrogenic activity) or 0.02-50 μ M flutamide (to assess anti-androgenic activity). In these assays, 2,4-D was stated to be negative for agonist and antagonist activity at both the

human estrogen receptor and human androgen receptor. No 2,4-D specific data were presented, however.

The YES and YAS assays were well run with appropriate positive controls and with both negative and vehicle controls. The primary limitation of this assay reported by Orton *et al.* involves the use of a yeast-based assay system, as discussed in the note above. The lack of specific data for 2,4-D is also considered a weakness.

In addition to the yeast assays discussed above, Orton *et al.* also evaluated the effects of pesticide exposure on the frog ovarian production of progesterone, testosterone and estradiol *in vitro* and on the number of oocytes ovulated during the exposure period. Ovaries were removed from sexually mature female *Xenopus laevis*, sliced into 10 fragments, and cultured in Modified Barth's Media in 24-well plates (2 fragments per well). After 20 hrs incubation (incubation temperature not reported) with varying concentrations of human chorionic gonadotropin (hCG) or media only, fragments were incubated for 20 hrs in either control media, a concentration of hCG causing a 60% maximum ovulatory response, or 6.25 or 62.5 μM pesticide plus hCG. Samples of media were taken after both the initial and subsequent incubation periods for evaluation of progesterone, testosterone, and estradiol concentrations via radioimmunoassay. After incubation, oocytes were tested for viability using trypan blue and fixed with trichloroacetic acid. The number of ovulated oocytes (determined via dissociation from the follicular tissues and the presence of a "white spot") were counted. In each assay run, six replicate samples were examined for each treatment and the assay was run three separate times.

There are several reservations regarding the performance of this assay. The stage of the collected oocytes used by Orton *et al.* appears fairly random, although a higher proportion of stage V-VI ovaries were collected (for the 10 fragments/ovary, 7 or 8 were stage V/VI and 2 or 3 stage I-IV). Two fragments were used for each test well, but it does not appear that distribution of fragments to the wells was systematic regarding the stages distributed. Sretarugsa and Wallace (1997) report considerable differences for steroid hormone production from frog oocytes at different stages, with optimum steroidogenesis at stages IV and VI but a marked shift in the ratio of progesterone to estradiol production over this development window. This suggests that chance assortment of relatively more or less mature oocytes could have significantly altered the assay outcomes. Further, only two concentrations of 2,4-D were tested.

2,4-D treatment had no effect on ovulation or on steroidogenesis under the conditions of this assay; but the reservations regarding the assay methods should be noted.

Petit *et al.*, 1997

Forty-nine chemicals, including 2,4-D, were evaluated for estrogenic activity in three different *in vitro* assays. The ability of test compounds to induce vitellogenin mRNA expression was evaluated using primary hepatocyte cultures derived from male rainbow trout (250-750 grams each). Liver cells were harvested and $1-2 \times 10^7$ cells per petri dish cultured in DMEM/Ham's F-12 (without phenol red) plus additional supplements. Once aggregates formed, cells were treated for 48 hours with either E2 (10^{-6}

M) or test compound (10^{-4} M or concentration that gave maximal β -galactosidase activity in yeast system); mRNA was then harvested and vitellogenin mRNA expression assessed by slot blot assay (normalizing against β -actin expression). Reported results represent the mean values of 3 to 16 independent experiments. In the case of 2,4-D, vitellogenin expression was slightly greater than that induced by solvent alone; however, the study authors felt that the level of induction was too close to control levels to determine if it was ineffective or weakly estrogenic. Weaknesses include that only a single concentration of 2,4-D was tested.

ER transactivation was evaluated using a stably transfected yeast system expressing both the rainbow trout ER and a β -galactosidase reporter gene downstream of two tandem copies of the consensus ERE. Yeast cells in liquid culture were incubated for 4 hours at 30°C to 10^{-8} – 10^{-4} M of test chemicals or E2, then assayed for β -galactosidase activity using a spectrophotometer. Reported results were based on mean values from at least three separate experiments. Treatment with 2,4-D (purity and form not reported) did not increase β -galactosidase activity significantly above solvent control levels, indicating the absence of rainbow trout ER transactivation.

Rainbow trout ER binding was assessed using the same yeast system described above. Whole yeast extracts were incubated for 16 hours at 4°C to 20 nM ^3H -E2 in the presence or absence of increasing concentrations (2×10^{-6} – 2×10^{-3} M) of radio-inert test compounds, after which samples were treated with dextran-coated charcoal to separate free from bound ^3H -E2. Radio-inert E2 (2×10^{-9} – 2×10^{-5} M) was run as a positive reference chemical. Results were based on mean values from duplicate samples tested at each concentration. 2,4-D did not competitively bind to the rainbow trout ER. The primary limitation of the second and third assays reported by Petit *et al.* involves the use of a yeast-based assay system, as discussed in the note above.

Soto *et al.*, 1995

Various chemicals, including 2,4-D, were examined for estrogenicity using the E-SCREEN assay. MCF-7 cells were seeded into DMEM supplemented with 5% FBS at a density of 2×10^5 cells per well. After 24 hours, the medium was removed, replaced with phenol red-free DMEM supplemented with 5% charcoal-dextran treated human serum, and a range of test compound concentrations added (exact concentrations not reported). After six days of incubation, the cells were lysed and nuclei counted as a measure of cell number. Both the relative proliferative potency (ratio between minimal concentration of E2 need for maximum cell yield and the minimal dose of test compound for a similar effect) and the relative proliferative effect (100 times the ratio between the highest cell yield obtained with the chemical and with E2) were determined for each chemical. Reported results were the mean of at least five separate experiments using duplicate samples in each. In this study, 2,4-D (purity not reported) was considered non-estrogenic. Additional studies of ER and progesterone receptor (PR) processing, induction of pS2 expression (an estrogen responsive gene), and binding to the ER were conducted; however, these studies appear to have only been done on those compounds that were positive in the E-SCREEN, hence 2,4-D was not tested in these additional assays.

Vonier et al., 1996

Various pesticides, including 2,4-D, were evaluated for the ability to competitively bind to the ER and progesterone receptor (PR) extracted from oviduct tissue of adult female alligators (*Alligator mississippiensis*) captured in central Florida. Frozen tissues were thawed on ice, homogenized in buffer, centrifuged, and supernatants taken for subsequent receptor binding assays. To assess ER binding, protein extracts were incubated with 2.5 nM ³H-E2 for 1 hour at 25°C, after which any free ³H-E2 was removed using a charcoal-dextran mixture. Samples were then incubated with increasing concentrations of radio-inert test chemicals (exact concentrations not reported) for 1 hour at 25°C. Free ³H-E2 was again removed using the charcoal-dextran mixture and samples analyzed for ³H-E2 binding. To evaluate PR binding, samples were incubated at 4°C for 12 hours with ³H-R5020 in the presence or absence of increasing concentrations of radio-inert test chemicals. Free ³H-R5020 was removed using a charcoal-dextran mixture and samples analyzed for ³H-R5020 binding. Findings for both receptor binding assays are the mean results of at least three independent experiments using three replicates each. 2,4-D did not bind the ER or PR to any significant extent in these experiments.

This study was conducted using a different model system (alligator ER) than that specified in the EPA guidelines for the ER binding assay (EPA, 2009a) and the exact range of test compound concentrations examined were not reported (Klimisch 3). Nevertheless, this study provides additional support indicating that 2,4-D does not bind to the ER. Additionally, it shows that 2,4-D does not act at the PR either.

Supplementary Appendix VI

Ecotoxicological Studies with Klimisch Score of 3 or 4⁵

Amphibians

Aronzon et al. (2011)

Aronzon *et al.* evaluated acute and short term toxicity of 2,4-D DBE (99% purity) and a commercial formulation (CF) to South American Toad, *Rhinella arenarum*, embryos. Assays were conducted with the AMPHIEMB protocol for early life stage toxicity and AMPHISHORT for seven-day toxicity. Duplicate batches of 10 embryos were treated with either 2,4-D DBE or with CF. Experiments were repeated four times.

Embryos were exposed

- (i) continuously from 2–4 blastomere stage up to the end of embryonic development in concentrations ranging from 1 to 15 mg/L 2,4-D DBE and from 0.25 to 3.85 mg/L of CF,

⁵ See Publication **Table 7** for Klimisch scores and comments for published ecotoxicological studies

- (ii) in pulses during the blastula, gastrula, tail bud, muscular activity, gill circulation, open mouth, opercular folds stages for 24 hours with doses ranging from 1 to 26 mg/L for 2,4-D DBE and 0.1 to 10 mg/L for CF, and
- (iii) continuously from the end of embryonic development for 168 hours to concentrations ranging from 8 to 20 mg/L for 2,4-D DBE and 2 to 5 mg/L for CF.

Adverse effects were analyzed for lethality, malformations, stage-dependent susceptibility, and ultrastructural features. The CF was significantly more toxic than 2,4-D DBE in all cases where embryos were continuously exposed from the blastula stage onwards (2.68 ± 0.05), in all pulse treatment experiments, and in post-embryonic development continuous exposure studies (4.45 ± 0.37). The authors stated both compounds were teratogenic. However, it appears no controls were run for these experiments and therefore it is not possible to interpret the results.

Heggstrom, 2009

Wood frog tadpoles (*Rana sylvatica*) were exposed to nominal concentrations of 0.1, 1.0, and 100 µg/L 2,4-D dimethylamine (99% pure) in microcosms from shortly after the time of hatch until metamorphic climax (approximately 30- 56 days depending on tadpole development). Forty tadpoles were placed in each microcosm and five replicate microcosms were used per treatment level. Survival, deformities, time to metamorphic climax and morphometric measures including total length, snout-vent length, and wet weight were analyzed. In addition, circulating levels of cortisol were determined from successful blood collections. There were no 2,4-D dimethylamine treatment-related effects on any of the measured parameters, except for total length, which was observed to be significantly decreased in the 0.1 µg/L treatment group (but not the 1.0 or 100 µg/L treatment groups) relative to controls. The biological significance of this finding is unknown, since this endpoint was reported to be highly variable at metamorphic completion. The absence of dose response suggests this is not an exposure-related effect. Additionally, high mortality among the tadpoles was observed in the microcosms across all treatment groups, possibly due to viral infections or latex toxicity (due to lab gloves), thus the interpretation of this microcosm study in regard to 2,4-D dimethylamine toxicity is problematic.

The same basic experimental design and measurement endpoints were used with wood frog tadpoles in field ponds (forested, agricultural, agricultural + 10 µg/L 2,4-D dimethylamine application). The results from the field pond studies indicated that tadpoles in the agricultural ponds were larger than the tadpoles in the forested ponds. Additionally, tadpoles in the agricultural ponds had increased cortisol levels compared to tadpoles in the forested pond. The relevance of 2,4-D exposure on the observed endpoints cannot be ascertained from the field pond study since tadpoles from the agricultural pond in the absence of 2,4-D dimethylamine and tadpoles from the agricultural pond applied with 10 µg/L 2,4-D dimethylamine gave similar results in the measured endpoints.

