Received: 15 November 2023 | Revised: 5 January 2024 | Accepted: 21 January 2024

DOI: 10.1002/bdr2.2311

Birth Defects Research | Society for Birth Defects Research & Prevention | WILEY

SHORT COMMUNICATION

HESI workshop summary: Interpretation of developmental and reproductive toxicity endpoints and the impact on data interpretation of adverse events

M. L. Green[1] | **A. Kluever**[2] | **Connie Chen**[3] | **S. Dobreniecki**[4] | **Wendy Halpern**[5] | **Bethany Hannas**[6] | **Alan Hoberman**[7] | **M. E. McNerney**[8] | **S. Mitchell-Ryan**[3] | **T. J. Shafer**[9] | **Steven Van Cruchten**[10] | **Tacey White**[11]

[1]Hurley Consulting Associates, Ltd., Summit, New Jersey, USA

[2]Office of Environmental Management, Department of Energy, Germantown, Maryland, USA

[3]Health and Environmental Sciences Institute, Washington, DC, USA

[4]U.S. Environmental Protection Agency, Office of Pesticide Programs and Office of Research and Development, Washington, DC, USA

[5]Genentech, A Member of the Roche Group, South San Francisco, California, USA

[6]Eli Lilly and Company, Indianapolis, Indiana, USA

[7]Charles River Laboratories, Inc., Horsham, Pennsylvania, USA

[8]U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Office of New Drugs, Office of Rare Diseases, Pediatrics, Urologic and Reproductive Medicine, Division of Pharmacology-Toxicology for Rare Diseases, Pediatrics, Urologic and Reproductive Medicine/Specialty Medicine, Silver Spring, Maryland, USA, Silver Spring, Maryland, USA

[9]Biomolecular and Computational Toxicology Division, Center for Computational Toxicology and Exposure, Research Triangle Park, North Carolina, USA

[10]Comparative Perinatal Development, University of Antwerp, Antwerp, Belgium

[11]Tacey White Toxicology Consultant, LLC, Glenside, Pennsylvania, USA

Correspondence
M. L. Green, Hurley Consulting Associates Ltd., Summit, NJ, USA.
Email: mgreen@hurleyconsulting.com

Funding information
Health and Environmental Sciences Institute

Abstract

The Health and Environmental Sciences Institute Developmental and Reproductive Toxicology (HESI-DART) group held a hybrid in-person and virtual workshop in Washington, DC, in 2022. The workshop was entitled, "Interpretation of DART in Regulatory Contexts and Frameworks." There were 154 participants (37 in person and 117 virtual) across 9 countries. The purpose of the workshop was to capture key consensus approaches used to assess DART risks associated with chemical product exposure when a nonclinical finding is identified. The decision-making process for determining whether a DART endpoint

The findings and conclusions in this communication do not represent and should not be construed to represent any agency determination or policy. This report has been reviewed and cleared by the US EPA, Office of Pesticide Programs and Office of Research and Development. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

is considered adverse is critical because the outcome may have downstream implications (e.g., increased animal usage, modifications to reproductive classification and pregnancy labeling, impact on enrollment in clinical trials and value chains). The workshop included a series of webinar modules to train and engage in discussions with federal and international regulators, clinicians, academic investigators, nongovernmental organizations, contract research organization scientists, and private sector scientists on the best practices and principles of interpreting DART and new approach methodologies in the context of regulatory requirements and processes. Despite the differences in regulatory frameworks between the chemical and pharmaceutical sectors, the same foundational principles for data interpretation should be applied. The discussions led to the categorization of principles, which offer guidance for the systematic interpretation of data. Step 1 entails identifying any hazard by closely analyzing the data at the study endpoint level, while Step 2 involves assessing risk using weight of evidence. These guiding principles were derived from the collective outcomes of the workshop deliberations.

KEYWORDS

adverse event, developmental and reproductive toxicology, developmental neurotoxicity, hazard identification, new approach methodologies, risk assessment

1 | INTRODUCTION

Developmental and reproductive toxicology (DART) is a highly specialized subfield of toxicology. Experts with fluency in interpretation of DART data rely on years of graduate, postgraduate, and applied professional experience to elucidate and discern adverse effects of substance exposure in developing organisms, including embryos, fetuses, infants, and children. Hence, it is essential to perform a careful analysis (e.g., rigorous WoE methodology, data quality evaluation, relevant weighting of endpoint data) to identify adverse DART endpoints because this may impact regulatory decision-making.

Before the workshop, participants were polled regarding their comfort level in interpreting DART data. Approximately 50% of the participants selected "Not comfortable" or "Not comfortable at all." This highlights the need for additional training to expand expertise in analyzing DART data. As an added level of complexity, the advent of new approach methodologies (NAMs) which rely on data from in silico, in vitro, and alternative species are increasingly being incorporated into DART assessments, particularly for developmental neurotoxicity. Hence, this information sharing workshop was organized to discuss and reach a consensus on guiding principles for how to systematically approach and improve interpretation of nonclinical DART safety data, including NAMs data, related to adverse outcomes, and how to use this information to assess risk for humans.

