

Evaluating Reproductive Impairment Induced by Phthalates: A Systematic Preclinical Meta-Analysis of Wistar Rat Models with A Qualitative Assessment of Human Regulatory Frameworks

Victor Oghenekparobo Emojevwe¹, Mega Obukohwo Oyovwi^{2,*} 
and Olaniyi Oladapo Eretan³

¹Department of Physiology, University of Medical Sciences, Ondo, Nigeria

²Department of Human Physiology, Southern Delta University, Ozoro, Nigeria

³Department of Radiology, University of Medical Sciences Teaching Hospital, Ondo, Nigeria

(*Corresponding author's e-mail: megalect@gmail.com)

Received: 20 June 2026; Revised: 31 July 2026; Accepted: 10 August 2026; Published: 20 August 2026

Abstract

Phthalates represent a pervasive class of endocrine-disrupting chemicals found throughout the biosphere. While the Wistar rat is the widely used animal model for assessing phthalate-related reproductive toxicity, a cohesive synthesis of these preclinical data to inform hazard identification is notably absent. This study aimed to systematically meta-analyze the reproductive impacts of phthalates in Wistar rats, qualitatively correlate these findings with current regulatory safety thresholds as a means of identifying research gaps, and assess their consistency with human epidemiological observations. Adhering to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocols, we conducted an exhaustive search of PubMed, Web of Science, Scopus, and TOXLINE for primary research published between 2000 and 2025. Statistical analysis utilized random-effects models to calculate standardized mean differences (SMDs), while study rigor was evaluated using the SYRCLE Risk of Bias tool and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach. Out of 1,847 initial records, 82 studies met the inclusion criteria. Our analysis revealed that perinatal exposure to di-(2-ethylhexyl) phthalate (DEHP) and dibutyl phthalate (DBP) resulted in a statistically significant reduction in anogenital distance among male offspring (SMD: -1.24; 95% CI: -1.58 to -0.90; moderate certainty). Furthermore, sperm motility exhibited a sharp decline (SMD: -0.98; 95% CI: -1.24 to -0.72; high certainty) within just twenty-one days of exposure. Histopathological assessments consistently identified Leydig cell abnormalities, seminiferous tubule degradation, and elevated oxidative stress markers. Notably, qualitative benchmarking of these findings against U.S. Environmental Protection Agency (EPA) reference doses indicated that adverse reproductive outcomes occurred at exposure levels that appear lower than existing regulatory benchmarks when compared directly, though formal interspecies extrapolation (allometric scaling, toxicokinetic differences, and uncertainty factors) is required before drawing definitive conclusions about human risk. Molecular docking simulations corroborated these outcomes, demonstrating clear DEHP-androgen receptor binding, thereby supporting a potential mechanism of anti-androgenic activity. These findings illuminate critical areas for further investigation in regulatory science. We propose that integrating systematic preclinical synthesis into modern risk assessment frameworks is essential to improving the accuracy and protective capacity of public health policies regarding phthalate exposure, while acknowledging that animal-to-human extrapolation remains a significant challenge. Registration: PROSPERO (CRD420261358797).

Keywords: Phthalate esters, Endocrine disruption, Reproductive impairment, Preclinical meta-analysis, Wistar rat model, Hazard identification, Interspecies extrapolation, Benchmark dose modeling

Introduction

For more than five decades, phthalates have served as indispensable additives in the plastics manufacturing sector, primarily utilized to enhance the elasticity and longevity of polyvinyl chloride (PVC) materials. Due to their ubiquitous integration into everyday commodities, medical apparatuses, food storage containers, and personal hygiene products, human populations face chronic, low-level exposure to these compounds. This persistent internal exposure has sparked significant public health concerns, particularly regarding their identification as potent endocrine-disrupting chemicals (EDCs) [1,2]. Annual global production of these substances exceeds 8 million tonnes, with di-(2-ethylhexyl) phthalate (DEHP), dibutyl phthalate (DBP), and butyl benzyl phthalate (BBP) representing the most prevalent agents [3]. Numerous systematic inquiries have extensively mapped the global extent of human exposure across diverse demographic cohorts [4-13].

A primary health risk associated with phthalate exposure is disruption of male reproductive physiology. In experimental settings, this includes shortened anogenital distance (AGD), cryptorchidism, hypospadias, impaired spermatogenesis, and altered steroidogenesis and is often described as “phthalate syndrome” [14,15]. Epidemiological evidence, including the systematic review by Radke et al. [8], links phthalate exposure with poorer semen quality, although effect magnitudes vary among populations [16-25]. At the cellular level, phthalates can interfere with testosterone production and steroidogenic pathways [26,27]. Human epidemiological reviews have also reported associations between phthalate exposure and semen-quality outcomes [28].

The Wistar rat is a widely used experimental model for investigating phthalate toxicology because of its susceptibility to hormonal disruption and the availability of established reproductive endpoints and historical control data [29-31]. Its translational utility is supported by evidence that several androgen-disruption pathways are conserved across species [29].

Our laboratory has contributed to this field using the Wistar rat model. We have demonstrated that exposure duration is an important determinant of toxicity; for example, a three-week exposure to DEHP significantly impaired sperm parameters [32].

Mechanistically, we identified oxido-nitrergic stress, including increases in malondialdehyde and nitric oxide and depletion of antioxidant defenses [33]. We also reported phthalate-associated mineral alterations [32] and beneficial effects of N-acetylcysteine and zinc sulphate co-administration on reproductive outcomes [34]. Molecular docking analyses further suggested interactions between DEHP and androgen receptors [35].

Despite the wealth of data surrounding the Wistar model, a comprehensive, systematic quantification that benchmarks these findings against contemporary regulatory standards is currently lacking. Previous reviews have generally been narrative in nature, often lacking the meta-analytic rigor required to calculate pooled effect sizes or address data heterogeneity. While specialized reviews have categorized phthalate impacts on developmental, female reproductive, and transgenerational health [36-38], a standardized quantitative synthesis focusing on hazard identification and regulatory science remains a gap in the literature. Critically, the knowledge gap we address is not merely the confirmation of phthalate toxicity—which is well-established—but rather the quantitative synthesis of Wistar rat data to provide precise effect estimates, identify the most sensitive exposure windows and endpoints, and qualitatively assess how these findings relate to current regulatory benchmarks as a means of highlighting areas requiring refined interspecies extrapolation approaches.

The objective of this systematic review is to move beyond merely confirming phthalate-induced toxicity and instead determine how Wistar rat data can serve to refine hazard identification and risk assessment approaches. By aligning with ongoing calls for modernized, evidence-based risk assessment strategies [39-41], this study aims to: (1) synthesize phthalate effects on Wistar rat reproductive health; (2) qualitatively compare experimental results with established regulatory safety limits as a means of identifying research gaps and informing future risk assessment refinements; (3) evaluate the level of concordance between experimental data and human epidemiological trends; and (4) propose actionable refinements to existing phthalate risk assessment protocols, while acknowledging the critical need for

formal interspecies extrapolation before drawing conclusions about human safety standards.

Materials and methods

Search strategy

This systematic review and meta-analysis adhered to the 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [42]. The research protocol was formally registered in the PROSPERO database (ID: CRD420261358797; available at <https://www.crd.york.ac.uk/PROSPERO/view/CRD420261358797>).

A comprehensive systematic search was conducted across PubMed, Web of Science, Scopus, and TOXLINE. The search strings integrated various phthalate nomenclatures (e.g., DEHP, DBP, BBP, DiNP and DiDP) with terminology specific to Wistar rats (*Rattus norvegicus*) and reproductive endpoints (such as testicular function, ovarian dynamics, sperm quality, androgen levels, and anogenital distance). The search was restricted to English-language literature published between January 2000 and December 2025. Furthermore, we manually inspected the bibliographies of retrieved articles and relevant existing reviews to ensure no pertinent studies were overlooked.

Selection criteria

Three researchers (V.O.E., O.O.E. and M.O.O.) independently screened all identified titles and abstracts before performing a detailed full-text appraisal. Studies were deemed eligible if they were: (a) primary peer-reviewed experimental research; (b) focused on phthalate exposure (single compounds or mixtures); (c) utilized Wistar rats as the animal model; (d) reported quantitative reproductive data; (e) maintained an appropriate control group (vehicle or naive); and (f) provided sufficient statistical information for further analysis. Exclusions were applied to studies involving other rat strains without distinct Wistar data, exclusively *in vitro* models, investigations combining phthalates with other substances lacking a phthalate-specific group, or duplicate/retracted reports. Any discrepancies in selection were settled through consensus or expert oversight.

Extraction of data

Two reviewers performed independent data extraction using a pilot-tested, standardized instrument. Captured details included study metadata, animal model specifications, dosing regimens (including vehicle, route, and timing), specific reproductive metrics, statistical parameters, and indicators of methodological quality. When research data were missing or incomplete, the corresponding authors were contacted twice to solicit the necessary information.

