

Systematic Review Workflow Can Already Incorporate NAMs in Evidence Syntheses

Andrew Rooney, PhD

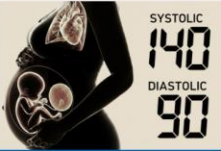
Acting Chief, Integrative Health Assessments Branch
Division of Translational Toxicology
National Institute of Environmental Health Sciences

DTT Integrative Health Assessments Branch

Links previous Office of Health Assessment and Translation (OHAT)
and Report on Carcinogens (RoC)

- **Critically assess the scientific evidence**
that environmental substances cause adverse health effects
- **Conclusions to inform**
public, government, scientific, and medical communities
- **Promote understanding and advance methodologies**
using the highest standards in rigor and reproducibility
- **Develop innovative informatic approaches**
anchored in systematic review requirements to support
decision making in health effects research and policy

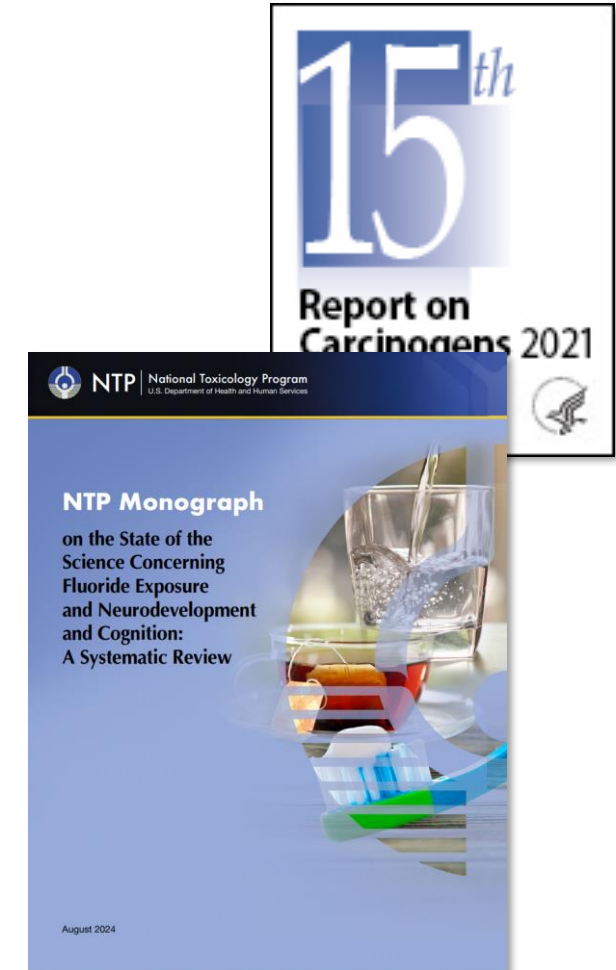
IHAB Projects and Tools

 <u>A Web-Based Tool for Autism Research and the Environment (aWARE)</u> The aWARE tool will be a resource for the public and the research community to help make research on ASD and environmental factors more findable, accessible, and ultimately more useful.	 <u>Cardiovascular Diseases: Physical and Social Environment Risk Factors</u> Researchers in DTT developed an interactive tool to help the public and research community explore the scientific literature on physical and social environmental factors and cardiovascular disease (CVD). CVD is the leading cause of death in the United States and worldwide.	 <u>Class-Based Approach to Assessing Organohalogen Flame Retardants</u> Organohalogen flame retardants are a large group of chemicals often divided into subclasses. Researchers in DTT are evaluating the health effects evidence for subclasses of these chemicals collectively, considering both structural similarity and biological mechanisms of effects.
 <u>Dextr: Semi-Automated Data Extraction Tool</u> Dextr is a web-based tool that supports automated data extraction of published research for literature reviews. Workflow enables users to verify entities, connect related data, leverage controlled vocabularies, and extract data from text or tables.	 <u>Hypertensive Disorders of Pregnancy (HDP)</u> DTT researchers are developing a systematic evidence map to characterize biomarkers associated with HDP. This interactive tool will enable users to explore the data, advance understanding of HDP biomarkers, and assess the utility of rodent models to environmental exposures.	 <u>Personal Care Products and Their Chemical Constituents</u> A growing body of research suggests that some chemicals found in personal care products may affect human health, such as impacting the timing of puberty or altering fetal growth. DTT is using a "class-based" approach to identify and assess scientific evidence for groups of related chemicals.