DART studies are generally required by international agencies that regulate chemical and pharmaceutical product development. DART studies are intended to evaluate effects on the full reproductive cycle, including effects on fertility, embryo-fetal development, and postnatal development into adulthood, including behavior and reproductive performance of offspring. Regulatory agencies have developed guidelines to facilitate the evaluation of DART endpoints, provide approaches for integrating endpoints across different study types, and determine the potential increased risk of adverse human DART outcomes associated with human exposure to chemicals or pharmaceuticals. Valuable resources include ECETOC, 2002; EPA, 1991; EPA, 1996; FDA, 2011; and OECD, 2018a. However, determining the adversity of effects in DART studies is not standardized; inconsistent DART data interpretation has been evident in published manuscripts, differences in product labeling, and differences in requests for additional study data by regulatory authorities. These inconsistencies can have a global impact, leading to differences in how substances are regulated in different regions of the world. This report highlights the workshop's guiding principles for consistent DART data interpretation.

2 | METHODS

The workshop featured presentations and discussions with DART experts. Attendees also engaged in breakout group

discussions, guided by DART experts, to assess endpoint adversity in case studies. The case studies involved reviewing complex datasets, including NAMs data for developmental neurotoxicity, and provided hands-on experience with immediate feedback from subject matter experts. The workshop concluded with a brainstorming session to discuss lessons learned, including areas of consensus, divergence, and recommendations.

3 | DISCUSSION SUMMARIES

3.1 | Overview

The workshop began with Dr. Kluever and Dr. Green introducing the workshop and discussing the complexities in deciding adverse DART endpoints. Dr. McNerney shared a regulatory approach to addressing adversity determination in DART studies. Dr. Hoberman outlined key considerations for interpreting common DART endpoint scenarios. Dr. Halpern discussed interpreting adversity from a pathology standpoint. Dr. Shafer and Dr. Dobreniecki presented a case study on using NAMs to provide weight of evidence (WOE) in supporting regulatory decision-making.

3.2 | Data integration—M. McNerney

DART testing is required for development and registration of substances as per guidelines (ICH S5R3, ICH M3, OECD 422, 414, 416, 443) (ICH, 2009, 2020; OECD, 1996, 2018a, 2018b). The appropriate design, conduct, and interpretation of DART studies are essential for interpretation and communication. Identification of adverse DART effects is essential to protect human health and environmental safety.

Dr. McNerney presented a regulatory perspective on how to approach the problem of determining adversity in DART studies. Various factors are considered in this decision, and decision trees and workflows (FDA, 2011) have been published to guide the process. However, advances in technologies and changes to regulatory guidelines have led to an inconsistent approach to interpretation of DART data. A modernized perspective on determining adversity and assessing data in DART studies was presented. Examples of considerations required for Step 2 were provided.

3.3 | Debatable DART adverse outcomes—A. Hoberman

Before Dr. Hoberman's presentation, workshop participants were asked to identify which DART endpoints were most difficult to interpret. Of the 35 respondents, the most challenging endpoints were identified as maternal toxicity (MT) versus developmental toxicity (DT), body weight change, interpretation of fetal abnormalities (i.e., % of fetuses vs. % litter mean, single or low incidence), rare abnormalities, fetal variations, pup mortality, neurobehavior data, gravid uterine weight, and relevance of fetal and maternal hormone levels to DT, as well as how to incorporate in vitro DART endpoints to inform on adversity.

While there is some distinction (e.g., study design; exposure-based risk assessment) between the chemical and pharmaceutical sectors, the fundamental principles of data interpretation remain unchanged. For instance, endpoints for developmental and reproductive toxicity were ranked by sensitivity (Stump et al., 2012), and mode of action may determine which measure is more sensitive. The critical importance of examining patterns of effects, and reconciliation of biological plausibility of the overall effect, cannot be overemphasized. Some points-to-consider when interpreting common scenarios for DART endpoints are listed in Table 1.

3.4 | Interpretation of adversity: A pathology perspective—W. Halpern

When polled, workshop participants had a consensus view that it is challenging to determine when to consider histopathology endpoints of the endocrine system (e.g., thyroid and male and female reproductive tissues) adverse. Dr. Halpern's presentation captured common challenges encountered when determining adversity to the test article. A table of considerations and examples of DART-relevant pathology endpoints are listed in Table 2. While reproductive tract tissues are routinely evaluated in general toxicity studies, histopathology endpoints are often optional in DART studies. They may be included if specific concerns warrant pathology data for better understanding. Also, tissues can be collected prospectively for possible future microscopic assessment, should other endpoints indicate a potential issue. Only a few specialized studies, such as the EPA pubertal assays (OCSPP Test Guideline Series 890), incorporate specific histopathology endpoints.

