

**Adult Organophosphate and Carbamate Insecticide Exposure and Sperm
Concentration: A Systematic Review and Meta-Analysis of the
Epidemiological Evidence**

Protocol and Amendments

Table of Contents

INTRODUCTION.....	3
Rationale.....	3
Review Question and Objectives	3
METHODS	4
Systematic Review Methodology	4
Review Team	4
Protocol Development and Protocol Amendments	5
Search Methods	5
Study Selection Criteria.....	6
Study Screening	7
Data Extraction	8
Risk of Bias	9
Statistical Analysis.....	10
Quality and Strength of Evidence Assessment	14
APPENDIX.....	15
Appendix I. Scientific database search terms.	15
Appendix II. Non-governmental organization website search strategies.	16
Appendix III. Title-abstract and full-text screening criteria.	17
Appendix IV. Data extraction form and guidance.....	19
Study Characteristics	Error! Bookmark not defined.
Meta-Analysis Eligibility	20
Study Results	21
Appendix V. Low risk of bias designation criteria for each domain of bias.....	21
Appendix VI. Meta-analysis screening criteria.	25

INTRODUCTION

Rationale

This review is motivated by the potential impact of widespread global pesticide use on male reproductive health, particularly sperm concentration, which evidence demonstrates has been in decline for the last near-century.¹⁻⁴ Sperm concentration is one of three conventional semen quality parameters used to diagnose idiopathic male factor infertility in clinical settings,⁵ in addition to being a broader indicator of general male health.⁶ Pesticides have been shown to directly damage the male testes and disrupt the body's endocrine system in many ways, leading to declines in sperm concentration.⁷

Given recent evidence of declining sperm concentration across the world, understanding the relationship between adult exposure to pesticides and sperm concentration is paramount. Existing systematic reviews published on the topic of pesticide exposure and sperm quality parameters (e.g., concentration, motility, and morphology) are qualitative in nature and spread out over time.⁸⁻¹⁰ Thus, a comprehensive systematic review and meta-analysis is needed to assess the qualitative and quantitative strength of epidemiological evidence of the association between pesticide exposure and sperm concentration.

Amendments made during the review

A supplemental rationale regarding the contemporary use and reproductive harms of organophosphate (OP) and N-methyl carbamate (NMC) insecticides was developed to support the review's ultimate focus on these compounds.

Review Question and Objectives

Research question: What is the association between adult exposure to pesticides and sperm concentration?

Objectives:

1. Identify observational studies conducted in humans concerning the association between adult pesticide exposure and sperm concentration;
2. Evaluate the evidence of an association across studies, including by conducting a meta-analysis;
3. Assess the risk of bias of individual studies and assess its impact on the observed associations; and
4. Rate the quality and strength of human evidence of an association between adult pesticide exposure and sperm concentration.

Amendments made during the review

In May 2021, the review team narrowed the scope of the research question from all pesticides to only organophosphate (OP) and N-methyl carbamate (NMC) insecticides given the high volume of studies deemed relevant when considering all pesticides and the desire to conduct a more focused and thorough analysis. OP and NMC insecticides were assessed together because (1)

both are contemporary use insecticides, (2) there is evidence of shared exposure measurement through cholinesterase monitoring, and (3) there is a wealth of studies available to assess.

METHODS

Systematic Review Methodology

This systematic review will be performed according to the Navigation Guide, a transparent and validated systematic review methodology designed to evaluate the quality and strength of evidence in an environmental and reproductive health context.^{11–13} Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 reporting guidelines will also be followed.¹⁴

Amendments made during the review

None.

Review Team & Contributors

A review team with expertise in pesticide exposure, reproductive epidemiology, toxicology, biostatistics, and systematic review methodology was established in October 2020. The review team consists mostly of members of Dr. Melissa Perry's research lab focused on pesticide exposure and reproductive epidemiology housed within the Department of Environmental and Occupational Health at the Milken Institute School of Public Health at the George Washington University in Washington, DC, USA. Dr. Daniele Mandrioli, Director of the Cesare Maltoni Cancer Research Center at the Ramazzini Institute in Bologna, Italy, joined the review team to provide expertise and guidance in pesticide toxicology and systematic review methodology. Dr. Daria Sgargi, a biostatistician with the Ramazzini Institute, joined the review team to provide expertise and guidance in meta-analysis.

Review team members and affiliations are listed below:

- **Lauren Ellis**, MPH, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington, DC, USA
- **C. Rebecca Robbins**, PA-C, MSc, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington, DC, USA
- **Marlaina Freisthler**, JD, MPH, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington, DC, USA
- **Daria Sgargi**, PhD, Cesare Maltoni Cancer Research Center, Ramazzini Institute, Bologna, Italy
- **Daniele Mandrioli**, MD, Cesare Maltoni Cancer Research Center, Ramazzini Institute, Bologna, Italy; Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy

- **Melissa J. Perry**, ScD, MHS, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington, DC, USA

Other review contributors are acknowledged below:

- **Danielly P. Magalhães**, PhD, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington, DC, USA
- **Suzanne Arrington**, MPH, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington, DC, USA

The first author (LBE) will perform the literature search(es), screen studies, extract data, and conduct statistical analyses. Three review authors (LBE, CRR, and MF) and two review contributors [one post-doctoral research fellow (DM) and one research assistant (SA)] will assess individual studies for risk of bias. Review contributors (DM and SA) will be acknowledged for their contributions to the risk of bias assessments in this review but will not be members of the publishing review team. All members of the review team (LBE, CRR, MF, DM, MJP) played a role in developing the protocol and will participate in evaluating the quality and strength of evidence, as well as writing and editing the manuscript.

Amendments made during the review

In May 2021, an additional review author [Karen Molina, affiliated with the George Washington University (GW)] joined the review team to conduct an independent secondary screen for and extract data from each study, helping to ensure the reproducibility of the systematic review methods and reduce bias in the review process. Also in May 2021, an additional review contributor (Erin Gomez, affiliated with GW) was recruited from Dr. Melissa Perry's lab to assist with the risk of bias assessments. In August 2022, Dr. Melissa Perry started a new position as Dean of the College of Public Health at George Mason University in Fairfax, Virginia, USA.

Protocol Development

This protocol is adapted from the protocol for Navigation Guide Case Study #5 on PBDEs and human neurodevelopment.¹⁵ Throughout the duration of the review, amendments to the protocol and rationales will be documented within each section of the protocol.

Amendments made during the review

None.

Search Methods

Data sources

Three scientific databases (PubMed, SCOPUS, and Web of Science) and five non-governmental organization (NGO) websites (Beyond Pesticides, National Pesticide Information Center, Pesticide Action Network North America, Collaborative on Health and the Environment, and

Environmental Working Group) will be searched. Database search results will be imported into Covidence, a systematic review management software,¹⁶ for screening. In contrast, the lead review author (LBE) will “hand-search”^a NGO website search results for unique records (titles-abstracts) prior to importation. To capture studies that may be missed in the search, the lead review author (LBE) will also hand-search the reference lists of all reports (full-text articles) sought for retrieval, including secondary literature sources. Records deemed potentially relevant during hand-searching will be checked against the existing Covidence report database and unique records will be imported into Covidence for screening.

