

RESEARCH

Open Access



Systems for rating bodies of evidence used in systematic reviews of air pollution exposure and reproductive and children's health: a methodological survey

Sophie K. F. Michel^{1*} , Aishwarya Atmakuri²  and Ondine S. von Ehrenstein^{1,3} 

Abstract

Background Translating findings from systematic reviews assessing associations between environmental exposures and reproductive and children's health into policy recommendations requires valid and transparent evidence grading.

Methods We aimed to evaluate systems for grading bodies of evidence used in systematic reviews of environmental exposures and reproductive/ children's health outcomes, by conducting a methodological survey of air pollution research, comprising a comprehensive search for and assessment of all relevant systematic reviews. To evaluate the frameworks used for rating the internal validity of primary studies and for grading bodies of evidence (multiple studies), we considered whether and how specific criteria or domains were operationalized to address reproductive/ children's environmental health, e.g., whether the timing of exposure assessment was evaluated with regard to vulnerable developmental stages.

Results Eighteen out of 177 (9.8%) systematic reviews used formal systems for rating the body of evidence; 15 distinct internal validity assessment tools for primary studies, and nine different grading systems for bodies of evidence were used, with multiple modifications applied to the cited approaches. The Newcastle Ottawa Scale (NOS) and the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) framework, neither developed specifically for this field, were the most commonly used approaches for rating individual studies and bodies of evidence, respectively. Overall, the identified approaches were highly heterogeneous in both their comprehensiveness and their applicability to reproductive/children's environmental health research.

Conclusion Establishing the wider use of more appropriate evidence grading methods is instrumental both for strengthening systematic review methodologies, and for the effective development and implementation of environmental public health policies, particularly for protecting pregnant persons and children.

Keywords Air pollution, Reproductive and children's health, Evidence grading, Body of evidence, Systematic reviews, Methodological survey

*Correspondence:

Sophie K. F. Michel
sophiemichel@g.ucla.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

A range of detrimental impacts of air pollution exposure on reproductive and children's health have been established [1–5]. However, air quality regulatory efforts, and especially those accounting for the specific vulnerabilities inherent to reproductive and children's health, have yet to be effectively implemented on a larger scale [6–8]. Formally assessing the quality of the body of evidence, meaning the collection of available individual studies, has been identified as central to translating research into policy [9]. In fact, grading the quality of the body of evidence has become an integral part of the systematic review process [10], reflected in recent additions to the revised Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 guidelines, recommending authors explicitly report their approach to the process of rating the body of evidence [11]. Evidence grading approaches were developed predominantly for clinical questions, including well-established guidelines such as the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) criteria [10, 12].

However, the field of reproductive and children's environmental health, including research on air pollutant exposure, is affected by characteristics that may complicate the critical evaluation of primary studies and bodies of evidence:

- i) The predominantly observational nature of available studies means that, due to inherent differences in study design compared to experimental studies, a different approach is required for identifying and addressing potential confounding and other biases [13–15]. Specific aspects of epidemiologic studies of air pollutant exposure and reproductive/ children's health outcomes that may result in confounding (e.g., frequent use of spatial rather than temporal comparators, lack of covariate information from birth records or other sources), have been described [16, 17]. However, both the default ranking of experimental studies above observational studies, as well as the practice of rating primary studies based on how well they emulate a "hypothetical target RCT" have been criticized [18–23].
- ii) Highly heterogeneous and dynamic population characteristics that define the field of reproductive and children's health (e.g., vulnerabilities related to developmental stages, rapid changes in health-related behaviors) require a lifestage-specific approach. Profound physiological and developmental differences between children and adults impact the toxicity and adverse biological implications of chemical exposures, based on variations in metabolic rates, (de-)toxification processes, and vulnerability during specific developmental windows [17, 24–26].
- iii) Further aspects specific to reproductive and children's health, including generally longer expected lifespans and long latency periods, life course perspectives (e.g., developmental origins of disease), trans-generational effects, among others, necessitate a tailored approach [24].
- iv) Challenges related to exposure assessments are generally an issue in observational vs. experimental studies, where exposures are not controlled by investigators, and in particular, in environmental health studies [14, 27]. Exposure assessments regarding air pollution are characterized by specific challenges (e.g., differences in the availability of air monitoring data, seasonal variations in exposure patterns, etc.) [17, 27], potentially increasing misclassification, also with regard to relevant developmental periods, such as gestational trimesters. Also, there are additional considerations with regard to reproductive and children's health: Due to differences in body size and behaviors, among others, exposure patterns are different for developing fetuses, children, and pregnant persons vs. non-pregnant adults (e.g., relative exposure doses, exposure routes and settings, timing and duration of exposure in relation to windows of susceptibility) [24, 27–29]. For example, children have different breathing zones (due to shorter stature) and oxygen consumption patterns, affecting their individual exposure to air pollution [25].
- v) The co-exposure to mixes of pollutants reflects the real-world risks faced by the global population, which may include additive/ synergistic effects between chemicals, and while modeling impacts of multiple pollutants jointly could provide more valid results, there are challenges such as collinearity and high dimensionality, among others [17, 27, 30–32].
- vi) Further, the context of decision-making in environmental health research differs: Unlike in the clinical setting, environmental exposures are often assessed for risks only after exposure -often wide-spread and long-term- has already occurred in the population [28]. Also, environmental health studies are focused on protecting, rather than improving health [28]. Therefore, while clinical research is primarily concerned with demonstrating a desired treatment effect, reproductive/ children's environmental health should, arguably, be concerned with demonstrating the *absence* of adverse effects: For the former, the burden of proof lies in demonstrating an association or effect, while for the latter, it would lie in demonstrating no association or effect, in essence, safety [33–36]. Statistical methods for testing for the

absence of effects (e.g., equivalence tests) are available, and in addition to providing evidence regarding the equivalence of different exposure scenarios, may also help to reduce publication bias [37–39].

Methodological weaknesses specific to assessing evidence related to environment exposures [40, 41], and specifically ambient air pollution [42], and pregnancy outcomes [43, 44], were previously identified among systematic reviews, particularly related to assessing internal validity and a lack of transparent evidence grading methodologies. Because systematic review methodologies were primarily developed for clinical trials, their suitability for evaluating evidence from observational/ environmental health, and how these methods can best be adapted, has been debated [14]. Further, certain aspects of existing approaches, including the aforementioned default ranking of evidence from randomized controlled trials (RCT) above that from observational studies, have previously been criticized in the context of environmental health [18, 22, 23].

In this methodological survey we aimed to evaluate frameworks for critically assessing bodies of evidence, applied in systematic reviews of epidemiological studies of environmental exposures and adverse reproductive/ child health outcomes, using research on air pollution exposure as a case-study. Air pollutant exposure was chosen based on the comparability of approaches within this research area, and the large body of available systematic reviews [45]. Based on this, we exemplify and discuss challenges and recommendations for evidence grading in the context of reproductive/ children's environmental health.

Methods

As the unit of analysis of this work was systematic reviews, we adhered to the Preferred Reporting Items for Overviews of Reviews (PRIOR) guidelines (Supplemental Material S1, PRIOR checklist) [46], and further relevant guidance [47–51]. Two reviewers independently completed all steps of the systematic process, including screening for eligible references, extracting data, and assessing risk of bias (SM and AA). Discrepancies were resolved by discussing or by consulting with the third reviewer (OVE).

Eligibility criteria and review selection

The inclusion criteria are presented and explained in Table 1.

As highlighted in Table 1, we identified systematic reviews explicitly employing published criteria or

guidelines for assessing or rating the quality of the body of evidence, among the collection of systematic reviews of studies of air pollutant exposure and adverse reproductive and child health outcomes.

Titles, abstracts, and full-texts of the identified publications were consecutively screened, and included in the subsequent screening step, unless there was explicit indication that the publication did not meet our inclusion criteria.

Data sources and search strategy

For identifying systematic reviews, PubMed and Epistemonikos have been identified as the database combination with the highest inclusion rate [54], and we additionally searched the database Embase. For identifying systematic reviews, in favor of built-in filters, we developed a hedge combining searches of text words, filters, and publication types, based on current recommendations for achieving maximum sensitivity [54–57].

Controlled vocabulary terms and keywords were employed to combine the concepts “air pollution”, “childhood”, and “systematic review” (Supplemental Material S2: Full electronic search strategies). We used the PubMed Reminer tool [58], and the SearchRefiner tool from the Systematic Review Accelerator website [59], to develop and assess the sensitivity and specificity of our search strategy.

On December 9, 2020, we conducted the initial systematic search of the electronic databases, without language or publication status restrictions. All searches of electronic databases were performed by SM and updated until April 07, 2023.

In addition, supplementary searches were performed using the search engines Google and Google Scholar. Search engines are used supplementarily, as these allow limited insights into how search results are produced [60]. Further, we manually performed backward and forward citation searching.

Data extraction

Data on systematic review characteristics were extracted using a standardized data extraction form. For extracting information pertaining to the evidence grading systems, descriptions reported in the original articles, as well as cited guidance documents and further related references (i.e., organization websites, etc.) were consulted. Also, we considered the versions of the approaches used within the identified systematic reviews, although in some cases, newer versions exist. If necessary, we attempted to contact systematic review authors to identify or clarify missing or unclear information.

Table 1 Inclusion criteria

Domain	Inclusion criteria	Exclusion criteria	Elaboration
Population	Human studies Reproductive and child health was defined as any timepoint from conception until 18 years of age, including studies of fertility	Reviews additionally examining other populations (i.e., unrelated adult populations) in addition that of interest were not included	This criterion was introduced to ensure the applicability and comparability of identified evidence grading approaches to studies of reproductive/ child health
Exposure	All outdoor or indoor air pollutants (PM ₁₀ , PM _{2.5} , SO ₂ , NO ₂ , air toxics, biofuels, etc.)	Tobacco smoke, mold, wildfires, or allergens. Reviews examining other exposures in addition to air pollution were not eligible	To maximize the applicability and comparability of identified evidence grading approaches to studies of air pollution exposure, we considered only systematic reviews focusing solely in air pollution
Outcome	All adverse reproductive or children's health outcomes	Exposure levels, individual behaviors (e.g., physical activity) were not eligible endpoints	To ensure the comparability of systematic reviews, only those with an adverse health outcome as a primary endpoint were included
Review design	Systematic reviews of observational studies. Only reviews considered "systematic," based on the following reporting criteria were included [52]: (i) review question (ii) reproducible search strategy (e.g., naming of databases and search platforms/engines, search dates and complete search strategy) (iii) inclusion and exclusion criteria (iv) selection (screening) methods (v) critically appraises and reports the quality/risk of bias of the included studies (vi) information about data analysis and synthesis that allows the reproducibility of the results (vii) critically appraises and reports the quality of the body of evidence	Non-systematic literature reviews, scoping reviews, pooled studies, and meta-analyses not based on formal systematic review	Non-systematic reviews may be highly dissimilar in their objectives and methods compared to systematic reviews
System for rating the body of evidence	Systematic reviews that explicitly used a published tool or framework to rate the quality of the body of evidence	Systematic reviews that did not explicitly use a tool, framework, or other published guidance for rating the body of evidence	For purposes of comparability, systematic reviews that did not rate the quality of the body of evidence, or that rated the quality of the body of evidence only as part of their discussion or in another informal manner were not included
Timeframe	Articles published from 1995 onward were considered,	Published to any of the included sources before 1995	Evidence rating was initially proposed as a stage of research synthesis in 1995 [10, 53]
Language	English or German	All other languages	While systematic reviews in other languages may be relevant, we were not able to consider other languages due to resource constraints