LaChapelle et al., 2007

Xenopus oocytes were harvested from anesthetized adult female frogs and exposed to 2,4-D (purity not reported) for 48 hours to determine the mechanism of 2,4-D inhibition of oocyte maturation. The exposure concentration was 2.43 g/L (10 mM). Oocytes were microinjected with radiolabelled *in vitro*

transcripts of Mos RNA and exposed to progesterone and 2,4-D to decipher the mechanism of 2,4-D dysfunction of the meiotic signaling mechanism. 2,4-D induced irreversible dysfunction of the meiotic signaling mechanism and this was linked to inhibition in the cascade of signaling events from membrane-bound progesterone receptor binding. The investigators indicated the results show that 2,4-D may inhibit progesterone induced gamete maturation. This study was a mechanistic study and tested only a single extremely high concentration of 2,4-D. The concentration of 2,4-D evaluated in this study was orders of magnitude greater (approximately 1,000 x greater) than an environmentally relevant dose. Furthermore, a dose of this level would be expected to result in overt toxicity (lethality) *in vivo*, based on the LC50 for 2,4-D ranging from 225 to 359 mg/L among various species of frogs (Palmer and Krueger, 1997; Morgan et al., 1996). It is not considered relevant to any realistic exposure scenario.

Morgan et al., 1996

The teratogenic potential of a “commercial formulation” of 2,4-D (Aldrich Chemical Company, 99% pure) was evaluated in the frog embryo teratogenic assay-Xenopus (FETAX). The FETAX test is a 96-hour static renewal bioassay used to determine the teratogenic potential of chemicals. Toxicity of 2,4-D toward *Xenopus* embryos was assessed in FETAX assay buffer and in natural water. The natural water used in this study is of unknown composition, and may be lacking essential ingredients for normal development, thus the results of the evaluation in FETAX assay buffer (supplemented with essential cations and anions for proper embryo development) is discussed. Fertilized eggs were harvested from breeding adult *Xenopus laevis* frogs and de-jellied. The de-jellied embryos were placed in 100 mm x 15 mm plastic petri dishes (20 embryos/plate and four replicate plates per treatment) and exposed to waterborne concentrations of 2,4-D for 96 hours with daily renewals of test solutions. Nominal 2,4-D concentrations were 180, 190, 210, 220, 230, 240, 250, 260, and 270 mg/L. The LOEC, EC50 (malformations), and LC50 for 2,4-D were determined to be 226, 245, and 254 mg/L, respectively. Both mild and severe edema of the gut were noted among tadpoles exposed to 2,4-D. The results of this study indicate that 2,4-D is teratogenic to *Xenopus laevis* embryos only at high concentrations, similar to those concentrations causing lethality. This study is not considered useful for the weight-of-the evidence evaluation.

Stebbins-Boaz et al., 2004

Xenopus oocytes were harvested from anesthetized adult female frogs and exposed *in vitro* to 2.5 - 10 mM 2,4-D sodium salt (0.6 – 2.43 g/L) at pH 6.8 for variable periods of time (1-10 hours). At the concentrations of 2,4-D evaluated, germinal vesicle breakdown (GVBD) was blocked in the presence of progesterone (via inhibition of the Mos protein via blocking post translational polyadenylation of mRNA) and also microtubule depolymerization was observed.. The results suggested that 2,4-D requires microfilaments (actin) to initiate observed alterations in oocyte morphology. Also, 2,4-D induced MAPK activation. There was no discussion of the environmental relevance of these findings; rather the paper was mechanistic in focus. The concentrations of 2,4-D used in this *in vitro* study were orders of magnitude above expected environmental concentrations and were also above concentrations that are associated with acute lethality in various species of frogs. It is not considered useful for the weight of evidence evaluation.

Vardia et al., 1984

Indian toad tadpoles (*Bufo melanostictus*) were exposed to waterborne concentrations of 2,4-D (source unknown; assumed to be commercially available formulation) over 96 hours in a static renewal exposure system. The concentrations of 2,4-D used were 0, 7.5, 10 and 11 mg/L. One liter beakers were used as test vessels. The temperature during experimentation was around 25 °C, the pH was around 8.3, and total alkalinity and hardness were 210 and 220 mg/L CaCO₃, respectively. The 96-hour LC50 value for 2,4-D was reported to be 8.05 mg/L. Little experimental detail is discussed in this publication and the actual form of 2,4-D tested is unknown. This study lacks sufficient information to be useful for assessment of potential endocrine effects.

Fish

Xie et al., 2005

In this investigation juvenile rainbow trout (*Oncorhynchus mykiss*) of unknown sex (standard length 11.5 ± 2.2 cm) were exposed to 2,4-D dimethylamine (Nufarm Co St. Joseph, MO, USA; purity not stated) for 7 days in a daily static renewal system. Several different exposure scenarios with 2,4-D dimethylamine were conducted. Rainbow trout were exposed to one “worst case” concentration of 1.64 mg/L 2,4-D dimethylamine (measured), and also to 2,4-D dimethylamine at concentrations of 0, 0.00164, 0.0164, 0.164, and 1.64 mg/L. Rainbow trout were exposed to 2,4-D dimethylamine (at the same nominal concentrations) in combination with the surfactants R-11 and Target Prospreader Activator (TPA). Each test concentration level was composed of three replicate tanks with two fish in each tank, in a static test system.

Following exposure, blood was collected from each fish, and vitellogenin levels were measured in the blood plasma with the use of a commercially available rainbow trout ELISA kit (Biosense, Bergen, Norway). In the first exposure to the “worst case” concentration of 2,4-D (1.64 mg/L), vitellogenin levels were significantly greater than control levels. In combination with the surfactant R-11, the same concentration of 2,4-D, however, did not result in vitellogenin levels that were significantly greater than controls. In combination with the surfactant TPA, however, the same concentration of 2,4-D did result in vitellogenin levels that were significantly greater than controls. In the dose-response study with 2,4-D, concentrations ≥ 0.164 mg/L resulted in significantly greater concentrations of vitellogenin compared to controls. Thus the NOEC and LOEC for vitellogenin induction in the rainbow trout were reported as 0.0164 and 0.164 mg/L, respectively. These values were altered based on co-treatment with both R11 and TPA surfactants. The NOEC and LOEC for 2,4-D in the presence of various concentrations of TPA were lowered to 0.00164 and 0.0164 mg/L, respectively. The NOEC and LOEC for 2,4-D in the presence of various concentrations of R-11 were increased to 0.164 and 1.64 mg/L, respectively.

The sample size is small, especially for measuring inherently variable endocrine parameters such as vitellogenin concentrations. Although not specifically reported, the variability around these measured vitellogenin concentrations was notably high (by visual inspection of the figures). For example, control vitellogenin responses ranged from near zero to as high as approximately 10 ng/mg total plasma protein; such variation confounds clear identification of NOEC and LOEC values reported in this study. Xie et al. also noted that a combined treatment with 1.64 mg/L 2,4-D and 1.46 mg/L of the surfactant R-

11 resulted in approximately 16-18 ng/mg vitellogenin expression. A subsequent experiment, however, using a similar treatment period and 2,4-D concentration (1.64 mg/L) but a lower concentration of R-11 (0.89 mg/L) resulted in much higher vitellogenin expression (approximately 50 ng/mg).

The sex and age of the juvenile fish were not reported in this study. Schlenck et al (2008) in a rebuttal to an unpublished critique (Kramer *et al.*) on the Xie *et al.* study indicated that the fish were young juveniles and sex would therefore not influence vitellogenin levels. However, detectable VTG has been reported in immature female trout (Bon *et al.*, 1997). It is possible therefore that endogenous VTG in any female fish confounds interpretation of the study results.

No effects on vitellogenin were observed in the 2,4-D FSTR assay (Coady *et al.*, 2013), in contrast to the reported findings by Xie *et al.* This may be due to species differences (the FSTRA was in fathead minnows, versus the Xie *et al.* study in rainbow trout). Schlenk *et al.* (2008) postulated the observed effects in the Xie *et al.* study might be due to a 2,4-D metabolite; this does not however explain the absence of effects on vitellogenin in the Coady *et al.* (2013) FSTRA study which exposed more fish to 2,4-D for a longer duration.

As the Xie *et al.* test system was static it is also possible 2,4-dichlorophenol (2,4-DCP) accumulated as an aquatic 2,4-D degradate (Coady *et al.* used a flow through system). 2,4-DCP is a photolysis/hydrolysis environmental metabolite of 2,4-D, and exhibits relatively weak *in vitro* estrogen receptor binding activity (Li *et al.*, 2012; Nishihara *et al.*, 2000). Ma *et al.* (2012), reported adverse effects on zebrafish reproduction in fish exposed to 2,4-DCP at 0.3 mg/L; it is unclear whether these effects are causally linked to an estrogenic mode of action, or whether decreased reproduction is related to general systemic toxicity, since the acute toxicity for 2,4-DCP exposed zebrafish is only approximately 10-fold greater (LC50 of 3.9 mg/L (ECOTOX database)). Ma *et al.* (2012) concluded that there was potential evidence of changes in steroidogenesis that might affect reproduction. The Ma *et al.* study appeared to be reasonably well conducted, although hormonal and reproductive data were variable and the concentrations of 2,4-DCP evaluated were above what would be expected in the environment.

Thus, the relatively small sample size, variable nature of the reported responses and unknown age and sex of the fish studied by Xie *et al.* (2005) requires further investigation and replication before inferring either fish or mammalian environmental health risks.

Holcombe *et al.*, 1995

Japanese medaka (*Oryzias latipes*) were exposed to 2,4-D acid (CAS # 94757; 99% pure) in 96- hour acute and two 28-day larval survival and growth tests. In the chronic test, exposure to 2,4-D acid was discontinued after 28 days at which time live fish weights were taken and fish were then transferred to clean water for an additional 5 months. At the end of a five month period, fish tissues were assessed pathologically for incidences of tumors (data not reported in this publication). 2,4-D stock solutions were pH adjusted with NaOH and HNO₃. A continuous-flow mini diluter exposure system with a flow rate of 25 ml/min was used for both the acute and chronic tests. Water chemistry data including dissolved oxygen, pH, temperature, hardness, and alkalinity were measured during the exposure and were within acceptable ranges for medaka. The medaka used for all tests were obtained from the

Environmental Research Laboratory-Duluth culture unit. Twenty medaka (10 per replicate) were exposed per concentration in the acute tests. Mean measured concentrations of 2,4-D acid in the acute study were < detection limit (control), 567, 1100, 2250, 4560, and 8970 mg/L. The 96-hour LC50 for medaka was determined to be 2780 mg/L 2,4-D Acid. Sixty medaka larvae were stocked in each test vessel (18.5 x 14.0 x 13.0 cm deep) in the chronic studies. Two replicate vessels per test concentration were used in both of the 28-day chronic toxicity tests. Mean measured concentrations of 2,4-D acid in the first 28-day study were < detection limit (control), 27.2, 56.5, 113, 221, 425 mg/L. Mean measured concentrations of 2,4-D acid in the second 28-day study were < detection limit (control), 2.37, 5.73, 13.5, 30, 60.2 mg/L. Survival and growth of medaka were both significantly reduced by 2,4-D concentrations of 56.5 mg/L and 60.2 mg/L during test #1 and test #2, respectively. Although the study quality is adequate, there are no specific endpoints related to endocrine activity in this study.