Amendments made during the review

In May 2021, the review team added two government databases (NIOSHTIC-2, Science.gov) to the literature search to be as comprehensive as possible. Rather than importing all search results from government databases, the lead author (LBE) hand-searched government database search results and imported unique relevant records into Covidence for screening.

Search terms

Search terms from Knapke et al., 2022¹⁷ will be used to search the three scientific databases (PubMed, SCOPUS, and Web of Science).^b Search strategies used for NGO websites will be developed based on the scientific database search terms. Database search terms are presented in [Appendix I](#) and NGO website search strategies are presented in [Appendix II](#).

Amendments made during the review

In May 2021, following the addition of two government data sources, the review team developed government databases search terms based on the existing scientific database search terms and NGO website search strategies.

Study Selection

Study selection criteria and PECO framework

Primary epidemiology studies that examined a relationship between adult pesticide exposure and sperm concentration will be considered eligible for inclusion in this review. Studies published in or after 1991 will be considered eligible for inclusion in this review. Studies published in any language other than English will be considered eligible for inclusion in this review and translated to English using Google Translate prior to screening.

The following PECO (Population, Exposure, Comparator, and Outcomes) framework will aid in the selection of studies eligible for inclusion in the review:

Population: Adult (18 years or older) human males

^a According to [Chapter 5 of the Cochrane Handbook](#), “handsearching” refers to “page by page searches of a variety of sources including journals and supplements, conference proceedings and abstracts, and correspondence” to identify potentially relevant reports.

^b Knapke et al., 2022 was a publication of Dr. Melissa Perry’s research lab that was under development during the development of this protocol. The Medical Subject Headings (MeSH) database was used to find synonyms for pesticides and semen quality parameters, in addition to specific terms for different chemical groups and individual pesticides.

Exposure: Any type of pesticide

- Occupation-based, self-report, proxy, and bio-monitored exposure definitions;
- Documented, measured, or estimated exposure directly in study participants;
- Environmental and occupational exposure settings; and
- Non-acute exposure context

Comparator: Unexposed or less-exposed adult human males

- Irrespective of the statistical significance of examined differences in exposure levels between exposure and comparator groups, if applicable

Outcome: Sperm concentration

- Operationally defined as millions of sperm per milliliter of semen (million/mL)
- Measured on a continuous (million/mL) or dichotomous scale (e.g., clinical diagnosis of low sperm concentration, or oligospermia)

Amendments made during the review

In May 2021, the review team changed the exposure inclusion criterion from any type of pesticide to any type of OP or NMC insecticide based on the decision to narrow the research question. In April 2022, the review team removed the study publication date eligibility criterion after discovering relevant studies conducted prior to 1991; this change was enabled by additional author capacity to expand the search at the time of the decision.

Study Screening

Screening software and duplicates

Records (titles-abstracts) and reports (full-text articles) will be screened in Covidence, a systematic review management software.¹⁶ Duplicate records will be automatically identified and removed by the software, as well as manually marked as a duplicate and removed as needed.

Amendments made during the review

None.

Screening stages and processes

Stage 1: Title-abstract screening criteria

The lead review author (LBE) will independently screen each record (title-abstract) for eligibility according to prespecified eligibility criteria ([Appendix III](#)). Other review authors (CRR, MF, MJP) will be consulted as needed. If a record is inaccessible, the lead review author (LBE) will complete a brief online search before screening according to the title, erring on the side of over-inclusion at the title-abstract screening stage.

Records will be tagged as either “Include” or “Exclude” and proceed to full-text screening accordingly. Relevant secondary literature (e.g., reviews, book chapters, conference materials, commentaries) will be included at the title-abstract screening stage to reflect the fact that the respective full-text articles will indeed be screened and hand-searched for relevant primary studies that may have been missed in the search.

Stage 2: Full-text screening criteria

The lead review author (LBE) will independently screen each report (full-text article) for eligibility according to prespecified eligibility criteria ([Appendix III](#)). Other review authors (CRR, MF, MJP) will be consulted as needed. If a report is inaccessible, the lead review author (LBE) will attempt to contact the corresponding study author(s) via email to obtain the full-text article.

Full-text articles will be tagged as either “Include” (eligible) or “Exclude” (ineligible) according to one of the following exclusion reasons:

- Wrong population
- Wrong exposure
- Wrong comparator
- Wrong outcome
- Wrong study design

Only one exclusion tag will be applied to each study excluded at the full-text screening stage, based on the order in which the tags appear above (PECO+S hierarchy). Studies excluded as “Wrong study design” will be additionally categorized according to the type of secondary literature.

Amendments made during the review

After joining the review team in May 2021, KM began to independently screen all records and reports according to the pre-determined screening criteria. This allowed for duplicate screening and reduced the potential for personal biases to impact the results of the screening effort. Discrepancies were resolved via screener (LBE and KM) discussion; other review authors (CRR, MF, MJP) were consulted as needed. Also in May 2021, screeners tagged each report assessed for eligibility according to its pesticide class in the Covidence screening platform and then performed an additional third screening stage (OP/NMC screen) in which all reports included during full-text screening were further screened based on their respective pesticide class tags. Only studies that met the full-text screening criteria *and* examined an OP and/or NMC insecticide exposure were ultimately included in the review. In April 2022, following the review team’s decision to eliminate the publication date eligibility criterion, the following actions were taken to ensure that all studies, not just those published after 1991, were captured: (1) searched for studies published prior to 1991, and (2) revised the title-abstract and full-text screening criteria by removing the publication date criterion, and (3) re-screened excluded records and reports using the revised screening criteria. The revised screening criteria with no publication date eligibility criterion was applied prospectively moving forward. Final full-text and OP/NMC screening criteria is available in the supplemental material of the final publication (Table S4-Word & Table S5-Word)

Data Extraction

Multiple reports (full-text articles) of the same study, defined as having overlapping study participants and investigation of the same exposure, will be consolidated prior to data extraction and analysis. The lead review author (LBE) will extract data from included studies in three

stages according to prespecified data extraction criteria and guidance. Other review authors (CRR, MF, MJP) will be consulted as needed.

First, study characteristics ([Appendix IV](#)) will be extracted. Second, meta-analysis eligibility factors ([Appendix IV](#)) will be extracted to enable the identification of the most commonly reported effect size index (mean difference, odds ratio, correlation) across studies, which will directly inform the meta-analysis screening process. Third, following identification of the most commonly reported type of effect size across studies and meta-analysis inclusion screening, quantitative study results will be extracted from studies included in the meta-analysis ([Appendix IV](#)). Results from all relevant exposure definitions, doses, and sampling time periods will be included. If a study reports multiple effect size indices, the most adjusted index will be extracted. If a study reports multiple effect size indices with the same level of adjustment, the most commonly reported index across studies will be extracted to reduce the need for statistical transformations. If a study does not report results data needed for statistical analysis, the lead review author (LBE) will request those data from the corresponding study author(s) by email.