Rigorous methods to support decision making

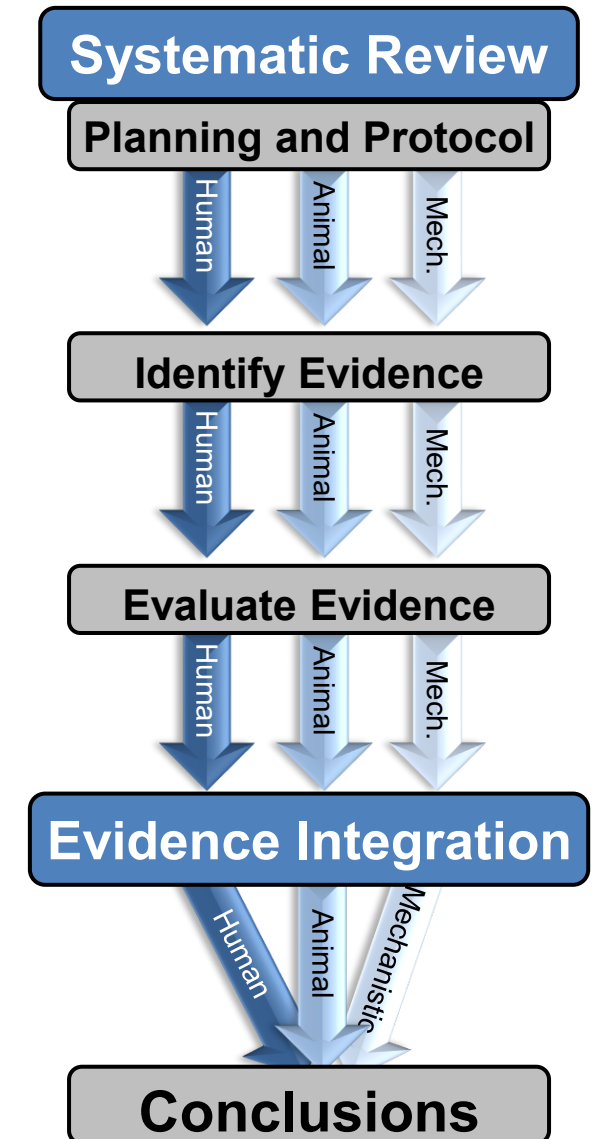
Rigor, reproducibility, transparency

- Systematic reviews are essential for evidence-based assessments of potential human health effects from chemical or environmental exposures
- These methods minimize bias and ensure transparency in identifying, capturing, critically assessing, and synthesizing research to answer specific questions



Systematic reviews can assess all data types

- **Established workflows**
 - Evaluating evidence across studies of the same or similar endpoints (apples to apples)
 - Growing experience with mechanistic data and NAMs
- **Evidence Integration**
 - Consideration of factors that could increase or decrease confidence in a body of data
 - Includes assessment of consistency both within and across evidence streams
 - Could be used to transparently evaluate available data to anchor NAMs data



Example consideration of in vitro data

- **Objective:** to develop hazard conclusions on the association between exposure to PFOA or PFOS and immunotoxicity
- **Health effects:**
 - Strongest and sensitive data were on immunosuppression
 - Suppression of the antibody response
- Data from all three evidence streams
 - Human studies
 - Animal studies
 - In vitro studies



National Toxicology Program
U.S. Department of Health and Human Services

What We Study ▾ How We Work ▾ Data & Resources Publications Who We Are ▾

Home > What We Study > Health Effects Assessments > Noncancer Health Effects > Completed Evaluations > PFOA & PFOS

Immunotoxicity Associated with Exposure to Perfluorooctanoic Acid (PFOA) or Perfluorooctane Sulfonate (PFOS)