When determining adversity, typically a WoE approach is used; histopathology, if evaluated, can be one part of that assessment. However, not all histopathology findings of reproductive tissues are necessarily linked to the study treatment or are definitively adverse. Some findings occur sporadically as background (e.g., cystic follicles, luteal cysts, segmental atrophy of seminiferous tubules, vacuolation of epididymal epithelium, and

TABLE 1 Basic principles used to interpret complex scenarios for DART endpoints.

Endpoints	Examples
Male Fertility • Rodents have a large sperm reserve and can tolerate some effects on sperm (count and motility) without influencing fertility; however, the same is not true for humans. • For potential risk on sperm parameters in men, determine the exposure margins at all dose levels in animals compared with humans in the clinic.	If male fertility is affected at the high dose and sperm parameters are affected at the mid- and high dose, consider the relationship between other fertility parameters at the mid-dose, which may be relevant for humans who do not have the same excess capacity of sperm.
Mating and Fertility Indices and Reproductive Outcome • Mating index is generally considered reliable, although it is an indirect measure of normal sexual behavior and libido. ○ Mating is confirmed by a copulatory plug or sperm in the vaginal lavage. ○ HC: Female Mating Index Mean[a] = 98.1%, *SD* = 3.21 (*N* = 291 studies) • Fertility index is the primary measure describing the outcome of a mating trial. ○ In isolation, fertility index is generally considered a poor predictor of perturbations in the reproductive system. ○ Female Fertility Index Mean[a] = 93.4%, *SD* = 6.6 (*N* = 313 studies) • Viable Litter Size is a very stable index on reproductive toxicity studies ○ Historical Control Mean[a] = 14.1, *SD* = 0.90 (*N* = 375 studies) ○ Frequently the most sensitive measure of reproductive toxicity ○ Decreases in live litter size can arise from changes in: ○ Ovulatory rate ○ Tubal transport timing ○ Implantation rate ○ Postimplantation survival ○ Sperm parameters	Decrement of ≥1.5 pups/litter is generally considered biologically relevant. Live litter size is considered a more sensitive indicator of reproductive insult than mating and fertility indices.
Neonatal Growth • Adverse effects on neonatal growth manifest as reduced birth weights caused by growth retardation in utero and/or by reduced body weight gain following birth. • HC: PND 1a: Male Mean[a] = 7.1 g, *SD* = 0.24 (*N* = 199 studies) • HC: PND 21a: Male Mean[a] = 49.2 g, *SD* = 4.66 (*N* = 126 studies) • Evaluation of the neonatal growth curve, in conjunction with litter size, is important to control for the confounding effects of within-litter competition.	Mean pup body weight differences of ≥5% will typically show statistical significance.
Offspring Survival • Neonatal survival, in conjunction with pup body weights, is often used to gauge disturbances in postnatal health, growth, and development • Most frequently, adverse effects on pup survival occur during the period prior to litter standardization (culling) on PND 4 • Litters are standardized on PND 4 to 4/sex • Consider any incidence of cannibalism or food restriction (e.g., no milk lines in the pup's stomach). • HC: Birth to PND 4 (pre-cull)[a]: Mean = 98.3%, *SD* = 2.14 (*N* = 194 studies) • HC: PND 4 (post-cull) to PND[a] 21*: Mean = 97.9%, *SD* = 2.18 (*N* = 119 studies)	Less than 1% of control dams that deliver have a total litter loss. Therefore, a treatment group (20–30 animals) with just two total litter losses would be considered biologically relevant.

TABLE 1 (Continued)