Koc and Akbulut, 2012

Koc and Akbulut exposed zebrafish for 5 days to 2,4-D (purity not defined, possibly a formulation) at concentrations of 0.1-1 mg/L. No dose analyses were done. They reported ovarian histopathological changes and atretic follicles with a dose-related increase in severity. Selected slides of ovarian tissue were presented. The control slide was at a different level of magnification than slides from exposure fish, making direct comparison difficult. Additionally, some of the slides show different staining (suggesting excessive heat during tissue processing), and sectioning artifacts (characteristic of a dull microtome blade). Despite these problems there were some marked differences in the appearance of follicles between the control and exposed groups. In the absence of information regarding the test material and in view of possible processing and sectioning artifacts in the ovarian histology, this study is not considered of sufficient quality to include in the WoE.

Padilla et al., 2012

As part of the Computational Toxicology Research Program of the U.S. EPA, the toxicity of the 309 ToxCast™ Phase I chemicals was assessed using a zebrafish screen for developmental toxicity. Exposure (immersion) was from 6–8 hours post fertilization to 5 days post fertilization. 2,4-D (purity greater than 90%) in aqueous DMSO was tested in a single concentration study at a relatively high nominal dose (80 µM). Five days post fertilization larvae were assessed for death and overt structural defects. 2,4-D was reported as “positive” in the single high concentration study, although the degree of response was weak. Further details of the response were not specified. 2,4-D was then tested in a multiple concentration study, with concentrations ranging from 0.001 to 80 µM. 2,4-D was negative in this study and no AC50 was calculated. Although this study was considered relatively reliable (Klimisch score of 2) the findings reported in this study are too non-specific to provide evidence of potential endocrine pathway interactions (and further were not replicated).

Rehwold et al., 1977

Acute toxicity tests (96 hour) with 2,4-D (no form or purity information given) were conducted with striped bass (*Marone saxatillis*), the banded killifish (*Fundulus diaphanous*), pumpkinseed, white perch (*Roccus americanus*), American eel (*Anguilla rostrata*), carp (*Cyprinus carpio*) and guppy (*Libistes*

reticulatus). All fish species, except for the guppies, were collected from the Hudson River. Guppies were purchased from a pet store. 2,4-D was obtained from Analabs, Inc. The experimental design and test concentrations were not described in detail in the publication. 96-hour LC50 values ranged from 26.7 mg/L (banded killifish) to 300.6 mg/L (American eel). Chronic exposure studies with fish species exposed to 0.1 mg/L 2,4-D were conducted for 10 months, and resulted in no observable physiological symptoms, however no histopathological investigation was performed in these studies and no detailed methods were available to assess the validity of these studies. Breeding experiments were conducted with the guppies, with no observable effects for those exposed to 2,4-D at 0.1 mg/L for 10 months. This data is not considered valid for use in a WoE for potential endocrine effects since, among other deficiencies; the majority of fish species used in the studies were field collected from the Hudson River and presumably may have been exposed to multiple environmental contaminants. In addition, there were no experimental details provided in the publication, thus the validity of these studies could not be adequately assessed (Klimisch 4).

Avian

Somers et al., 1974

Aqueous solutions of 2,4-D (amine form; no purity information given) at recommended (2.8 - 3.4 kg/ha) and 20x field application rates (44.8 kg/ha) were sprayed on fertilized hen's eggs (*Gallus domesticus*; single comb while leghorn chickens) prior to incubation. (The number of eggs tested in the study was not reported.) Mortalities *in ovo*, at pip, and at hatch were assessed in addition to weight gain of the chicks 3-4 weeks subsequent to hatch. Vent sexing was performed on all chicks that successfully hatched. No adverse effects on any of the measured parameters were noted with respect to 2,4-D application. Residue analysis indicated that internal contents of the egg showed entry of 2,4-D to be far less than proportionate. This study is not relevant because of the unusual method of application, demonstrated low 2,4-D absorption and the lack of information on the nature of the test substance.

Somers et al., 1978a

Hens and cockerels (*Gallus domesticus*) that came from fertile eggs, were subjected to spray contamination with a "10x recommendation dosage" of 2,4-D (commercial formulation "Esteron 99"; purity not reported), and were evaluated for various aspects of reproduction including egg production, egg weight, shell porosity and strength, sperm counts, testes weights, gross appearance of the testes, embryo mortality, hatching success, and chick mortality and malformations, and weight gain. Two incubation experiments were conducted: the first evaluated reproductive success of the offspring as a function of parental spray treatment, the second involved an egg re-treatment with spray contamination restricted to the pre-incubation period. In general, there was no definitive evidence that 2,4-D had any significant effects on chicken reproduction through one generation and into the second. This study is not relevant for the weight of the evidence evaluation because of the lack of dosimetry information and the lack of information on the test substance.

Somers et al., 1978b

Eggs were obtained daily from a commercial strain of Single Comb White Leghorn hens (*Gallus domesticus*). A total of 432 eggs over 12 replicates per treatment for each of the 3 stages of embryonic development were used in the study design. Commercially available formulation of 2,4-D (Trade name: Esteron 99, purity not reported) was applied via spray to the fertilized eggs at a rate of 10x the “recommended rate (11.2 kg/ha).” Late dead germs, pipped dead germs and all chicks were examined for gross abnormalities. Live performances of a subset of successful hatches were assessed until males were 20 weeks of age and females were 16 weeks of age. During this time body weights and mortality were assessed. (Numbers of subjects and experimental design were not reported in detail.) A statistical evaluation of all data associated with incubation failed to disclose a significant interaction involving spray treatment and the stage of embryonic development at application. Furthermore, 2,4-D application had no effect on hatching success, weight gain, or mortality even up to 16-20 weeks of age. This study is not relevant for the weight of the evidence evaluation because of the lack of information on the test substance and uncertain penetration of the test material through the shell. It suggests, however, that there should be little concern for bird eggs inadvertently exposed to formulated 2,4-D.

Mixed Species

Relyea, 2005

Pesticide treatments were replicated six times in 1000 L outdoor microcosms containing algal and 25 animal species including dragonflies, cladocerans, copepods, snails, salamanders, four species of anuran amphibian tadpoles, and various other water bugs. The commercial form of 2,4-D (44.5% active ingredient) was added to microcosms at the manufacturer’s recommended application rate (0.117 mL/m² for 2,4-D). After 13 days of exposure, the microcosms were examined for community diversity, survival, and biomass. Species richness, biomass, and survival of the exposed animal species were not significantly different between control and 2,4-D microcosms. Furthermore, the survival of four tadpole species (Leopard frog, American toad, Gray tree frog, Spring peeper) were unaffected by 2,4-D application. The other insecticides and herbicides evaluated in this test system had marked effects on species richness. There is no specific endpoints related to endocrine activity in this microcosm study, and it is not included in the WoE. The results indicate however, that 2,4-D, when applied appropriately, has no effect on community species richness or tadpole survival.

Supplementary Appendix VII

Mammalian Toxicological Studies with Klimisch Score of 3 or 4⁶

Developmental Studies

Bage et al., 1973

In this study, pregnant NMRI mice were dosed subcutaneously with formulated herbicides containing either 2,4,5-T or a mixture of 2,4,5-T and 2,4-D from days 6-14 of gestation. Doses were 110 mg/kg/day 2,4,5-T; 50 mg/kg/day 2,4,5-T; 110 mg/kg/day 2,4-D/2,4,5-T (2:1); 50 mg/kg/day 2,4-D/2,4,5-T (2:1); and vehicle, for groups A, B, C, D, and E, respectively. The vehicle control group received an equivalent volume of DMSO with 19 mg/mL of the formulated product petroleum distillate mixture. The mice were cesarean-sectioned at gestation day (GD) 18, and the fetuses were evaluated.

Maternal toxicity was not evaluated in this study. There was a marked increase in fetal resorptions, particularly in the high-dose groups. The fetal incidence of cleft palate was also increased in these groups. Decreased mean fetal body weights were also observed in these groups, and the lower body weight fetuses were described as showing retarded skeletal development. The percentage of fetuses with skeletal malformations was also increased, particularly in the 2,4,5-T high-dose group.

This study provides results that are generally considered consistent with other high-dose developmental studies of phenoxy acids at doses that exceed renal clearance of 2,4-D (Garabrant and Philbert, 2002). Because of study design and reporting deficiencies, however, including a route of administration inappropriate for risk assessment, the use of formulated product and absence of a group testing 2,4-D alone, the study is not considered of high quality. This study does not provide any specific information regarding the endocrine-disrupting potential of 2,4-D and is not considered useful for the WoE.

Cavieres et al., 2002

Cavieres *et al.* (2002) tested the effects of a commercial formulation of 2,4-D (7.59%), mecoprop (3.66%), dicamba (0.84 %) and excipients, administered in drinking water, on implantation, litter size, pup weights, and crown rump length in ND4 mice. This study combined several studies conducted with two different dosing regimens and smaller dose groups, one from GD 0-15, and one from GD 6-15. Nominal doses of 2,4-D were 0, 0.01, 0.10, 20 and 100 mg/kg/day. Cavieres *et al.* (2002) reported a decrease in live litter size and implantations at all doses of 2,4-D. They also interpreted their data as showing a selective low dose effect of 2,4-D. However, there are a number of significant problems with the study designs and data interpretation that are the bases for these conclusions.

The data were analyzed in depth by Lamb *et al.* (2003), who reviewed the data in the publication in detail, as well as reviewing Cavieres' dissertation, Cavieres (2001) which appeared to serve as the basis

⁶ See publication **Table 11** for Klimisch scores and summary comments for published mammalian toxicological studies.

for much of the published information. Lamb *et al.* (2003) submitted a letter to the journal editor with the conclusions from their review. Ashby *et al.* (2003) also submitted an independent critique of the Cavieres *et al.* (2002) conclusions. The major flaws in the Cavieres *et al.* research stem from combining data from multiple experiments, including an experimental design that cannot correctly assess potential effects on implantation, as implantation in mice was completed prior to the start of dosing in several of the studies that were combined by Cavieres *et al.* (2002) for analysis. Cavieres *et al.* (2002) could not have accurately assessed *corpora lutea*, as dams were not sacrificed until weaning of the litters, and *corpora lutea* in rodents will regress during the lactation period. Therefore, the number of potential pups per litter is unknown. Further, Cavieres *et al.* (2002) did not assess litter size immediately after birth, so the possibility that pups were cannibalized and not counted cannot be excluded. Additionally, implantation sites were not counted for all mice. Review of the dissertation further showed data discrepancies between the dissertation and the publication which acted to enhance the statistical significance of the reported effects in some groups of animals. The study authors responded (Cavieres *et al.*, 2003), but do not appear to have addressed all of the critical questions regarding the original publication. In view of testing of a formulation with unknown excipients, the questionable design and inappropriate merging of data from different studies, Cavieres *et al.* (2002) is not included in the WoE.

Collins and Williams, 1971

Syrian golden hamsters were dosed (7-12 litters/group) with test compound from one of three different lots of technical 2,4-D on GD 6-10, inclusive. Doses ranged from 20-100 mg/kg/day. (Six lots of 2,4,5-T were also tested; this phase of the study is not discussed further.) It appears that controls were pooled, so comparison to concurrent control groups is not possible. Dams were caesarean sectioned on gestation day 14, and the uterus examined for resorptions. Fetuses were removed, weighed and examined externally. Approximately 1/3 of fetuses were processed for skeletal evaluation, with the remaining fetuses split into two groups, one for dissection and one for histopathological evaluation.