Amendments made during the review

A second review team member (KM) extracted data from included studies, so that all studies were extracted in duplicate. Discrepancies were resolved via extractor (LBE and KM) discussion; other review authors (CRR, MF, MJP) were consulted as needed. To facilitate data extraction consistency, a pilot data extraction process was created in which KM screened three studies to address nuances in the data extraction form. The review team also created a quality control step in which two review authors (LBE and KM) independently performed a quality control check by reviewing the final extracted data (including resolved discrepancies) for accuracy. Review authors were instructed to record any changes made during the quality control check; any changes made by review authors were finalized via discussion. Amendments and clarifications made to the data extraction form were recorded and reported in the supplemental material of the final publication (Table S6-Word). The review team considered attempts to contact study authors for data as unsuccessful if study author contact information was unavailable or led to an error message, or study authors did not respond to requests after being contacted through two email messages over the course of the review.

Risk of Bias

Risk of bias domains and criteria

Consistent with Navigation Guide systematic review methodology,¹² seven risk of bias domains will be evaluated for each study according to prespecified risk of bias criteria ([Appendix V](#)):

- (1) recruitment strategy,
- (2) blinding,
- (3) exposure assessment,
- (4) outcome assessment,
- (5) missing outcome data,
- (6) selective outcome reporting, and
- (7) conflict of interest.

Amendments made during the review

None.

Risk of bias assessment process

Two of five possible risk of bias assessors [three review authors (LBE, CRR, MF) and two review contributors (DM and SA)] will independently rate the risk of bias for each study included in the review according to prespecified risk of bias criteria ([Appendix V](#)). Discrepancies in risk of bias domain ratings will be resolved by a third independent assessor (one of five possible assessors listed above) and discussed as a group as needed. A training session will be held before initiating the risk of bias assessments to review and clarify nuances in the risk of bias criteria. Following the training, all risk of bias assessors will complete three pilot studies before independently assessing studies for risk of bias. The prespecified risk of bias criteria will be available as a reference to assessors during risk of bias assessments in an effort to improve inter-rater consistency.

Amendments made during the review

In May 2021, one additional review author (KM) and one additional review contributor (EG) joined the risk of bias assessment effort. Also in May 2021, the review team created a quality control step in which two review authors (LBE and KM) independently performed a quality control check by reviewing the final risk of bias assessment ratings (including resolved discrepancies) for accuracy. Any inaccuracies identified during the quality control check were recorded and discussed before finalizing the data.

Statistical Analysis

Effect size index for synthesis

The most commonly reported effect size index (e.g., mean difference, odds ratio, correlation) across included studies will be extracted and synthesized. The review team chose to synthesize the most commonly reported effect size index to include as many studies as possible and reduce the number of statistical transformations that may be needed.

Amendments made during the review

None.

Meta-analysis eligibility

Studies will be excluded from the meta-analysis if the effect size index chosen for synthesis is not reported or calculable (either due to selective outcome reporting or lack of relevant analysis) or cannot not be back-transformed from a statistically-transformed scale to raw scale. The complete meta-analysis screening criteria is presented in [Appendix VI](#).

Amendments made during the review

None.

Data preparation

If not directly reported, means and/or standard deviations will be calculated from reported data (e.g., median, range, interquartile range) according to Wan et al., 2014.¹⁸ The conversion

formulas presented in Bornstein & Hedges, 2009 will be used to convert between effect sizes indices,¹⁹ where possible.

For studies that report an adjusted linear beta coefficient from a regression model with continuous exposure and outcome variables (“continuous beta coefficient”) as the most adjusted effect size index, continuous beta coefficients will be homogenized to represent an absolute change in sperm concentration for every one-unit increase in exposure, as measured by the study, using methods presented in Rodriguez-Barranca et al., 2017.²⁰ After homogenizing beta coefficients, a partial regression coefficient will be calculated based on the homogenized t-statistic using methods from Aloe & Thompson, 2013.²¹

For studies that report an adjusted mean difference or a linear beta coefficient from a regression model with a binary exposure variable (“binary beta coefficient”) as the most adjusted effect size index, a “partial” standardized mean difference will be calculated according to the formulas presented in Aloe et al., 2022.²² While the review team will attempt to obtain unadjusted standard deviations or regression model R^2 values to remove the impact of covariates on partial (or adjusted) effect sizes, such data will not be required to include adjusted effect sizes in the meta-analysis. Log scale means and mean differences will be transformed to raw scale using methods presented in Higgins et al., 2008.²³ The R code used to calculate and convert between effect sizes will be made available to the public via GitHub at the time of publication.

Amendments made during the review

Where data was not available to impute missing means and/or standard deviations, missing data was imputed with the average value across relevant extracted results.

Meta-analysis model

Meta-analyses will be performed using three R packages: *metafor*,²⁴ *clubSandwich*,²⁵ and *dmatar*.²⁶ A three-level, multivariate random-effect meta-analysis model will be employed to account for both hierarchical and correlational dependencies in the data. The review team will assume a strong correlation ($r=0.8$) between dependent effect sizes originating from fully or partially overlapping study participants. Meta-analysis study weights will be calculated using the generic inverse variance method. A cluster-robust variance estimator with small sample size adjustments will also be employed to reduce the chance of a Type I error in hypothesis testing.²⁷

Heterogeneity statistics will be estimated using restricted maximum likelihood estimation (REML) for each of the three levels of random-effects in the model – (1) sampling error, (2) within-study population heterogeneity, and (3) between-study population heterogeneity.²⁸ Statistical significance of heterogeneity will be tested using Q-test statistic (Q_E) based on a chi-square distribution.²⁸ It should be noted that the REML heterogeneity point estimate (τ^2) is imprecise and can influence the pooled effect estimate, so a Q-profile confidence interval around the heterogeneity point estimates will be estimated.²⁹ Corresponding I^2 statistics will be calculated using functions available in *dmatar*,²⁶ representing the proportion of total variation that can be attributed to each level of random effect.

Amendments made during the review

None.