Endpoints	Examples
Malformation and Variations • Malformations are permanent structural deviations that generally are incompatible with or severely detrimental to normal postnatal development or survival. • Variations are structural changes that do not impact viability, development, or function (e.g., delays in skeletal bone ossification) which can be reversible and are found in the normal population under investigation. • There is no generally accepted classification of malformations and variations; other terms that are often used include anomalies, deformations, and aberrations. • Distinguishing between variations and malformations is difficult since there exists a continuum of responses from the normal to the extremely deviant. • Makris et al., 2009 provides guidance for DART terminology, but this may not fit all situations such as when fetuses are recorded with multiple alterations.	Variations are reversible or minor manifestations of developmental toxicity (e.g., skeletal variations) by themselves are of minimal concern from a risk assessment perspective. An increased incidence of variations can influence the interpretation of an equivocal increase in related malformations. The extent of concern will be influenced by other factors (e.g., exposure multiple at which the findings occurred; cross-species concordance).
Pre-and Postnatal Development • Sexual maturation may be delayed in males and/or females. • Body weights in the affected groups may be correspondingly lower than controls.	A delay in sexual maturation may be due to delayed growth.
Dose–Response Relationships • Expect a dose–response—"The dose makes the poison" is the most basic principle of toxicology. • When we see a "U" shaped or flat dose–response curve consider: ○ Eliminate all extraneous reasons—it may not be due to the test material ○ Distribution of animals to groups, ○ Group sizes, replicates. ○ Understand the kinetics. ○ Longer gestation days (in utero) can influence pups' growth and development to make it appear accelerated.	If there are one or more rare fetal malformation(s) at low incidences in one or more dose groups and the malformation(s) does not occur in a dose–response manner and/or there is a single occurrence of a malformation at the high dose, compare the concurrent and HCD incidence. Then, the scientific information should be evaluated for a better biological understanding of how each type of toxicity or response occurs; the understanding of how the toxicity is caused is called the MOA. Determine whether the MOA is a non-linear or linear dose–response assessment. Is there biological plausibility?
Historical Control Data (HCD) • Genetic drift is a common occurrence in rodents and rabbits. It is suggested to use historical data obtained from the same test facility and from studies performed ≤5 years before the study in question (Keenan, Elmore, Francke-Carroll, Kemp, Kerlin, Peddada, Pletcher, Rinke, Schmidt, Taylor, & Douglas, 2009, Keenan, Elmore, Francke-Carroll, Kemp, Kerlin, Peddada, Pletcher, Rinke, Schmidt, Taylor, & Wolf, 2009). • Consistent presentation of data. • Differences in environment, nutrition, and practices will differ between testing facilities; therefore, it is important to use HCD from the performing laboratory for a particular study. • Terminology for fetal malformations and variations has changed over time. Readers are referred to publications by Makris et al. (2009), to aid interpretation.	Abnormalities such as "small eye" may also be termed "microphthalmia" in HCD.
Maternal Toxicity • Dose selection principles: ○ Limit dose of 1000 mg/kg, toxicokinetic saturation or try to achieve "adequate" maternal toxicity. ○ Maternal toxicity above thresholds of 25-fold (defined in ICH S5(R3)) is likely not relevant to human exposures.	Reduced food consumption in rabbits is often a sign of maternal toxicity and can lead to abortions, if it is severe and sustained (Fleeman et al., 2005). Vaginal discharge often observed when resorptions are occurring.

(Continues)

TABLE 1 (Continued)

Endpoints	Examples
• Look for fetal effects at highest possible dose • If dose is too high resorptions, abortions (rabbits), or spurious malformations may occur. • Difficult to discern maternal toxicity vs compound effects • Must interpret data considering maternal toxicity • Rodents do not abort but will resorb their conceptuses.	

Fetal Morphology
• Maternal toxicity
  ○ Evident as decreased weight gain and food consumption
  ○ Argument that sensitive periods for abortion is later in organogenesis
  ○ Outbred populations drift overtime—how long is HCD relevant?
  ○ Maternal nutrition and environment; pair housing
• Developmental Toxicity:
  ○ Variants, Malformation, embryo-fetal death
  ○ Impaired growth vs structural changes
  ○ Minor alterations considered of less biological relevance.
  ○ Cervical ribs, wavy ribs, thoracolumbar ribs, and bifid vertebral centrum are often representative of delayed ossification and are observed due the timing of Caesarean section which stops development of the fetus allowing for delays in development to be observed
  ○ Bent long bones and bent scapulae often resolve postnatally
  ○ Consistent nomenclature and presentation of data
	Maternal toxicity should be at a suitable level – ICH guidelines look for multiples of clinical exposure; OECD looking for adequate levels of maternal effects.
	Claims of secondary effects need to be substantiated with clear correlations and a possible mechanism.

Statistics or Defined Thresholds • Statistics are a tool to aid in the overall assessment of toxicity • Group sizes in DART studies are larger, 20/sex as opposed to 10/sex in general toxicology studies • Larger group sizes mean smaller differences are statistically significant • Statistical changes in adult body weights, body weight gains, and food consumption are often not observed in smaller group sizes. • Not all DART data are normally distributed—uterine contents, fetal anomalies—non-parametric statistics may be more appropriate • Reductions in litter size and/or fetal weight can be reflected in maternal body weight.	Mean fetal body weight of >5% below control is generally acceptable for biological significance.

New Approach Methods for Developmental Neurotoxicity (DNT)
• Identify the DART data gap (e.g., DNT, mechanistic, biological pathway) that will be addressed.
• Describe how the relationship between the DART endpoints and NAM data will be used (e.g., screening; WoE) and incorporated into the regulatory framework by using an Integrated Approaches to Testing and Assessment (IATA) problem formulation to drive collection of appropriate data.
• In vitro DNT NAMs Coverage of Common Neurodevelopmental Processes
  ○ Neurite proliferation
  ○ Migration
  ○ Differentiation
  ○ Apoptosis
  ○ Neurite growth
	WOE for Decision on In Vivo DNT Waiver: An Example using DNT NAMs for Isomers of Glufosinate
	• Guideline DNT for DL-glufosinate: adverse effects on motor activity and brain morphometrics in pups
	• **Question:** Is the guideline DNT for DL-glufosinate sufficient to inform decisions for L-glufosinate isomers or is additional in vivo data needed for the isomers?
	  ○ Literature
	  ○ Glufosinate is neurotoxic
	  ○ In vitro effects on network activity following acute exposure
	  ○ Similar in vivo endpoints for DL- and L- glufosinate (e.g., increased motor activity and decreased body weights in pups)
	  ○ Morphometric data not available for L- isomers
	  ○ Selected neurite outgrowth and network formation as relevant in vitro assays