Abbreviations: NO₂ Nitrogen dioxide, PM_{2.5} Fine particulate matter, PM₁₀ Coarse particulate matter, SO₂ Sulfur dioxide

Table 2 Considered characteristics regarding reproductive/children's environmental health

Considerations applied to risk of bias or quality ^a assessment tools for individual studies with regard to reproductive/ children's environmental health	Considerations applied to systems for rating bodies of evidence with regard to reproductive/ children's environmental health
<ul style="list-style-type: none"> • Is the exposure assessment method evaluated? [14, 17, 24, 25, 27–29] • Are co-exposures assessed? [17, 27, 30, 31] • Is confounding considered? [13–17] 	<ul style="list-style-type: none"> • Are ratings assigned based on a hierarchy of study designs (i.e., experimental vs. observational studies)? [18–23] • Are developmental stages, child physiology or behaviors, or child-specific health outcomes explicitly considered in evaluating the applicability of the evidence, heterogeneity of results, or potential confounding/ biases? [17, 24–26] • As part of the directness or other domain, was the adequacy of the timing of exposure assessment and the length of follow-up considered? [24, 28, 29, 67] • How is evidence for absence of an association assessed? [28, 33–36]

^a Risk of bias and quality assessments of individual studies were considered jointly herein

Risk of bias assessment (ROBIS)

Risk of bias in systematic reviews was evaluated using the ROBIS tool, based on 1) the appropriateness of study eligibility criteria, 2) methods for identifying and selection of studies, 3) data extraction and quality appraisal methods, and 4) appropriateness of data synthesis, and 5) overall risk of bias [61].

Qualitative analysis/synthesis

We calculated the proportion of systematic reviews explicitly employing formal evidence grading frameworks. The main characteristics of these reviews, including both the main objectives and findings, as well as the systematic review methods, were synthesized in descriptive and tabular format. Methodological characteristics, specifically the guidelines and approaches used for grading bodies of evidence were reviewed. Notably, because approaches used for assessing a body of evidence are partially based on preceding assessments of the quality or risk of bias among primary/ individual studies, both of these types of assessments in the systematic review processes were distinctly considered herein.

With regard to individual studies, quality versus risk of bias or internal validity are related but distinct concepts, concerned with the critical assessment of individual studies. Risk of bias, refers to aspects of study design, conduct, or analysis that could give rise to systematic error in study results, and can be used synonymously with internal validity, which is the extent to which bias has been prevented through methodological aspects [62]. Study quality, on the other hand, may refer to (a) reporting quality; (b) internal validity or risk of bias; and (c) external validity or directness and applicability, among others [15]. However, while risk of bias vs. study quality assessments are truly distinct concepts, they are often interchanged or merged in research practice [63]. For this reason, in this

methodological survey, these approaches were considered jointly.

The quality of/ certainty in the body of evidence, on the other hand, is assessed based on strengths and limitations of a collection of individual studies, and incorporates results from preceding risk of bias assessments, as well as aspects of directness/ applicability of the identified primary studies with regard to the review question, heterogeneity/ inconsistency across studies, the magnitude and precision of effect estimates, potential publication biases, and further criteria [12, 15]. Sometimes this step is followed by subsequent ratings regarding the strength or levels of evidence, or hazard identification, across study types, outcomes, or species [15, 64].

Certain criteria are applied differently when assessing the internal validity of individual studies versus the body of evidence. For example, while an identified risk of confounding will result in a lower internal validity score for an individual study, a body of evidence may receive a higher quality rating, if all plausible confounding “would reduce a demonstrated effect, or suggest a spurious effect when results show no effect”, as noted by multiple guidelines [64–66].

We considered characteristics of frameworks for rating risk of bias in individual studies, and for grading the body of evidence, specifically as they relate to reproductive/ children's environmental health, as discussed earlier (see Table 2).

Results

Review selection process

The selection process of systematic reviews is shown in Fig. 1. After screening 10,241 titles, 1,030 abstracts and 423 full texts, 177 systematic reviews were found to assess the association between exposure to air pollution and adverse reproductive/ children's health outcomes.

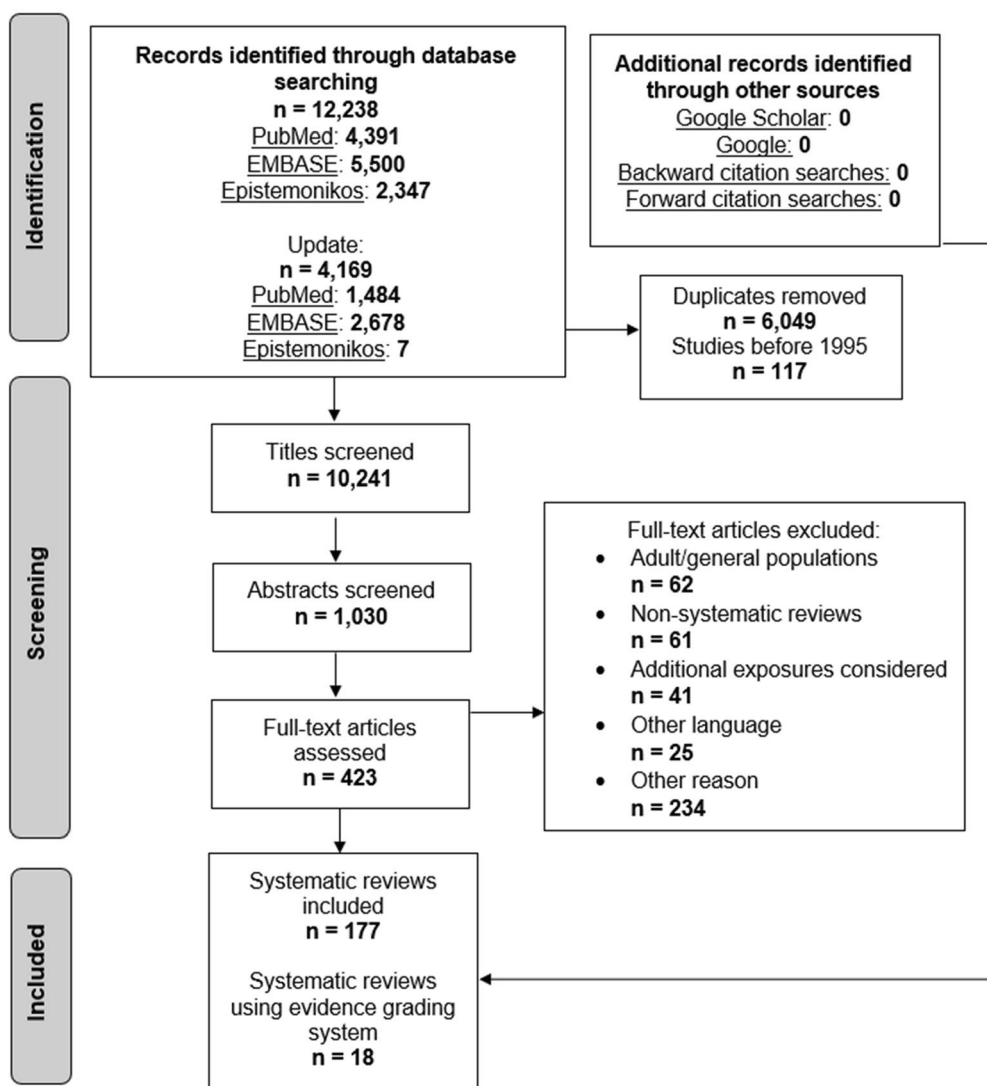


Fig. 1 Screening process for systematic reviews

The most common reasons for exclusions of full texts were that the reviews considered adult or general populations ($n=62$), and that reviews were non-systematic ($n=61$). Out of the 177 eligible systematic reviews, 18 articles (9.8%) explicitly reported using evidence grading systems [5, 68–84]. The proportion of systematic reviews using evidence grading systems appeared to increase over time (see Fig. 2).

Systematic review characteristics

General characteristics of the 18 systematic reviews that used formal evidence grading systems are summarized in Table 3. These reviews were published between 2015 and 2023; and outcomes assessed were: spontaneous abortion [80], gestational diabetes mellitus [83], fetal growth [72], preterm birth [73, 79], birth weight [76], term birth weight

[77], congenital anomalies [74], upper respiratory tract infections [81], bronchiolitis in infants [71], sleep-disordered breathing [70], blood pressure in children and adolescents [84], neuropsychological development [68, 78], autism spectrum disorder [5, 69], academic performance [75], or all child health outcomes [82].

None of the included reviews specified inclusion criteria related to the method of exposure assessment (e.g., modeling vs. monitoring approaches) (see Table 3). Two reviews considered both intervention and observational studies [70, 81], while the others included only observational studies. Between 7 and 84 studies were included by the individual reviews (Table 3) [79, 80]. One review included only studies using air monitoring stations’ data [74], while others reported a variety of exposure assessment methods and data sources.