Primary meta-analysis

A primary meta-analysis will be performed across all eligible studies. Outlier effects will be identified via the Tukey's Fences method.³⁰ A prediction interval around the pooled effect estimate will be calculated to examine where the true effects would be expected for 95% of similar studies that may be performed in the future.³¹

Amendments made during the review

In addition to identifying outlier effects via the Tukey's Fences method, the review team decided to identify influential effects in the primary meta-analysis, defined as effects with a Cook's distance at least three times the mean Cook's distance across studies included in the meta-analysis.^{28,32} Cook's distance estimates represent the scaled change in fitted values resulting from the removal of each study (or other unit of analysis) from the model fitting.³²

Secondary meta-analyses

The following series of "data" sensitivity analyses will be performed to assess the robustness of meta-analytic model parameters, statistical imputations, and statistical transformations:

- *Rho Estimate*: Fix different rho values ranging from 0.2 to 0.95 to validate the robustness of the meta-analytic model to the assumed correlation between dependent effect sizes.
- *Heterogeneity Estimate*: For each level of random effect, fix the corresponding heterogeneity estimate to reflect the lower and upper bounds of the Q-profile confidence interval around the point estimate to validate the robustness of the meta-analytic model to the heterogeneity estimate.
- *Imputing Missing Mean/SD Data*: Remove results calculated from imputed mean or standard deviation data to examine the potential impact of such imputations on the pooled effect estimate.
- *Log to Raw Scale Transformation*: Remove results that been back-transformed from log scale to examine the potential impact of such transformations on the pooled effect estimate.
- *Effect Size Transformation*: Remove transformed effect sizes to examine the potential impact of such transformations on the pooled effect estimate.

The following series of "moderator" sensitivity analyses will be performed to explore potential sources of heterogeneity and aid in our interpretation and rating of the quality of evidence:

- *Risk of Bias*: Remove studies with high or probably high risk of bias in the three most affected domains, in addition to confounding (regardless of whether this domain showed a substantial degree of high or probably high risk of bias across studies), to examine the potential impact of such biases on the pooled effect estimate; the assessment of confounding will also shed light on the impact of pooling unadjusted and adjusted results in the same synthesis.
- *Medical Risk Factors*: Remove studies that did not control for medical risk factors for low sperm concentration, such as reproductive disorders or experience with medications known to adversely impact sperm, to assess the impact of this variable on the pooled effect estimate.
- *Abstinence Time*: Remove studies that did not control for abstinence time, a variable known to impact sperm concentration, to assess the impact of this variable on the pooled effect estimate.

- *Co-exposures*: Remove studies that did not control for known exposures to other pesticide or non-pesticide reproductive toxicants to examine the potential impact of co-exposures, or “indirectness” in the exposure definition, on the pooled effect estimate.

The following series of subgroup analyses will be performed to explore potential sources of heterogeneity and to test the hypothesis that the pooled effect estimate significantly differs by the following subgroups:

- *Pesticide Class*: Test the hypothesis that the pooled association significantly differs across different classes of pesticides with different mechanisms of toxicities.
- *Exposure Setting*: Test the hypothesis that men exposed in occupational settings may experience a greater adverse impact than men exposed in the environment, presumably due to higher exposure levels in occupational settings.
- *Recruitment Setting*: Test the hypothesis that men recruited from infertility clinics will show a greater association than men recruited from the general population, possibly due to higher rates of reproductive ailments that may impact sperm concentration.

Subgroup analysis rather than meta-regression will be employed based on the assumption that heterogeneity varies between subgroups.^{28,33} The review team will consider a given meta-analysis as having sufficient statistical power if there is at least four degrees of freedom, after Satterwhite-adjustment ($df_{\text{Satt}} > 4$), as simulations show Type I error rates are of concern below this cut-off.²⁷ If individual subgroup meta-analyses are statistically independent and have sufficient statistical power, the review team will compare the pooled effect estimates across subgroups using a fixed-effect meta-analysis model and a standard Wald-type X^2 test.²⁸ If these conditions are not met, the review team will qualitatively examine any differences in the pooled effect estimate across subgroups.

Amendments made during the review

The review team decided to test the impact of removing outlier and influential effects through an additional data sensitivity analysis. Following the decision to focus on OP and NMC studies in May 2021, the “pesticide class” subgroup analysis referred to the following categories: OP, NMC, and OP and/or NMC. Regarding the subgroup analysis, an F test rather than a standard Wald-type X^2 test was chosen to compare subgroup estimates as appropriate.³³

Publication bias

A modified Egger’s meta-regression test³⁴ with cluster-robust variance estimation will be used to estimate the impact of study precision on study-reported effect sizes, one source of publication bias. Statistical adjustments for publication bias will not be conducted because methods such as trim and fill are not yet available for multi-level, multivariate meta-analyses.

Amendments made during the review

A sample size-based precision estimate ($1/\sqrt{n}$) was chosen as the predictor rather than standard error to avoid known distortions in funnel plot assessments of standardized mean differences and standard errors.³⁶

Quality and Strength of Evidence Assessment

Quality and strength of evidence factors

Quality and strength of evidence will be assessed according to Navigation Guide systematic methodology.^{12,13} As such, the baseline quality of human evidence will be assumed to be “moderate.” This “moderate” quality rating may be modified down to “low” or up to “high” according to eight quality factors. The quality of evidence may be downgraded either one or two levels if the following factors are relevant: (1) substantial risk of bias across studies, (2) indirectness across studies, (3) inconsistency or imprecision in effect sizes, and/or (4) publication bias. The quality of evidence may be upgraded either one or two levels if the following factors are relevant: (1) large magnitude of effect, (2) consistent dose-response relationships, and/or (3) minimizing effect of confounding. If the quality factors do not apply, no change will be made from the original quality rating of “moderate.”

Strength of evidence will be assessed according to four considerations: (1) quality of the body of evidence, (2) direction of the pooled effect estimate, (3) confidence in the pooled effect estimate, and (4) other compelling attributes of the data that may influence certainty. Strength of evidence definitions developed by the Navigation Guide will be applied to reflect the level of certainty in relationship between the exposure and outcome under investigation.

Amendments made during the review

Specific criteria and considerations for each quality and strength of evidence factor were developed prior to the quality and strength of evidence assessment (see Table S9-Word & Table S11-Word of the supplemental material of the final publication).

Quality and strength of evidence assessment process

The lead review author (LBE) will rate the quality and strength of evidence. Next, entire review (LBE, CRR, MF, DM, MJP) team will meet to review, discuss, and finalize the quality and strength of evidence ratings and rationales. Individual and collective rating rationales will be recorded throughout the deliberation process.

Amendments made during the review

In May 2021, the review team modified the quality and strength of evidence assessment process so that assessments would be performed in duplicate (independently assessed by both LBE and KM).

APPENDIX

Appendix I. Scientific database search terms.