 **Topic Overview**
CASRN: 335-67-1; 1763-23-1
Status: Evaluation completed

— BACKGROUND INFORMATION —

Perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS) are extremely persistent chemicals that are widely distributed in the environment as a result of high chemical stability under normal environmental conditions and extensive use over the last 50 years in commercial and industrial applications including fluoropolymer manufacturing, food packaging, lubricants, water-resistant coating, and fire-fighting foams. PFOS was phased out of production and use in 2002, and U.S. manufacturers eliminated PFOA emissions and product content at the end of 2015. Although emissions have been dramatically reduced in the United States and Western Europe, it is not clear if global production has changed as there has been a shift in productions to Asia. Some published studies of PFOA and PFOS raised concerns about potential immune system health effects and NTP received nominations to conduct a review of immune effects for these chemicals.

NTP conducted a systematic review to evaluate the evidence on exposure to PFOA or PFOS and immune-related health effects to determine whether exposure to either chemical is associated with immunotoxicity for humans. NTP concludes that both PFOA and PFOS are presumed to be an immune hazard to humans based on a high level of evidence from animal studies that PFOA and PFOS suppressed the antibody response and a moderate level of evidence from studies in humans. The evidence that these chemicals affect multiple aspects of the immune system supports the overall conclusion that both PFOA and PFOS alter immune functions in humans.