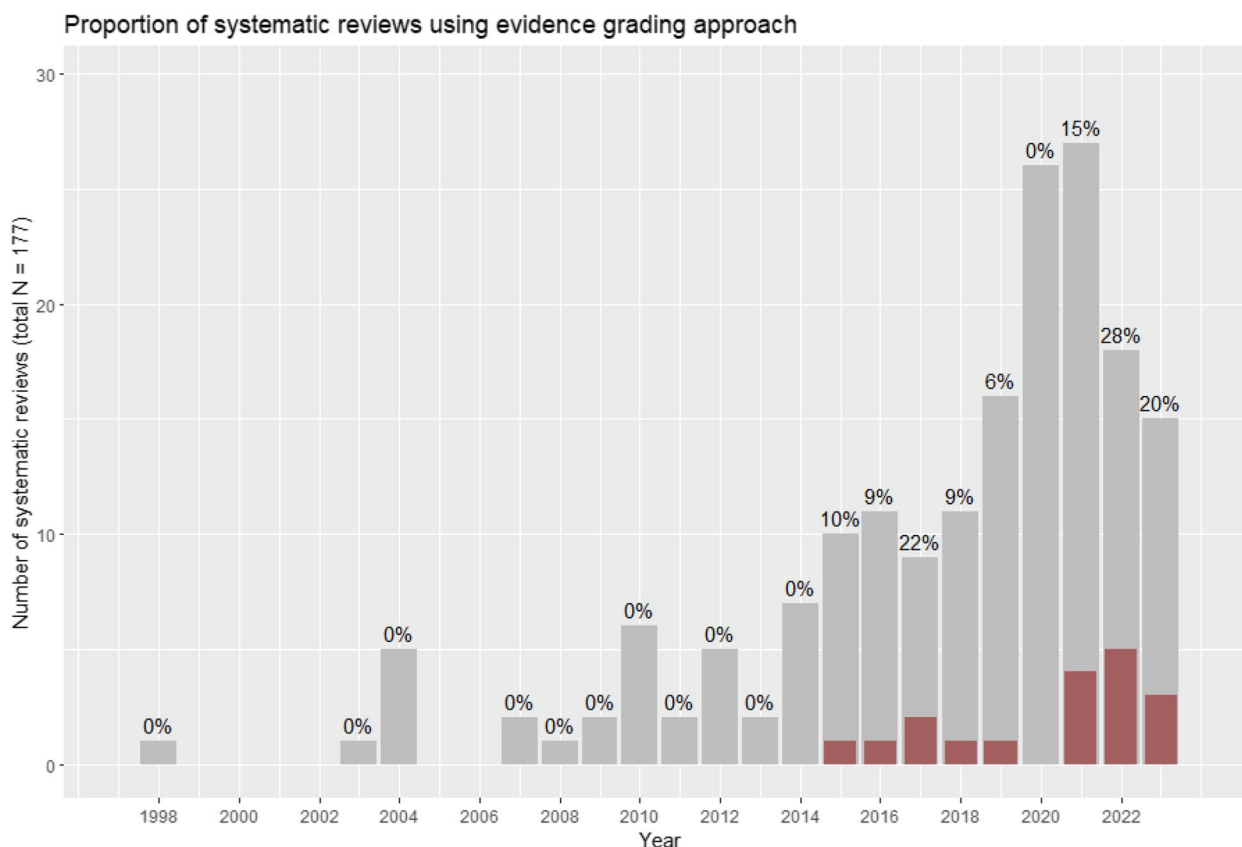


Fig. 2 Systematic reviews per year, and proportion using evidence grading approach (designed using R software). Includes publications up until April 2023

Individual-level measures of exposure (e.g., adducts in cord blood, backpack for individual monitoring) were reported for few studies included in the systematic reviews [68, 72, 76, 77].

The majority of systematic reviews included fixed or random effects meta-analyses, while five refrained for statistical pooling and synthesized their findings in narrative form [68–71, 75]. All meta-analyses included adjusted effect estimates; several reported only considering single-pollutant models.

ROBIS assessment results

Four of the included systematic reviews were rated at a low risk of bias [5, 71, 75, 76], four at a high risk of bias [68–70, 74], and the remaining ten at an unclear risk of bias. The most critical concerns related to methods used to search for primary studies, synthesis approaches, and insufficient reporting (Fig. 3). Between one and eight databases were searched by the various review teams [69, 75]. Six groups made no additional efforts to identify published or unpublished literature [68, 69, 71, 79, 81, 82], while eight additionally screened the reference lists of included studies and/ or those of

relevant reviews [70, 72, 74–76, 80, 83, 84], in some cases additionally searching relevant reports [73], using web search engines [78], and one further searched grey literature databases and relevant websites, performed forward citation searches, and contacted experts in the field (Supplemental Materials S3 and S4: Details of ROBIS assessment) [5]. Methods used for primary study appraisal, synthesis, and evidence grading are described further below.

Methodological characteristics- Methods for assessing risk of bias/quality in primary studies

The 18 included systematic reviews used 15 distinct approaches for assessing risk of bias/ quality/ internal validity among primary studies (Tables 3 and 4). The Newcastle–Ottawa Scale (NOS) was the most commonly cited tool (*n*=9 reviews) [86], with an additional four reviews using modified NOS versions, followed by the Office of Health Assessment and Translation (OHAT) approach (*n*=4 reviews)[66]. However, multiple reviews reported using multiple tools, in order to assess quality and risk of bias separately, as well as to address various study designs (e.g., cohort vs. cross-sectional studies)

Table 3 Characteristics of included systematic reviews (sorted by publication year)

Author, year, synthesis type	PEOS of review	Number and type of studies included	Main findings	Risk of bias / quality assessment of individual studies	Risk of bias findings	Body of evidence assessment method	Body of evidence assessment findings
Suades-González, E., et al. (2015) [68], Systematic review	<p>Population: Children 0–18 years old</p> <p>Exposure: Outdoor air pollution during pregnancy, around birth or during childhood, using direct or indirect assessment methods</p> <p>Outcome: Neuropsychological development, cognition, behavior, neurodevelopmental disorders, psychomotor outcomes</p> <p>Studies: Cohort, case-control, or cross-sectional design, "original research articles" Published after 2012 in English language</p>	20 cohort, 6 case-control, and 6 cross-sectional studies	Pre- or postnatal exposure to PAH was associated with a lower global IQ, pre- or postnatal exposure to PM _{2.5} was associated with an increased risk of ASD, and prenatal exposure to NO _x was associated with higher risk of ASD. All other associations showed mixed findings	Own criteria: study design, sample size, exposure and outcome assessment, confounder control	25 out of 32 included studies were rated as a "good quality studies"	Modified IARC (2006) – authors differentiate between "inadequate" and "insufficient" evidence based on whether or not studies "reported an association"	Pre- or postnatal exposure: PAH and global IQ, PM _{2.5} and ASD: Sufficient evidence Prenatal exposure to NO _x and ASD: Limited evidence All other associations: Inadequate or insufficient evidence, due to few studies, low quality, or inconsistency
Lam, J., et al. (2016) [5], Systematic review and meta-analysis	<p>Population: Humans</p> <p>Exposure: Indoor/outdoor air pollution exposure during any developmental period (maternal or paternal, in proximity to conception, during pregnancy, or childhood), prior to outcome assessment</p> <p>Outcome: Any clinical diagnosis or other continuous or dichotomous scale assessment of ASD, based on ICD 9/ 10, or DSM 4/ 5 criteria</p> <p>Studies: Original data</p>	17 case-control studies, 4 ecological, 2 cohort studies	Per 10 µg/m ³ increase in PM ₁₀ : (SOR: 1.07, 95% CI: 1.06–1.08, n=6 studies) and PM _{2.5} (SOR: 2.32, 95% CI: 2.15–2.51, n=3 studies) All included studies generally showed increased risk of ASD with increasing exposure to air pollution, but with some inconsistency across chemical components	Modified instrument, developed based on the Cochrane Collaboration's tool and the AHRQ domains (i.e., selection bias, confounding, performance bias, attrition bias, detection bias, and reporting bias). Also, an approach for rating exposure assessment methods was newly developed	Majority of studies were rated as "low" or "probably low" risk of bias, besides for domains confounding and exposure assessment	Navigation Guide	Moderate quality of evidence across all air pollutants, due to small number of studies in meta-analysis and unexplained statistical heterogeneity

Table 3 (continued)

Author, year, synthesis type	PEOS of review	Number and type of studies included	Main findings	Risk of bias / quality assessment of individual studies	Risk of bias findings	Body of evidence assessment method	Body of evidence assessment findings
Morales-Suárez-Varela, M., et al. (2017) [69], Systematic review	<p>Population: Humans</p> <p>Exposure: PM_{2.5}, PM₁₀ or diesel PM exposure during pregnancy/early childhood</p> <p>Outcome: Measures of ASD symptoms or diagnosis</p> <p>Studies: Published after 2005 in English</p>	4 cohort and 9 case-control studies	9 out of 13 studies suggested positive associations during specific exposure windows. Authors conclude there is an increased risk of ASD due to PM exposure, with varying magnitude according to the particle size and composition, with the association with PM _{2.5} and diesel PM being largest.	Scottish Intercollegiate Guideline Network (SIGN) used to assign levels of evidence based on study design. Additional considerations were sample size, specified inclusion and assessment criteria, exposure and outcome assessment, confounder control	5 studies at a high risk of confounding or bias and a significant risk that the relationship is not causal, 8 studies at a low risk of bias and a moderate probability of causal association	SIGN levels of recommendation	Level of recommendation: C-D C: A body of evidence of studies at a low risk of bias, directly applicable to the target population, with consistent results D: Extrapolated evidence from studies at a low risk of bias An association between PM exposure and ASD cannot be ruled out. Data is insufficient to reach a consensus about ASD risk
Tenero, L., et al. (2017) [70], Systematic review	<p>Population: Children</p> <p>Exposure: Indoor or outdoor air pollution</p> <p>Outcome: Sleep disordered breathing, sleep apnea, excluding asthma or SIDS</p> <p>Studies: Observational and intervention studies included, published in English language</p>	4 cohort studies, 2 cross-sectional studies, 2 "prospective surveys" / intervention studies	Results suggest an involvement of environmental pollution in the worsening of sleep-disordered breathing in children	Centre for Evidence Based Medicine guidelines (2009, 2011)	Evidence level 3B for all studies	Centre for Evidence Based Medicine guidelines (2009, 2011)	Grade C ("cohort or case-control studies")

Table 3 (continued)

Author, year, synthesis type	PEOS of review	Number and type of studies included	Main findings	Risk of bias / quality assessment of individual studies	Risk of bias findings	Body of evidence assessment method	Body of evidence assessment findings
King, C., et al. (2018) [71], Systematic review	<p>Population: Children < 3 years old</p> <p>Exposure: Criteria air pollutants^a, at any time period before hospitalization, categorized as acute (less than 7 days), sub-chronic (1 month prior), or lifetime exposure</p> <p>Outcome: Hospital admission, emergency department visits, unscheduled primary care visits, or critical care admission for bronchiolitis</p> <p>Studies: Cohort, case-control, time series, case-crossover designs included, ecological designs excluded</p>	4 case-control and 4 case-crossover studies	Long term exposure to PM _{2.5} and acute exposure to SO ₂ and NO ₂ may be associated with increased risk of hospitalization for bronchiolitis. Results for other pollutants were inconsistent	<p>NOS, in addition to considerations of specific aspects of: selection bias, exposure and outcome assessment, adjustment for confounders. Studies were rated at low risk of bias if they adjusted for at least two pre-specified confounders. Further, studies were classified as higher quality if infants were < 2 years old</p>	NOS score 7–8 (good quality) 2 studies rated as “unclear/ high” risk of bias across all domains, 6 others were rated as “low” across all domains	GRADE	Low to moderate, frequent downgrading for inconsistency, upgrading for precision, and up- or downgrading for study quality
Fu, L., et al. (2019) [72], Systematic review and meta-analysis	<p>Population: Fetuses, infants at birth</p> <p>Exposure: Criteria air pollutants^a</p> <p>Outcome: Fetal growth indicators during pregnancy and anthropometric measurements at birth. Studies reporting exposure windows, sample size, and partial regression coefficient with 95% CI or standard errors were included</p> <p>Studies: English or Chinese language</p>	11 prospective and 4 retrospective studies	Higher PM _{2.5} exposure during entire pregnancy negatively associated with head circumference at birth (-0.30 cm, 95% CI: -0.49 to -0.10), and NO ₂ exposure during entire pregnancy associated with shorter length at birth (-0.03 cm, 95% CI: -0.05 to -0.02). All other associations were not significant or inconclusive	ACROBAT-NRS1, with modifications	7 studies rated at low, 5 at moderate, and 3 at high risk of bias. Common concerns were exposure or outcome assessment methods, lack of confounder control, or study design (retrospective studies)	GRADE	PM _{2.5} and head circumference: Moderate PM ₁₀ and head circumference: Low NO ₂ and birth length: Low NO ₂ and head circumference: Very low Downgraded for risk of bias and inconsistency, upgraded for dose-response gradient