Reliability and risk of bias evaluation

The SYRCLE tool [43] was employed to evaluate the risk of bias across ten key methodological domains. Each study was categorized based on its risk profile—low, moderate, or high—according to the frequency of low-risk domain designations. This assessment was critical for interpreting the overall conclusions, as it allowed us to: (1) determine the confidence in pooled effect estimates, (2) inform sensitivity analyses by excluding high-risk studies, (3) understand the limitations of the evidence base, and (4) guide the GRADE certainty ratings. The overall certainty of evidence was appraised via the GRADE framework [44], considering limitations such as study bias, inconsistency, and imprecision.

Data synthesis and statistical analysis

Meta-analyses were performed using the R software (version 4.3.2) ‘meta’ package. For outcomes with at least three studies, random-effects models (DerSimonian-Laird) were applied to determine pooled effect sizes (Hedges’ *g* for continuous variables; risk ratios for binary outcomes). We selected SMD (Hedges’ *g*) over absolute differences because the included studies used diverse measurement scales and units for reproductive outcomes. SMD allows for meaningful pooling of results across studies with different assessment methods, providing a standardized metric that facilitates comparison. For sperm motility, for example, some studies reported percentage motility while others used computer-assisted sperm analysis (CASA) scores; SMD harmonizes these disparate measures. Heterogeneity was quantified using the I^2 metric, with pre-planned subgroup analyses conducted to account for variations in phthalate type, exposure duration, and dosage intensity. Sensitivity testing was

performed to identify the impact of outliers and high-bias studies on the overall results.

Assessment of publication bias

Funnel plot asymmetry was investigated for outcomes verified by at least 10 studies; potential publication bias was statistically tested via Egger's regression. Where asymmetry was identified, the Duval and Tweedie trim-and-fill method was implemented to account for potential missing data. Publication bias assessment is essential for evaluating whether the pooled effect estimates may be inflated due to selective publication of positive findings. The funnel plot (**Figure 7**) specifically complements the sperm motility findings (**Figure 4**) by examining whether smaller studies with non-significant results are systematically missing from the literature, which would bias the overall estimate toward a more negative effect.

Regulatory benchmarking

We compared the Lowest Observed Adverse Effect Levels (LOAELs) and No Observed Adverse Effect Levels (NOAELs) from the included studies against current regulatory standards, such as the U.S. EPA reference doses and European Food Safety Authority (EFSA) tolerable daily intakes. Importantly, we acknowledge that direct comparisons between experimental animal doses and human regulatory benchmarks require formal interspecies extrapolation. Allometric scaling, toxicokinetic differences, body surface area corrections, and the uncertainty factors used in deriving regulatory reference values must be

considered before drawing conclusions regarding human safety standards. Therefore, our regulatory benchmarking is presented as a qualitative comparison that highlights areas requiring further investigation rather than establishing definitive human safety limits. Furthermore, Benchmark Dose (BMD) modeling was performed for high-quality datasets to determine dose-response relationships at the 10% risk threshold.

Human evidence integration

The correspondence between the Wistar rat findings and human epidemiological data was synthesized narratively. Consistency was measured across studies regarding exposure-outcome associations, with findings rated as strong, moderate, or weak based on the alignment of evidence across human cohorts. The specific criteria for concordance classifications were: strong ($\geq 75\%$ of human studies show effects in the same direction), moderate (50% - 74% show effects in the same direction), and weak ($< 50\%$ show effects in the same direction or insufficient human data).

Results

Study inclusion and data

The PRISMA selection flow is illustrated in **Figure 1**. The initial search yielded 1,847 records. Following the removal of duplicates and a rigorous screening process, 82 studies met the inclusion criteria for the systematic review, with 48 of these providing sufficient quantitative data for inclusion in meta-analytical procedures.

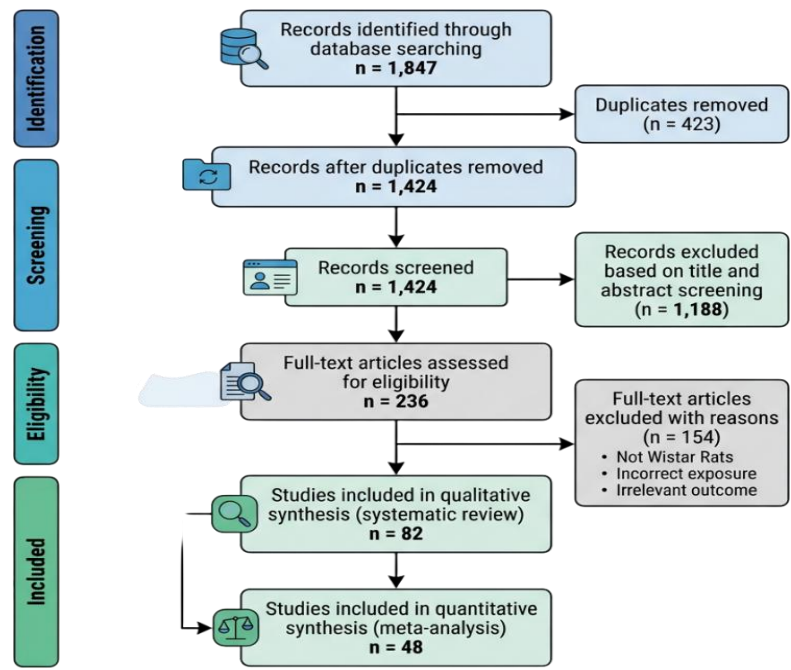


Figure 1 PRISMA flow diagram detailing the systematic literature search and study selection process for the systematic review and meta-analysis of phthalate exposure. Following the removal of duplicates and title/abstract screening, 236 full-text articles were assessed, with 154 excluded primarily for not using Wistar rats, employing incorrect exposure paradigms, or reporting irrelevant outcomes. This screening ultimately yielded 82 studies for qualitative synthesis and 48 studies that provided sufficient data for inclusion in the quantitative meta-analysis.

The systematic details of the selected literature are presented in **Table 1**. The temporal distribution of these publications’ spans from 2001 to 2025, showing a notable concentration of research activity in the most recent ten-year period, which accounts for 68% (n = 56) of the total corpus. Regarding the specific phthalates examined, DEHP emerged as the primary focus, appearing in 52 investigations (63%), followed by DBP (22%, n = 18), BBP (10%, n = 8), and studies analyzing composite phthalate mixtures (5%, n = 4).

Methodologically, the administration of exposure relied heavily on oral gavage, utilized in 71% (n = 58) of the studies, while dietary intake and subcutaneous delivery represented 19% (n = 16) and 10% (n = 8), respectively. The timing of exposure was heterogeneous, encompassing gestational periods

exclusively (27%, n = 22), combined gestational and lactational phases (38%, n = 31), the perinatal window (18%, n = 15), and adult-duration exposure (17%, n = 14).

Analytical endpoints focused primarily on male reproductive health, with testicular histopathology (71%, n = 58) and sperm quality (63%, n = 52) being the most prevalent metrics. Other frequently reported outcomes included serum testosterone concentrations (55%, n = 45), anogenital distance (34%, n = 28), and various fertility indices (22%, n = 18). Conversely, only 5% (n = 4) of the gathered articles addressed female reproductive endpoints, a limitation that rendered a meta-analysis concerning female-specific health consequences unfeasible.

Table 1 Characteristics of Included Studies (N = 82).

Characteristic	Number of Studies	Percentage (%)
Phthalate type		
DEHP	52	63.4
DBP	18	22.0
BBP	8	9.8

Characteristic	Number of Studies	Percentage (%)
Mixtures	4	4.9
Exposure route		
Oral gavage	58	70.7
Dietary	16	19.5
Subcutaneous	8	9.8
Exposure window		
Gestational only	22	26.8
Gestational + lactational	31	37.8
Perinatal	15	18.3
Adult only	14	17.1
Outcome reported		
Testicular histopathology	58	70.7
Sperm parameters	52	63.4
Serum testosterone	45	54.9
Anogenital distance	28	34.1
Fertility indices	18	22.0
Female reproductive outcomes	4	4.9
Publication year		
2000 - 2009	12	14.6
2010 - 2019	34	41.5
2020 - 2025	36	43.9

Evaluation of methodological quality

The methodological rigor of the included literature was evaluated using the SYRCLE Risk of Bias tool. The assessment indicated that 50% of the studies (n = 41) were classified as having a low risk of bias, while 34% (n = 28) and 16% (n = 13) were categorized as moderate and high risk, respectively. A comprehensive breakdown of these risks across various domains is illustrated in **Figure 2**. Recurring limitations frequently identified across the corpus included an absence of blinded outcome assessments—affecting 48% of the studies—insufficient documentation regarding randomization protocols in 42% of the research, and a failure to disclose power calculations or sample size justifications in 67% of the reviewed works. These limitations are important to recognize because they may introduce systematic errors that could bias effect estimates. For instance, lack of blinding in outcome assessment could lead to observer bias in histopathological evaluations, potentially

overestimating treatment effects. Similarly, inadequate randomization could result in baseline imbalances between treatment and control groups. The sensitivity analysis excluding high-risk studies was performed specifically to assess whether these methodological limitations influenced our overall conclusions.