Slightly decreased fetal viability was observed in the mid- and high-dose compound lot C dams, although no strong dose response was present. There was no treatment-related increase in resorptions or effect on fetal body weight. Percent abnormalities per live litter (according to Table 1) were not statistically significantly higher than the control in any 2,4-D-exposed group. An increased incidence of specific anomalies was reported at doses of 60 mg/kg/day of 2,4-D and higher, although the fetal incidence was low, and dose response was present only for compound lot C. The findings in this study should be considered in the context that hamsters have an extremely high background incidence of malformations, such that they are seldom used in current developmental toxicity evaluations. Although this study has multiple weaknesses that limit its reliability including use of pooled controls and small group size, the results are generally considered consistent with those of other high-dose developmental studies of phenoxy acids at doses that would be expected to exceed renal clearance of 2,4-D. However, the weaknesses in this study are significant.

Dominant Lethal

Oakes et al., 2002a

Male Sprague-Dawley rats (12 weeks old, five/group) were dosed with Tordon 75D[®] (2,4-D and picloram in formulation) 5 days per week for 9 weeks at doses equivalent to 37.5, 75, and 150 mg/kg/day of 2,4-D, and 9.4, 18.7, and 37.5 mg/kg/day picloram for the low-, mid-, and high-dose groups, respectively. Untreated and positive control (cyclophosphamide [CP]) groups were also used, with a similar exposure regimen. Males were given the opportunity to mate with two untreated females during weeks 2 and 3, 4 and 5, and 8 and 9 of treatment. The males were then left untreated and unmated for 11 weeks, after which they were given the opportunity to mate with four females over a 2-week period. This study was done in two replicates over a 2-year period. In a small additional experiment, control, positive control and high-dose males (2/group) were mated with two untreated females, and the females were allowed to litter to give an F1 generation. At maturity pups from the F1 litters were mated (with offspring of the other female from the same dose group). For both studies, females were killed on gestation day 20, corpora lutea were counted, live and dead fetuses and visible resorptions were counted, and fetuses were weighed and evaluated for visceral and skeletal anomalies. Data were analyzed by analysis of variance (ANOVA) with the Fisher exact test. The litter was considered the unit of analyses for the reproductive studies.

Some rats (dose level unspecified) appeared lethargic shortly after dosing. There was slight body weight loss in the high-dose group through the first 4 weeks, and decreased body weight gain through the remainder of the dosing period. Body weights in the low and mid-dose were equivalent to the controls. There were no dose-related effects on fertility, post-implantation loss, mean litter size or weight, number of runts, incidence of malformations or skeletal variations in breeding results from any mating interval attributable to Tordon 75D[®]. Pre-implantation loss was increased in the high-dose group for both the 2- to 3-week and 4- to 5-week mating interval litters, (15.2% and 16.2%, respectively, compared to a range of 5.1%-11% in the various control groups); these increases are not statistically significantly different from control. Although the study authors consider these to be possibly treatment-related findings, we consider this unlikely. The incidence of pre-implantation loss is based on comparison of the total of fetuses and resorptions to the number of *corpora lutea*, which is inexact both because of the difficulty in detecting early resorptions (particularly if the uteri are not stained to facilitate this, which was not done in this study), and also because the *corpora lutea* count may be inexact. There was a statistically significant increased incidence of hydronephrosis in high-dose fetuses from the 2- to 3-week mating interval litters (5.8% compared to 0-1.9% in the various control groups). It cannot be determined from the data presented whether these are fetal or litter incidences. The authors do not discuss the treatment relationship of this finding. Hydronephrosis/dilated renal pelvis may be a subjective finding that tends usually to have a moderate background incidence in Sprague-Dawley rats. The incidence reported in this study for the Tordon 75D[®] high-dose level is within the historical control range reported for Sprague-Dawley rat fetuses reported by Middle Atlantic Reproduction and Teratology Association (MARTA), which examined the incidence of malformations in 223 studies conducted from 1989-1992 (reported in *Handbook of Developmental Toxicity*, R. Hood, ed., CRC Press, 1996). Overall,

Tordon 75D® showed minimal treatment-related effects (if any), and none of the findings characteristic of dominant lethal or male mediated toxicity.

In contrast, the positive control (CP) dosed rats showed increased pre-implantation loss (16.9% and 19.8% for the 2- to 3-week and 8- to 9-week mating interval litters, respectively; an increased number of female fetuses, for the 4- to 5-week mating interval; increased post-implantation losses at all three mating intervals (18.5%-60.7%) and correspondingly decreased litter sizes; and increased malformations for the 4- to 5-week mating interval litters. These results are as would be expected for a positive dominant lethal response.

In the small scale study breeding F1 animals, there were no apparent differences from control in the F2 generation fetuses for either the high-dose Tordon 75D® or the CP groups. The latter study is clearly limited by the very small group size evaluated. The authors conclude that these studies indicate “within the limitations of the power of the study, that herbicide mixtures of 2,4-D and picloram are unlikely to be male reproductive toxicants.” The major problems in this study are that 2,4-D was tested only as part of a formulation mixture, and the group size tested was relatively small.

Male Reproductive Toxicity

Blakley, et al., 1989

In this study, male CD-1 mice were dosed via drinking water for 60 days with Tordon 202c and mated to untreated females. Calculated compound intake were 5, 10, and 20 mg/kg/day picloram and 84, 168, and 336 mg/kg/day 2,4-D for the low, medium, and high concentration groups, respectively. Reduced water consumption and high mortality was seen in the adult males at the highest concentration. Females were cesarean sectioned at gestation day 18 and fetuses evaluated. The numbers of implantations, live or dead fetuses, and resorptions were not affected. The investigators note that fetal weight was decreased at the high dose, and that there were increased variations and/or malformations at all doses tested. No pattern of specific malformations was identifiable, and no dose response was present. Variations were reported primarily as the decreased ossification typical of developmental delay, correlating with the decreased mean fetal weight noted in the high-dose group. There are several deficiencies in the study design and reporting that make the results difficult to interpret with respect to developmental toxicity of 2,4-D. First, 2,4-D was not tested separately, but only in conjunction with picloram and with other formulation excipients (not identified). Second, the group size was small (5-14 litters). Third, fetal and litter incidence of specific types of malformations and/or variations were not reported. The study findings are considered inconclusive and not attributable to endocrine disruption. Although it is possible that some toxicity to sperm occurred in the high-dose group, it should be noted that this group was dehydrated and showed high mortality. This study is not included in the WoE evaluation.

Stoker et al., 2007

Stoker *et al.* reported an evaluation of the effects of 2,4-D on pubertal development and thyroid function in the juvenile male Wistar rat. This study was reported in an abstract only, with limited supplementary information presented in a book chapter by Stoker and Zorrilla, 2010; information important for interpreting the results such as impact on body weight of the dosed animals was not reported. The rats were exposed to 0, 3, 30, 100 or 200 mg/kg/day of 2,4-D by oral gavage from postnatal day (PND) 23 to 53, with the two high dose levels markedly exceeding the TSRC. Preputial separation (PPS) was evaluated beginning on PND 33. The rats were necropsied on PND 53 and tissue weights recorded. Serum was collected at necropsy and analyzed for thyroid, testicular and pituitary hormone concentrations. PPS was significantly delayed for 2.7 days in the 200 mg/kg/day group compared to controls. However, no effect on PPS was observed at 100 mg/kg/day. The mean weight of the ventral prostate and levator ani plus bulbocavernosus muscles (LABC) was reported to be significantly decreased in the 200 mg/kg/day group. Testosterone and androstenedione were reported to be significantly decreased following exposure to the highest dose of 2,4-D, but luteinizing hormone and prolactin were not altered at either dose. Triiodothyronine (T3) and thyroxine (T4) were significantly decreased at both 100 and 200 mg/kg of 2,4-D, but not at 30 mg/kg/day. TSH was not evaluated, and no histopathological evaluation of the thyroids was done, so the biological significance of this finding cannot be assessed, but it is not inconsistent with findings in subchronic studies at similar doses exceeding the TSRC. Information on this study is limited, the purity of the test material is not defined, and key information such as body weight effects are not reported. Unfortunately limited details are available regarding the conduct of this study, which otherwise supports that thyroid and male reproductive hormone changes are only evident at doses exceeding the TSRC in rats and are hence not relevant for human risk assessment.

Kim et al., 2002

This study used the Hershberger assay to evaluate potential androgenicity of 2,4-D (50 mg/kg/day) and its metabolite, 2-4-dichlorophenol (DCP; 100 mg/kg/day) in castrated CD rats. Male Sprague-Dawley rats castrated on postnatal day (PND) 42 were treated with 2,4-D or DCP from PND 57-66. Consistent with the OECD and EPA test guidelines, there were 6 animals per treatment group. The authors reported statistically significant increases in ventral prostate, Cowper's gland and glans penis weights with both 2,4-D and DCP treatment, suggesting androgenicity. Furthermore, the authors reported that these compounds potentiated increases in accessory sex tissue weights by testosterone.

It is difficult to fully review the manuscript by Kim *et al.* (2002) as the publication is written in Korean. Still, a few observations on study procedures and results are possible. First, it is unclear what dose of testosterone was used in this study (i.e., the abstract states "1 g/kg", whereas the Hershberger assay description states "1 ug/kg" and Figure 2 on Hershberger body weights states "1 mg/kg/day"). The Hershberger test guidelines (OPPTS 890.1400; OECD 441) recommend doses of 0.2-0.4 mg/kg/day testosterone propionate for the anti-androgenic assessment. In the Kim study, only a single dose level

of 2,4-D was examined (50 mg/kg/day, exceeding the TSRC), whereas a minimum of 2 dose levels are required in the test guidelines to assess androgenicity, and 3 dose levels to assess anti-androgenicity.

The relative potency of 2,4-D relative to testosterone reported in the Kim study also poses questions. 2,4-D was reported to significantly increase seminal vesicle and Cowper's gland weights and interestingly, the magnitude of increases in accessory sex gland weights with 2,4-D were reported to be similar to the increases seen with testosterone treatment. Similarly, DCP at 100 mg/kg/day also were reported to have had effects similar in magnitude to testosterone. A later publication indicates "In a previous study, we showed that these compounds exhibited synergistic androgenic effects by co-treatment with testosterone in the Hershberger assay." The problem is that the testosterone supplemented assessment in the Hershberger is designed to measure potential anti-androgenic activity, not androgenic activity, and it never has been validated for the latter purpose.

On closer inspection, the organ weights reported by Kim *et al.* (2002) differ from other studies using testosterone-supplemented male Sprague-Dawley rats castrated at ~6 weeks of age and with similar terminal body weights. For example, control organ weights in the Kim *et al.* study were 7.66 mg ventral prostate; 65.83 mg seminal vesicle; 151.18 mg levator ani-bulbocavernosus muscle; 29.47 mg glans penis and 2.28 mg Cowper's glands; whereas mean control organ weights in the validation study by Yamasaki *et al.* (2003) were 18.0 mg ventral prostate (range 12.6-21.1); 45.4 mg seminal vesicles (range 39.7-52.5); 209.1 mg levator ani-bulbocavernosus muscle (192.3-236.3); 54.0 mg glans penis (range 48.9-62.6) and 6.9 mg Cowper's glands (range 6.5-8.1). Using the same test system (CD rats castrated at 6 weeks of age), Kang *et al.* (2004) reported accessory sex tissue weights similar to Yamasaki *et al.* (2003).