TABLE 1 (Continued)

Endpoints	Examples
○ Synaptogenesis ○ Neural network formation and function	• **Data** ○ Lack of effect on neurite outgrowth and network formation, but positive and negative controls performed as expected ○ Effects on acute network activity replicated published reports ○ In vitro-to-in vivo extrapolation • WOE **Evaluation** ○ Similar effects (in vitro and in vivo) for DL- and L-isomers ○ Concentrations tested in vitro exceeded PODs selected for L-glufosinate risk assessment ○ Calculated risks for dietary and non-dietary exposures were not of concern ○ Additional in vivo data would not likely identify a lower POD or more sensitive endpoint for risk assessment **Decision:** Waivers recommended for in vivo DNT guideline studies for L-isomers *Source*: Dobreniecki et al., 2022

Note: Fertility index = Number of females confirmed gravid/number of females used for mating × 100.
Abbreviations: DNT, developmental neurotoxicity; GD, gestation day; HC, historical control; MOA, mode of action; *N*, number of animals; NAM, new approach methods; PND, postnatal day; POD, point of departure; *SD*, standard deviation; WoE, weight of evidence.
[a]The mean and standard deviation values were from the historical control (HC) database from Sprague–Dawley rats for embryo-fetal developmental or fertility studies generated at Charles River Laboratories, Inc. (Horsham, PA).

atrophy of the prostate gland epithelium), making it essential to discern whether a change in an endpoint is related to the test article. Typically, lab-specific historical control data, as well as consideration of dose relationships for incidence and severity of findings, can be used to provide appropriate context and interpretation for a given study. Additionally, a determination of adversity in nonclinical studies should consider whether the histologic finding was impactful to the test system (Kerlin et al., 2016). Some examples can be found in Table 2. Due to the uncertainty in translation, some risk is often assumed when there are reproductive tract histopathology findings. The WoE approach should incorporate all available data and potential contributing mechanisms for a comprehensive risk assessment. For additional details, see comprehensive reviews (Creasy et al., 2012 and Dixon et al., 2014).

3.5 | Use of NAMs to inform potential adverse DART endpoints—T. J. Shafer and S. Dobreniecki

Participants were polled and asked the question, "How familiar are you with the developmental neurotoxicology (DNT) NAMs test battery?" prior to beginning Day 2 of the workshop. Responses were as follows: six participants selected "I don't know anything about them," six participants selected "I've read about them," and zero participants selected "I've used them." Drs. Timothy J Shafer and Sarah Dobreniecki presented (Step 2, see Section 3.4) the background and a case study involving the use of data from DNT NAMs as part of the WoE for a regulatory decision.

NAMs enable the assessment of adverse DART effects of pharmaceuticals and other chemicals. In the pharmaceutical sector, the updated ICH (S5)R3 guideline now includes the use of alternative assays as part of an integrated testing strategy. However, the case study presented was specific for the chemical sector. Within DART, DNT is a subdiscipline, which poses unique challenges in determining whether observed endpoints are adverse, especially in the chemical sector. In vivo DNT studies for chemicals used for regulatory purposes (OECD DNT test guideline, EPA DNT test guideline) or studies that include DNT cohorts (Extended One-Generation Reproductive Toxicity Study) are some of the most expensive and resource-intensive studies in terms of cost, time, and animal lives. It is well recognized within the regulatory community that testing every chemical for potential DNT is impossible (Bal-Price et al., 2012; Crofton et al., 2012; Lein et al., 2007; NAFTA, 2006). Hence, DNT is an area of DART for which NAMs are critically needed to address data gaps, both for hazard identification as well as study selection and design.

DNT NAMs were designed to cover specific critical neurodevelopmental processes, such as neuroprogenitor proliferation, synaptogenesis, myelination, and neural network formation and function (Sachana et al., 2021). Using NAMs within

TABLE 2 Basic principles used to interpret complex DART pathology endpoints.