— DOCUMENTS —

⬆ BACK TO TOP

<https://ntp.niehs.nih.gov/go/749926>

Apples to apples and mechanisms

We are not as good at
assessing study quality or
risk of bias for NAMs studies

Antibody Response and mechanisms



Human Data

- antibodies to vaccines
- anti-rubella IgM
- anti-rubella IgG



Experimental Animal Data

- antibodies to T-cell antigens
- anti-SRBC IgM
- anti-SRBC IgG



In vitro and NAMs Data

- *in vitro* IgM
- mechanisms of antibody production/response

NAMs



NAMs

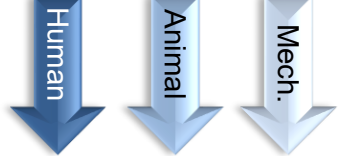


NAMs

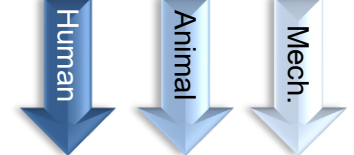


Systematic Review

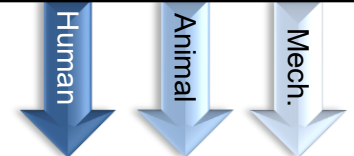
Plan and Protocol



Identify Evidence



Evaluate Evidence

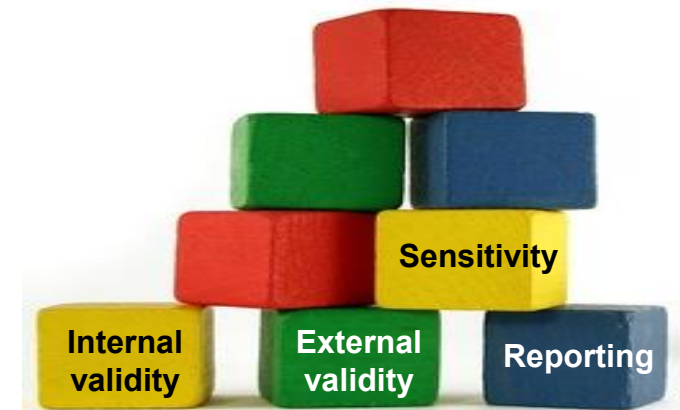


Evidence Integration



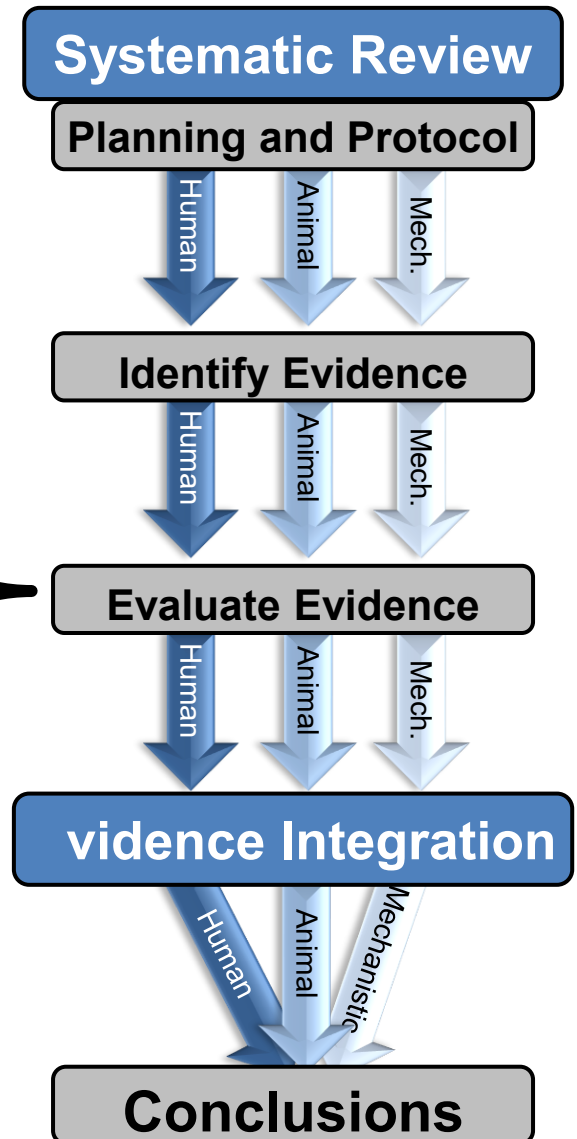
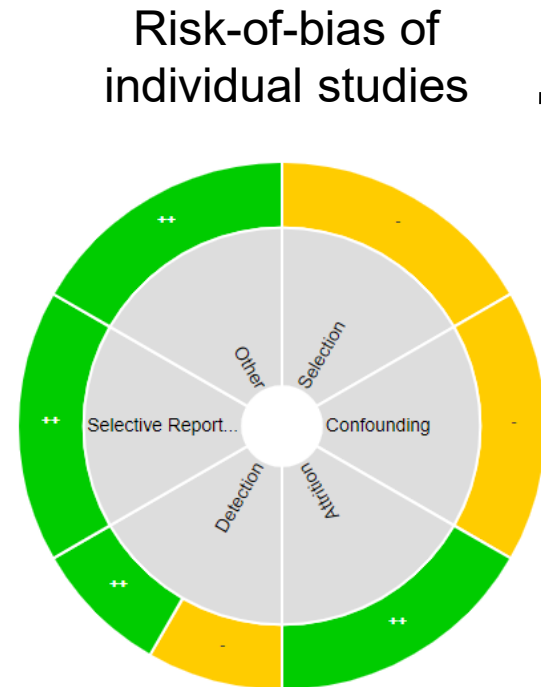
Risk of bias in context

- **Reporting quality**
 - How well was the study reported?
- **Risk of bias (also called internal validity)**
 - How credible are the findings based on study design and conduct?
 - Methodological characteristics of a study that may lead to systematic errors in magnitude or direction of the results
- **Directness and Applicability (also called external validity)**
 - How well does the study address the topic under review?
 - Population of interest? Upstream indicator of primary health outcome
 - Route, timing/duration of exposure and health outcome assessment
 - Relevance of animal model for human health



Why assess risk-of-bias (RoB) in systematic reviews?