It also seems unlikely that 2,4-D could have similar androgenic potency to testosterone and not be detected in other androgen receptor binding assays or in *in vivo* studies including androgen-sensitive endpoints. For example, 2,4-D failed to show evidence of androgenic activity in the EOGRT study (Marty *et al.*, 2010), wherein male reproductive organ weights were not increased in parental or F1 males, age at puberty onset was not decreased, anogenital distance was not altered, and sperm parameters and reproductive organ histopathology were unaffected. Because only a single dose group was used by Kim *et al.* it is not possible to evaluate a dose response. Because of these flaws and the lack of consistency of reported findings with other 2,4-D studies this study is excluded from the WoE.

This publication also described an *in vitro* experiment with 2,4-D. It appears these data were also published (in English) in Kim *et al.*, 2005.

Oakes *et al.*, 2002b

Male SD rats (16 weeks old, 5/group) were dosed with Tordon 75D® (2,4-D and picloram in formulation) at 0.125 ml/kg/day, 0.25 ml/kg/day or 0.5 ml/kg 5 days per week for 9 weeks (2,4-D doses of 37.5, 75 and 150 mg/kg/day), and then held unexposed for a 21-week recovery period. Untreated and surfactant control groups were also used, with a similar exposure regimen. At the end of the recovery period,

testes and epididymides were weighed for all groups. Additionally for the high dose group and the control groups, blood was sampled for plasma testosterone levels, and the testes were examined histopathologically. In a second study to evaluate the testicular findings, groups of high dose animals were evaluated after 1, 2, 4 and 9 weeks of treatment for plasma testosterone levels, testes and epididymides weights and testicular histopathology (findings compared to 9-week control groups).

High dose rats appeared lethargic shortly after dosing. There was slight body weight loss in the high dose group at the two-week interval; and it appeared there may have been slightly decreased weight gain during the remaining seven weeks of exposure (not statistically significant compared to controls); also see concluding comment re method deficiencies.

In the second (high dose) study, there were no treatment-related effects on plasma testosterone levels at any interval. Testicular absolute and relative weights were statistically significantly decreased compared to the controls in the high dose group at 9 weeks only. A duration-related increase in incidence and severity of testicular damage was evident histopathologically, characterized by germ cell depletion and shrunken seminiferous tubules. Generally Leydig cells appeared to be conserved. A single high dose rat at 4 weeks showed normal testis weight and histopathological findings in one testis, but enlargement of the other testis with both germinal cell depletion and Leydig cell hypoplasia. The relationship to treatment of the findings in this animal was unclear. One of 5 rats showed germ cell depletion 21 weeks post treatment, so reversibility of the high dose testicular effects was not established. However, similar findings (but unilateral) were made in one surfactant-exposed control rat (at 9 weeks), and one other surfactant exposed animal also showed a mixture of normal and depleted tubules. Certain formulation excipients may therefore have contributed to the observed findings.

There were no treatment-related changes in testicular weights in the mid and low dose groups (evaluated 21 weeks post exposure only); other parameters were not evaluated for these groups. The findings in this study cannot be clearly attributed to 2,4-D alone. The effects seen at the high dose are suggestive of direct cytotoxicity to the germinal cell epithelium rather than endocrine-related changes (as Leydig cells were generally preserved and there were no discernable effects on plasma testosterone levels, although the evaluation of the latter was limited by the small group size). The effects were characterized only at a very high dose of 2,4-D (150 mg/kg/day) in the Tordon®75 formulation, which would be a dose expected to exceed the renal clearance threshold for 2,4-D.

The study quality was generally poor. The description of experimental design in this study is difficult to follow. The group size is relatively small (5/group), which is considered insufficient to reliably evaluate testosterone. The methods do not state whether statistical comparisons were made with the untreated or the surfactant controls. (Recovery group comparisons were made to the untreated controls.)

The high dose study design appeared deficient in that there were no concurrent controls for testosterone, testes weights, and testicular histopathology for the 1, 2, and 4 week intervals. Further, the reporting of body weight data is incomplete (as control weight gain is reported for the 9-week

interval only) so that treatment-related effects on body weight gain cannot clearly be evaluated for the high dose study.

Body weight effects, if any existed, cannot be evaluated at all for the low and mid dose groups from the data presented (as weight change is shown only following the 21-week recovery period for these dose groups). The study design did not include histopathological evaluation of the testes for the low and mid dose groups, and did not evaluate the testes for these groups at all at 9 weeks (when the findings in the high dose pathogenesis study appeared to be the most severe). Therefore, it is not possible to establish a NOAEL for this study.

Importantly 2,4-D was not tested in isolation; and it is unknown whether picloram or the other formulation components contributed to the observed effect.

Studies of Female Reproductive Health, Lactation, Nursing Behavior

Stürtz et al., 2010

The authors evaluated the effects on lactation and the control thereof after oral dietary exposure to 2,4-dichlorophenoxyacetic acid (2,4-D) in lactating Wistar rats. The authors reported decreased pup weight gain in all treated groups (2.5, 5, 15, 25, 50, and 70 mg 2,4-D/kg bw) during the first 16 days *postpartum* which they contend is due to interference with hormonal control of lactation at the level of the hypothalamo-pituitary axis. (Note that the dosing units are assumed based on other publications by the same group of authors; these are not clearly specified in the paper.) There are, however, several limitations in the study's design and documentation that weaken their conclusions.

First, it is not clear how the doses to the animals that received 2,4-D in their feed were calculated/verified. Diets were formulated in a manner corresponding to that described below in Stürtz *et al.*, 2006. Apparently individually dosed feed was prepared for each animal every second day by spraying the feed with a solution of 2,4-D. Homogeneity was not reported. It was not reported if the feed was available *ad lib*, or how or when feed consumption was measured. It was stated that analytical verification of target concentrations was conducted, but no results are shown. It was not indicated how excursions from target doses were handled.

Second, it was not indicated that pups were kept away from the feed. Most rat pups begin to eat some amount of chow beginning around PND 13 or 14 (Tyl *et al.*, 2008). This could seriously confound the interpretation of the results. Third, no bedding or nesting materials are mentioned. Female rats about to litter exhibit nesting behaviors and stress to the dams could result from unavailability of bedding or nesting materials. Fourth, because the dams are the exposed individuals, the litter should have been the experimental unit for statistical evaluations. A mean of pup body weights should have been calculated for each litter and these litter means (N=8) should have been the N used for evaluation of pup body weights. Ideally male and female pup weights should be analyzed separately. The tables in the paper appear to present means of individual pup body weights (N = ~64) which will inflate erroneously the power of the study and could lead to incorrect statements of statistical significance. Further,

although litters were culled to result in approximately equal sex ratios, any skewing because of a shortage of males or females could influence the group mean weight.

Fifth, the thrust of this paper appears to be that there are important changes that occur in pups when there are no grossly observable alterations in dams. Nevertheless, the authors note that because there were no differences in general condition, body weight gain, or mortality, the naïve and vehicle control groups were combined. This belies the premise of the study and serves to artificially increase the power of the study by increasing the size of the control group.

Sixth, it is not clear how (or if) the tissues of the anterior pituitary or arcuate nucleus were pooled for analyses related to monoamines and NOS. A table of wet weights for these tissues was not presented, although the authors claim the weights of the tissues were not substantially different. Using wet tissue weights for very small tissue samples is also a potential source of error as small tissue samples are affected greatly by amounts of hydration/dehydration that can occur prior to weighing.

Lastly, the authors evaluated milk production by taking total litter weights before and after previously-starved litters were allowed to suckle for 15 minutes. This is a creative idea but there limited control data and no historical control data for comparison. In particular, the authors' major point is made in an experiment that compared milk production in control animals to those exposed to 50 mg 2,4-D/kg ip and to those exposed to 2,4-D/kg ip plus 100mU oxytocin. The results are dramatic, but they are inconclusive because oxytocin is a strong signal for release of milk from the mammae. A more complete picture would have been gained had there been a fourth group (control plus 100mU oxytocin) to evaluate the impact of oxytocin on non-exposed mammae.

Based on data generated from their procedures mentioned above, the authors reported compound-related changes in pup weight gain, milk production, and hypothalamic concentration of several neurochemicals which they use in constructing a complex argument to explain a potential effect of 2,4-D on milk production in rats. Due to the lack of rigor in obtaining and analyzing the data, their theoretical construct remains speculative at best. One interesting point to note is that there were marked effects on pup body weight, and similar to the Sturtz *et al.* study a lack of clinical signs of toxicity, in the range-finding study of 2,4-D conducted by Shakil *et al.*, 2008 but only at far higher dietary dose levels; supporting the questions regarding diet preparation and the actual doses administered in the Sturtz *et al.* study.

Stürtz *et al.*, 2008

Stürtz *et al.* (2008) exposed lactating Wistar dams to dietary 2,4-D (15, 25, or 50 mg/kg body weight/day) from lactation day (LD) 1 to 7 and maternal behaviors were assessed. Maternal nest building was not altered at any dose of 2,4-D; however, the authors reported altered dam-pup interactions, particularly with respect to nursing. Retrieval, crouching and licking of pups were absent or reduced during the 10-minute monitoring period in 2,4-D-exposed dams. Latency to retrieval and latency for maternal crouching for pup nursing was increased with 2,4-D treatment. The authors also reported that dams consumed more food during the light phase and had high self-grooming. 2,4-D-

treated dams were reported to have increased catecholamine levels and decreased indolamine levels in the arcuate nucleus. Serum prolactin levels were decreased 62-70% in dams given 2,4-D.

One major limitation of this study is that the authors did not appear to use the litter as the unit of analysis.

Rats were dosed via the same procedures described in the Stürtz *et al.* (2006) study; thus, the same issues with the dosing regimen apply. The authors state that dietary concentrations were analytically verified, but these results are not presented, nor do the authors present data on feed consumption or the frequency and formulas used for adjusting dietary concentrations during lactation. Homogeneity was not reported. It was not indicated how excursions from target doses were handled. The dose levels used in this study were high such that the mid and high dose dams would have exhibited nonlinear toxicokinetics (Saghir *et al.*, 2008a), making these data irrelevant for 2,4-D risk assessment.

For maternal care-giving behaviors, dams were tested for maternal behaviors in their home cages; thus, there is no indication that rats were counterbalanced across test sessions or cage locations. The monitoring period for maternal behaviors was only 10 minutes; thus, the significance of altered maternal behaviors over such a short period of time is questionable, particularly when there were no effects on pup mortality or pup body weight gains. For example, the authors report an absence of maternal anogenital licking on PND 5 and 7 in 2,4-D-treated dams; however, it is not possible for these dams to have discontinued anogenital licking in pups for a prolonged period as this stimulation is needed during the first 3 weeks of life for pups to initiate urination and defecation (Brouette-Lahlou *et al.*, 1999). Without maternal anogenital licking, pups would die, which was not reported. Further, pups suffering from maternal neglect are usually not found nested, rather pups are isolated, and often cold and unresponsive. These types of observations were not reported.

Data from vehicle-treated control dams were combined with data from untreated dams, which serves to artificially increase the power of the study by increasing the size of the control group. Prolactin levels are highly variable and subject to change by numerous factors, including stress (cage transport, euthanasia procedures, diurnal variation, etc.). It is unclear if euthanasia was stratified across dose groups, which also may contribute to prolactin effects. Based on inconsistencies in the dosing regimen, high exposure levels, and the lack of corresponding pup effects to support limited monitoring of maternal care, the Stürtz *et al.* study is considered limited and the findings are difficult to interpret.