Reproductive System Assessment in Mature Rodents
• Most general toxicity studies of at least 2 weeks duration will be conducted in sexually mature rodents, so extensive pathology data will often be available. • Sexual maturity is relatively easy to predict in rodents based on age, but both age at study start and age at time of necropsy need to be considered. • Microscopic assessment of reversibility can be confounded by senescence. • Standard pathology endpoints relevant to DART typically include organ weights, as well as macroscopic and microscopic evaluation of reproductive and endocrine tissues. • Optional DART endpoints (typically for cause) in general toxicity studies could include assessment of sperm parameters (males) or monitoring of estrous cycle (females).
Examples of Reproductive System Pathology in Males:
• Decreased testosterone → degeneration of pachytene spermatocytes and round stage VII/VIII spermatids; if severe, Leydig cell atrophy, tubular vacuolation • Inhibition of androgen receptor → Leydig cell hypertrophy/hyperplasia • Decreased thyroid hormones → Sertoli cell degeneration → generalized atrophy • Stress → HPG axis pathology and decreased body, epididymal, and prostate gland weights • Additional considerations: ○ Degeneration/atrophy of seminiferous tubules is often a "final common pathway," but initiating cause can be hard to discern in longer studies. ○ Segmental atrophy may not result in decreased fertility in rodents due to large reserve of the testis. ○ In short-term studies, effects on epididymal transit and/or sperm maturation can appear to be worst in recovery cohorts, as the damage can take time to develop. ○ Disruption of hormone production or balance leading to a fertility effect may not be identified in microscopic pathology, but accessory sex organ weights can be sensitive.
Examples of Reproductive Pathology in Females:
• Hormone disruption (thyroid, LH, androgens, prolactin) → arrested cycle progression with large atretic or cystic follicles in ovaries and few corpora lutea. • Pro-estrogenic compounds → persistent estrus with large anovulatory follicles • Persistent, low prolactin → retained, nonfunctional CLs • Increased progesterone or PGR inhibition → luteal cysts • Additional consideration: Effects of disrupted cycling can be evident across reproductive tissues, so assessment of ovaries, uterus, vagina, and mammary gland for individual animals is recommended
Thyroid Pathology Relevant to DART Testing:
• Histopathology of the thyroid gland is the most sensitive endpoint for detection of thyrotoxicosis. Parallel changes can be found in the hippocampus and pituitary gland, but are often more subtle; in male juvenile/pubertal rats, thyrotoxicosis is associated with delayed testis maturation • Chemicals: Historically are based on rat pubertal assay with defined study design and endpoints; assessment integrates thyroid weight, follicular epithelial height, and amount of colloid, as well as thyroid hormones and liver evaluation. Currently, thyroid evaluations are incorporated into DART studies. • Pharmaceuticals: Thyroid weight/histology, but not follicle & colloid scores or thyroid hormones in routine studies. Investigational studies (typically for cause) may expand endpoints
• Example: Decreased thyroid function → histologic changes such as increased follicular epithelial height along with increased thyroid weight, increased serum cholesterol, and decreased growth (in young animals), but relative increase in body weight; may see decreased thyroid hormones if evaluated. • Additional considerations: ○ Critical roles for neurodevelopment/cognitive success, growth, and puberty ○ Sexual dimorphism in follicular epithelial height (males > females)

Abbreviations: CL, corpora lutea; HPG, hypothalamic–pituitary–gonadal axis; LH, luteinizing hormone; PGR, progesterone receptor; T3, triiodothyronine; T4, thyroxine.

appropriate contexts can confer advantages to hazard assessments. For example, the DNT NAMs presented could be used to support screening level or WoE assessments (Table 1).

3.6 | Steps and guiding principles

The outcome of the workshop was the identification of a consensus of guiding principles to aid in DART data

interpretation, which were categorized as either Step 1 or Step 2 in the process.

- *Step 1: Use standard principles to identify hazards based on data analysis in each individual study.*
- *Step 2: Determine the risk for an adverse effect in humans based on an integration of data across studies, including systemic exposure and application of regulatory paradigms.*

Some common data scenarios with associated guiding principles follow. These examples were areas of consensus and are listed to highlight common approaches used to identify adverse effects in the context of DART data interpretation.

3.6.1 | Step 1. Guiding principles (individual study data)

1. *Dose–response relationship*: Determine substance-related DART endpoint changes, assess for a dose–response, and differentiate high dose findings as random or biologically relevant using historical control and a WoE assessment.
2. *Biological significance*: Consider the potential impact of data on longevity or quality of life, in addition to statistical significance and dose response. Statistically significant endpoints may not always have biological significance.
3. *Relationships between endpoints*: Assess relationships between DART endpoints, including maternal body weight, food consumption, number of conceptuses, fetal growth, and developmental toxicity. Determine if changes in these endpoints are related and if there is a progression in severity of developmental toxicity as the dose increases.