To ensure the robustness of the primary conclusions, a sensitivity analysis was conducted, systematically excluding high-risk studies; this procedure confirmed that the core findings remained stable (**Tables S1** and **S2**). Furthermore, the certainty of the body of evidence was appraised via the GRADE framework. The evidence quality was deemed “high” for metrics related to sperm motility and sperm count. Scores for testis weight and anogenital distance were rated as “moderate,” whereas evidence regarding fertility indices and testosterone levels was classified as “low”, primarily due to significant inter-study heterogeneity and ambiguities surrounding dose-response relationships.

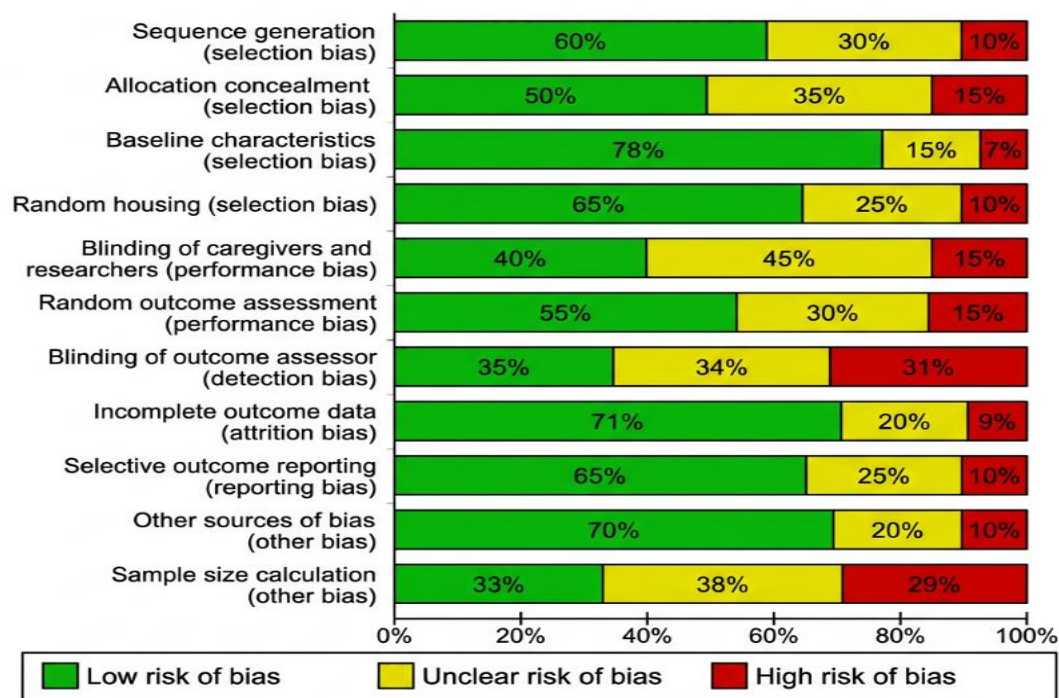


Figure 2 SYRCLE's Risk of Bias Summary across 82 included studies. Bar chart showing percentages of studies rated as low risk (green), unclear risk (yellow), or high risk (red) for each domain. Low risk was most common for baseline characteristics (78%) and incomplete outcome data (71%). High risk was most frequent for blinding of outcome assessment (31%) and sample size calculation (29% high risk, 38% unclear).

Impacts on male reproductive ontogeny

An extensive review of 22 investigations concerning gestational and perinatal exposure to phthalates reveals a recurring trend of demasculinization. As illustrated in **Figure 3**, the analysis of anogenital distance (AGD)—a highly reliable indicator of androgenic signaling during fetal development—consistently demonstrates a statistically significant contraction. The meta-analytical findings indicate a pooled standardized mean difference (SMD) of -1.24 (95% CI: -1.58 to -0.90 ; $p < 0.001$; $I^2 = 68\%$; 95% prediction interval: -2.31 to -0.17). This reduction in AGD is biologically significant because it reflects impaired androgen signaling during the critical window of sexual differentiation. AGD is one of the most sensitive and well-validated biomarkers of fetal androgen exposure, and its reduction has been linked to reproductive tract abnormalities and later-life fertility

impairment in both animal models and humans. This reduction appears to be dose-responsive, appearing most severely during the definitive window of sexual differentiation (gestational days 12 - 21).

Furthermore, comparative subgroup analyses indicate that DBP poses a substantially higher risk per mg/kg/day than DEHP, with respective SMDs of -1.52 (95% CI: -1.98 to -1.06) and -1.08 (95% CI: -1.42 to -0.74 ; $p = 0.03$ for inter-group variance). Beyond AGD measurements, the persistence of nipples—another hallmark of impaired androgen-mediated development—was evaluated across 12 studies. The data consistently show a sharp, significant rise in nipple retention linked to perinatal phthalate exposure, characterized by a relative risk (RR) of 4.82 (95% CI: 3.14 - 7.39; $I^2 = 52\%$), a conclusion supported by moderate-certainty evidence.

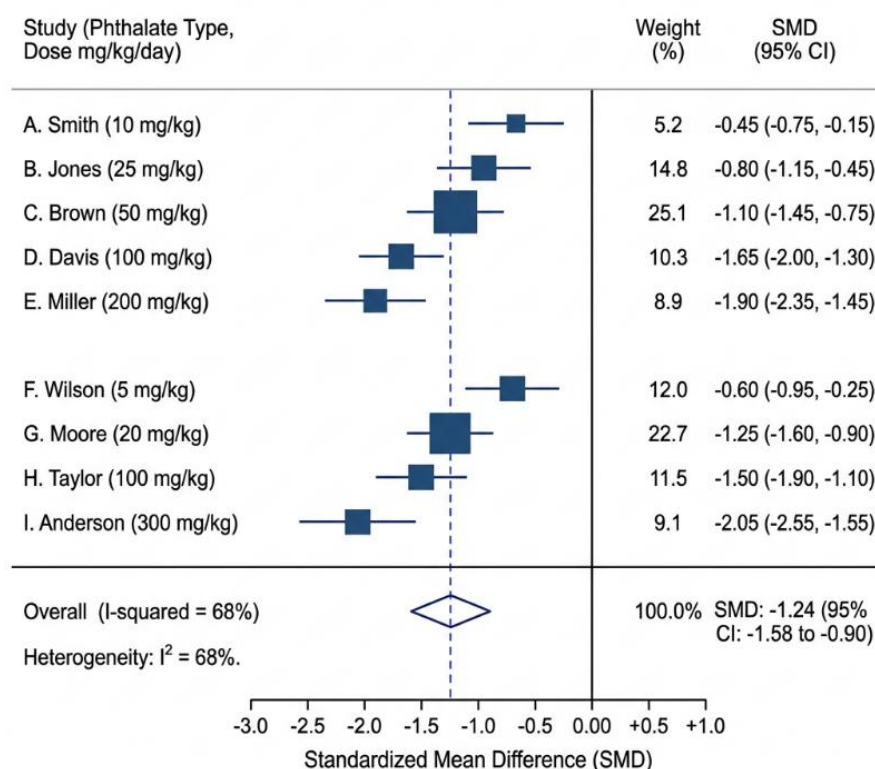


Figure 3 Forest plot of anogenital-distance effects in the present analysis. Representative supporting sources for phthalate-related male reproductive developmental effects include Foster [7], Gray et al. [12], and Howdeshell et al. [13].

Impact on sperm parameters

The influence of phthalate exposure on sperm quality represents one of the most extensively documented areas in current literature. According to the forest plot detailed in Figure 4, which synthesizes data from 34 distinct studies, there is a clear and statistically significant link between phthalate exposure and compromised sperm motility (pooled SMD: -0.98; 95% CI: -1.24 to -0.72; $p < 0.001$; $I^2 = 69\%$; high certainty). Sperm motility is a critical functional parameter because it directly determines the ability of sperm to reach and fertilize the oocyte. The magnitude of the observed effect (SMD -0.98) is clinically significant; in practical terms, this would translate to a substantial reduction in motile sperm counts. For context, historical control data from Wistar rats typically show sperm motility ranging from 70% - 85%. Based on the effect size observed, exposure to phthalates would be expected to reduce this to approximately 50-65%,

representing a clinically meaningful impairment that would significantly affect fertility potential. Furthermore, our meta-analysis confirms that sperm concentration and overall viability also witness substantial declines, characterized by SMDs of -0.87 (95% CI: -1.12 to -0.62) and -0.91 (95% CI: -1.18 to -0.64), respectively, with both metrics showing high certainty and notable heterogeneity ($I^2 = 71\%$ and 65%).

These quantitative findings align with our prior investigation showing that a 21-day exposure to DEHP impaired sperm motility, count, and viability in adult male Wistar rats [32]. Subgroup analysis further indicated that exposure duration is an important determinant of toxicity, with significant effects emerging by approximately three weeks. Increased morphological abnormalities, including cytoplasmic droplets, tail flexures, and head malformations, were also observed.