Database	Search terms
PubMed	((pesticides [MeSH] OR acaricides [MeSH] OR chemosterilants [MeSH] OR fungicides [MeSH] OR herbicides [MeSH] OR "insect repellents" [MeSH] OR insecticides [MeSH] OR organophosphate [tw] OR organochlorine [tw] OR carbamate [tw] OR pyrethroid [tw] OR triazine [tw] OR azole [tw] OR DDT [MeSH] OR DDT [tw] OR endosulfan [tw] OR mancozeb [tw] OR chlorpyrifos [tw] OR atrazine [tw] OR simazine [tw] OR penconazole [tw] OR permethrin [tw] OR deltamethrin [tw] OR bifenthrin [tw] OR "pirimiphos-methyl" [tw] OR parathion [tw] OR dichlorvos [tw] OR "azinphos-methyl" [tw] OR malathion [tw]) AND (sperm [tw] OR sperm quality [tw] OR sperm parameters [tw] OR spermatozoa [MeSH] OR semen quality [MeSH] OR semen quality analysis [MeSH] OR sperm count [MeSH] OR sperm motility [MeSH]))
SCOPUS	KEY (pesticides OR acaricides OR chemosterilants OR fungicides OR herbicides OR "insect repellent" OR insecticides OR organophosphate OR organochlorine OR carbamate OR pyrethroid OR triazine OR azole OR endosulfan OR mancozeb OR chlorpyrifos OR atrazine OR simazine OR penconazole OR permethrin OR deltamethrin OR DDT OR bifenthrin OR "pirimiphos-methyl" OR parathion OR dichlorvos OR "azinphos-methyl" OR malathion) AND KEY ("sperm quality" OR "sperm parameters" OR spermatozoa OR "semen quality" OR "semen quality analysis" OR "sperm count" OR "sperm motility")
Web of Science	ALL=((((pesticides OR acaricides OR chemosterilants OR fungicides OR herbicides OR "insect repellent" OR insecticides OR organophosphate OR organochlorine OR carbamate OR pyrethroid OR triazine OR azole OR endosulfan OR mancozeb OR chlorpyrifos OR atrazine OR simazine OR penconazole OR DDT OR permethrin OR deltamethrin OR bifenthrin OR "pirimiphos-methyl" OR parathion OR dichlorvos OR "azinphos-methyl" OR malathion) AND ("sperm quality" OR "sperm parameters" OR spermatozoa OR "semen quality" OR "semen quality analysis" OR "sperm count" OR "sperm motility"))))

Note: Scientific databases and search terms were adopted from Knapke et al. (2021). Search dates and filters were recorded and reported in the supplemental material of the final publication (Table S2-Word).

Appendix II. Non-governmental organization website search strategies.

NGO – website	Search strategy
Beyond Pesticides – https://www.beyondpesticides.org	<ol style="list-style-type: none"> 1. Click on the following web-tabs: Resources > Pesticide Induced Diseases Database > Sexual and Reproductive Dysfunction 2. Hand-search topic-specific reference list on pesticide exposure and sexual and reproductive dysfunction [here] <ul style="list-style-type: none"> • Command F “sperm” – search • Command F “semen” – search 3. Check potentially relevant records against Covidence database before importing
National Pesticide Information Center/Mother To Baby – http://npic.orst.edu	<ol style="list-style-type: none"> 1. Click on the following web-tabs: Health & Environment > Human Health > Pesticides and Pregnancy 2. On the Pesticides and Pregnancy webpage, find “Additional resources: Paternal Exposures and Pregnancy (OTIS)” 3. Link will redirect to “Mother To Baby factsheet” 4. Scroll down and click “Please click here to view references” 5. Hand-search topic-specific reference list on paternal exposures [here] 6. Check potentially relevant records against Covidence database before importing
Pesticide Action Network North America – http://www.panna.org	<ol style="list-style-type: none"> 1. Click on the following web-tabs: The Pesticide Problem > Human Health Harms > Reproductive Health > Pesticides: A key culprit 2. Hand-search embedded hyperlinks related to sperm outcomes on the ‘Pesticides: A key culprit’ webpage [here] 3. Check potentially relevant records against Covidence database before importing
Collaborative on Health and the Environment – https://www.healthandenvironment.org	<ol style="list-style-type: none"> 1. Click on the following web-tabs: Our Work > Toxicant and Disease Database 2. Filter the database by “Disease: Abnormal sperm (morphology, motility and sperm count)” 3. Click on the following relevant hyperlinks for all pesticide exposures, including individual compounds and general “pesticides” 4. Hand-search embedded reference lists 5. Check potentially relevant records against Covidence database before importing
Environmental Working Group – https://www.ewg.org	<ol style="list-style-type: none"> 1. Search “pesticides and sperm” in the search bar at the top right corner of the Home page 2. Hand-search reference list 3. Check potentially relevant records against Covidence database before importing

Appendix III. Title-abstract and full-text screening criteria.

PECO+S Domain	Screening stage and criteria
Population	<p><u>TITLE-ASBTRACT & FULL-TEXT</u></p> <p>Include: Humans; adult men; 18 years or older</p> <p>Exclude: Animals; women; men younger than 18 years old</p>
Exposures	<p><u>TITLE-ASBTRACT & FULL-TEXT</u></p> <p>Include: Exposure to any type of pesticide; exposure measured directly in each study participant; exposure status determined by occupation, self-report, proxy, environmental monitoring or bio-monitoring; environmental or occupational exposure setting; non-acute exposure events</p> <ul style="list-style-type: none"> Pesticide exposure must be isolated from other exposures, if other exposures are known or measured. For example, if a study measures pesticides and other chemicals in serum, pesticides must be isolated from other chemicals as an independent variable. <p>Exclude: Exposure to non-pesticide chemicals; group-level exposure definitions (e.g., ecological studies); exposure measured in someone other than study participant (e.g., mother or father), exposure that occurred in early life (i.e., in-utero, pre/neo/post-natal, or childhood), acute exposure events (e.g., industrial accidents or acts of warfare)</p>
Comparator	<p><u>TITLE-ABSTRACT & FULL-TEXT</u></p> <p>Include: Unexposed or less exposed adult men that <i>do not have</i> exclusive known or measured exposures to other reproductive toxicants; between-population comparison groups</p> <ul style="list-style-type: none"> “Exclusive” refers to exposures only known to impact the comparator group (i.e., not known or measured in exposed group) <p>Exclude: Unexposed or less exposed adult men that <i>have</i> exclusive known or measured exposures to other reproductive toxicants; within-population comparison groups (i.e., comparison of the same study participants at a different time point)</p>
Outcomes	<p><u>TITLE-ABSTRACT & FULL-TEXT</u></p> <p>Include: Sperm concentration (operational definition: million sperm / mL of ejaculate)</p> <ul style="list-style-type: none"> If the study <i>may</i> have examined sperm concentration, but does not specifically mention so in the title/abstract, tag as include (since many studies report semen analysis results, including sperm concentration, when looking at other sperm outcomes) to screen in more detail at the next stage <p>Exclude: Any other operational definition of the sperm concentration; other sperm parameters; reproductive hormone levels; sperm DNA integrity; chromosomal aberrations); pregnancy/birth outcomes; general male reproductive or fertility problems</p>