Table 3 (continued)

Author, year, synthesis type	PEOS of review	Number and type of studies included	Main findings	Risk of bias / quality assessment of individual studies	Risk of bias findings	Body of evidence assessment method	Body of evidence assessment findings
Rappazzo, KM., et al. (2021) [73], Systematic review and meta-analysis	<p>Population: Any population capable of becoming pregnant, including those at increased risk of preterm birth</p> <p>Exposure: O₃ exposure during 1st or 2nd trimester. Exposure assessment must have covered entire trimester. Only studies reporting a continuous exposure contrast included</p> <p>Outcome: Preterm birth</p> <p>Studies: Cohort, case-control studies. Reviews and abstract-only references excluded. Only English language articles included</p>	16 cohort, 3 case control studies	Increased risk of PTB from exposure during 1st trimester (SOR per 10 ppb increase in O ₃ : 1.06, 95 CI: 1.03–1.10), and during 2nd trimester (SOR per 10 ppb: 1.05, 95% CI: 1.02– 1.08)	Modified OHAT framework, domains included "participant selection, outcome, exposure, confounding (consideration of co-pollutants), analysis, selective reporting, sensitivity, and overall quality"	High $n = 1$, medium $n = 9$, and low $n = 9$ confidence. Common concerns included insufficient reporting or exposure assessment methods	OHAT	Moderate (no up- or downgrading in any domain)
Ravindra, K., et al. (2021) [74], Systematic review and meta-analysis	<p>Population: Children < 5 years old. Live births, stillbirths, and terminations eligible</p> <p>Exposure: Outdoor or indoor air pollutants</p> <p>Outcome: Congenital anomalies, meaning anomalies of prenatal origin present at birth, Neurological defects, ASD, ADHD included</p> <p>Studies: Case-control, cohort, or ecological studies included. Published after 1950 in English</p>	16 case-control, 9 cohort and 1 ecological study	N/A (due to serious concern with the data synthesis methods, results not reported here)	Risk of bias tool modified for prevalence studies from tool by Leboeuf-Yde and Lauritsen (Hoy et al. 2012), and ROBINS-E preliminary tool	Hoy et al.: Moderate ($n = 4$) to low ($n = 22$) ROBINS-E: Moderate ($n = 11$), low ($n = 13$), high ($n = 1$), unclear ($n = 1$)	Navigation Guide	NO ₂ and atrial septal defects, ventricular septal defects: <i>High</i> All other associations <i>Low to very low</i> Downgrading was due to risk of bias, imprecision, or inconsistency

Table 3 (continued)

Author, year, synthesis type	PEOS of review	Number and type of studies included	Main findings	Risk of bias / quality assessment of individual studies	Risk of bias findings	Body of evidence assessment method	Body of evidence assessment findings
Stenson, C., et al. (2021) [75], Systematic review	<p>Population: Children and adolescents</p> <p>Exposure: TRAP pollutants exposure during pregnancy, childhood, or adolescence</p> <p>Outcome: Academic performance measured using standardized tests: exam results, GPA, other, (but not tests of cognitive function)</p> <p>Studies: Cross-sectional, ecological, prospective or retrospective cohort, panel and case-control studies. Only peer reviewed articles in English included. Abstract-only or conference materials excluded</p>	7 cross-sectional designs, 2 retrospective cohort studies, and 1 longitudinal study	9 studies reported link between higher TRAP exposure and poorer student academic performance. Effect sizes were generally small	Adapted version of the OHAT tool. Key confounders and co-exposures were pre-determined, and adjustment for these was considered	Serious concerns about insufficient reporting and exposure assessments	OHAT approach. Also, an overall data assessment visualization table was developed to assess potential publication bias	Low Due to serious risk of bias, imprecision, and potential publication bias. 9 papers stated an absence of conflict of interest or reported funding sources which did not raise concerns
Uwak, I., et al. (2021) [76], Systematic review and meta-analysis	<p>Population: Pregnant women</p> <p>Exposure: Prenatal exposure to ambient particulate air pollution (PM_{2.5}, PM₁₀, PM_{2.5-10})</p> <p>Outcome: Birth weight measured as a continuous variable. Studies reporting birth weight as z-scores were excluded</p> <p>Studies: N/R</p>	51 cohort, 2 cross-sectional studies	Increased risk observed for PM _{2.5} exposure in the 2nd or 3rd trimester: -5.69 g (95% CI: -10.58, -0.79, I ² : 68%) and -10.67 g birth weight (95% CI: -20.91, -0.43, I ² : 84%), respectively. Over the entire pregnancy: -27.55 g (95% CI: -48.45, -6.65, I ² : 94%) PM ₁₀ exposure in the 3rd trimester and the entire pregnancy: -6.57 g (95% CI: -10.66, -2.48, I ² : 0%) and -8.65 g (95% CI: -16.83, -0.48, I ² : 84%), respectively	Navigation Guide: recruitment strategy, blinding, confounding, exposure assessment, incomplete outcome data, selective reporting, conflicts of interest, other concerns	Risk of bias generally "low"/"probably low" for most domains. 43% of studies at "probably high" risk of bias for exposure assessment method (reliance on county-level monitoring data without adequate temporal coverage/ spatial resolution)	Navigation Guide	PM _{2.5} : Low, due to imprecision and/or unexplained heterogeneity PM ₁₀ : Low (imprecision) to moderate PM _{2.5-10} : Very low/low (risk of bias, inconsistency, and imprecision)

Table 3 (continued)

Author, year, synthesis type	PEOS of review	Number and type of studies included	Main findings	Risk of bias / quality assessment of individual studies	Risk of bias findings	Body of evidence assessment method	Body of evidence assessment findings
Gong, C., et al. (2022) [77], Systematic review and meta-analysis	<p>Population: Full-term, singleton neonates</p> <p>Exposure: Ambient PM_{2.5}</p> <p>Outcome: Change in grams of term birth weight (≥ 37 weeks of gestation) as a continuous variable</p> <p>Studies: Epidemiological studies</p>	61 cohort, 1 case-control	Per 10 $\mu\text{g}/\text{m}^3$ increase in PM _{2.5} : -16.54 g (95% CI: -20.07 g, -13.02 g, I ² = 96%)	NOS (only performed for studies included in meta-analysis)	Most studies (n = 22) obtained high NOS scores (7 + stars). Others (n = 9) obtained fair/low scores (5–6 stars)	GRADE (only for studies included in meta-analysis)	Overall: <i>Very low</i> (downgraded for high heterogeneity) Subgroup of studies using LUR-models: <i>Moderate</i> , upgraded for reasonable residual confounding which could reduce the effect estimates
Lin, LZ., et al. (2022) [78], Systematic review and meta-analysis	<p>Population: Children 0–18 years old</p> <p>Exposure: Ambient PM exposure (pre-conception, prenatal, postnatal)</p> <p>Outcome: Neurodevelopmental disorders (ASD, ADHD, dyslexia, etc.). Confirmed by clinician or structured interview</p> <p>Studies: Case-control, cross-sectional, cohort, case-crossover, time-series, and panel studies. Full-text must be available; in English; did not include abstracts, reviews, conference materials</p>	16 case-control, 13 cohort, 2 cross-sectional studies	Increased risk of ASD linked to exposures to PM _{2.5} during prenatal periods (SOR: 1.32, 95%CI, 1.03–1.69), 1st year after birth (SOR: 1.62, 95%CI, 1.22–2.15) and 2nd year after birth (SOR: 3.13, 95%CI, 1.47–6.67). Inconsistent evidence found for other types of PM and neurodevelopmental outcomes. Some heterogeneity explained by exposure assessment period and study country	NOS, AHRQ, and Cochrane ROB tool	NOS score for case-control and cohort studies: 6–8. AHRQ score for cross-sectional studies: moderate and low quality. Cochrane ROB tool for all studies: 25 studies rated at low risk of bias, 5 as unclear, 1 as high	The Best Evidence Synthesis (BES) System for synthesis without meta-analysis. GRADE for meta-analytical results	BES: PM _{2.5} exposure and the risk of ASD in 1st year after birth: <i>Strong evidence</i> PM _{2.5} exposure and the risk of ASD in 3rd year after birth: <i>Moderate evidence</i> 63 other associations: <i>insufficient or no evidence</i> GRADE: PM _{2.5} exposure and ASD in 2nd year after birth: <i>Moderate</i> (upgraded for magnitude of effect), 12 other associations: <i>low</i> (n = 3) or <i>very/low</i> (n = 9) (due to heterogeneity)

Table 3 (continued)

Author, year, synthesis type	PEOS of review	Number and type of studies included	Main findings	Risk of bias / quality assessment of individual studies	Risk of bias findings	Body of evidence assessment method	Body of evidence assessment findings
Yu, Z., et al. (2022) [79], Systematic review and meta-analysis	<p>Population: Pregnant women without specific medical conditions</p> <p>Exposure: Ambient PM_{2.5} short- and long-term exposures.</p> <p>Studies using proxies for exposure (e.g., traffic density) excluded</p> <p>Outcome: PTB diagnosed by clinical standardization or definite medical records</p> <p>Studies: Case-control, cross-sectional, time-series, or cohort studies. Excluded articles without full-text</p>	84 case-control and cohort studies	<p>Long-term exposure to PM_{2.5} and PM₁₀ during entire pregnancy (SOR per 10 µg/m³: 1.084 (95% CI: 1.055–1.113) and 1.034 (95% CI: 1.018–1.049), respectively. Positive associations were also found between PM_{2.5} exposure in 2nd trimester and PTB subtypes</p> <p>Short-term exposure to PM_{2.5} (SOR per 10 µg/m³: 1.003 (1.001–1.004, 12: 65%) and 1.003 (1.001–1.005, 12: 77%) and PM₁₀ (SOR: 1.001 (95% CI: 1.000–1.001), PM₁₀ exposure in 2 weeks prior to birth also increased PTB risk</p>	<p>Risk of bias: Tailored OHAT approach</p> <p>Quality: NOS and Mustafic et al. (2012) for case-cross-over and time-series studies</p>	<p>Most studies found to have moderate to high quality. Concerns were related to exposure assessment, confounding, and exclusion bias (missing data)</p>	Navigation Guide	Moderate (Concerns regarding inconsistency, but still reached an overall rating of "moderate")
Zhu, W., et al. (2022) [80], Systematic review and meta-analysis	<p>Population: Pregnant women that conceived naturally (no IVF/ET)</p> <p>Exposure: Chronic PM exposure (preconception to prenatal)</p> <p>Outcome: Spontaneous abortion, defined as a loss of fetus within 180 days gestation</p> <p>Studies: Only English language articles considered. Conference abstracts and reviews not included</p>	4 case-control/ case-crossover studies, 3 cohort studies	<p>Increased risk after PM_{2.5} and PM₁₀ exposure (SOR per 10 µg/m³: 1.20 (95%CI: 1.01–1.40) and 1.09 (95%CI: 1.02–1.15), respectively</p>	NOS	All studies were rated as "high quality" (score ≥ 7)	GRADE (GRADEpro App)	Association with PM _{2.5} and PM ₁₀ <i>Moderate</i>