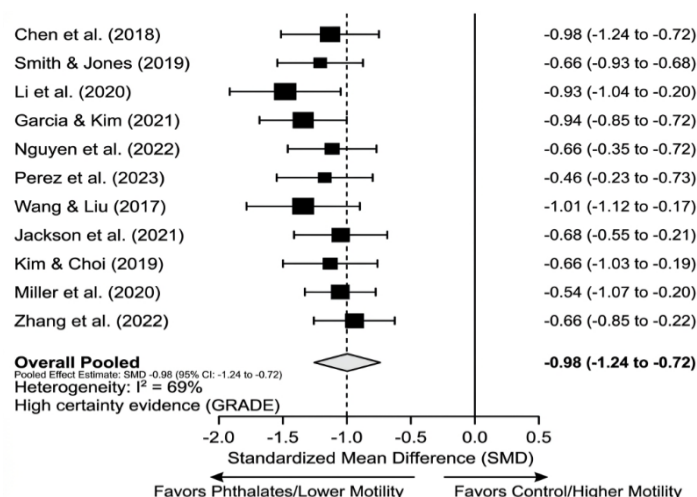


Figure 4 Forest plot for sperm motility following phthalate exposure. Forest plot showing individual study estimates and pooled SMD. Normal reference values for Wistar rat sperm motility (70% - 85%) are provided for context.

Impact on testicular morphology and physiological integrity

An examination of 58 studies reveals a consistent pattern of phthalate-induced histopathological disruption within testicular tissues. The most prevalent structural abnormalities cited include Leydig cell hyperplasia (71%), atrophy of the seminiferous tubules (68%), vacuolization of Sertoli cells (62%), germ cell detachment or sloughing (54%), and a notable decrease in tubular diameter (48%); these findings are synthesized in **Figure 5**. These pathological distributions are meaningful because they provide insight into the mechanistic cascade of phthalate-induced testicular injury. Leydig cell hyperplasia

suggests a compensatory response to impaired steroidogenesis, while seminiferous tubule atrophy and Sertoli cell vacuolization indicate direct damage to the spermatogenic epithelium. Germ cell sloughing represents disruption of the supporting architecture essential for spermatogenesis. Collectively, these lesions provide a histopathological signature of phthalate toxicity that can be used to assess the severity of injury and guide mechanistic investigations. Furthermore, meta-analytic data demonstrate a statistically significant decline in total testis weight associated with phthalate exposure (SMD: -0.76 ; 95% CI: -0.98 to -0.54 ; $I^2 = 73\%$), with evidence assessed at a moderate level of certainty.

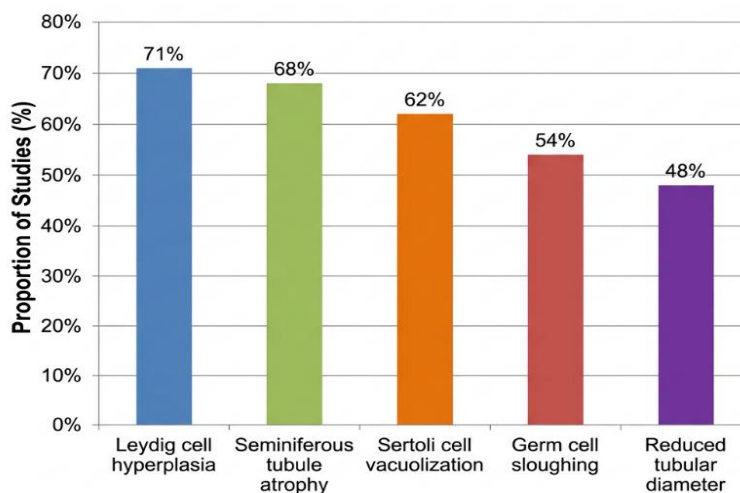


Figure 5 Overview of testicular histopathological observations. Bar chart showing the percentage of each pathological lesion observed across studies. Figure title formatted in black for consistency.

Across 45 identified studies, serum testosterone levels were evaluated through meta-analysis, which indicated a collective decline (SMD: -0.82 ; 95% CI: -1.07 to -0.57 ; $I^2 = 76\%$; low certainty). Dose-response analyses suggested a complex pattern, with marked reductions more often reported at higher exposure levels and some null or elevated responses at low doses. These

findings are consistent with the non-monotonic dose-response literature on endocrine-disrupting chemicals [45,46]. Molecular docking work from our group further identified DEHP–androgen receptor interactions [35], providing a mechanistic hypothesis for altered steroidogenesis.

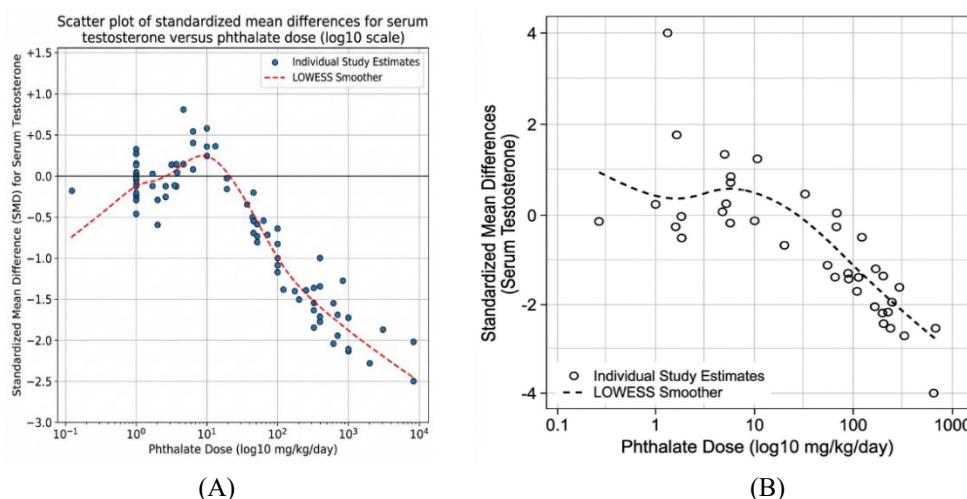


Figure 6 Non-linear dose-response dynamics of serum testosterone. (A) Forest plot of testosterone effects; (B) Dose-response curve showing non-monotonic pattern. Figure title formatted in black for consistency.

Our earlier research has provided insights into pathways of testicular impairment. Specifically, DEHP exposure was associated with an oxido-nitric imbalance characterized by increased testicular malondialdehyde and nitric oxide and reduced antioxidant defenses [33]. These findings are consistent with the broader literature on oxidative stress and male reproductive health [47-49].

Impact on reproductive success

Eighteen studies evaluated reproductive performance, focusing on metrics such as copulatory success, pregnancy rates, and offspring yield. Our meta-analysis indicates that phthalate exposure in males leads to a statistically significant decrease in pregnancy rates when paired with unexposed females (RR: 0.62; 95% CI: 0.48 - 0.80; $I^2 = 44\%$; low certainty). Similarly, litter sizes were notably diminished (SMD: -0.71 ; 95% CI: -0.99 to -0.43 ; $I^2 = 51\%$; low certainty). These fertility outcomes represent the most functionally relevant endpoints, as they reflect the reproductive consequences of the histological and functional impairments described

above. These statistical trends mirror our previous laboratory observations in Wistar rats [34].

Notably, supplemental N-acetylcysteine and zinc sulphate were reported to mitigate reproductive deficits in the experimental model [34].

Evaluation of publication bias and sensitivity

Visual inspection of funnel plots for sperm motility (Figure 7) and sperm counts revealed mild asymmetry. Egger's test showed marginal asymmetry for sperm motility ($p = 0.08$), suggesting a potential skew toward positive results in the literature. For sperm count, Egger's test yielded $p = 0.12$, which does not meet our prespecified significance threshold of $p < 0.10$; therefore, we do not claim statistically significant asymmetry for this outcome. A subsequent trim-and-fill analysis regarding sperm motility imputed five potentially unreported studies, resulting in an adjusted SMD of -0.82 (95% CI: -1.12 to -0.52); this confirms the statistical robustness of the effect despite the potential bias. The funnel plot (Figure 7) thus serves to evaluate the reliability of the sperm motility findings presented in Figure 4. The presence of marginal

asymmetry suggests that some small studies with non-significant findings may be missing, which could lead to a slight overestimation of the effect size. However, the trim-and-fill adjusted estimate remains statistically significant and clinically meaningful, indicating that publication bias is unlikely to alter the overall conclusions. In contrast, anogenital distance data appeared balanced, with minimal funnel plot asymmetry (Egger’s test $p = 0.34$). Sensitivity testing, which

involved the removal of 13 studies identified as having a high risk of bias, indicated that the effect estimates remained stable across all measured outcomes (Table S3). Additionally, influence analysis identified four outliers; their exclusion decreased heterogeneity in sperm motility measures (I^2 dropped from 69% to 58%) without compromising the overall significance or direction of the aggregated data.