Stürtz *et al.* (2006)

Stürtz *et al.* (2006) exposed lactating Wistar dams to dietary 2,4-D (15, 25, 50 or 70 mg/kg body weight) from lactation day (LD) 1 to 16 and measured systemic toxicity and toxicokinetic (TK) parameters. 2,4-D did not alter maternal body weight gains, but significantly decreased pup body weight gains at all dose levels. Milk components were analyzed and the authors reported that 2,4-D decreased total lipid content of milk, altered fatty acid content, and resulted in the loss of 3 minor milk proteins. 2,4-D levels in dam serum, dam milk and pup serum also were measured and are discussed below.

One major limitation of this study is that the authors did not use the litter as the unit of analysis.

Furthermore, the Stürtz *et al.* paper lacked detail on critical study procedures. For example, with the limited information presented, the precise dose administered to the animals is difficult to determine. To begin, diet preparations were unusual in the Stürtz *et al.* study as these investigators dissolved 2,4-D in an alcohol solution, sprayed it over the food, and dried it to prepare test diets. The authors state that dietary concentrations were analytically verified, but these results are not presented, nor do the authors present data on feed consumption or the frequency and formulas used for adjusting dietary concentrations during lactation.

The potential issues with the dietary dose administration in this study are further revealed when TK data are examined. The TK data for LD 16 Wistar rats presented in Stürtz *et al.* (2006) can be compared with TK data generated in Saghir *et al.* (2008) using LD 14 CD[®] rats. The Saghir *et al.* (2008a) study was a GLP study with verified purity of test material, homogeneity, stability and dose concentration analyses. For 2,4-D, marked rat strain differences in TK have not been reported and in fact, Fischer 344 and CD[®] rats show similar TK with 2,4-D exposures (e.g., Van Ravenzwaay *et al.*, 2003; Saghir *et al.*, 2006; Saghir *et al.*, 2008a). The primary mechanism for 2,4-D clearance is via renal organic acid transporter 1 (OAT-1; Hasegawa *et al.*, 2003), which also is expressed in Wistar rats and shows similar sex steroid regulation across rat strains (Kudo *et al.* 2002, Buist *et al.*, 2002). In a comparison of the Stürtz *et al.* (2006) and Saghir *et al.* (2008a) TK data, marked differences in lactational TK were observed. At 15 mg/kg/day 2,4-D in the diet, Stürtz *et al.* (2006) reported dam serum levels of 26 µg 2,4-D/ml, whereas at 15 mg/kg/day, Saghir *et al.* (2008a) reported dam plasma levels of 3.5 µg 2,4-D/g; thus, Stürtz *et al.* reported 7.4-fold higher dam serum concentrations of 2,4-D with putatively similar dietary exposures (typically, the differential between plasma and serum values for 2,4-D would be minor).

Furthermore, 2,4-D TK exhibited markedly different TK behaviors between these two studies. At 70 mg/kg/day dietary 2,4-D, dam serum levels were 74 µg 2,4-D/ml in the Stürtz *et al.* (2006) study, which represented only 60% of the predicted serum level based on linear TK using serum 2,4-D concentrations at 15 mg/kg/day (26 µg 2,4-D/ml); thus, sublinear increases in serum levels were observed at 70 mg/kg/day. In contrast, Saghir *et al.* (2008a) reported dam plasma levels of 35.3 µg 2,4-D/g at 58 mg/kg/day, which represented 270% of the predicted plasma level based on linear TK using plasma 2,4-D concentrations at 15 mg/kg/day (3.5 µg 2,4-D/g); thus, supralinear increases in plasma levels were observed at 58 mg/kg/day due to saturation of renal clearance in the dams at this dose level. The non-linear TK performance observed in the Saghir *et al.* (2008a) study was consistent with toxicokinetic phenomenon reported previously for 2,4-D (e.g., Gorzinski *et al.*, 1987; Van Ravenzwaay *et al.*, 2003), thus calling into doubt the opposite findings in the Stürtz *et al.* (2006) study.

These TK differences were further reflected in milk levels of 2,4-D, where Stürtz *et al.* (2006) reported milk 2,4-D levels of 29 µg/ml at 15 mg/kg/day, whereas Saghir *et al.* (2008a) reported 2.35 µg 2,4-D/ml at 16 mg/kg/day (milk:serum partitioning of approximately 1 in Stürtz *et al.* compared with milk:plasma partitioning of 0.7 in Saghir *et al.*). At 204 mg/kg/day, the Saghir *et al.* study reported comparable milk 2,4-D concentrations (128 µg/ml) to dams given 74 mg/kg/day 2,4-D in the Stürtz *et al.* study (115

µg/ml). Pup serum levels of 2,4-D also differed (6 µg 2,4-D/ml serum in the Stürtz *et al.* study at 15 mg/kg/day dam dietary dose compared with 2.8 µg 2,4-D/g plasma at 15 mg/kg/day dam dietary dose in Saghir *et al.*). At 70 mg/kg/day, pup serum levels of 2,4-D were less than predicted based on linear TK at 15 mg/kg/day dose in the Stürtz *et al.* study. In the Saghir *et al.* study, pup plasma levels of 2,4-D were markedly increased at doses >57.5 mg/kg/day, again reflecting nonlinear TK and saturation of renal clearance.

Clearly, the TK data indicate the dosing regimens in these two studies differed, which also is reflected in the pup body weight effects. In the Stürtz *et al.* study, pup body weight gains were decreased at all doses of 2,4-D beginning at LD 6 (mean body weight differences of 4-5 g in 2,4-D treated pups by PND 16), whereas in the Saghir *et al.* study, pup body weights were not significantly affected at 100 and 400 ppm (dam doses of 10-15 mg/kg/day and 37-58 mg/kg/day over LD 4-14; dietary concentrations were not adjusted during lactation). Based on differences in dose administration and systemic toxicity between these studies, the Stürtz *et al.* (2006) study is considered limited and the findings are difficult to interpret. Furthermore, the data in the Stürtz *et al.* study do not represent direct endocrine-mediated endpoints, particularly with respect to the activities of concern (estrogen, androgen, thyroid).

Stürtz *et al.*, 2000

Lactating Wistar dams (unclear how many dams/litters were used) were exposed to 50, 70 or 100 mg/kg 2,4-D in DMSO via intraperitoneal (ip) injection. Dams were injected every two days such that pups were exposed to 2,4-D through maternal milk during lactation. Dams were exposed until the offspring reached PND 9 or 15. Pup body weights were significantly decreased at all doses of 2,4-D. Twenty-four hours after the last dose, pups were euthanized; weights of the pups' stomach contents (milk), brain and kidney were decreased with 2,4-D treatment. 2,4-D residue levels were measured in these tissues and blood.

In this study, dams were exposed to 2,4-D by ip injection, which is not a relevant route of exposure. Furthermore, dose levels of 2,4-D were high and likely in the nonlinear TK range. Based on the lack of a relevant route of exposure and the large margin of exposure for 2,4-D (i.e., making findings at nonlinear TK doses irrelevant), these data cannot be used for human risk assessment. On PND 16, 2,4-D-treated pups weighed 40, 65 and 70% of control pup body weights at 100, 70 and 50 mg/kg/day, respectively, indicating that the maximum tolerated dose (MTD) was exceeded at all three dose levels tested. Given the substantial decreases in pup body weights, it is not surprising that some tissue weights (i.e., kidney, stomach contents, brain) differed from the control values. During analytical experiments to determine 2,4-D concentrations in tissues, 2,4-D recovery from spiked samples was only 50-70%. It does not appear as if the litter was the unit of statistical analysis in this study.

Duffard et al., 1995

In an abstract, Duffard *et al.* (1995) described 2,4-D effects on prolactin, progesterone and estrous cycle alterations in F1 generation females only at a dose level (70 mg/kg/day) to dams well above the TSRC and thus irrelevant for risk assessment. Changes were reported in hormone levels for both the control and exposed rats; actual data were not presented but the reporting is considered unreliable. Interestingly the abstract was titled developmental neurotoxicity; however, no data relevant to assessment of neurotoxicity were reported. The Duffard *et al.* study is not considered reliable because of the limited data presented (abstract only) and questionable presentation of findings.

Other Studies

Kobal et al., 2000

Wistar rats (8/sex/dose group) were gavaged with an aqueous 2,4-D formulation (purity or excipients unknown) at doses of 11 or 110 mg/kg/day for 10 days. The high dose exceeds the TSRC in rats. Blood samples were taken 3 days before the start of dosing and on days 6 and 13. Serum T3 and T4 were determined by radioimmunoassay. There were clear decreases in serum T4 in the high-dose group male and females and decreased T3 in high-dose males on the 6th day of the experiment. Although the investigators report that statistically significant differences were observed in the low-dose group (increases in females and decreases in males), examination of the figures suggest that these changes were not biologically significant and may represent the inherent variability of the assay parameters. Data were portrayed graphically only. No other thyroid-related parameters (such as thyroid stimulating hormone [TSH], thyroid weights or thyroid histopathology) were assessed in this study, and the biological significance of the findings cannot be assessed. This study is limited by the unknown nature of the test material and the relatively few parameters assessed.

Pochettino et al. (2010)

Pochettino *et al.* (2010) studied the effects of prenatal and postnatal exposure to 2,4-D on oxidative stress in ventral prostate, ovary, and mammary tissue in rats by measuring both the levels of reactive oxygen species (ROS) and the activities of antioxidant enzymes (Tables X and X2). Pregnant rats (6 dosed and 6 control) were exposed to dietary doses of 70 mg/kg/day of 2,4-D daily from 16 days of gestation up to postnatal day (PND) 23. Concentrations were adjusted based on the dams most recent body weight to achieve the approximate mg/kg/day dose, however doses to pups, whose direct feed intake is approximately 2x-3x the dams between PND 17 and 23 would have been exposed to much higher doses. Neither 2,4-D purity nor the specific source of 2,4-D was identified in the study. Dose analyses were either not conducted or the results not reported. 70 mg/kg/day far exceeds the TSRC and has shown significant toxicity in prior rat studies.

F1 young adults (six/sex/dose, reported to be taken from different litters) were sacrificed at PND 45, 60, or 90 at which point ventral prostates, mammary tissue, and ovaries were removed and assayed for ROS levels or antioxidant enzyme activities. To evaluate ROS stress, hydroxyl radical levels, total carbonyl

and thiol levels, and lipid peroxidation (LPO) activity levels were measured (**Table S9**). Antioxidant response was assessed by measuring activities of antioxidant enzymes glutathione s-transferase (GST), catalase (CAT), selenio-glutathione peroxidase (Se-GPx), glutathione reductase (GR), and superoxide dismutases (SODs) (**Table S10**).

2,4-D exposure significantly reduced weight gain in both males and females.

Most of the changes in ROS and anti-oxidant enzyme activity were in the 1-2.5 fold range compared to control. Similar fold differences were seen between control groups at different ages for many of the parameters measured, although the variability in the controls was not statistically analyzed or discussed. Although limited by the single high dose administered and not considered a high quality study, this study provides some evidence of oxidative stress which may explain some of the high dose findings above the TSRC, not relevant to hazard or risk assessment, that otherwise might suggest anti-androgenic activity or adverse effects on steroidogenesis.