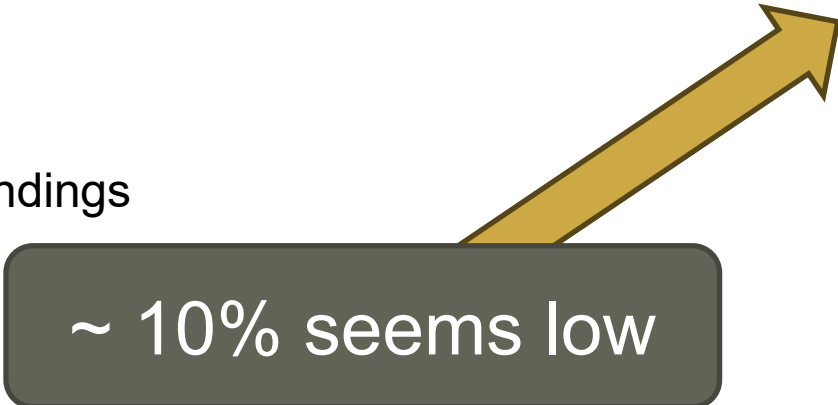
- Critical, transparent, and consistent evaluation of the body of evidence is required for a systematic review
 - Absence of RoB severely limits utility
 - Is it a systematic review without RoB?
- Characterization of the RoB of individual studies informs confidence in the body of evidence
 - Conclusions, confidence, and any discussion of uncertainty should consider entire body of evidence, and
 - With particular attention to the RoB in the best evidence



Risk of bias
is a critical aspect for
systematic reviews in
toxicology and public health,
but assessing it takes too long...

Estimated percent of time for each step

Step	Percent Time
Problem Formulation and Protocol Development	8 %
Identify the Evidence (<i>primarily split screening and extracting</i>)	39 %
Evaluate the Evidence	9 %
Summarize and Synthesize Data	5 %
Integrate Evidence and Report Findings	20 %
<i>Project Management</i>	19 %



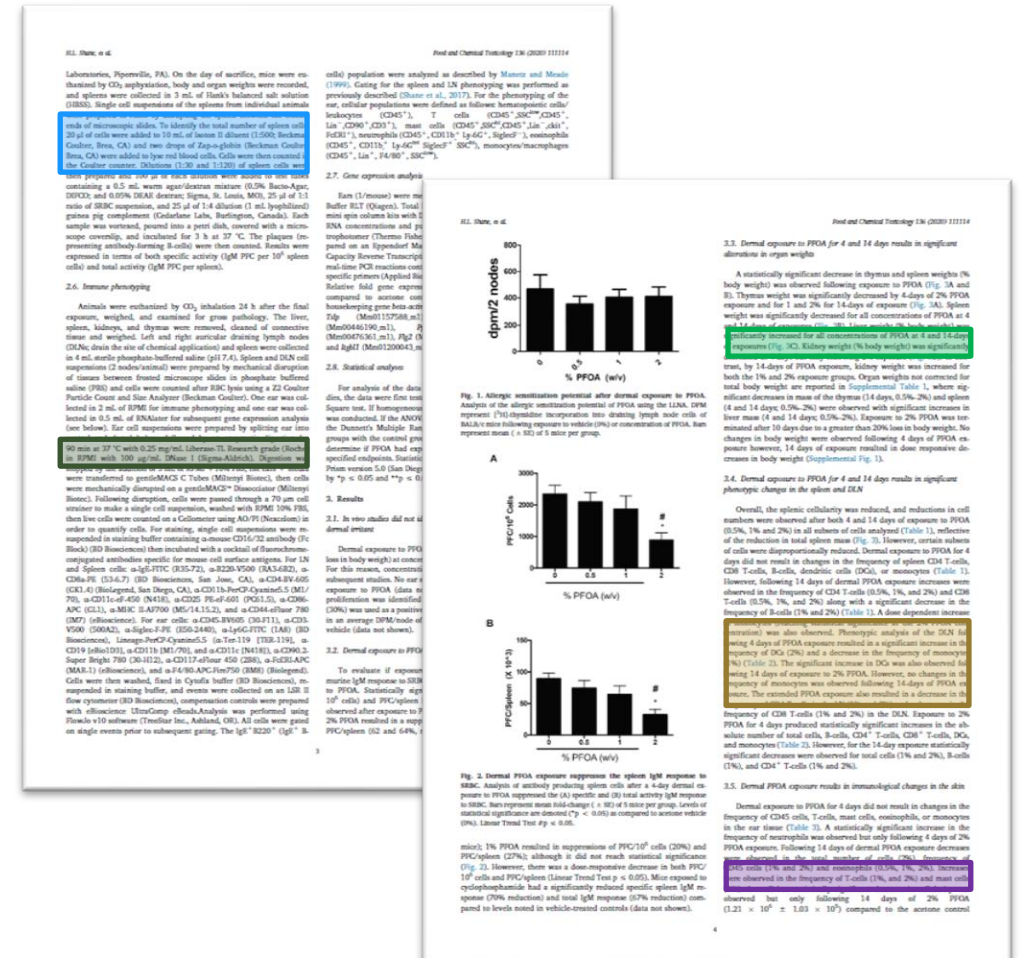
~ 10% seems low

Opportunities for ML, LLM, and AI - what about Risk of Bias?