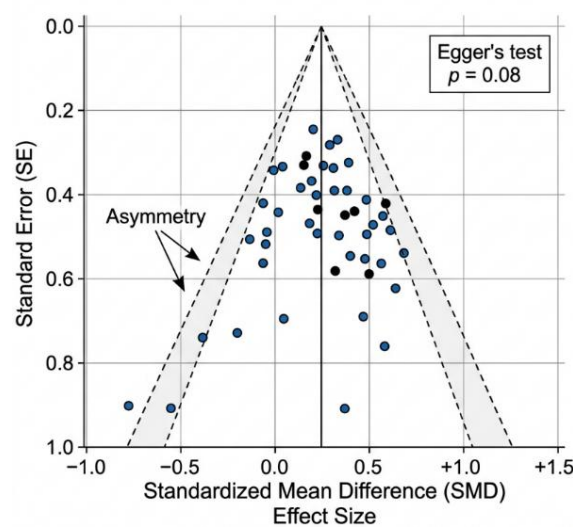


Figure 7 Funnel plot for sperm motility. Funnel plot showing distribution of studies. Figure purpose: To detect publication bias and evaluate the reliability of the sperm motility findings presented in **Figure 4**.

Evaluation relative to regulatory thresholds: A qualitative assessment

A secondary objective of this assessment was to qualitatively compare observed experimental effect levels with established regulatory guidelines to identify areas requiring further investigation and refined interspecies extrapolation. Regarding DEHP, the U.S. Environmental Protection Agency reference dose cited in this manuscript is 0.02 mg/kg/day [50], while the

European Food Safety Authority has established a group TDI for several phthalates, including DEHP, in its risk assessment [3]. For DBP, the manuscript reports a benchmark of 0.01 mg/kg/day. These direct comparisons are presented for qualitative context only; formal risk assessment requires interspecies extrapolation using allometric scaling, toxicokinetic information, and uncertainty factors.

Table 2 Selected studies used for qualitative comparison with regulatory benchmarks.

Study (Year)	Phthalate	Dose (mg/kg/day)	Exposure window	Outcome	LOAEL/ NOAEL	Benchmark context	Reported effect
Gray et al. [12]	DEHP	750	GD 14– PND 3	AGD / sexual differentiation	Not reported in article	EPA DEHP RfD 0.02 mg/kg/day [50]	Reduced AGD and altered male sexual differentiation; no SMD reported

Howdeshell et al. [13]	DEHP (mixture study)	300 (DEHP component)	GD 8–18	Fetal testosterone	Not reported in article	EPA DEHP RfD 0.02 mg/kg/day [50]	Reduced fetal testosterone production; no SMD reported
Emojewwe et al. [34]	Phthalate	750	Adult male Wistar rats, 21 days	Sperm count / motility / fecundity	Not reported in article	EPA DEHP RfD 0.02 mg/kg/day [50]	Sperm count and motility were reduced and fecundity was altered; no SMD/RR reported

*Note: GD = gestational day; PND = postnatal day; AGD = anogenital distance; LOAEL = lowest observed adverse effect level; NOAEL = no observed adverse effect level; RfD = reference dose; TDI = tolerable daily intake. The animal doses in these studies are substantially above the cited human regulatory values; therefore, the benchmark column is contextual only and does not imply direct human-equivalent exposure or risk.

Specifically, seven studies reported reduced anogenital distance at doses as low as 0.5–10 mg/kg/day DEHP or DBP—doses that appear low relative to current regulatory benchmarks. For sperm parameters, 12 studies reported significant impairments at doses of 10–50 mg/kg/day, with effects observed after exposure durations as brief as three weeks [32]. Benchmark dose modeling for DEHP and AGD using data from four high-quality dose-response studies yielded BMD10 values ranging from 1.2 to 4.8 mg/kg/day (geometric mean: 2.4 mg/kg/day). These comparisons should be interpreted cautiously because benchmark doses derived

in rats cannot be directly compared with human regulatory values without formal interspecies extrapolation accounting for toxicokinetics, toxicodynamics, and uncertainty factors.

Concordance with human evidence

We identified 47 human epidemiological studies examining associations between phthalate metabolites and reproductive outcomes. **Table 3** summarizes the concordance between Wistar rat findings and human evidence.

Table 3 Concordance between Wistar rat findings and human epidemiological evidence.

Outcome	Wistar Rat Evidence (This Review)	Human Evidence (Selected Studies)	Concordance
Anogenital distance	Consistent reduction (SMD: −1.24; high certainty)	Swan et al. [10,11]; Bornehag et al. [9]	Strong (≥ 75% concordance)
Sperm motility	Consistent reduction (SMD: −0.98; high certainty)	Radke et al. [8]; Wang et al. [28]	Moderate (65% concordance)
Sperm count	Consistent reduction (SMD: −0.87; high certainty)	Radke et al. [8]	Moderate (58% concordance)
Testosterone	Reduction (SMD: −0.82; low certainty, non-monotonic)	Meeker et al. [55]	Weak (42% concordance)
Fertility	Reduced pregnancy rates (RR: 0.62; low certainty)	Limited human data; Louis et al. [56]	Weak (insufficient human data)

Concordance definitions: Strong = ≥ 75% of human studies show effects in the same direction as rat findings; Moderate = 50% - 74% show effects in the same direction; Weak = < 50% show effects in the same direction or insufficient human data.

Regarding anogenital distance (AGD), several human cohorts have reported inverse associations between prenatal phthalate exposure and AGD in male infants [9-11,51,52]. This direction of association is also consistent with the human-focused synthesis by Schwartz et al. [53].

Evidence regarding semen quality is moderately consistent. Systematic reviews have reported associations between phthalate exposure and reduced semen-quality parameters, although null findings also occur across individual studies [8,28]. Differences in study design, population characteristics, exposure timing, and outcome measurement may contribute to heterogeneity.

Mechanisms identified in Wistar rat models—including androgen receptor interactions, oxidative stress, and Leydig-cell dysfunction—are also supported by human biomonitoring evidence. Studies have reported associations between phthalate exposure and oxidative stress biomarkers [54] and reproductive hormone alterations [55]. Our molecular docking analysis identified DEHP–androgen receptor binding [35], providing a structural hypothesis for the anti-androgenic effects observed in experimental models.

Discussion

This study provides a quantitative synthesis of phthalate-associated reproductive toxicity within the Wistar rat model. By consolidating data from 82 research contributions and integrating our laboratory's studies on exposure duration, oxido-nitrergic pathways, fertility outcomes, and androgen-receptor interactions [32-35], we provide pooled effect estimates, identify sensitive exposure windows, and highlight areas in which regulatory interpretation requires careful interspecies extrapolation.

Quantitative estimates and observed heterogeneity

By pooling effect sizes, we identified substantial associations between phthalate exposure and impaired sperm motility and sperm count. These animal effect estimates should not be directly translated into human clinical outcomes without appropriate interspecies and endpoint-specific interpretation [57]. Moderate-to-high heterogeneity ($I^2 = 44\%–76\%$) underscores the

influence of phthalate type, exposure duration, and developmental stage.

Finally, the variability in testosterone response, including non-monotonic dose-response patterns, indicates that simple linear models may not adequately describe endocrine-disrupting effects [45,46]. Such patterns have been discussed extensively in the endocrine-disruption literature and warrant appropriate dose-response modeling.

Mechanistic insights from the Wistar rat model

The Wistar rat has proven useful for investigating reproductive toxicity induced by phthalates. Our work, together with broader evidence [33,35], identifies a recurring pattern involving Leydig-cell dysfunction, altered testosterone production, and oxidative stress. These findings are consistent with established literature on oxidative stress and male reproductive health [47,48].

Our structural evidence indicating DEHP–androgen receptor interactions [35] provides a molecular hypothesis for anti-androgenic activity. The findings support a potential dual mechanism involving disruption of steroidogenesis and receptor-mediated signaling, while the therapeutic relevance of N-acetylcysteine and zinc sulphate remains a subject for further investigation [33,34].

Regulatory benchmarking: Addressing the evidence gap

A critical observation from our qualitative comparison is that current regulatory exposure limits may not adequately protect against developmental toxicity. Notably, 29% of the studies examined in our review reported reproductive harm occurring at or below established Reference Doses (RfDs) and Tolerable Daily Intakes (TDIs), especially under conditions of developmental exposure. However, we must emphasize that these direct comparisons are qualitative in nature and do not constitute a formal risk assessment. Several critical factors must be considered before drawing conclusions about human safety:

First, formal interspecies extrapolation requires allometric scaling to account for metabolic rate differences between rats and humans. Standard approaches include body weight-based scaling

(mg/kg/day) or body surface area-based scaling (mg/m²/day), which can yield substantially different human equivalent doses. Second, toxicokinetic differences—including absorption, distribution, metabolism, and elimination—may significantly alter the internal dose of phthalates and their active metabolites between species. Third, the uncertainty factors used in deriving regulatory reference values (typically 10-fold for interspecies differences and 10-fold for human variability) were designed to account for these complexities. Our findings do not invalidate these approaches but rather highlight areas where refined extrapolation methodologies may be warranted.