Table 3 (continued)

Author, year, synthesis type	PEOS of review	Number and type of studies included	Main findings	Risk of bias / quality assessment of individual studies	Risk of bias findings	Body of evidence assessment method	Body of evidence assessment findings
Ziou, M., et al. (2022) [81], Systematic review and meta-analysis	<p>Population: Children/adolescents < 19 years old. Studies exclusively conducted in asthmatic, allergic, or immunocompromised populations excluded</p> <p>Exposure: Ambient PM (short- or long-term exposure)</p> <p>Outcome: Upper respiratory tract infections, including otitis</p> <p>Studies: English, French, Spanish language articles from peer-reviewed journals included. Abstracts, reviews, case studies excluded</p>	19 time-series, 4 cohort, 4 case-crossover, 4 meta-analyses, 2 cross-sectional, and 1 interventional study	Both PM _{2.5} and PM ₁₀ associated with hospital presentations for URTIs (SRR: 1.010, 95%CI: 1.007–1.014), SRR: 1.016, 95%CI: 1.011–1.021) in meta-analyses. Narrative analysis found total suspended particulates to be associated with URTIs, but mixed results were found for PM _{2.5} and PM ₁₀	Quality of studies: NOS for cohort studies, modified NOS (Modesti et al. for cross-sectional studies, and Mustafaic et al. for case-crossover/time-series) Risk of bias: Mixture of modified OHAT tool and Navigation Guide, as reported by previous reviews	Quality and risk of bias ratings ranged from low to high. Concerns included outcome and exposure assessment methods, missing data, recall bias, confounding, and selection bias	OHAT, with modifications to address crossover/time-series studies	PM _{2.5} and PM ₁₀ : Moderate (upgraded for confounding, as studies were conducted in hospital or ambulatory settings, where only a minority of children with URTIs present. Thus, there may be a bias towards null) Total suspended particles/ PM unspecified. Low (no change)
Blanc, N., et al. (2023) [82], Systematic review and meta-analysis	<p>Population: Humans</p> <p>Exposure: Maternal or paternal exposure to outdoor air pollutants in the pre-conception period. Preconception period defined as 1 month to 1 year before conception</p> <p>Outcome: All child health outcomes</p> <p>Studies: English language articles</p>	16 cohort and 6 case-control	Exposure to outdoor air pollutants during maternal preconception period were associated with various health outcomes, of which birth defects showed the most consistent findings Large sample sizes, however inconsistencies observed in air pollutant levels and reported associations	NOS	All studies rated as high to moderate quality	Modified GRADE (starting observational studies at low) (Moragan 2016). Available human and experimental animal data and mechanistic data additionally considered Publication bias not assessed. Study funding sources also considered	Preconception exposure to PM ₁₀ and PM _{2.5} and birth defects Moderate PM _{2.5} and birth-weight: Moderate All other associations: Low

Table 3 (continued)

Author, year, synthesis type	PEOS of review	Number and type of studies included	Main findings	Risk of bias / quality assessment of individual studies	Risk of bias findings	Body of evidence assessment method	Body of evidence assessment findings
Liang, W., et al. (2023) [83], Systematic review and meta-analysis	Population: Pregnant women Exposure: Ambient air pollutants. Excluded studies assessing exposure exclusively to indoor pollutants, traffic, or extreme natural environments. Categorical comparisons not included Outcome: Risk of GDM, with explicit diagnostic criteria reported Studies: Cohort studies included. Conference abstracts and studies without accessible data excluded	15 prospective, 16 retrospective cohort studies	Exposure to NO ₂ , SO ₂ , PM _{2.5} , PM ₁₀ , BC, and nitrate may significantly increase the risk of GDM. PM _{2.5} exposure had largest effect on GDM risk	NOS	All studies rated as high quality (7–9 stars)	Modified GRADE for air pollution studies by WHO (starting observational studies at moderate)	NO ₂ exposure during entire pregnancy, and BC exposure in first trimester: Low All other pollutants and exposure periods (32 associations): Moderate (14) to high (18) Main reasons for downgrading were high risk of bias and wide 80% prediction intervals
Tandon, S., et al. (2023) [84], Systematic review and meta-analysis	Population: 10–19-year-old children/ adolescents Exposure: Short- or long-term exposure to ambient air pollutants Outcome: Blood pressure (BP) Studies: Longitudinal (cohort, panel) or cross-sectional studies	5 cohort, 3 cross-sectional studies	Non-significant associations were observed for cohort studies assessing long-term exposure to PM ₁₀ , PM _{2.5} , and NO ₂ . Significant positive associations were observed for cross-sectional studies assessing long-term exposure to PM ₁₀ (0.34 mmHg, 95% CI: 0.19, 0.50) and NO ₂ on diastolic BP (0.40 mmHg, 95% CI: 0.09, 0.71), and PM ₁₀ on systolic (0.48 mmHg, 95% CI: 0.19, 0.77) BP	NOS for cohort studies, adapted NOS for cross-sectional studies (from Herzog et al)	3 studies rated as good quality, 4 as fair quality, 1 as poor quality	GRADE	Low to very low Reasons for downgrading: risk of bias (due to potential selection bias, loss of follow-up, issues with exposure and outcome assessment methods), and inconsistency

Abbreviations: ACROBAT-NRS1 A Cochrane Risk Of Bias Assessment Tool for Non-Randomized Studies of Interventions, ADHD Attention deficit hyperactivity disorder, AHRQ Agency for Healthcare Research and Quality, ASD Autism spectrum disorder, BC Black carbon, BES Best Evidence Synthesis, BP Blood pressure, CI Confidence interval, DSM Diagnostic and Statistical Manual of Mental Disorders, GDM Gestational diabetes mellitus, GPA Grade point average, GRADE Grading of Recommendations, Assessment, Development, and Evaluations, β^2 Heterogeneity statistic, IARC International Agency for Research on Cancer, ICD International classification of diseases, IQ Intelligence quotient, IVF/ET In vitro fertilization or embryo transfer, MOS Newcastle Ottawa Scale, NO₂ Nitrogen dioxide, O₃ Ozone, OHAT Office of Health Assessment and Translation, PAH Polycyclic aromatic hydrocarbons, PM Particulate matter, PM_{2.5} Fine particulate matter, PM₁₀ Coarse particulate matter, PTB preterm birth, ROB Risk of bias, ROBINS-E Risk of bias in non-randomized studies of exposure, SIDS Sudden infant death syndrome, SIGN Scottish Intercollegiate Guidelines Network, SO₂ Sulfur dioxide, SOR Summary odds ratio, SSR Summary risk ratio, TRAP Traffic-related air pollution, URTI Upper respiratory tract infection, WHO World Health Organization

^a Criteria air pollutants: PM₁₀, PM_{2.5}, NO₂, SO₂, O₃, CO, lead

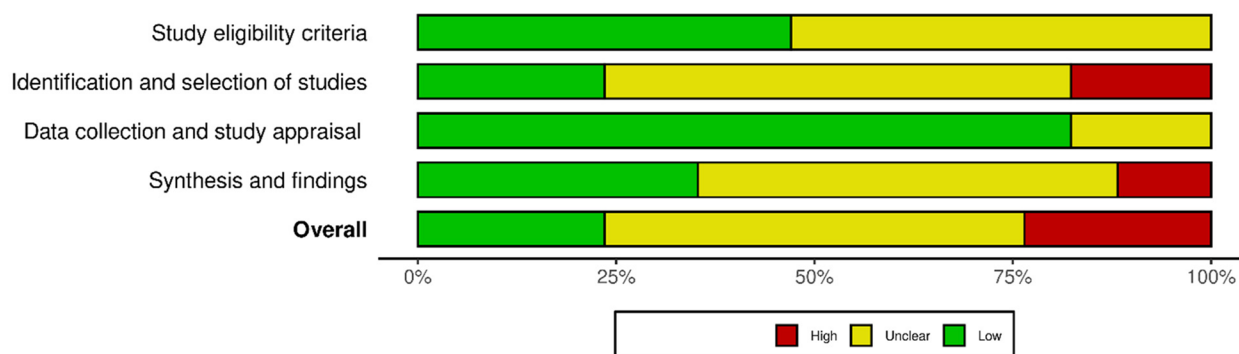


Fig. 3 Summary of risk of bias assessment. Designed using the *robvis* tool [85]

included within reviews [74, 78, 79, 81, 84]. Further, six reviews modified/ tailored the selected tools themselves [69, 71–73, 75, 81], while four reviews used tools as modified by preceding systematic reviews [74, 79, 81, 84].

The tools originated from a wide range of research fields (see Table 4), and only the Navigation Guide and OHAT approaches, used by six reviews, were developed specifically for environmental health research [64, 66]. The Risk of Bias in Non-randomized Studies of Exposure (ROBINS-E) tool was developed for studies of non-randomized studies of exposures and used in one review in its preliminary version [94]. Two reviews newly developed their own criteria for assessing primary study quality/ risk of bias [5, 68]. Notably, Lam et al. further developed the Navigation Guide risk of bias tool with expert input, as part of their application of the Navigation Guide methodology. This included developing an approach for rating exposure assessment methods for different air pollutants/ chemical classes [5]. This approach was subsequently adopted by other identified reviews [76, 81].

Exposure assessment methods in general were evaluated in all but two out of fifteen approaches [92, 93], although we considered only four tools applicable to environmental/ air pollution exposures in this regard [64, 66, 87, 94]. Co-exposures were explicitly considered by five tools [15, 62, 87, 91, 94], while all but one tool assessed confounder control [93]. However, review authors modified existing tools in some cases, for example adding considerations of sample size, selection bias, exposure assessment method, and confounder adjustment [69, 71]. Another review group used subgroup analyses to explore the effect of different exposure assessment methods [77].

Methodological characteristics- Evidence assessment methods

As stated above, 18 out of 177 systematic reviews used formal systems for assessing the quality/certainty of the

body of evidence, and nine different approaches were used f by these 18 reviews (see Table 5), including published modifications of existing tools. The majority of reviews ($n=8$) used the GRADE system [12], followed by modified versions of GRADE, namely the Navigation Guide ($n=4$) [64], an approach developed by the World Health Organization (WHO) for air pollution research ($n=1$) [98], and a modified version for environmental health research ($n=1$) [99]. Other approaches were adopted from OHAT ($n=3$) [15], and the International Agency for Research on Cancer’s (IARC) preamble for monographs (2006) [100], among others (see Table 5) [95–97, 101]. Modifications to and deviations from the frameworks were noted [68, 69, 71, 81].