Subchronic toxicity in Mammals other than Rodents

Obidike et al. (2012)

Obidike *et al.* studied the effects of a 2,4-D formulation in male West African Dwarf goats. Goats were treated by gavage with 75 mg/kg, 100 mg/kg, or 125 mg/kg of 720 g/L 2,4-D (unspecified purity) every 72 hours for a period of 112 days. Control goats were administered no 2,4-D. On the 112th day, all goats were sacrificed and testes and epididymis were removed to investigate the testicular morphology and gonadal (testicular) and extra-gonadal (epididymal) sperm reserves.

There were no significant differences in combined (left and right) testicular sperm reserve at 75 mg/kg. A 1.96-fold and a 2.36-fold decrease were observed at 100 and 125 mg/kg, respectively. Total epididymal sperm reserve in caput, corpus and cauda segments combined was decreased significantly at 75, 100 and 125 mg/kg (1.68-fold, 2.38-fold, and 2.28-fold, respectively). Authors noted that goats dosed with 2,4-D showed hyperemia and edema of the stroma and decreased Sertoli cells. There was no consistent change in the spermatogenic cell population of seminiferous tubules. This study has some significant limitations. First, it is not possible to tell what the administered dose of 2,4-D was, i.e., was the low dose 75 mg/kg 2,4-D or 75 mg/kg of the 720g/L formulation? Second, the nature of the formulation is not specified and excipients are unknown. Only 5 goats per group were evaluated, which is a limited number for assessment of sperm parameters which are inherently variable. The age variation of the goats was significant (greater than 5 months); it is unknown whether younger goats may have differences in testicular structure and function from older goats but this seems likely based on extrapolation from other species. Goats came from different sources; although a single variety of goats was tested there is no assurance the control group was adequately representative. Further, goats were treated during the study for endo- and ecto-parasites; the potential male reproductive toxicity associated with these treatments was not discussed. Little is known about saturation kinetics in goats or other ruminants; based on analogous testicular findings in rats at doses greater than 100 mg/kg/day the

doses in this study are likely to have markedly exceeded the threshold for saturation kinetics. Information in this study is considered too limited to use in the weight-of-the evidence analyses.

Rawlings et al., 1998

In this study, mature female ewes were dosed via gelatin capsules administered directly to the rumen for 43 days (10 mg/kg bw, 3x/week). Serum concentrations of LH, FSH, progesterone, estradiol, thyroxine, cortisol, and insulin were measured, and histopathological examinations were completed of the thyroid, adrenals, pancreas, pituitary, ovaries, oviducts, uterus, and other organs for any treatment-related lesions. 2,4-D exposure resulted in a statistically significant decrease in serum concentrations of thyroxine at 36 days; however, there was no correlating exposure-related effect on thyroid histopathology, and the finding is not considered adverse. Likewise, there was no effect on other hormone concentrations or on the endocrine organs or other organs evaluated. Overall, the evaluation of effects on the thyroid, and on other hormonal concentrations, is too limited to place credence in the reported outcomes. This lack of confidence is based on the limited group size (n=6) and the testing of a single dose level, as well as the unconventional choice of a ruminant species for evaluation, with compound administration directly into the microbial rich rumen. Information in this study is considered too limited to use in the weight-of-the evidence analyses.

References to Appendices (includes references only cited in Appendices; references cited in both Appendices and primary study are referenced in primary study)

Ashby J, Pate I, Tinwell H. (2003) Reported seasonal dependence of herbicide developmental toxicity in mice. [Letter]. *Environ Health Perspect.* 111:A450-A451.

Bélanger B, Bélanger A, Labrie F, Dupont A, Cusan L, Monfette G. (1989) Comparison of residual C-19 steroids in plasma and prostatic tissue of human, rat and guinea pig after castration: Unique importance of extratesticular androgens in men. *J Steroid Biochem.* 32(5):695–698.

Behnam-Rassoli M, Herbert LC, Howard V, Pharoah PO, Stanisstreet M. (1991) Effect of propylthiouracil treatment during prenatal and early postnatal development on the neocortex of rat pups. *Neuroendocrin.* 53(4):321-7.

Bon E, Barbe U, Nunez Rodriguez J, Cuisset B, Pelissero C, Sumpter JP, LeMenn F. (1997) Plasma vitellogenin levels during the annual reproductive cycle of the female Rainbow Trout (*Oncorhynchus mykiss*): establishment and validation of an ELISA. *Comp Biochem Physiol.* 117B, 1:75–84.

Brouette-Lahlou I, Godinot F, Vernet-Maury E. (1999) The mother rat's vomeronasal organ is involved in detection of dodecyl propionate, the pup's preputial gland pheromone. *Physiol Behav.* 66:427-236.

Buist SC, Cherrington NJ, Choudhuri S, Hartley DP, Klaassen CD. (2002) Gender-specific and developmental influences on the expression of rat organic anion transporters. *J Pharm Exp Ther.* 301(1):145-51.

Cavieres MF, Jaeger J, Porter W. (2003) Developmental effects of herbicides in mice. [Letter] *Environ Health Perspect.* 111(14):A748.

Cavieres MF. (2001) Reproductive and developmental toxicity of a commercial herbicide formulation in mice. [PhD Thesis]. Madison, WI: University of Wisconsin-Madison.

Dantzer WH, Wright SH. (2003) The molecular and cellular physiology of basolateral organic anion transport in mammalian renal tubules. *Biochim et Biophys Acta - Biomembranes* 1618, 2, 185–193.

Gaido KW, Leonard LS, Lovell S, Gould JC, Babai D, Portier CJ and McDonnell DP. (1997) Evaluation of chemicals with endocrine modulating activity in a yeast-based steroid hormone receptor gene transcription assay. *Toxicol Appl Pharmacol.* 143:205-212.

Hanley Jr TR, Watanabe, PG. (1985) Measurement of solid feed consumption patterns in neonatal rats by ¹⁴¹Ce-radiolabeled microspheres. *Toxicol Appl Pharmacol.* 77:496-500.

Kang IH, Kim HS, Shin JH, Kim TS, Moon HJ, Kim IY, Choi KS, Kil KS, Park YI, Dong MS, Han SY. (2004) Comparison of anti-androgenic activity of flutamide, vinclozolin, procymidone, linuron, and p, p'-DDE in rodent 10-day Hershberger assay. *Toxicol.* 199:145-159.

Kudo N, Katakura M, Sato Y, Kawashima Y. (2002) Sex hormone-regulated renal transport of perfluorooctanoic acid. *Chem Biol Interact.* 139:301-316.

Lamb JC, Neal BH, Ginevan ME, Bus JS, Mahlborg WM. (2003) Herbicide effects on embryo implantation and litter size. [Letter re Cavieres et al.]. *Environ Health Perspect.* 111:A450.

Larsen PR, Ingbar SH. (1992) The thyroid gland. Williams Textbook on Endocrinology. (Wilson, J.D. and Foster, D.W., Eds.) W.B. Saunders Company, Philadelphia, Pennsylvania, pp. 357-487.

Lavado-Autric R, Auso E, Garcia-Velasco JV, del Carmen Arufe M, Escobar del Rey F, Berbel P, Morreale de Escobar G. (2003) Early maternal hypothyroxinemia alters histogenesis and cerebral cortex cytoarchitecture of the progeny. *J. Clin. Invest.* 111:1073-1082.

Li Z, Zhang H, Gibson M, Li J. (2012) An evaluation on combination effects of phenolic endocrine disruptors by estrogen receptor binding assay. *Toxicol In Vitro.* 26(6):769-74.

Ma Y, Han J, Guo Y, Lam PK, Wu RS, Giesy JP, Zhang X, Zhou B. (2012) Disruption of endocrine function in in vitro H295R cell-based and in in vivo assay in zebrafish by 2,4-dichlorophenol. *Aquat Toxicol.* 106-107: 173-81.

Mak P, McDonnell DP, Weigel NL, Schrader WT, O'Malley BW. (1989) Expression of functional chicken oviduct progesterone receptors in yeast (*Saccharomyces cerevisiae*). *J Biol Chem.* 265:21613–21618.

O'Connor JC, Cook JC, Marty MS, Davis LG, Kaplan AM, Carney EW. (2002) Evaluation of Tier I screening approaches for detecting endocrine-active compounds (EACs). *Crit Rev Toxicol.* 32:521-549.

Palmer SJ, Krueger HO. (1997) 2,4-D (2,4-dichlorophenoxy acetic acid): A 96 hour static acute toxicity test with the leopard frog tadpoles (*Rana pipiens*). Unpublished study by Wildlife International, Ltd., Easton MD. Project No. 467A-102.

Pham TA, Elliston JF, Nawaz Z, McDonnell DP, Tsai MJ, O'Malley BW. (1991) Antiestrogen can establish nonproductive receptor complexes and alter chromatin structure at target enhancers. *Proc Natl Acad Sci.* 88:3125-3129.

Pham TA, Hwung YP, Santiso-Mere D, McDonnell DP, O'Malley BW. (1992) Ligand-dependent and -independent function of the transactivation regions of the human estrogen receptor in yeast. *Mol Endocrinol.* 6:1043-1050.

Rehm S. (2000) Spontaneous testicular lesions in purpose-bred beagle dogs. *Toxicol Pathol.* 28(6):782-7.

Routledge EJ, Sumpter JP. (1996) Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. *Environ Toxicol Chem.* 15:241–248.

Sandberg JA, Duhart HM, Lipe G, Binienda Z, Slikker W Jr, Kim CS. (1996) Distribution of 2,4-dichlorophenoxyacetic acid (2,4-D) in maternal and fetal rabbits. *J Toxicol Environ Health.* 49(5):497-509.

Schlenk D. (2008) Letter to the Editor: Response to: Comments on "Evaluation of Estrogenic Activities of Aquatic Herbicides and Surfactants Using a Rainbow Trout Vitellogenin Assay." *Toxicol Sci.* 104(1):231–233.

Schoonover CM, Seibel MM, Jolson DM, Stack MJ, Rahman RJ, Jones SA, Mariash CN, Anderson GW. (2004) Thyroid hormone regulates oligodendrocyte accumulation in developing rat brain white matter tracts. *Endocrin.* 145:5013-5020.

Shibutani M, Woo GH, Fujimoto H, Saegusa Y, Takahashi M, Inoue K, Hirose M, Nishikawa A. (2009) Assessment of developmental effects of hypothyroidism in rats from in utero and lactation exposure to anti-thyroid agents. *Reprod Toxicol.* 28:297-307.

Sretarugsa P, Wallace RA. (1997) The developing *Xenopus* oocyte specifies the type of gonadotropin-stimulated steroidogenesis performed by its associated follicle cells. *Dev Growth Differ.* 39(1):87-97.

Tyl RW, Crofton K, Moretto A, Moser V, Sheets LP, Sobotka TJ. (2008) Identification and interpretation of developmental neurotoxicity effects: A report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. *Neurotoxicol Teratol.* 30:349-381.

Yamasaki K, Sawaki M, Ohta R, Okuda H, Katayama S, Yamada T, Ohta T, Kosaka T, Owens W. (2003) OECD validation of the Hershberger assay in Japan: Phase 2 dose response of methyltestosterone, vinclozolin, and p,p'-DDE. *Environ Health Perspect.* 111:1912-1919.