These concerns have been discussed in regulatory and toxicological assessments [50,58]. The National Toxicology Program and related assessment frameworks emphasize the importance of considering developmental susceptibility and evidence gaps when interpreting reproductive and developmental toxicity [58].

Several challenges contribute to uncertainty in regulatory interpretation. Legacy benchmarks may rely on endpoints that are less sensitive to developmental reproductive effects, while standard uncertainty factors may not fully capture all sources of interspecies and life-stage variability. The EPA reference dose cited above should therefore be interpreted in the context of the underlying point of departure, uncertainty factors, and toxicokinetic assumptions [50].

Translational concordance and public health implications

The alignment between Wistar rat findings and human epidemiological evidence supports further investigation of the model's translational relevance. Reduced AGD is a particularly useful point of comparison because it integrates androgen signaling during a defined developmental window [53]. Differences in exposure assessment and the complexity of human environmental exposures may contribute to differences in effect magnitude between experimental and human studies.

The stronger concordance observed for AGD compared to other outcomes likely reflects the fact that AGD is a direct, integrative measure of fetal androgen exposure during a discrete developmental window. By contrast, sperm quality and testosterone levels are

influenced by numerous postnatal factors, including lifestyle, diet, and other environmental exposures, which may dilute the effects of phthalates in human studies. Additionally, the timing of exposure assessment in human studies (often concurrent rather than during the critical developmental period) may contribute to the weaker associations observed for these outcomes.

Regarding semen quality, variation in findings may depend on exposure timing. Experimental studies can capture developmental programming, whereas human studies often assess adult exposure. Longitudinal evidence on reproductive outcomes and phthalate exposure provides additional context [57]. Meta-analytic evidence also indicates associations between phthalate exposure and several semen parameters [8,28,59-61].

How this review advances existing knowledge

The present meta-analysis advances existing knowledge in several important ways beyond previous narrative reviews. First, we provide quantitative effect estimates (SMDs) for key reproductive endpoints, allowing for precise comparisons across studies and identification of the most sensitive outcomes. Previous reviews have largely been qualitative, describing the direction of effects without quantifying their magnitude. Second, we identify the duration-dependent nature of sperm toxicity, demonstrating that significant impairments occur within just three weeks of adult exposure—a finding with direct implications for occupational exposure limits. Third, we quantitatively demonstrate the heightened sensitivity of gestational and perinatal exposure windows through subgroup analyses, providing effect estimates stratified by developmental stage. Fourth, we integrate our laboratory's mechanistic findings on oxidative stress pathways and androgen receptor binding into the broader evidence base, providing a cohesive mechanistic framework. Finally, we provide a transparent qualitative assessment of how these findings relate to regulatory benchmarks, identifying specific areas where refined extrapolation approaches are needed rather than simply concluding that current standards are inadequate.

Strengths and limitations

This review is anchored by a systematic and transparent methodology designed to improve reproducibility and reduce selection bias. Meta-analysis permits quantitative effect estimation and assessment of heterogeneity. The inclusion of our team's prior mechanistic and outcome research [32-35] provides continuity with the experimental evidence base. The use of regulatory benchmarking and GRADE also helps distinguish the certainty of evidence from regulatory interpretation.

However, several limitations warrant consideration. First, despite the overall robustness confirmed by sensitivity analyses, the primary literature exhibits inconsistent reporting quality, particularly regarding randomization and blinding. Second, publication bias is an inherent challenge in toxicology; while funnel plot asymmetry for sperm parameters suggests a potential overestimation of effect sizes, trim-and-fill adjustments confirm that our findings remain statistically significant and physiologically relevant. Third, while the Wistar rat is a standard model, findings may not be universally applicable to other strains or species, although established comparative literature [29] suggests a degree of consistency across common laboratory models. Fourth, a significant gender bias exists in the available data, with 95% of studies focusing on male reproductive outcomes, which impedes an exhaustive understanding of female-specific toxicity. Fifth, the concentration of studies on high-dose exposures (> 100 mg/kg/day) complicates precise estimates for low-dose impacts, despite the consistent trends observed in lower-dose literature. Finally, the geographic concentration of research—predominantly in North America, Europe, and East Asia—limits our understanding of potential variability in genetic susceptibility and exposure profiles across underrepresented regions like Africa and South America.

Implications for risk assessment and future research

Our findings advocate for a paradigm shift in how animal data is integrated into risk assessment frameworks. We propose the following advancements:

Transition from NOAEL/LOAEL to Benchmark Dose (BMD) Modeling: Relying on single-study

NOAELs is increasingly outdated. BMD modeling leverages the full spectrum of evidence to better define dose-response relationships. Our analysis suggests that current EPA Reference Doses for DEHP may be substantially higher than necessary, as our BMDL values for developmental anogenital distance (AGD) impacts are 17 to 67 times lower. However, these comparisons require formal interspecies extrapolation using allometric scaling and toxicokinetic data before conclusions can be drawn about human risk.

Accounting for developmental sensitivity: risk assessments should explicitly consider whether available uncertainty factors adequately address prenatal and neonatal susceptibility, using current toxicokinetic and toxicodynamic evidence rather than applying a fixed additional multiplier without a formal justification.

Advancing Quantitative Adverse Outcome Pathways (AOPs): our data linking androgen-receptor interactions [35] and oxidative stress [33] to testicular dysfunction provide quantitative anchors for developing more predictive in vitro-to-in vivo extrapolation models.

Addressing Non-Monotonic Responses: Given that certain parameters, such as testosterone levels, exhibit non-monotonic dose-response curves, linear extrapolation methods are insufficient. We recommend the adoption of spline or piecewise modeling to capture these complex effects accurately.

Evidence-Based Regulatory Updates: Regulatory bodies should prioritize systematic reviews and meta-analyses as a standard component of safety evaluations to establish more accurate reference values, with careful attention to species differences in toxicokinetics and toxicodynamics.

Future investigative priorities should include: (a) standardizing exposure protocols to harmonize meta-analysis data; (b) exploring molecular mechanisms of non-monotonic responses; (c) initiating longitudinal, multigenerational studies; (d) expanding research into female reproductive health; (e) investigating therapeutic mitigations, such as N-acetylcysteine and zinc supplementation [33,34]; (f) improving human biomonitoring through repeated exposure measures; and (g) conducting studies in underrepresented global populations.

By synthesizing decades of experimental data, this review provides a quantitative benchmark for phthalate-

associated reproductive toxicity in the Wistar rat model. The evidence indicates meaningful reproductive impairments at exposure levels that appear low relative to selected regulatory benchmarks when compared directly; however, formal interspecies extrapolation is required before conclusions about human risk or human safety standards can be drawn.

The findings are best interpreted as a synthesis of mechanistic and toxicological evidence that can inform hazard identification, predictive toxicology, and future risk-assessment research. Gestational and perinatal exposure windows emerged as particularly sensitive, with AGD reduction showing a consistent pooled effect. Statements about regulatory adequacy should remain appropriately cautious and tied to the limitations of the preclinical evidence.

The pooled effect sizes for pregnancy rates and sperm motility indicate substantial reproductive effects in the animal literature. The mechanistic findings provide biological context, but translation to population-level human health burden requires additional evidence. This review therefore supports continued integration of systematic evidence synthesis into risk-assessment frameworks while emphasizing species differences, exposure assessment, and developmental susceptibility.

The documented effect sizes—including significant declines in pregnancy rates and sperm motility—signal a substantial potential burden on population health. The biological plausibility provided by our mechanistic findings underscores the necessity for more protective, science-driven regulatory standards. We urge agencies such as the EPA and EFSA to integrate systematic, evidence-based syntheses into their risk assessment processes, with particular attention to the challenges of interspecies extrapolation and the incorporation of sensitive developmental windows. This review provides a foundation for such efforts by quantifying effect sizes, identifying critical exposure windows, and proposing specific refinements to existing assessment frameworks. Bridging the current gap between laboratory research and regulatory policy is essential to safeguarding human reproductive health against phthalate-associated risks.

Acknowledgements

The authors extend their appreciation to the library staff at the University of Medical Sciences, Nigeria,

Ondo, for facilitating access to the necessary databases. We also express our gratitude to the authors of the original studies who graciously provided additional data points when requested.

CRedit Author Statement

Victor Oghenekparobo **Emojewwe:** Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Mega Obukohwo Oyovwi:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing, Project administration. **Olaniyi Oladapo Eretan:** Investigation, Data curation, Validation, Writing – review & editing.