3.6.2 | Step 2. Guiding principles (integrated assessments)

1. *Study preparation*: Design studies well. Adhere to regulatory requirements to capture all endpoints and avoid unnecessary animal testing. Incorporate NAMs into the testing paradigm for enhanced scientific interpretation.
2. *Data integration*: Differentiate adverse effects from adaptive responses using WoE, including mechanism, pharmacology, and range-finding data. Utilize diverse information sources, including epidemiological, clinical, and NAM data in appropriate contexts and robust study designs.
3. *Data interpretation*: Evaluate NAMs data with the same scientific rigor as in vivo DART data, including quality assessment and mechanistic insight. Maintain open communication with regulatory authorities and thorough documentation, particularly in data-limited chemical contexts, for improved accuracy in interpretations.
4. *Health and safety*: Determine potential concerns for human and environmental health and safety by considering the demographics of the anticipated population and the extrapolation of nonclinical data to human pregnancy outcomes. Safety margins (i.e., pharmaceuticals) or exposure-based decisions at the no observed adverse effect level and lowest observed adverse effect level should be evaluated (Andrews et al., 2019).
5. *Communication*: Use concise language and suitable formats. Describe adverse effects in the context of maternal toxicity, dose levels, and organ systems clearly and coherently. For NAMs results, specify the covered endpoints and their role in WoE. Ensure interpretations tell a cohesive and logical story, contextualizing decisions and considering differing interpretations to aid understanding of the decision-making process.

4 | CONCLUSIONS

Key consensus themes emerged, which led to the formulation of the guiding principles. The guiding principles aim to promote consistent interpretation of DART and NAM adverse endpoints through use of a WoE risk assessment and to facilitate effective communication among stakeholders. The participants emphasized the need to further address discrepancies in interpreting DART data, especially with fertility, juvenile toxicity, and NAM endpoints.

FUNDING INFORMATION

Funding for the Workshop was provided by the ILSI Health and Environmental Sciences Institute. Author employers are as listed. HESI, HESI-DART Working Group, Case study providers, break-out session leaders, presenters, notetakers, reviewers: Christopher Bowman, Kimberly Brannen, Natasha Catlin, Pragati Coder, Richard Curie, Jennifer Foreman, Katie Goyak, Thomas Knudsen, Kazushige Maki, Sue Makris, Pallavi McElroy, Fumito Mikashima, Daniel Mink, Stephanie Powlin, Nicole Principato, Anthony Scialli, Dinesh Stanislaus, Vicki Sutherland, Belen Tornesi.

CONFLICT OF INTEREST STATEMENT

The authors M. L. Green, A. Kluever, C. Chen, S. Dobreniecki, A. Hoberman, B. Hannas, M. E. McNerney, S. Mitchell-Ryan, T. J. Shafer, T. White, S. Van Cruchten have no conflicts of interest to disclose. The author S. Dobreniecki declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article. The author W. Halpern is employed by Genentech and owns stock in Roche.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

M. L. Green https://orcid.org/0000-0002-8397-2299
Steven Van Cruchten https://orcid.org/0000-0002-3144-1840

REFERENCES

Andrews, P. A., Blanset, D., Costa, P. L., Green, M., Green, M. L., Jacobs, A., Kadaba, R., Lebron, J. A., Mattson, B., McNerney, M. E., Minck, D., Oliveira, L. C., Theunissen, P. T., & DeGeorge, J. J. (2019). Analysis of exposure margins in DT studies for detection of human teratogens. *Regulatory Toxicology and Pharmacology*, *105*, 62–68.

Bal-Price, A. K., Coecke, S., Costa, L., Crofton, K. M., Fritsche, E., Goldberg, A., Grandjean, P., Lein, P. J., Li, A., Lucchini, R., Mundy, W. R., Padilla, S., Persico, A. M., Seiler, A. E., & Kreysa, J. (2012). Advancing the science of developmental neurotoxicity (DNT): Testing for better safety evaluation. *ALTEX*, *29*(2), 202–215.

Creasy, D., Bube, A., de Rijk, E., Kandori, H., Kuwahara, M., Masson, R., Nolte, T., Reams, R., Regan, K., Rehm, S., Rogerson, P., & Whitney, K. (2012). Proliferative and nonproliferative lesions of the rat and mouse male reproductive system. *Toxicologic Pathology*, *40*(6 suppl), 40S–121S.

Crofton, K. M., Mundy, W. R., & Shafer, T. J. (2012). Developmental neurotoxicity testing: A path forward. *Congenit Anom*, *52*(3), 140–146.

Dixon, D., Alison, R., Bach, U., Colman, K., Foley, G. L., Harleman, J. H., Haworth, R., Herbert, R., Heuser, A., Long, G., Mirsky, M., Regan, K., van Esch, E., Westwood, F. R., Vidal, J., & Yoshida, M. (2014). Nonproliferative and proliferative lesions of the rat and mouse female reproductive system. *Journal of Toxicologic Pathology*, *27*(3-4 suppl), 1S–107S.

Dobreniecki, S., Mendez, E., Lowit, A., Freudenrich, T. M., Wallace, K., Carpenter, A., Wetmore, B. A., Kreutz, A., Korol-Bexell, E., Friedman, K. P., & Shafer, T. J. (2022). Integration of toxicodynamic and toxicokinetic new approach methods into a weight-of-evidence analysis for pesticide developmental neurotoxicity assessment: A case-study with DL- and L-glufosinate. *Regulatory Toxicology and Pharmacology*, *131*, 105167.