Methods for using AI in literature search and screening are increasingly common

... including data extraction tools

- Machine learning (ML) approaches have been used for years
- The emergence of large language models (LLMs) provides opportunities to move forward with less effort on training data sets
- These approaches can identify specific text or sections of text in published research



AI assisted extraction

Dextr

- Web-based user interface that processes full text PDFs using structure data extraction forms and AI-assisted models.
- Researcher verifies (or edits) model predictions to confirm extraction element.
- QC Module enables dual extractors or AI-assisted verification.

Dextr User Interface

ecting testis function in adult male rats

Extraction (default) ▾

Find field

+ Entry

+ ○ Chemical

Bisphenol AF (BPAF)

Chemical	Bisphenol AF (BPAF)	✓	✗
CAS	80945	✓	✗
Chemical Source	Tokyo Chemical Industry (TCI, Tokyo, Jap...	✓	✗
Purity Available	True	✓	✗
Purity	99%	✓	✗
Purity Qualifier	=	✓	✗
Vehicle	corn oil with 4% ethanol	✓	✗

User Verification
“Human-in-the-loop”

+ ○ Sprague

Name *	Sprague-Dawley rat	✓	✗
Species *	Rat	✓	✗
Strain *	Sprague-Dawley	✓	✗
Sex *	Male	✓	✗
Animal Source	Academy of Military Medical Sciences, Be...	✓	✗
Lifestage Exposed	Juvenile	✓	✗
Lifestage Assessed	Juvenile	✓	✗
Siblings			
Generation	Select		
Parents			
Diet	Food and water ad libitum	✓	✗
Comments	Housed one per metabolic cage; 12-h lig...	✓	✗

PDF

© 2012 Elsevier Ireland Ltd. All rights reserved.

Dextr Integrates Multiple Models

NLP, regular expressions and LLMs
identify data extraction elements

In this study, adult male rats were dosed via oral gavage with 2, 10, 50 and 200 mg/kg/d BPAF for 14 days. The toxicity was evaluated by testicular BPAF concentration, serum levels of LH, FSH, and testosterone, as well as serum total cholesterol. In addition, the mRNA levels of nuclear steroid receptors were detected. The involvement of steroidogenesis in the toxicity was investigated at both gene and protein levels.

model predictions
identified in text

2.1. Animals

Male Sprague-Dawley (SD) rats aged 7 weeks were obtained from the Academy of Military Medical Sciences, Beijing, China. Six animals were assigned to treatment and control group. Animals were housed one per metabolic cage and maintained in a mass air displacement room with a 12-h light-dark cycle and a relative humidity of 30–70%. Animals had access to food and water ad libitum. All rats were acclimatized for one week before experiments were begun.

User verifies

Extraction (default) Find field +

+ Animal Group

Sprague-Dawley rat

Name *

Sprague-Dawley rat

Animals had access to food and water ad libitum.

Animals were housed one per metabolic cage and maintained in a mass air displacement room with a 12-h light-dark cycle and a relative humidity of 30–70%.

All rats were acclimatized for one week before experiments were begun.

Type your comment here

Species *	Rat	✓	✗
Strain *	Sprague-Dawley	✓	✗
Sex *	Male	✓	✗
Animal Source	Academy of Military Medical Sciences, Beijing, China	✓	✗
Lifestage Exposed	Juvenile	✓	✗
Lifestage Assessed	Juvenile	✓	✗
Siblings			
Generation	Select		

Risk of bias includes simple and complex concepts

Randomization

Relatively easy to identify:

- random, randomize, etc.
- or how animals were assigned to treatment

Outcome Assessment

Which methods are reliable?