The identified approaches for evidence grading were originally developed either for clinical practice [95–97, 101], or for research on environmental exposures [15, 64, 66, 98, 99, 104], including air pollution [98], and were characterized by highly heterogeneous methodologies. The original GRADE system assigns an initial rating based on study type, where RCTs begin at a “high” quality rating, while observational studies begin as “low”, before considering various criteria (e.g., consistency between studies), to reach a final rating of the body of evidence [106, 107]. The GRADE system as modified by the WHO (for air pollution studies) and the Navigation Guide (developed for environmental health studies, partly based on the U.S. Environmental Protection Agency’s (EPA) criteria for reproductive and developmental toxicity [28]) differ from the original version in that observational studies are initially rated as “moderate” quality, rather than “low”, among other distinguishing features (e.g., additionally calculating 80% prediction intervals to assess heterogeneity) [64, 98, 108].

While the OHAT approach is based on the GRADE system, the initial rating is based on the number of present study-design features, rather than the study type. These include: controlled exposure, exposure prior to outcome, individual outcome data, and comparison group used.

Table 4 Risk of bias/ quality assessment tools for primary studies used by included systematic reviews

Tool used	Number of reviews using approach	Number of reviews that used modifications	Originally developed for	Exposure assessment (original version)	Co-exposures (original version)	Confounding (original version)
Newcastle–Ottawa Scale (NOS) [86]	9 [71, 77–84]	2 [71, 81]	Evaluating non-randomized studies in systematic reviews	Yes (general) ^a	No	Yes
Mustafic et al. (modified NOS) [87]	2 [79, 81]	0	Time-series and case-crossover studies of air pollution exposure	Yes	Possible (“long-term trends”)	Yes
Herzog et al. (modified NOS) [88]	1 [84]	0	Cross-sectional studies (developed for vaccine-related knowledge, attitude, and behavior)	Yes (general) ^a	No	Yes
Modesti et al. (modified NOS) [89]	1 [81]	0	Cross-sectional studies (developed for studies of blood pressure)	Yes (general) ^a	No	Yes
Office of Health Assessment and Translation (OHAT) [15, 66, 90]	4 [73, 75, 79, 81]	3 [73, 75, 81]	Systematic reviews and evidence integrations of environmental health research	Yes	Yes	Yes
Navigation Guide [5, 64]	2 [76, 81]	0	Systematic reviews of human studies in environmental health	Yes	No [§]	Yes
ACROBAT-NRSI [91] (later ROBINS-I) [97]	1 [72]	1 [72]	Non-randomized studies of interventions	Yes (“measurement of intervention”)	Yes (“co-interventions”)	Yes
Agency for Healthcare Research and Quality (AHRQ) [62]	1 [78]	0	Systematic reviews of studies of healthcare interventions	Yes (general) ^a	Yes (“co-interventions”)	Yes
Cochrane tool for RCTs (ROB 1) [92]	1 [78]	0	Systematic reviews of individual RCTs	No	No	Yes (randomization)
Hoy et al. [93]	1 [74]	0	Prevalence studies of low back and neck pain	No	No	No
ROBINS-E (preliminary version) [94]	1 [74]	0	Non-randomized studies of exposures	Yes	Yes	Yes
Centre for Evidence Based-Medicine (CEBM) [95, 96]	1 [70]	Unclear (insufficiently reported)	Clinical decision-making	Yes (general) ^a	No	Yes
Scottish Intercollegiate Guidelines Network (SIGN) [97]	1 [69]	1 [69]	Guideline development of clinical care (all study designs)	Yes (general) ^a	No	Yes
Own criteria [5, 68]	2 [5, 68]	N/A	N/A	Yes	No	Yes

Abbreviations: ACROBAT-NRSI A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions, AHRQ Agency for Healthcare Research and Quality, CEBM Centre for Evidence Based-Medicine, NOS Newcastle Ottawa Scale, OHAT Office of Health Assessment and Translation, ROB Risk of bias, ROBINS-E Risk of bias in non-randomized studies of exposure, ROBINS-I Risk of bias in non-randomized studies of interventions, SIGN Scottish Intercollegiate Guidelines Network

^a Yes (general) refers to tools that include a criterion relating to the validity of the exposure assessment method, without clear relevance to environmental exposures

[§] In other case studies (e.g., on flame retardant exposure) other pollutants were considered under the confounding domain, but not in the version considered herein

Therefore, evidence from observational studies, due to a lack of controlled exposure, will never start higher than “moderate” . Unlike the GRADE system, upgrades

may additionally be given for consistency across different study designs, species, or dissimilar populations, and for “other” reasons [15]. Guidance for subsequently

Table 5 Evidence grading tools used by included systematic reviews

Tool used	Number of reviews using approach	Number of reviews that used published modified versions	Number of reviews that modified/deviated from guidelines	Originally developed for	Initial grading based on study type	Main considerations	Reproductive/children's health-related considerations of directness, heterogeneity, confounding	Timing of exposure and outcome assessment considered?	Considerations of absence of effects
GRADE [12, 98, 99, 102, 103]	8 [71, 72, 77, 78, 80, 82–84]	2 [82, 83]	1 [71]	Assessing quality of evidence in clinical practice	Yes (observational studies start at low, or moderate in modified version by WHO)	Study design, risk of bias among primary studies, inconsistency, imprecision, indirectness, publication bias, magnitude, dose–response association, residual opposing confounding (see modified versions for additional considerations) [99, 100]	Evidence in adults vs. children mentioned under “indirectness” [67] Provided child-specific examples regarding confounding	Timing of outcome assessment (general) considered under “indirectness” [67]	No explicit guidance provided
Navigation Guide [64, 104]	4 [5, 74, 76, 79]	0	0	Systematic reviews in environmental health	Yes (observational studies start at moderate)	Study design, risk of bias among primary studies, inconsistency, imprecision, indirectness, publication bias, magnitude, dose–response association, residual opposing confounding)	Provided child-specific examples regarding confounding	Yes, both considered as part of rating quality and strength of evidence	Yes, in terms of publication bias and in terms of “strength of evidence (i.e., evidence for no effect)”

Table 5 (continued)

Tool used	Number of reviews using approach	Number of reviews that used published modified versions	Number of reviews that modified/deviated from guidelines	Originally developed for	Initial grading based on study type	Main considerations	Reproductive/children's health-related considerations of directness, heterogeneity, confounding	Timing of exposure and outcome assessment considered?	Considerations of absence of effects
Office of Health Assessment and Translation (OHAT) [15, 66]	3 [73, 75, 81]	0	1 [81]	Literature-based risk evaluations of environmental substances	Presence of study-design features considered (Controlled exposures, exposure prior to outcome, individual outcome data, comparison groups used)	Study-design features, risk of bias, inconsistency, imprecision, indirectness, publication bias, magnitude of effect, dose-response, residual confounding, consistency (across species, populations, study designs), other	For inconsistency, the lifestage at exposure and assessment considered	Timing of exposure and outcome mentioned (in general). For inconsistency and indirectness, exposure duration and timing relative to outcome considered	Yes, in translating quality to level of evidence: "evidence of no health effect"
Preamble for monographs by the International Agency for Research on Cancer (IARC) (2006) [100]	1 [68]	0	1 [68]	Identification and evaluation of potential carcinogens (etiology)	No, but limitations of different study designs are considered in relation to research question	Assignment based on study design, quantity, and quality of included studies, statistical power, and consistency between studies Preceded by considerations including exposure assessment methods, temporal effects of exposure, use of biomarkers, criteria for causality (based on Bradford Hill criteria) [105]. In the most recent version from 2019, this is replaced by "considerations for assessing the body of epidemiological evidence" [13, 21, 105]	No	Temporal effects and latency considered prior to evidence grading	Yes (assigned for several unbiased, consistent, precise study results based on exposures covering the full expected range in humans, with adequate length of follow-up available)

Table 5 (continued)

Tool used	Number of reviews using approach	Number of reviews that used published modified versions	Number of reviews that modified/deviated from guidelines	Originally developed for	Initial grading based on study type	Main considerations	Reproductive/children's health-related considerations of directness, heterogeneity, confounding	Timing of exposure and outcome assessment considered?	Considerations of absence of effects
Centre for Evidence-Based Medicine (CEBM) guidelines (2009, 2011) [95, 96]	1 [70]	0	0	Grading quality of evidence for use in clinical decision-making	Yes (experimental studies are rated higher than observational ones)	Study design and quality, precision, consistency, directness, magnitude of effect	No	Length of follow-up considered (in general)	No
Scottish Intercollegiate Guidelines Network (SIGN) (2011) [97]	1 [69]	0	1 [69]	Guideline development in clinical care research	Yes (observational studies cannot score higher than "B")	Previously determined level of evidence (based on study design, risk of bias, and likelihood that "relationship is causal"), consistency of studies, and applicability of the evidence to the target population	No	No	No
The Best Evidence Synthesis (BES) System [101]	1 [78]	0	0	Clinical management guidelines for acute lower back problems	No	Number, relevance, and quality of available studies	No	No	No

Abbreviations: BES Best Evidence Synthesis, CEBM Centre for Evidence Based-Medicine, GRADE Grading of Recommendations, Assessment, Development, and Evaluations, IARC International agency for research on cancer, OHAT Office of Health Assessment and Translation, SIGN Scottish Intercollegiate Guidelines Network

considering quality of evidence across multiple exposures encourages considerations across the entire body of evidence.

In the IARC approach, no initial rating is assigned based on study type, although the appropriateness of different study designs in relation to the research question are considered [100]. Further criteria include study quantity and quality, statistical power, and consistency of findings. This is preceded by considerations including exposure assessment methods, temporality, use of biomarkers, and Hill's criteria for causality [105]. In the most recent version, this is replaced by "considerations for assessing the body of epidemiological evidence" [13, 21, 105].

The Centre for Evidence Based-Medicine (CEBM) and Scottish Intercollegiate Guidelines Network (SIGN) systems again use previously assigned ratings of each included primary study, based on study type and quality, in addition to a subset of the same criteria as GRADE, but with markedly less specific guidance and explanation, compared to the aforementioned systems. The updated version of the SIGN handbook from 2019 now recommends using the GRADE system for grading evidence. The Best Evidence Synthesis (BES) system, developed for research on lower back problems, does not explicitly rate the study type as a criterion, instead presenting a highly abbreviated approach of considering merely the number, relevance, and quality of available studies [101].