ECETOC. (2002). Guidance on evaluation of reproductive toxicity data. Monograph no. 31. https://www.ecetoc.org/wp-content/uploads/2014/08/MON-031.pdf

EPA. (1991). Guidelines for developmental toxicity risk assessment. https://www.epa.gov/risk/guidelines-developmental-toxicity-risk-assessment

EPA. (1996). Guidelines for reproductive toxicity risk assessment. https://www.epa.gov/sites/default/files/2014-11/documents/guidelines_repro_toxicity.pdf

FDA. (2011). Guidance for industry reproductive and developmental toxicities — Integrating study results to assess concerns (Docket Number: FDA-1999-N-0082). https://www.fda.gov/media/72231/download

Fleeman, T. L., Cappon, G. D., Chapin, R. E., & Hurtt, M. E. (2005). The effects of feed restriction during organogenesis on embryo-fetal development in the rat. *Birth Defects Research. Part B, Developmental and Reproductive Toxicology*, *74*(5), 442–449.

ICH. (2009). M3(R2) Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. https://database.ich.org/sites/default/files/M3_R2__Guideline.pdf

ICH. (2020). S5(R3) Guidance on detection of reproductive and developmental toxicity for human pharmaceuticals. https://database.ich.org/sites/default/files/S5-R3_Step4_Guideline_2020_

Keenan, C., Elmore, S., Francke-Carroll, S., Kemp, R., Kerlin, R., Peddada, S., Pletcher, J., Rinke, M., Schmidt, S. P., Taylor, I., & Wolf, D. C. (2009). Best practices for use of historical control data of proliferative rodent lesions. *Toxicologic Pathology*, *37*(5), 679–693.

Keenan, C., Elmore, S., Francke-Carroll, S., Kemp, R., Kerlin, R., Peddada, S., Pletcher, J., Rinke, M., Schmidt, S. P., Taylor, I., & Douglas, C. (2009). Wolf best practices for use of historical control data of proliferative rodent lesions. *Toxicologic Pathology*, *37*(5), 679–693.

Kerlin, R., Bolon, B., Burkhardt, J., Francke, S., Greaves, P., Meador, V., & Popp, J. (2016). Scientific and regulatory policy committee: Recommended ("best") practices for determining, communicating, and using adverse effect data from nonclinical studies. *Toxicologic Pathology*, *44*(2), 147–162.

Lein, P., Locke, P., & Goldberg, A. (2007). Meeting report: Alternatives for developmental neurotoxicity testing. *Environmental Health Perspectives*, *115*(5), 764–768.

Makris, S. L., Solomon, H. M., Clark, R., Shiota, K., Barbellion, S., Buschmann, J., Ema, M., Fujiwara, M., Grote, K., Hazelden, K. P., Hew, K. W., Horimoto, M., Ooshima, Y., Parkinson, M., & Wise, L. D. (2009). Terminology of developmental abnormalities in common laboratory mammals (version 2). *Reproductive Toxicology*, *28*(3), 371–434.

NAFTA. (2006). Technical working group (TWG) on pesticides: Developmental neurotoxicity study (DNT) guidance document. https://www.epa.gov/sites/default/files/2017-02/documents/developmental_neurotoxicity_study_internal_guidance_document_final_0.pdf

OECD. (1996). *Test No. 422: Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test*. OECD Publishing.

OECD. (2018a). *Test No. 414: Prenatal developmental toxicity study, OECD guidelines for the testing of chemicals, Section 4*. OECD Publishing.

OECD. (2018b). *Test No. 443: Extended one-generation reproductive toxicity study, OECD guidelines for the testing of chemicals, Section 4*. OECD Publishing.

Sachana, M., Shafer, T. J., & Terron, A. (2021). Toward a better testing paradigm for developmental neurotoxicity: OECD efforts and regulatory considerations. *Biology*, *10*(2), 86.

Stump, D. G., Nemec, M. D., Parker, G. A., Coder, P. S., Sloter, E. D., & Varsho, B. J. (2012). Significance, reliability, and interpretation of developmental and reproductive toxicity study findings. In R. D. Hood (Ed.), *Developmental and reproductive toxicology: A practical approach* (3rd ed., p. 73). Taylor and Francis Group.

How to cite this article: Green, M. L., Kluever, A., Chen, C., Dobreniecki, S., Halpern, W., Hannas, B., Hoberman, A., McNerney, M. E., Mitchell-Ryan, S., Shafer, T. J., Van Cruchten, S., & White, T. (2024). HESI workshop summary: Interpretation of developmental and reproductive toxicity endpoints and the impact on data interpretation of adverse events. *Birth Defects Research*, *116*(2), e2311. https://doi.org/10.1002/bdr2.2311