Study Design/ Publication Date	<p><u>TITLE-ABSTRACT</u></p> <p>Include: Primary epidemiology studies (e.g., cross-sectional, case-control or cohort); <i>secondary literature (e.g., reviews, meta-analyses, conference materials, response articles, comment articles, letters to the editor)</i></p> <ul style="list-style-type: none"> • To enable hand-searching of secondary literature, secondary literature (e.g., reviews, meta-analyses,) will be included during title-abstract screening, and then reviewed, categorized, and excluded during full-text screening based on “Wrong study design” (see below) <ul style="list-style-type: none"> ○ Will hand-search the reference lists of secondary literature excluded at the full-text screening stage to identify potentially relevant primary studies not captured in the literature search <p>Exclude: Toxicology/animal studies; in-vitro/mechanistic studies; controlled human experiments/interventions; ecological observational studies</p> <p><u>FULL-TEXT</u></p> <p>Include: Primary epidemiology studies (e.g., cross-sectional, case-control or cohort)</p> <p>Exclude: Toxicology/animal studies; in-vitro/mechanistic studies; controlled human experiments/interventions; ecological observational studies; <i>secondary literature (e.g., reviews, meta-analyses, conference materials, response articles, comment articles, letters to the editor)</i></p> <ul style="list-style-type: none"> • Categorize the study according to the type of secondary literature (using tags) and then exclude as “Wrong study design”
-----------------------------------	---

Note: Screening criteria is based on the PECO framework, and adds study design and publication date eligibility criteria. If a study meets two different exclusion criteria, will be excluded according to PECO hierarchy. Clarifications made during the screening effort were recorded and reported in the supplemental material of the final publication (Table S4-Word).

Appendix IV. Data extraction form and guidance.

Fields are free-form except where choices (in italics) are shown. “NR” indicates “not reported.” Study characteristics and meta-analysis eligibility factors will be extracted from all studies included in the systematic review, while study results will only be extracted for studies included in the meta-analysis.

STUDY CHARACTERISTICS

- **Short citation**
 - Format: First author’s last name (publication date)
- **Study design**: *Case-control, Cross-sectional, Cohort, Nested*
 - Nested indicates that the study took place within a larger cohort; if applicable, write name of cohort in parentheses following the study design
- **Case definition**
 - Format: “Case: <insert definition>; Control: <insert definition>”
 - Only applicable to case-control studies; write “NA” if not a case-control study
- **Country**
 - If multiple countries are represented, include all (separated by commas)
- **Study participants**
 - Format: [Number of total study participants] [descriptor] men
 - Extract the number of participants successfully recruited into or included in the study followed by a brief descriptor
- **Mean age [age range]**
 - Format: Mean age of total study population, or mean age of exposure or case groups separated by a slash (in order of exposure, high to low) [age range of total study population]
 - Example (one group): “33 [20-50]”
 - Example (two groups): “32/34 [20-50]”
- **Was abstinence time accounted for?** *Yes-Stats, Yes-Design, No*
 - Design = exclusion criterion or insignificant differences between exposure or case groups
 - Stats = tested for model change or included in the statistical model
- **Were medical risk factors accounted for?** *Yes-Stats, Yes-Design, No*
 - Design = exclusion criterion or insignificant differences between exposure or case groups
 - Stats = tested for model change or included in the statistical model
 - Examples of medical risk factors include: varicocele, cryptorchidism, genetic or endocrine disorders, urogenital tract infection, vasectomy, and medication associated with male fertility impairment (e.g., hormone treatments, radiotherapy, chemotherapy);
 - Not all examples have to be accounted for to be considered a “yes;” rely on author’s characterization if unclear; accounting for vasectomy alone is not sufficient for a “yes”
- **Medical risk factor rationale**
 - Quote directly from the study when possible
- **Place of recruitment**
 - Format: [Place of recruitment (population or clinic)] in [city, state, country]
 - Example: “[A large pesticide manufacturing plant and nearby textile factory] in [Anhui, China]” or “[College campus] in [Rochester, NY]”
- **Recruitment setting**: *Population, Clinic*
 - Clinic refers to infertility clinics; prenatal clinic is a population-based recruitment setting
- **Sampling period**
- **Power calculation**

- **Format:** Copy/paste quote directly from paper with quotes; if not reported, write “NR”
- **Pesticide class**
- **Exposure setting:** *Environmental, Occupational*
 - If both environmental and occupational exposures are present, categorize as occupational
- **Most advanced exposure assessment method:** *Self-report, Proxy, Biomonitoring*
 - Code according to the single most advanced assessment method utilized in the study, regardless of the exposure variable included in the analysis between exposure and outcome; it should match the exposure assessment method used to determine risk of exposure assessment bias
 - Exposure assessment methods in order of most to least advanced, for purposes of this review: biomonitoring, proxy, self-report
- **Corresponding exposure matrix(ces):** *Occupation, Questionnaire, Job-exposure-matrix (JEM), Food frequency questionnaire (FFQ), Dermal, Indoor air, Personal air, Urine, Blood, Serum, Plasma, Seminal fluid, Seminal plasma*
 - Corresponds to the most advanced exposure assessment method utilized in the study
- **Corresponding exposure definition(s) [LOD; LOQ; DF]**
 - Corresponds to the most advanced exposure assessment method and related matrix(ces) utilized in the study, regardless of whether the exposure(s) was(ere) assessed in relation to the outcome of interest (although the methods should lay out such intention)

META-ANALYSIS ELIGIBILITY

- **Most adjusted effect size index(ices):** *Means, Adj-Means, MD, Adj-MD, Beta-MD, OR, Beta-OR, R, Beta-Cont*
 - *Beta-MD:* Beta coefficient from a linear regression model that examined a binary exposure variable (group membership, yes/no) and a continuous outcome variable
 - *Beta-Cont:* Beta coefficient from a linear regression model that examined a continuous exposure variable (e.g., metabolite concentration) and a continuous outcome variable
 - *Beta-OR:* Beta coefficient from a logistic regression model that examined either a binary or continuous exposure variable and a binary outcome variable
- **Was the insecticide exposure variable statistically transformed?** *Yes-[type of transformation], No*
- **Was sperm concentration statistically transformed?** *Yes-[type of transformation], No*
- **Statistical transformation notes**
- **Is it possible to back-transform (Means/MD)?**
 - If results are already on raw scale, choose from the following options:
 - *NA*
 - *NA--no transformation*
 - *NA--results presented on raw scale*
 - If results are not on raw scale, choose from the following options:
 - *No--[insert type of transformation other than log or ln]*
 - *No--type of transformation NR*
 - *No--group-level or total log/geometric mean sperm concentration NR*
 - *Yes--group-level and/or total log mean sperm concentration reported*
- **Is it possible to back-transform (R/Beta-Cont)?**
 - If results are already on raw scale, choose from the following options:
 - *NA*
 - *NA--no transformation*
 - *NA--results presented on raw scale*

- If results are not on raw scale, choose from the following options:
 - *No--[insert type of transformation other than log or ln]*
 - *No--type of transformation NR*
 - *No--[group-level and/or total] arithmetic mean [exposure and/or sperm concentration] NR*
 - *Yes--[group-level and/or total] arithmetic mean [exposure and/or sperm concentration] reported*

STUDY RESULTS

Rules on choosing which effect size(s) to extract from each study, in priority order:

1. If a study reports multiple effect size indices (mean difference, odds ratio, correlation):
 - Extract the most adjusted index
 - The level of adjustment in the extracted effect size index must reflect the level of adjustment used to rate risk of confounding bias.
2. If a study reports multiple effect size indices with the same level of adjustment:
 - Extract the most commonly reported index across studies.
3. If there is more than one relevant exposure definition for the effect size index being extracted:
 - Extract effect sizes for all relevant exposure definitions, and
 - Extract effect sizes for all time periods and dose groups within a given exposure definition.