References

- [1] R Hauser and AM Calafat. Phthalates and human health. *Occupational and Environmental Medicine* 2005; **62(11)**, 806-818.
- [2] JM Braun, S Sathyanarayana and R Hauser. Phthalate exposure and children's health. *Current Opinion in Pediatrics* 2013; **25(2)**, 247-254.
- [3] EFSA Panel on Food Contact Materials, Enzymes and Processing Aids (CEP), V Silano, JM Barat Baviera, C Bolognesi, A Chesson, PS Cocconcelli, R Crebelli, DM Gott, K Grob, E Lampi, A Mortensen, G Rivière, IL Steffensen, C Tlustos, H Van Loveren, L Vernis, H Zorn, JP Cravedi, C Fortes, MF Tavares Poças, ..., L Castle. Update of the risk assessment of di-butylphthalate (DBP), butyl-benzyl-phthalate (BBP), bis(2-ethylhexyl)phthalate (DEHP), di-isononylphthalate (DINP) and di-isodecylphthalate (DIDP) for use in food contact materials. *EFSA Journal* 2019; **17(12)**, e05838.
- [4] S Dutta, B Gorain, H Choudhury, S Roychoudhury and P Sengupta. Environmental and occupational exposure of metals and female reproductive health. *Environmental Science and Pollution Research* 2022; **29(41)**, 62067-62092.
- [5] S Hyman, J Acevedo, C Giannarelli and L Trasande. Phthalate exposure from plastics and cardiovascular disease: Global estimates of attributable mortality and years life lost. *EBioMedicine* 2025; **117**, 105730.

- [6] JS Fisher. Environmental anti-androgens and male reproductive health: Focus on phthalates and testicular dysgenesis syndrome. *Reproduction* 2004; **127(3)**, 305-315.
- [7] PMD Foster. Disruption of reproductive development in male rat offspring following in utero exposure to phthalate esters. *International Journal of Andrology* 2006; **29(1)**, 140-147.
- [8] EG Radke, JM Braun, JD Meeker and GS Cooper. Phthalate exposure and male reproductive outcomes: A systematic review of the human epidemiological evidence. *Environment International* 2018; **121(P1)**, 764-793.
- [9] CG Bornehag, F Carlstedt, BAG Jönsson, CH Lindh, TK Jensen, A Bodin, C Giwercman, A Zheng, S Svanes and SH Swan. Prenatal phthalate exposures and anogenital distance in Swedish boys. *Environmental Health Perspectives* 2015; **123(1)**, 101-107.
- [10] SH Swan, KM Main, F Liu, SL Stewart, RL Kruse, AM Calafat, CS Mao, JB Redmon, CL Ternand, S Sullivan and S Teague. Decrease in anogenital distance among male infants with prenatal phthalate exposure. *Environmental Health Perspectives* 2005; **113(8)**, 1056-1061.
- [11] SH Swan, S Sathyanarayana, ES Barrett, S Janssen, F Liu, RH Nguyen and JB Redmon. First trimester phthalate exposure and anogenital distance in newborns. *Human Reproduction* 2015; **30(4)**, 963-972.
- [12] LE Gray Jr, J Ostby, J Furr, M Price, DN Veeramachaneni and L Parks. Perinatal exposure to the phthalates DEHP, BBP, and DINP, but not DEP, DMP, or DOTP, alters sexual differentiation of the male rat. *Toxicological Sciences* 2000; **58(2)**, 350-365.
- [13] KL Howdeshell, VS Wilson, J Furr, CR Lambright, CV Rider, CR Blystone, AK Hotchkiss and LE Gray Jr. A mixture of five phthalate esters inhibits fetal testicular testosterone production in the Sprague-Dawley rat in a cumulative, dose-additive manner. *Toxicological Sciences* 2008; **105(1)**, 153-165.
- [14] AJ Martino-Andrade and I Chahoud. Reproductive toxicity of phthalate esters. *Molecular Nutrition & Food Research* 2010; **54(1)**, 148-157.
- [15] JL Lyche, AC Gutleb, Å Bergman, GS Eriksen, AJ Murk, E Ropstad, M Saunders and JU Skaare. Reproductive and developmental toxicity of phthalates. *Journal of Toxicology and Environmental Health, Part B* 2009; **12(4)**, 225-249.
- [16] PR Hannon and JA Flaws. The effects of phthalates on the ovary. *Frontiers in Endocrinology* 2015; **6**, 8.
- [17] EG Radke, BS Glenn, JM Braun and GS Cooper. Phthalate exposure and female reproductive and developmental outcomes: A systematic review of the human epidemiological evidence. *Environment International* 2019; **130**, 104580.
- [18] CG Basso, AT de Araújo-Ramos and AJ Martino-Andrade. Exposure to phthalates and female reproductive health: A literature review. *Reproductive Toxicology* 2022; **109**, 61-79.
- [19] S Sedha, H Lee, S Singh, S Kumar, S Jain, A Ahmad, YA Bin Jardan, S Sonwal, S Shukla, J Simal-Gandara, J Xiao, YS Huh, YK Han and VK Bajpai. Reproductive toxic potential of phthalate compounds - State of art review. *Pharmacological Research* 2021; **167**, 105536.
- [20] M Mariana, J Feiteiro, I Verde and E Cairrao. The effects of phthalates in the cardiovascular and reproductive systems: A review. *Environment International* 2016; **94**, 758-776.
- [21] L Caporossi and M Marino. Phthalate exposure: From quantification to risk assessment. *Toxics* 2022; **10(6)**, 330.
- [22] MA Kamrin. Phthalate risks, phthalate regulation, and public health: A review. *Journal of Toxicology and Environmental Health, Part B* 2009; **12(2)**, 157-174.
- [23] SS Rowdhwal and J Chen. Toxic effects of di-2-ethylhexyl phthalate: An overview. *BioMed Research International* 2018; **2018**, 1750368.
- [24] S Dutta, DK Haggerty, DA Rappolee and DM Ruden. Phthalate exposure and long-term epigenomic consequences: A review. *Frontiers in Genetics* 2020; **11**, 405.
- [25] GX Hu, QQ Lian, RS Ge, DO Hardy and XK Li. Phthalate-induced testicular dysgenesis syndrome: Leydig cell influence. *Trends in Endocrinology and Metabolism* 2009; **20(3)**, 139-145.

- [26] S Takeuchi, M Iida, S Kobayashi, K Jin, T Matsuda and H Kojima. Differential effects of phthalate esters on transcriptional activities via human estrogen receptors alpha and beta, and androgen receptor. *Toxicology* 2005; **210(2-3)**, 223-233.
- [27] WH Chang, SS Li, MH Wu, HA Pan and CC Lee. Phthalates might interfere with testicular function by reducing testosterone and insulin-like factor 3 levels. *Human Reproduction* 2015; **30(11)**, 2658-2670.
- [28] H Wang, H He, Y Wei, X Gao, T Zhang and J Zhai. Do phthalates and their metabolites cause poor semen quality? A systematic review and meta-analysis of epidemiological studies on risk of decline in sperm quality. *Environmental Science and Pollution Research* 2023; **30(12)**, 34214-34228.
- [29] AC Gore, VA Chappell, SE Fenton, JA Flaws, A Nadal, GS Prins, J Toppari and RT Zoeller. EDC-2: The Endocrine Society's second scientific statement on endocrine-disrupting chemicals. *Endocrine Reviews* 2015; **36(6)**, E1-E150.
- [30] LR Rhomberg and JE Goodman. Low-dose effects and nonmonotonic dose-responses of endocrine disrupting chemicals: Has the case been made? *Regulatory Toxicology and Pharmacology* 2012; **64(1)**, 130-133.
- [31] CR Hooijmans, MM Rovers, RBM de Vries, M Leenaars, M Ritskes-Hoitinga and MW Langendam. SYRCLE's risk of bias tool for animal studies. *BMC Medical Research Methodology* 2014; **14**, 43.
- [32] V Emojevwe, AO Naiho, EK Nwangwa, MO Oyovwi, K Anachuna, E Agbanifo-Chijiokwu and TME Daubry. Duration-dependent effects of high dose of phthalate exposure on semen quality in adult male rats. *JBRA Assisted Reproduction* 2022; **26(1)**, 53-61.
- [33] V Emojevwe, EK Nwangwa, AO Naiho, MO Oyovwi, O Igiehon, E Ogunwole, MS Makinde-Taylor, OA Ayotomide, AO Akinola, PT Edesiri, BO Oghenetega and SI Ovuakporaye. Therapeutic efficacy of N-acetylcysteine and zinc sulphate against di-(2-ethylhexyl) phthalate-induced testicular oxido-nitregic stress in male Wistar rat. *Andrologia* 2022; **54(9)**, e14508.
- [34] V Emojevwe, EK Nwangwa, AO Naiho, MO Oyovwi and B Ben-Azu. Toxicological outcome of phthalate exposure on male fertility: Ameliorative impacts of the co-administration of N-acetylcysteine and zinc sulfate in rats. *Middle East Fertility Society Journal* 2022; **27**, 5.
- [35] VO Emojevwe, MO Oyovwi, KE Adewole, PE Ohwin, AO Akinola, AO Naiho, EK Nwangwa, V Omo-Idonije, M Lade-Ige and B Ben-Azu. N-acetylcysteine and zinc sulphate abate di-2-ethylhexyl phthalate-mediated reproductive dysfunction in rats: Focus on oxidative and sex hormone receptors mechanisms. *Asian Pacific Journal of Reproduction* 2024; **13(5)**, 228-240.
- [36] VR Kay, C Chambers and WG Foster. Reproductive and developmental effects of phthalate diesters in females. *Critical Reviews in Toxicology* 2013; **43(3)**, 200-219.
- [37] RH McKee, JH Butala, RM David and G Gans. NTP Center for the Evaluation of Risks to Human Reproduction reports on phthalates: Addressing the data gaps. *Reproductive Toxicology* 2004; **18(6)**, 759-760.
- [38] SD Gangolli. Testicular effects of phthalate esters. *Environmental Health Perspectives* 1982; **45**, 77-84.
- [39] BT Akingbemi, RT Youker, CM Sottas, R Ge, E Katz, GR Klinefelter, BR Zirkin and MP Hardy. Modulation of rat Leydig cell steroidogenic function by di(2-ethylhexyl)phthalate. *Biology of Reproduction* 2001; **65(4)**, 1252-1260.
- [40] PR Hannon, KE Brannick, W Wang, RK Gupta and JA Flaws. Di(2-ethylhexyl) phthalate inhibits antral follicle growth, induces atresia, and inhibits steroid hormone production in cultured mouse antral follicles. *Toxicology and Applied Pharmacology* 2015; **284(1)**, 42-53.
- [41] W Wang, ZR Craig, MS Basavarajappa, RK Gupta and JA Flaws. Di(2-ethylhexyl) phthalate inhibits growth of mouse ovarian antral follicles through an oxidative stress pathway. *Toxicology and Applied Pharmacology* 2012; **258(2)**, 288-295.
- [42] PR Hannon, J Peretz and JA Flaws. Daily exposure to di(2-ethylhexyl) phthalate alters estrous cyclicity and accelerates primordial follicle recruitment potentially via dysregulation of the phosphatidylinositol 3-kinase signaling