In terms of considering aspects of reproductive/ children's environmental health research in the "indirectness", "heterogeneity", or "confounding/ bias" domains, the Navigation Guide, GRADE approach, and OHAT framework all provide brief commentary, in the form of examples or general guidance, while the other tools make no specific reference to reproductive/ children's health (see Table 5). Besides the SIGN and the BES systems, all tools consider the timing of exposure and/ or outcome assessment, although only the Navigation Guide and OHAT approach explicitly address this aspect with regard to reproductive/ children's health research (e.g., developmental stages). Finally, only the Navigation Guide, OHAT approach, and IARC framework provide guidance on assessing "evidence for no effect". Notably, systematic review authors addressed some of these aspects outside of their application of the evidence grading frameworks, in their methods (e.g., by applying relevant inclusion criteria, or by conducting subgroup analyses of different pregnancy trimesters or age groups [73, 76, 78, 83, 84]), or in their discussions.

Discussion

This is to our knowledge the first methodological survey to systematically identify and describe evidence grading systems used in the area of air pollution exposure and

adverse reproductive/ child outcomes. Of note, this is not an overview of recommended, but of practiced methods in the field. Only 18 out of 177 systematic reviews (9.8%) were found to explicitly utilize formal rating systems for bodies of evidence. Such a small proportion suggests that this process is still not common in the field, although an increase was observed after 2015 (see Fig. 2), which is in line with previous findings on evidence grading approaches used in systematic reviews of air pollution exposure [42]. The inconsistency in the approaches used—15 different risk of bias assessment and 9 different evidence grading tools used across 18 reviews— plus the numerous modifications applied, reflect a lack of consensus. The NOS and GRADE system were the most commonly used tools for assessing internal validity and for grading evidence, respectively, discussed further below. It is noteworthy that multiple reviews "borrowed" tools originating from rather unrelated fields (e.g., clinical research on lower back problems), and there was marked heterogeneity in the comprehensiveness and relevance of the employed tools.

Further, numerous systematic reviews cited preceding reviews using the same approach, in reference to their own approach [5, 74, 76–81, 83]. This suggests a "propagated" methods adoption, where systematic review authors use preceding reviews for guidance, possibly leading to the uptake of inappropriate methods [109]. This implicates that the publication of worked examples, as those provided by the Navigation Guide group [110], are essential for further improving the methodological quality of systematic reviews.

Risk of bias assessment

Our findings indicate that systematic review authors use a wide range of approaches for assessing risk of bias/ quality among individual studies, in many cases originating from clinical or other less related fields. 13 reviews were found to use the original or a modified NOS version. The widespread use of the contested NOS may be one of the most "spectacular" examples of the risks of quotation errors and citation copying [109, 111]. Vandenberg et al. recently outlined how flawed exposure assessment methods put public health at risk [27], and this extends to a lack of appropriate and comprehensive evaluations of exposure assessment methods. The NOS includes only a cursory evaluation of exposures assessment methods that is arguably not applicable to environmental exposures. In general, risk of bias/ quality assessment tools have been criticized for focusing on mechanically determining the potential presence of biases, often based on how closely they emulate a hypothetical "target" RCT, rather than their likely direction,

magnitude, and relative importance [18, 112]. Rather than assigning ratings based on study design, assessments should identify the most probable and important biases in relation to the particular population, exposure, and outcome under investigation, rate each study on how effectively it addresses each potential bias, and differences in results across studies should be considered in relation to susceptibility to each bias [14, 112–114].

The iterative development of the ROBINS-E tool [94, 115], which in its preliminary version was criticized for being based on comparisons to the “ideal” RCT, among other limitations [116], but in its final version addressed many of these concerns, including a more nuanced approach to causal inference [117], demonstrates that continuous collaboration between experts and critical appraisal of developing tools is effective and desirable. Also, the WHO has introduced a risk of bias assessment tool for air pollution exposure studies in systematic reviews [118]. In addition, informative evaluations of additional risk of bias tools available for environmental health studies have been presented [119]. Useful interactive data visualization tools exist to facilitate comparison and selection of risk of bias/ methodological quality tools for observational studies of exposures [120], collated on the basis of a preceding systematic review [63].

Evidence grading approaches

In this methodological survey, 16 out of 18 reviews used evidence grading systems that provided higher scores to experimental (vs. non-experimental) studies or related study features. The practice of ranking evidence based on a crude hierarchy of study designs has been criticized [18–21, 23]. For one, experimental studies may be no better at reducing “intractable” confounding, and other approaches (e.g., difference-in-difference) may be much more effectual in addressing particular confounding scenarios [23]. Pluralistic approaches to causal inference, that extend beyond counterfactual and interventionist approaches, have been proposed [21, 22].

Six reviews were found to use the original GRADE system for rating bodies of evidence, for which we noted a lack of consideration with regard to heterogeneities across different developmental stages, a paucity of attention paid to the timing of exposure to environmental risks, and a lack of discussion of evidence for no association or effect, in addition to the default ranking of experimental studies above observational ones. The applicability of the GRADE approach to observational studies has previously been discussed [121, 122], and challenges with rating the body of evidence from observational studies have been reported [123–126], including rating evidence from non-randomized studies as “low” by default, difficulties in assessing complex bodies of evidence consisting of different study designs,

and limited applicability regarding research on etiology, among others [124, 127].

The GRADE working group has proposed the possibility of initially rating evidence from non-randomized studies as “high”, when used in conjunction with risk of bias assessment tools like ROBINS-I [94, 115, 128, 129]. The reasoning is that the lack of randomization will usually lead to rating down by at least two levels to “low”, so ultimately, evidence from observational studies will be rated as “low” with either method [115, 129], hence, this approach is again based on the principle that non-randomized studies are inherently inferior. Other suggestions have been made to start observational studies as “moderate”, as done in the Navigation Guide’s and WHO’s modified versions [64, 98], and expand criteria for upgrading [124]. In prognosis research, the GRADE system has been adapted to start observational studies at “high” [130]. Further developments of the GRADE system for environmental health research, including a recent exploration of how considerations of biological plausibility can be integrated into evidence grading [131], are in progress [99, 132].

Reproductive and children’s environmental health: specific guidance needed

While some of the identified frameworks were found to address selected aspects, concerns persist regarding reproductive/ children’s environmental health research: Firstly, the risk of bias assessment and evidence grading frameworks frequently used by existing systematic reviews often do not explicitly or comprehensively address important aspects, such as vulnerabilities related to developmental stages, considerations of exposure timing and relative dose, etc. [24, 25]. Also, only three evidence grading systems provide any guidance on assessing evidence for the absence of effects. Addressing these points would require considerations of how domains of current evidence grading frameworks are operationalized, including indirectness domains (e.g., timing of exposures, “worst-case” exposure scenarios [27], etc.), heterogeneity (disparities related to social determinants, diverse etiological mechanisms, etc.), and biases specific to research on pregnancy and childhood (e.g., live-birth bias). Some of the identified methodologies offer some insights into how existing frameworks may be adapted [66, 98, 104]: For example, considering null findings, in addition to positive ones, is advised by the Navigation Guide with regard to publication bias, meaning that an excess of null findings, especially from small or industry-sponsored studies, are also of concern [104]. With regard to subsequent assignments of levels of evidence, the OHAT approach notes that due to the intrinsic challenges of proving a negative, concluding “evidence for

no effect," requires high levels of confidence in evidence. Low/ moderate confidence should be considered "inadequate evidence" for absence of effects [15].

Failing to explicitly address the defining features and major characteristics of reproductive and children's environmental health as described above renders non-specific tools such as the NOS and GRADE inadequate for comprehensively evaluating the unique risks posed by environmental exposures during vulnerable developmental stages and across the lifespan. Failing to account for these complexities within evidence grading frameworks may result in an incomplete understanding of the risks posed by environmental exposures during crucial developmental stages. This lack of specification may give rise to invalid assessment results both at the level of primary studies, as well at the level of bodies of evidence, and thereby lead to erroneous conclusions about the certainty of the assessed evidence. This in turn may undermine the formulation of effective policies for protecting reproductive and children's health. Therefore, emphasizing the need for using more specialized frameworks (e.g., ROBINS-E, Navigation Guide, OHAT) for assessing studies on reproductive and children's environmental health is paramount for ensuring accurate findings and interpretations and, ultimately, safeguarding the health of future generations. Altogether, while the addition of new tools or domains may not be needed, further consensus and published direction on how exactly these can be operationalized in the context of reproductive/ children's environmental health may be useful. Providing explicit guidance and clear definitions, promoting the use of more applicable frameworks, and a continued refinement and tailoring of existing frameworks towards reproductive/ children's environmental health research is critical for improving current methodologies [133].

Further evidence grading systems and systematic review frameworks not utilized in the identified reviews

Additional evidence grading systems and systematic frameworks for environmental health research exist, but were not utilized by the identified reviews: In 2006, the EPA published a "Framework for Assessing Health Risks of Environmental Exposures to Children" [134], providing using a "lifestage" perspective. Developing specific assessment criteria during problem formulation is recommended. A weight-of-evidence approach is used, which places emphasis on higher quality studies for evidence grading [134]. Further systematic review frameworks developed for observational studies of etiology or environmental health and toxicology research include the COSMOS-E, COSTER, and SYRINA frameworks, among others. Notably, while provided guidance

on evidence grading generally reflect principles of the GRADE system, specific recommendations as to what tool or approach to use [113, 135], or whether to assign an initial rating based on study type [136], are avoided.

The existence of the approaches described above, as well as those with clear relevance to reproductive/ children's environmental health presented earlier (e.g., ROBINS-E, Navigation Guide, OHAT), together with the limited uptake we identified, suggest that the problem lies less in an absence of appropriate methods, but with their accessibility or implementation. Promoting simple, but not oversimplified, practicable, and specific guidance should be prioritized [109].

Also, calls for child-relevant extensions to the PRISMA checklist- "PRISMA-C" have been made [26, 137, 138], and are currently under development [139]. Specific recommendations regarding risk of bias and evidence assessments could be integrated herein.

Beyond evidence grading- linking evidence and triangulation

Different types of evidence (i.e., human and non-human studies) may be combined into integrated networks of evidences within systematic reviews of environmental health risks [18, 113]. In fact, the Navigation Guide, OHAT, and IARC methodologies provide guidance on integrating evidence from human, animal, and mechanist studies [15, 64, 66, 100, 104].

Further, triangulation (i.e., leveraging differences in evidence from diverse methodological approaches with different biases to strengthen causal inference) has been encouraged for environmental health research [22, 140]. However, guidelines are needed to help researchers integrate triangulation processes into systematic reviews effectively [140].

Implications for policy

Systematic review methods for environmental health research continue to evolve, including at the U.S. federal level, which may have a direct impact on policies to protect reproductive and children's health: Within the EPA, revisions are being made to current systematic review methodologies [141, 142], while proposed changes to the existing "weight-of-evidence" approach, which considers a plethora of different types of evidence, in favor of a "manipulative causation" framework, are being heavily contested [18, 143, 144], and probabilistic risk-specific dose distribution analyses are being piloted, to expand beyond previous threshold-based approaches [145]. This highlights that considerations of evidence assessment methodologies span scientific, political, and legal realms, and carry massive public health implications.

We hope this work can provide a comprehensive overview of the current state of practice in the field, and serve as a starting point for those working on the further refinement or promotion of evidence grading systems for reproductive/ children's environmental health research.