- Is formalin acceptable fixative for routine histology?
- What about for the spleen?

have been observed in the rat following oral [10,11] and dermal administration [12]. A direct interaction with the testicular components could not be excluded since lindane accumulates in testis [12,13].

process of sperm development can also be evaluated using the Sperm Chromatin Structure Assay (SCSA) [25]. SCSA allows detection of increased sperm DNA susceptibility to in situ denaturation, which indicates derangements of normal chromatin remodeling and packaging in the nucleus of the male gamete. The induction of aberrant sperm chromatin structure is a hallmark of germ cell damage [21,22].

In the present study, the possible delayed effects in the reproductive system of the male progeny of lindane-treated pregnant CD1 mice were investigated using conventional endpoints as well as FCM DNA analysis and SCSA. The effects of lindane were compared with those induced by diethylstilboestrol (DES) as a well-established model for estrogenic endocrine-disrupting chemical-mediated

the microdetermination of total protein and testicular creatinine, respectively.

The reagents used for FCM analysis are listed in the following detailed description of the methods.

For histology, the reagents used were Carnoy liquid, ethyl alcohol from Carlo Erba (Italy), 2-hydroxy-ethyl methacrylate Technovit 7100 from Heraeus Kulzer (Germany), Mayer's Haemalum and alcoholic eosin from Bio-optica (Italy).

2.2. Animals and treatments

Experiments were carried out in compliance with the ethical provisions enforced by the European Union and authorized by the National Committee of the Italian Ministry of Health on in vivo experimentation.

Mated CD1 mice (Harlan, Italy) were maintained one to a cage at a room temperature of $20 \pm 2^\circ\text{C}$ with $50 \pm 10\%$ humidity and a 12 h photoperiod. Water and food (open formula NIH

Lindane was administered from gestation day 10 to day 17 (0.5 mg/kg body weight) or 25 mg/kg body weight (0.5 mg/kg body weight) was administered. The lindane dose was based on the available literature [7]. The maternal toxicity or the DES dose level was based on a dose-range finding study. The control group was treated with vehicle (corn oil).

sodium hydroxide pellets (NaOH), sodium bicarbonate (NaHCO₃), potassium phosphate (K₂HPO₄), dipotassium hydrogen orthophosphate (K₂HPO₄), dipotassium hydrogen orthophosphate (K₂HPO₄), dipotassium hydrogen orthophosphate (K₂HPO₄).