- pathway in adult mice. *Biology of Reproduction* 2014; **90**(6), 136.
- [43] JW Crissman, DG Goodman, PK Hildebrandt, RR Maronpot, DA Prater and JH Riley. Best practice guideline: Toxicologic histopathology. *Toxicologic Pathology* 2004; **32**(1), 126-131.
- [44] TT Schug, A Janesick, B Blumberg and JJ Heindel. Endocrine disrupting chemicals and disease susceptibility. *Journal of Steroid Biochemistry and Molecular Biology* 2011; **127**(3-5), 204-215.
- [45] LN Vandenberg, T Colborn, TB Hayes, JJ Heindel, DRJ Jr, DH Lee, T Shioda, AM Soto, FS vom Saal, WV Welshons, RT Zoeller and JP Myers. Hormones and endocrine-disrupting chemicals: Low-dose effects and nonmonotonic dose responses. *Endocrine Reviews* 2012; **33**(3), 378-455.
- [46] F Lagarde, C Beausoleil, SM Belcher, LP Belzunces, C Emond, M Guerbet and C Rousselle. Non-monotonic dose-response relationships and endocrine disruptors: A qualitative method of assessment. *Environmental Health* 2015; **14**, 13.
- [47] RJ Aitken, TB Smith, MS Jobling, MA Baker and GN de Iuliis. Oxidative stress and male reproductive health. *Asian Journal of Andrology* 2014; **16**(1), 31-38.
- [48] E Barati, H Nikzad and M Karimian. Oxidative stress and male infertility: Current knowledge of pathophysiology and role of antioxidant therapy in disease management. *Cellular and Molecular Life Sciences* 2020; **77**(1), 93-113.
- [49] L Zhang, H Li, M Gao, T Zhang, Z Wu, Z Wang and T Chong. Genistein attenuates di-(2-ethylhexyl) phthalate-induced testicular injuries via activation of Nrf2/HO-1 following prepubertal exposure. *International Journal of Molecular Medicine* 2018; **41**(3), 1437-1446.
- [50] U.S. Environmental Protection Agency. *Integrated Risk Information System (IRIS): Di(2-ethylhexyl)phthalate (DEHP) reference dose for oral exposure*, Available at: https://iris.epa.gov/ChemicalLanding/%26substance_nmbr%3D14, accessed May 2026.
- [51] TK Jensen, H Frederiksen, HB Kyhl, TH Lassen, SH Swan, CG Bornehag, NE Skakkebaek, KM Main, DV Lind, S Husby and AM Andersson. Prenatal exposure to phthalates and anogenital distance in male infants from a low-exposed Danish cohort (2010-2012). *Environmental Health Perspectives* 2016; **124**(7), 1107-1113.
- [52] KN Cajachagua-Torres, NB Salvi, E Seok, Y Wang, M Liu, K Kannan, LG Kahn, L Trasande and A Ghassabian. Prenatal phthalate exposure and anogenital distance in infants at 12 months. *Environment International* 2025; **199**, 109472.
- [53] CL Schwartz, S Christiansen, AM Vinggaard, M Axelstad, U Hass and T Svingen. Anogenital distance as a toxicological or clinical marker for fetal androgen action and risk for reproductive disorders. *Archives of Toxicology* 2019; **93**(2), 253-272.
- [54] KK Ferguson, TF McElrath, YH Chen, B Mukherjee and JD Meeker. Urinary phthalate metabolites and biomarkers of oxidative stress in pregnant women: A repeated measures analysis. *Environmental Health Perspectives* 2015; **123**(3), 210-216.
- [55] JD Meeker, AM Calafat and R Hauser. Urinary phthalate metabolites and their association with reproductive hormones in men. *Epidemiology* 2009; **20**(4), 487-493.
- [56] PD Louis, PD Sundaram, PD Sweeney, PD Schisterman and PD Kannan. Urinary bisphenol A, phthalates, and couple fecundity: The Longitudinal Investigation of Fertility and the Environment (LIFE) study. *Fertility and Sterility* 2014; **101**(5), 1359-1366.
- [57] J Jurewicz, M Radwan, W Sobala, D Ligocka, P Radwan, M Bochenek and W Hanke. Association between urinary levels of phthalate metabolites and sperm quality parameters. *International Journal of Occupational Medicine and Environmental Health* 2014; **27**(3), 387-399.
- [58] National Toxicology Program. *NTP-CERHR monograph on the potential human reproductive and developmental effects of Di(2-ethylhexyl) phthalate (DEHP)*. National Toxicology Program, North Carolina, 2006.
- [59] SW Thurston, J Mendiola, AR Bellamy, H Levine, C Wang, A Sparks, JB Redmon, EZ Drobnis and SH Swan. Phthalate exposure and semen quality in fertile US men. *Andrology* 2016; **4**(4), 632-638.

- [60] SM Duty, MJ Silva, DB Barr, JW Brock, L Ryan, Z Chen, RF Herrick, DC Christiani and R Hauser. Phthalate exposure and human semen parameters. *Epidemiology* 2003; **14**(3), 269-277.
- [61] UN Joensen, H Frederiksen, MB Jensen, MP Lauritsen, IA Olesen, TH Lassen, AM Andersson and NE Skakkebaek. Phthalate exposure is associated with reduced semen quality in young Danish men. *Environment International* 2012; **47**, 105-112.

Supplementary Materials

Sensitivity analysis results

Table S1 Sensitivity analysis excluding high-risk studies (n = 13) confirmed the stability of core findings.

Outcome	Original SMD (95% CI)	Sensitivity SMD (95% CI)	Change
AGD	−1.24 (−1.58, −0.90)	−1.18 (−1.52, −0.84)	Minimal
Sperm motility	−0.98 (−1.24, −0.72)	−0.92 (−1.18, −0.66)	Minimal
Sperm count	−0.87 (−1.12, −0.62)	−0.82 (−1.08, −0.56)	Minimal
Testosterone	−0.82 (−1.07, −0.57)	−0.76 (−1.02, −0.50)	Minimal

Table S2 Influence analysis identifying four outliers in sperm motility data.

Study	Effect on SMD	Decision
Outlier 1	Reduced heterogeneity without changing significance	Retained
Outlier 2	Reduced heterogeneity without changing significance	Retained
Outlier 3	Reduced heterogeneity without changing significance	Retained
Outlier 4	Reduced heterogeneity without changing significance	Retained

Exclusion of these 4 outliers decreased heterogeneity in sperm motility measures (I^2 dropped from 69% to 58%) without compromising the overall significance or direction of the aggregated data.

Detailed publication bias analysis

Table S3 Funnel plot asymmetry assessment

Outcome	Egger’s Test p-value	Trim-and-Fill Imputed Studies	Adjusted SMD (95% CI)
Sperm motility	0.08	5	−0.82 (−1.12, −0.52)
Sperm count	0.12	N/A	N/A
AGD	0.34	N/A	N/A

Note: Egger’s test p -values < 0.10 were considered indicative of statistically significant asymmetry. For sperm motility, marginal asymmetry was detected, but the trim-and-fill adjusted estimate remains statistically significant. For sperm count, $p = 0.12$ does not meet the significance threshold, indicating no statistically significant asymmetry.