- **ES ID**
- **Study ID**
- **Study population ID**
- **Parent insecticide or metabolite**
 - If self-report or proxy, extract parent insecticide; if not reported, write “non-specific”; if there are multiple parent insecticides, write “varied”
 - If biomonitoring, extract metabolite assessed
- **Comparison groups**
 - Format: [More exposed definition]/[less exposed definition]
- **Exposure duration notes**
 - Note anything on magnitude or duration of exposure
- **Statistical test**
- **Effect size being extracted**
- **Effect size being extracted rationale**
- **Formula inputs**
 - Examples: Means-SDs, MD-T, Beta-OR, R, Beta-Cont-T
- **Statistical inputs**
 - Examples: Means-SEs, Beta-MD-SE, Beta-MD-95%CI, Beta-Cont-SE, Beta-Cont-95%CI, etc.
- **Is the effect size bivariate or multivariate?** *Multivar, Bivar*
- **No. of model predictors**
 - Covariates, plus the exposure we’re interested in isolating
- **Model covariates**

- **Effect size/variance estimates**
 - *Individual group sample sizes, means, SDs*
 - *Study-reported mean difference (MD), SE, 95%CI*
 - *Beta-MD, SE, 95%CI*
 - *Odds ratio, SE, 95%CI*
 - *Correlation coefficient, SE, 95%CI*
 - *Beta-Cont, SE, 95%CI*
- **T-stat, if applicable**
- **Model R², if applicable**
- **Total N (sample size)**
- **p-value (level of statistical significance)**
- **Notes**

Appendix V. Low risk of bias designation criteria for each domain of bias.

Risk of bias domain (rating options)	Low risk of bias designation criteria
Recruitment strategy (low, probably low, probably high, high)	<ul style="list-style-type: none"> • Participants were recruited from the same study population at the same time using the same study eligibility criteria • Reported overall or group-level participation rate(s) • Attempted to determine baseline differences in participants (i.e., those who met inclusion criteria and chose to participate) and non-participants; and any reported differences between participants and non-participants show no indication of selection bias in regards to exposure or outcome
Blinding (low, probably low, probably high, high)	<ul style="list-style-type: none"> • Knowledge of exposure/outcome was prevented when assessing outcome/exposure
Exposure assessment (low, probably low, probably high, high)	<ul style="list-style-type: none"> • Risk of misclassification was minimized through use of validated biomonitoring methods in which a pesticide metabolite is measured in an appropriate biological matrix • Accounted for temporal variability (i.e., collected more than one exposure measurement over time) • Applied appropriate quality assurance/quality control (QA/QC) protocols
Outcome assessment (low, probably low, probably high, high)	<ul style="list-style-type: none"> • Risk of misclassification was minimized through validated semen analysis methods (WHO, 2021) <ul style="list-style-type: none"> ◦ Collected on-site; if collected off-site (e.g., at home), participant should be instructed to instructions on maintain proper temperature and sample should be delivered within 50 minutes of collection ◦ Recorded time of analysis, sample spillage, and abstinence time • Accounted for temporal variability (i.e., collected more than one semen sample over time) • If more than one person/lab analyzed semen samples, examined inter-technician/inter-lab variation
Confounding (low, high)	<ul style="list-style-type: none"> • Known confounders – age and smoking – were controlled for, either by design (i.e., matching, stratifying or excluding men with certain traits) or statistical adjustments
Incomplete outcome data (NA, low, probably low, probably high, high)	<ul style="list-style-type: none"> • No missing outcome data, • Missing data was imputed using appropriate methods, or • Attrition or missing outcome data was balanced in numbers across exposure groups, with similar reasons for missing data across groups
Selective outcome reporting (low, probably low, high)	<ul style="list-style-type: none"> • All outcomes specified in study (abstract, introduction, and methods sections) were reported

Conflict of interest (low, probably low, high)	<ul style="list-style-type: none"> • Study free of support from individual or entity having financial interest in outcome of study, defined as: <ul style="list-style-type: none"> ○ Study was funded by government, academic and/or non-profit study ○ Study author declaration of no conflict of interest (not required for low rating, but could indicate a “low” rating is appropriate in the case that study funding is not reported)
--	--

Note: Low risk of bias designation criteria was adapted from Lam et al., 2015, Protocol, Appendix VI, Instructions for Making Risk of Bias Determinations. Clarifications made during the risk of bias assessment effort were recorded and reported in the supplemental material of the final publication (Table S7-Word).

Appendix VI. Meta-analysis screening criteria.

INCLUDE

1. Study provides enough information to directly extract or calculate a “relevant” effect size, defined as one of three families of effect sizes - SMD, R, and OR:
 - Standardized mean difference
 - Group-level means:
 - Measure of central tendency *in sperm concentration* in more and less exposed men (group-level mean or median)
 - Measure of variation or spread *in sperm concentration* in more and less exposed men (group-level SD, SE or 95%CI)
 - Mean difference:
 - Study-reported population MD and SE or 95%CI
 - Beta-MD: coefficient from linear regression models on a *binary exposure variable* (Beta-MD) and SE or 95%CI
 - Correlation:
 - Bivariate correlation coefficient (Spearman or Pearson)
 - Beta-Cont: The most adjusted effect size is a beta coefficient from a linear regression model on a continuous exposure variable
 - Odds ratio:
 - Bivariate odds ratio (calculated from a 2x2 table or study-reported)
 - Beta-OR: The most adjusted effect size is a beta coefficient from a logistic regression model on a binary outcome variable (Beta-OR from logistic regression model)
2. Relevant effect size is on raw scale, or can be converted to raw scale from a statistically transformed scale (e.g., log, ln)

EXCLUDE

1. *Relevant effect size NR or calculable*: The authors did not report enough information to calculate a relevant effect size, defined by the relevant effect size indices outlined in meta-analysis inclusion criteria 1 (see above).
2. *Cannot back-transform*: The authors did not report enough information to back-transform statistically-transformed data.
 - Immediately exclude square-root, cubic-root, etc. transformations (anything other than log-transformations), as there are no existing methods for these types of back-transformations
 - If log-transformed (mean difference), group-level or total geometric (or log) mean sperm concentration is needed to back-transform; if not reported, exclude
 - If log-transformed (beta-continuous), group-level or total arithmetic mean exposure and/or sperm concentration is needed to back-transform; if not reported, exclude
3. *Selective outcome reporting*: The authors determine a relevant effect size index but did not report all or insignificant quantitative results; may or may not explicitly discuss the insignificance of the missing quantitative results.
 - In the case of missing results data, the lead review author (LBE) will contact study authors; if no response, exclude