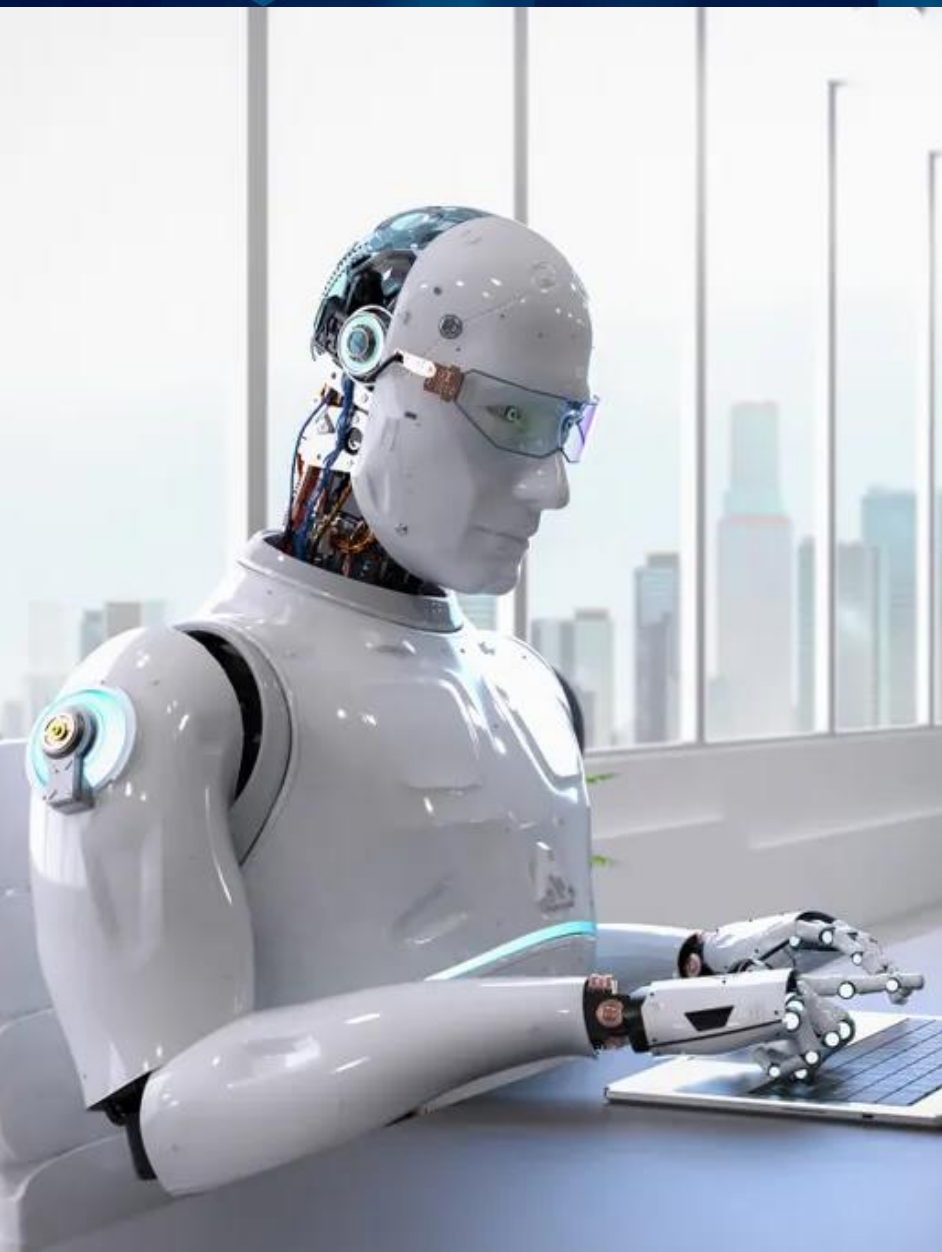
A kit from Sigma diagnostic and creatinine kit from Brainer-Mannheim GmbH were used for

Jump to		Type extracted value	+ Add	
Entries		✓ accept all	✗ reject all	
Exposure	+			
Study Design	+			
Outcome	+			
5-Author reported Endpoint	Sperm head count	✓	✗	
5-Author reported Endpoint	Sperm concentration	✓	✗	
5-Author reported Endpoint	Testicular weight	✓	✗	
5-Author reported Endpoint	Lactate dehydrogenase-C 4 (L...	✓	✗	
5-Author reported Endpoint	Sorbitol dehydrogenase (SDH)...	✓	✗	
5-Author reported Endpoint	Creatine concentration	✓	✗	
5-Author reported Endpoint	DNA content analysis	✓	✗	
5-Author reported Endpoint	Sperm Chromatin Structure As...	✓	✗	
5-Author reported Endpoint	Number of live pups/litter	✓	✗	
5-Author reported Endpoint	Stillbirth rate	✓	✗	
5-Author reported Endpoint	Survival to weaning	✓	✗	
5-Author reported Endpoint	Incidence of runs	✓	✗	
5-Author reported Endpoint	Sex ratio	✓	✗	
5-Author reported Endpoint	Birth weight	✓	✗	
5-Author reported Endpoint	Birth weight gain	✓	✗	
			+	+
			+	+
			+	+
Save final				

Accuracy, verification, and acceptability

Risk of bias is already highly debated

- **Current state:**
 - Subject matter experts are critical to manual work
 - Caution and perhaps fear of AI
 - Need for transparency, standards for AI use
- **Question:** How can AI help right now?
 - We can already do the easy stuff
 - Highlight key areas of text to speed expert review
 - “human-in-the-loop” approaches
- **Question:** How do we move forward?
 - Develop, test, validate the new approaches
 - Recognize we are facing a skeptical community





National Institute of
Environmental Health Sciences
Division of Translational Toxicology

Thank you