Zysk JR, Johnson B, Ozenberger BA, Bungham B, Gorski J. (1995) Selective uptake of estrogenic compounds in *Saccharomyces cerevisiae*: A mechanism for antiestrogen resistance in yeast expressing the mammalian estrogen receptor. In: Blair RM, Fang H, Branham WS, Haas BS, Dial SL, Moland CL, Tong W, Shi L, Perkins R, Sheehan DM. (2000) The estrogen receptor relative binding affinities of 188 natural and xenochemicals: Structural diversity of ligands. *Toxicol Sci.* 54:138-153.

Supplementary Tables

Table S1. Percent Specific Binding of [³H]-17β-Estradiol to the ER in Presence of 2,4-D (%)

2,4-D Conc. (M)	Run 1	Run 2	Run 3	Mean ± SEM ^a
10 ⁻¹¹	99.0	98.8	103.5	100.5 ± 1.59
10 ⁻¹⁰	95.4	99.1	103.6	99.4 ± 2.4
10 ⁻⁹	97.7	98.1	105.0	100.3 ± 2.4
10 ⁻⁸	99.0	102.0	103.5	101.5 ± 1.3
10 ⁻⁷	99.6	102.9	105.9	102.8 ± 1.8
10 ⁻⁶	99.5	102.7	104.0	102.0 ± 1.3
10 ⁻⁵	99.8	100.1	106.1	102.0 ± 2.0
10 ⁻⁴	101.1	98.4	108.7	102.7 ± 3.1

a. SEM Standard error of the mean.

Table S2. Percent Specific Binding of [³H]-R1881 to the AR in Presence of 2,4-D (%)^a

2,4-D Conc. (M)	Run 1	Run 2	Run 3	Mean ± SEM
10 ⁻¹¹	98.79	98.24	99.89	99.0 ± 0.5
10 ⁻¹⁰	102.44	99.72	99.55	100.6 ± 0.9
10 ⁻⁹	103.53	97.07	99.95	100.2 ± 1.9
10 ⁻⁸	103.22	99.22	104.25	102.2 ± 1.5
10 ⁻⁷	101.49	99.04	103.98	101.5 ± 1.4
10 ⁻⁶	103.17	98.32	104.51	102.0 ± 1.9
10 ⁻⁵	102.33	97.65	105.58	101.9 ± 2.3
10 ⁻⁴	105.64	99.8	106.32	103.9 ± 2.1

SEM Standard error of the mean.

Table S3. Mean (\pm SD) Hormone Concentrations Following Treatment with 2,4-D for 48 Hours in the Steroidogenesis Assay

Nominal Concentration (μ M)	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Mean	\pm SD	Statistical Significance
	Testosterone (pg/ml)			Fold Difference					
DMSO	842.0	735.0	778.0	1.0	1.0	1.0	1.0	0.0	---
0.0001	776.0	691.7	746.3	0.9	0.9	1.0	0.9	0.0	None
0.001	805.7	696.0	752.0	1.0	0.9	1.0	1.0	0.0	None
0.01	737.7	696.3	726.7	0.9	0.9	0.9	0.9	0.0	None
0.1	752.0	647.0	745.3	0.9	0.9	1.0	0.9	0.0	None
1	796.7	665.0	734.0	0.9	0.9	0.9	0.9	0.0	None
10	787.7	684.3	747.0	0.9	0.9	1.0	0.9	0.0	None
100	714.3	610.3	728.0	0.8	0.8	0.9	0.9	0.1	None
	Estradiol (pg/ml)			Fold Difference					
DMSO	29.8	35.7	37.0	1.0	1.0	1.0	1.0	0.0	---
0.0001	24.7	37.2	36.0	0.8	1.0	1.0	0.9	0.1	None
0.001	24.7	35.7	34.2	0.8	1.0	0.9	0.9	0.1	None
0.01	25.6	34.8	37.8	0.9	1.0	1.0	1.0	0.1	None
0.1	26.5	36.4	34.9	0.9	1.0	0.9	1.0	0.1	None
1	25.3	33.7	35.1	0.8	0.9	0.9	0.9	0.1	None
10	27.0	33.8	37.9	0.9	0.9	1.0	1.0	0.1	None
100	34.6	44.3	45.1	1.2	1.2	1.2	1.2	0.0	Yes, all 3 trials

Table S4. 2,4-D AMA Study Design

Nominal concentration, mg a.e./L	Mean measured concentration, mg a.e./L	Number of replicates	Total number exposed (Stage 51 at initiation)	Number examined at Day 7 ^a	Number examined at Day 21 ^a	Number examined for thyroid histopath at Day 21
Negative control	<LOQ	4	80	20	60	20
0.4	0.273	4	80	20	60	20
4.0	3.24	4	80	20	59	20
40	38.0	4	80	20	60	20
100	113	4	80	20	60	20

^a Examinations included NF developmental stage, hind-limb length, snout-to-vent length, and body weight.

Table S5. Study Design 2,4-D Fish Short Term Reproductive Toxicity Assay

Nominal concentration mg a.e./L	Mean measured concentration mg a.e./L	Number of replicates	Number males per replicate	Number females per replicate	Total number adult fish exposed
Negative control	<LOQ	4	2	4	24
0.4	0.245	4	2	4	24
4.0	3.14	4	2	4	24
40	34.0	4	2	4	24
100	96.5	4	2	4	24

Table S6. P1 Dose Groups for F1-Extended One-Generation Dietary Toxicity Study

Dietary Concentration (ppm)		Targeted Approximate Doses (mg/kg/day)		P No. Rats/sex/dose	Satellite Group Targeted No. Pregnant Females/Dose
Male	Female	Male	Female		
Control	Control	0	0	27	12
100	100	5	5	27	12
300	300	15	15	27	12
800	600	40	30	27	12

Table S7. 2,4-D Dietary Concentration Adjustments During Lactation and Post-weaning

Exposure Period	TMI Increase ^a	Unadjusted Concentration (ppm)	Adjustment Factor	Adjusted Concentration (ppm)
LD 7-14	3.1X	100, 300, 600	2	50, 150, 300
LD 14-21	3.8X	100, 300, 600	3	33, 100, 200
PND 21-28 ^b	2.4X	100, 300, 600	2	50, 150, 300
PND 28-35 ^c	1.9X	F: 100, 300, 600 M: 100, 300, 800	2	F: 50, 150, 300 M: 50, 150, 400

a Relative to non-pregnant adult females (from Saghir et al., 2008a).

b TMI = test material intake based on feed consumption data derived from male CRL:CD(SD) rats at PND 23-28 from Marty *et al.*, 2003.

c After PND 35, pups returned to unadjusted dietary concentrations.

Table S8. Body weight and Testicular Histopathology in the Subchronic Toxicity Study in Dogs with 2,4-D. Schulze, 1990 (Study No. 2184-115)

Dose (mg/kg/day)	Animal Number	Individual Body Weight (kg)		Histopathology Testes	Histopathology Prostate
		Week 0	Week 13		
Control	27494	6.5	10.2	-	NE
	27495	6.9	9.8	-	NE
	27496	5.6	8.3	-	NE
	27497	8.2	12.9	-	NE
	27498	6.5	10.2	-	NE
0.3	27504	6.3	8.1	-	NE
	27505	8.0	10.5	-	NE
	27506	5.8	7.8	-	NE
	27507	7.6	10.1	-	NE
	27508	7.3	11.0	-	NE
1.0	27514	6.4	9.0	-	NE
	27515	5.9	9.0	-	NE
	27516	7.7	- ¹	-	NE
	27517	8.2	10.8	-	NE
	27518	7.5	11.0	-	NE
3.0	27524	7.6	9.2	-	NE
	27525	5.7	7.7	Unremarkable "small" at necropsy	NE
	27526	6.4	7.8	-	NE
	27527	7.4	9.6	-	NE
	27528	6.1	8.6	-	NE
10.0	27534	6.7	*	Hypospermatogenesis-moderate; Giant cells-moderate	NE
	27535	6.0	8.1	Hypospermatogenesis-moderate; Giant cells-moderate	NE
	27536	6.6	8.5	Hypospermatogenesis-slight; Giant cells-minimal	NE
	27537	8.1	7.9	-	NE
	27538	5.8	*	Unremarkable "small" at necropsy	NE

*Dead animal

NE Not evaluated

¹ Number illegible in report copy

**Table S8 (Cont.). Body Weights and Testicular and Prostate Histopathology in the Second 13-Week Dietary Toxicity Study of 2,4-D in Dogs. (Dalgard, 1993a)
(Lab Project ID: HWA 2184-125)**

Dose (mg/kg/day)	Animal Number	Individual Body Weight (kg)		Histopathology testes	Histopathology prostate
		Week 1	Week 14		
Control	G29625	7.7	10.7	Hypospermatogenesis/juvenile testis-slight; Giant cells-minimal	
	G29626	7.9	10.9	-	Inactive/ juvenile prostate
	G29627	6.8	9.4	Giant cells-minimal	
	G29628	9.4	12.2	Giant cells-minimal	
0.5	G29633	9.0	11.4	-	Inactive/ juvenile prostate
	G29634	7.1	9.2	Giant cells-minimal	
	G29635	7.4	11.3	-	
	G29636	9.8	11.4	Giant cells-minimal	
1.0	G29641	9.3	12.8	Giant cells-minimal	
	G29642	7.8	11.4	-	
	G29643	6.8	10.1	-	Inactive/ juvenile prostate; inflammation
	G29644	8.3	10.1	Giant cells-minimal	
3.75	G29649	9.8	11.9	Giant cells-minimal	
	G29650	10.2	11.8	-	
	G29651	7.9	9.9	Hypospermatogenesis/juvenile testis-slight; Giant cells-slight	Inactive juvenile prostate
	G29652	8.7	8.9	-	
10/7.5	G29657	7.0	9.3	Giant cells-minimal	Inactive/Juvenile prostate
	G29658	8.2	9.9	Hypospermatogenesis/juvenile testis- minimal; Giant cells-minimal	Inactive/juvenile prostate
	G29659	8.2	9.4	Hypospermatogenesis/juvenile testis-slight; Giant cells-minimal	Inactive /juvenile prostate
	G29660	6.7	8.1	-	Inactive juvenile prostate

*Dead animal

Table S9. Changes in ROS levels upon 2,4-D exposure (70 mg/kg/day) (Pochettino *et al.* 2010)

	Hydroxyl radical			Carbonyl			Thiol			LPO		
	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90
Ventral prostate	↑	↑	↑	↑	↑	↑	-	-	-	↑	↑	↑
Ovary	-	-	↑	↑	-	-	-	-	-	↑	↑	↑
Mammary	-	-	-	-	-	-	-	↓	↓	↑	↑	↑

Table S10. Changes in anti-oxidant enzyme activity upon 2,4-D exposure (70 mg/kg/day) (Pochettino *et al.* 2010)

	GST activity			CAT activity			Se-GPx activity			GR activity		
	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90
Ventral prostate	↑	↑	↑	↑	↑	↑	↑	↑	-	-	-	↓
Ovary	↓	↓	-	-	↓	-	↑	↑	↑	-	-	↓
Mammary	↓	↓	↓	-	↓	↓	↓	↓	↓	-	-	-

a No changes were observed in SOD activity at any interval in the ventral prostate, ovary or breast.