References

1. Carlsen E, Giwercman A, Keiding N, Skakkebaek NE. Evidence for decreasing quality of semen during past 50 years. *BMJ*. 1992;305(6854):609-613. doi:10.1136/bmj.305.6854.609
2. Auger J, Kunstmann JM, Czyglik F, Jouannet P. Decline in semen quality among fertile men in Paris during the past 20 years. *N Engl J Med*. 1995;332(5):281-285. doi:10.1056/NEJM199502023320501
3. Swan SH, Elkin EP, Fenster L. The question of declining sperm density revisited: an analysis of 101 studies published 1934-1996. *Environ Health Perspect*. 2000;108(10):961-966. doi:10.1289/ehp.00108961
4. Levine H, Jørgensen N, Martino-Andrade A, et al. Temporal trends in sperm count: a systematic review and meta-regression analysis. *Hum Reprod Update*. 2017;23(6):646-659. doi:10.1093/humupd/dmx022
5. Kumar N, Singh AK. Trends of male factor infertility, an important cause of infertility: A review of literature. *J Hum Reprod Sci*. 2015;8(4):191-196. doi:10.4103/0974-1208.170370
6. Ferlin A, Garolla A, Ghezzi M, et al. Sperm Count and Hypogonadism as Markers of General Male Health. *Eur Urol Focus*. 2021;7(1):205-213. doi:10.1016/j.euf.2019.08.001
7. Frazier LM. Reproductive Disorders Associated with Pesticide Exposure. *J Agromedicine*. 2007;12(1):27-37. doi:10.1300/J096v12n01_04
8. Perry MJ. Effects of environmental and occupational pesticide exposure on human sperm: a systematic review. *Hum Reprod Update*. 2008;14(3):233-242. doi:10.1093/humupd/dmm039
9. Koureas M, Tsakalof A, Tsatsakis A, Hadjichristodoulou C. Systematic review of biomonitoring studies to determine the association between exposure to organophosphorus and pyrethroid insecticides and human health outcomes. *Toxicol Lett*. 2012;210(2):155-168. doi:10.1016/j.toxlet.2011.10.007
10. Martenies SE, Perry MJ. Environmental and occupational pesticide exposure and human sperm parameters: a systematic review. *Toxicology*. 2013;307:66-73. doi:10.1016/j.tox.2013.02.005
11. Woodruff TJ, Sutton P. An Evidence-Based Medicine Methodology To Bridge The Gap Between Clinical And Environmental Health Sciences. *Health Aff (Millwood)*. 2011;30(5):931-937. doi:10.1377/hlthaff.2010.1219
12. Woodruff TJ, Sutton P. The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes. *Environ Health Perspect*. 2014;122(10):1007-1014. doi:10.1289/ehp.1307175
13. Lam J, Lanphear BP, Bellinger D, et al. Developmental PBDE Exposure and IQ/ADHD in Childhood: A Systematic Review and Meta-analysis. *Environ Health Perspect*. 2017;125(8):086001. doi:10.1289/EHP1632
14. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160. doi:10.1136/bmj.n160

15. Lam J, Sutton P, McPartland J, et al. Applying the navigation guide systematic review methodology. Case study #5: association between developmental exposures to PBDEs and human neurodevelopment. PROSPERO 2015 CRD42015019753. Accessed September 10, 2022. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42015019753
16. Covidence systematic review software. Published online 2022. Accessed April 15, 2022. www.covidence.org
17. Knapke ET, Magalhaes D de P, Dalvie MA, Mandrioli D, Perry MJ. Environmental and occupational pesticide exposure and human sperm parameters: A Navigation Guide review. *Toxicology*. 2022;465:153017. doi:10.1016/j.tox.2021.153017
18. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14(1):135. doi:10.1186/1471-2288-14-135
19. Borenstein M, Hedges LV. Effect Sizes for Meta-Analysis. In: Hedges LV, Cooper H, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis*. 3rd ed. Russell Sage Foundation; 2019:207-244. doi:10.7758/9781610448864.14
20. Rodríguez-Barranco M, Tobías A, Redondo D, Molina-Portillo E, Sánchez MJ. Standardizing effect size from linear regression models with log-transformed variables for meta-analysis. *BMC Med Res Methodol*. 2017;17(1):44. doi:10.1186/s12874-017-0322-8
21. Aloe AM, Thompson CG. The Synthesis of Partial Effect Sizes. *J Soc Soc Work Res*. 2013;4(4):390-405. doi:10.5243/jsswr.2013.24
22. Aloe AM, Thompson CG, Liu Z, Lin L. Estimating Partial Standardized Mean Differences from Regression Models. *J Exp Educ*. 2022;90(4):898-915. doi:10.1080/00220973.2021.1966605
23. Higgins JPT, White IR, Anzueto-Cabrera J. Meta-analysis of skewed data: Combining results reported on log-transformed or raw scales. *Stat Med*. 2008;27(29):6072-6092. doi:10.1002/sim.3427
24. Viechtbauer W. metafor: Meta-Analysis Package for R. Published online August 26, 2022. Accessed September 10, 2022. <https://CRAN.R-project.org/package=metafor>
25. Pustejovsky J. clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. Published online August 15, 2022. Accessed September 10, 2022. <https://CRAN.R-project.org/package=clubSandwich>
26. Harrer M, Cuijpers P, Furukawa T, Ebert DD. dmetar: Companion R Package for the Guide “Doing Meta-Analysis in R.” Published online 2019. Accessed September 4, 2022. <http://dmetar.protectlab.org/>
27. Tipton E. Small sample adjustments for robust variance estimation with meta-regression. *Psychol Methods*. 2015;20(3):375-393. doi:10.1037/met0000011
28. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw*. 2010;36:1-48. doi:10.18637/jss.v036.i03

29. Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Stat Med*. 2007;26(1):37-52. doi:10.1002/sim.2514
30. Tukey J. *Exploratory Data Analysis*. Addison-Wesley; 1977.
31. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6(7):e010247. doi:10.1136/bmjopen-2015-010247
32. Cook RD. Detection of Influential Observation in Linear Regression. *Technometrics*. 1977;19(1):15-18. doi:10.2307/1268249
33. Rubio-Aparicio M, López-López JA, Viechtbauer W, Marín-Martínez F, Botella J, Sánchez-Meca J. Testing Categorical Moderators in Mixed-Effects Meta-analysis in the Presence of Heteroscedasticity. *J Exp Educ*. 2020;88(2):288-310. doi:10.1080/00220973.2018.1561404
34. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634. doi:10.1136/bmj.315.7109.629
35. Rodgers MA, Pustejovsky JE. Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychol Methods*. 2021;26(2):141-160. doi:10.1037/met0000300
36. Zwetsloot PP, Van Der Naald M, Sena ES, et al. Standardized mean differences cause funnel plot distortion in publication bias assessments. *eLife*. 2017;6:e24260. doi:10.7554/eLife.24260