Abbreviations

BES	Best Evidence Synthesis
CEBM	The Centre for Evidence Based-Medicine
EPA	Environmental Protection Agency
GRADE	Grading of Recommendations, Assessment, Development, and Evaluations
IARC	International Agency for Research on Cancer's
NOS	Newcastle Ottawa Scale
OHAT	Office of Health Assessment and Translation
PRIOR	Preferred Reporting Items for Overviews of Reviews
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
SIGN	Scottish Intercollegiate Guidelines Network
RCT	Randomized controlled trial
ROBIS	Risk of bias in systematic reviews
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12940-024-01069-z>.

Additional file 1: Material S1. Preferred Reporting Items for Overviews of Reviews (PRIOR) checklist. **Material S2.** Full electronic search strategies. **Material S3.** Details of ROBIS assessment: Risk of bias assessment results for each systematic review. **Material S4.** Details of ROBIS assessment: Responses to each question of the ROBIS tool.

Permission to reproduce materials from other sources

Not applicable.

Registration and protocol

A registration or protocol for this methodological survey was not prepared.

Authors' contributions

SM and OVE: Conceptualization; SM, AA, and OVE: Data curation and Formal analysis; SM and OVE: Investigation and Methodology; SM: Project administration; SM: Resources and Software; OVE: Supervision; SM, AA, and OVE: Validation; SM: Visualization; Roles/Writing – SM: original draft; SM, AA, and OVE: Writing—review & editing. All authors approved the final version of the article. SM is the guarantor of this work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles (UCLA), 650 Charles E Young Dr S, Los Angeles, CA 90095, USA. ²Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. ³Department of Community Health Sciences, Fielding School of Public Health, University of California, Los Angeles (UCLA), Los Angeles, CA, USA.

Received: 22 September 2023 Accepted: 5 March 2024

Published online: 28 March 2024

References

- Nyadanu SD, Dunne J, Tessema GA, Mullins B, Kumi-Boateng B, Lee Bell M, et al. Prenatal exposure to ambient air pollution and adverse birth outcomes: an umbrella review of 36 systematic reviews and meta-analyses. *Environ Pollut.* 2022;306:119465.
- Sun X, Luo X, Zhao C, Chung Ng RW, Lim CE, Zhang B, et al. The association between fine particulate matter exposure during pregnancy and preterm birth: a meta-analysis. *BMC Pregnancy Childbirth.* 2015;15:300.
- Xing Z, Zhang S, Jiang YT, Wang XX, Cui H. Association between prenatal air pollution exposure and risk of hypospadias in offspring: a systematic review and meta-analysis of observational studies. *Aging (Albany NY).* 2021;13(6):8865–79.
- Lin HC, Guo JM, Ge P, Ou P. Association between prenatal exposure to ambient particulate matter and risk of hypospadias in offspring: a systematic review and meta-analysis. *Environ Res.* 2021;192:110190.
- Lam J, Sutton P, Kalkbrenner A, Windham G, Halladay A, Koustas E, et al. A systematic review and meta-analysis of multiple airborne pollutants and autism spectrum disorder. *PLoS One.* 2016;11(9):e0161851.
- Yamineva Y, Romppanen S. Is law failing to address air pollution? Reflections on international and EU developments. *Rev Eur Comp Int Environ Law.* 2017;26(3):189–200.
- United Nations Environment Programme. Regulating air quality: the first global assessment of air pollution legislation. 2021.
- Neira M, Fletcher E, Brune-Drisse MN, Pfeiffer M, Adair-Rohani H, Dora C. Environmental health policies for women's, children's and adolescents' health. *Bull World Health Organ.* 2017;95(8):604–6.
- Malekinejad M, Horvath H, Snyder H, Brindis CD. The discordance between evidence and health policy in the United States: the science of translational research and the critical role of diverse stakeholders. *Health Res Policy Syst.* 2018;16(1):81.
- Movsisyan A, Dennis J, Rehfuess E, Grant S, Montgomery P. Rating the quality of a body of evidence on the effectiveness of health and social interventions: a systematic review and mapping of evidence domains. *Res Synth Methods.* 2018;9(2):224–42.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71.
- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336(7650):924–6.
- Rothman KJ, Greenland S, Lash TL. *Modern epidemiology.* Vol. 3. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
- Arroyave WD, Mehta SS, Guha N, Schwingl P, Taylor KW, Glenn B, et al. Challenges and recommendations on the conduct of systematic reviews of observational epidemiologic studies in environmental and occupational health. *J Expo Sci Environ Epidemiol.* 2021;31(1):21–30.
- Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. Systematic review and evidence integration for literature-based environmental health science assessments. *Environ Health Perspect.* 2014;122(7):711–8.
- Strickland MJ, Klein M, Darrow LA, Flanders WD, Correa A, Marcus M, et al. The issue of confounding in epidemiological studies of ambient air pollution and pregnancy outcomes. *J Epidemiol Community Health.* 2009;63(6):500–4.
- Woodruff TJ, Parker JD, Darrow LA, Slama R, Bell ML, Choi H, et al. Methodological issues in studies of air pollution and reproductive health. *Environ Res.* 2009;109(3):311–20.

18. Steenland K, Schubauer-Berigan MK, Vermeulen R, Lunn RM, Straif K, Zahm S, et al. Risk of bias assessments and evidence syntheses for observational epidemiologic studies of environmental and occupational exposures: strengths and limitations. *Environ Health Perspect*. 2020;128(9):095002.
19. Rothman KJ. Six Persistent research misconceptions. *J Gen Intern Med*. 2014;29(7):1060–4.
20. Wyr PC. From MARS to MAGIC: the remarkable journey through time and space of the grading of recommendations assessment, development and evaluation initiative. *J Eval Clin Pract*. 2018;24(5):1191–202.
21. Vandembroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol*. 2016;45(6):1776–86.
22. Pearce N, Vandembroucke JP, Lawlor DA. Causal inference in environmental epidemiology: old and new approaches. *Epidemiology*. 2019;30(3):311–6.
23. Pearce N, Vandembroucke JP. Are target trial emulations the gold standard for observational studies? *Epidemiology*. 2023;34(5):614–8.
24. National Research Council (US) Committee on Pesticides in the Diets of Infants and Children. *Pesticides in the diets of Infants and children*. Washington (DC): National Academies Press (US); 1993.
25. Bearer CF. Environmental health hazards: how children are different from adults. *Future Child*. 1995;5(2):11–26.
26. Farid-Kapadia M, Askie L, Hartling L, Contopoulos-Ioannidis D, Bhutta ZA, Soll R, et al. Do systematic reviews on pediatric topics need special methodological considerations? *BMC Pediatr*. 2017;17(1):57.
27. Vandenberg LN, Rayasam SDG, Axelrad DA, Bennett DH, Brown P, Carignan CC, et al. Addressing systemic problems with exposure assessments to protect the public's health. *Environ Health*. 2023;21(1):121.
28. 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):931–7.
29. United States Environmental Protection Agency. *Guidance on Selecting Age Groups for Monitoring and Assessing Childhood Exposures to Environmental Contaminants*. Washington: DC; 2005.
30. Pohl HR, Abadin HG. Chemical mixtures: evaluation of risk for child-specific exposures in a multi-stressor environment. *Toxicol Appl Pharmacol*. 2008;233(1):116–25.
31. Hamra GB, Buckley JP. Environmental exposure mixtures: questions and methods to address them. *Curr Epidemiol Rep*. 2018;5(2):160–5.
32. Goldberg MS, Baumgartner J, Chevrier J. Statistical adjustments of environmental pollutants arising from multiple sources in epidemiologic studies: the role of markers of complex mixtures. *Atmos Environ*. 2022;270:118788.
33. Aronson JK. When I use a word ... The precautionary principle: a definition. *BMJ*. 2021;375:n3111.
34. Kriebel D, Tickner J, Epstein P, Lemons J, Levins R, Loechler EL, et al. The precautionary principle in environmental science. *Environ Health Perspect*. 2001;109(9):871–6.
35. Altman DG, Bland JM. Statistics notes: absence of evidence is not evidence of absence. *BMJ*. 1995;311(7003):485.
36. Smith PRM, Ware L, Adams C, Chalmers I. Claims of 'no difference' or 'no effect' in Cochrane and other systematic reviews. *BMJ Evid Based Med*. 2021;26(3):118–20.
37. Quertemont E. How to statistically show the absence of an effect. *Psychol Belg*. 2011;51(2):109–27.
38. Lakens D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Soc Psychol Personal Sci*. 2017;8(4):355–62.
39. Campbell H, Gustafson P. Conditional equivalence testing: an alternative remedy for publication bias. *PLoS One*. 2018;13(4):e0195145.
40. Sheehan MC, Lam J. Use of systematic review and meta-analysis in environmental health epidemiology: a systematic review and comparison with guidelines. *Curr Environ Health Rep*. 2015;2(3):272–83.
41. Sutton P, Chartres N, Rayasam SDG, Daniels N, Lam J, Maghrbi E, et al. Reviews in environmental health: how systematic are they? *Environ Int*. 2021;152:106473.
42. Sheehan MC, Lam J, Navas-Acien A, Chang HH. Ambient air pollution epidemiology systematic review and meta-analysis: a review of reporting and methods practice. *Environ Int*. 2016;92–93:647–56.
43. Nieuwenhuijsen MJ, Dadvand P, Grollier J, Martinez D, Vrijheid M. Environmental risk factors of pregnancy outcomes: a summary of recent meta-analyses of epidemiological studies. *Environ Health*. 2013;12(1):6.
44. Michel S, Atmakuri A, von Ehrenstein OS. Prenatal exposure to ambient air pollutants and congenital heart defects: an umbrella review. *Environ Int*. 2023;178:108076.
45. Berrang-Ford L, Sietsma AJ, Callaghan M, Minx JC, Scheelbeek PFD, Haddaway NR, et al. Systematic mapping of global research on climate and health: a machine learning review. *Lancet Planet Health*. 2021;5(8):e514–25.
46. Gates M, Gates A, Pieper D, Fernandes RM, Tricco AC, Moher D, et al. Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement. *BMJ*. 2022;378:e070849.
47. Pollock M, Fernandes RM, Becker LA, Featherstone R, Hartling L. What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative meta-summary. *Syst Rev*. 2016;5(1):190.
48. Pollock M, Fernandes R, Becker L, Pieper D, Hartling L. Chapter V: Overviews of Reviews. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 61. Cochrane; 2020. updated September 2020.
49. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *Int J Evid Based Healthc*. 2015;13(3):132–40.
50. Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: planning a relevant research question and objective for an overview. *Syst Rev*. 2018;7(1):39.
51. Lunny C, Brennan SE, McDonald S, McKenzie JE. Toward a comprehensive evidence map of overview of systematic review methods: paper 1—purpose, eligibility, search and data extraction. *Syst Rev*. 2017;6(1):231.
52. KronicMartinic M, Pieper D, Glatt A, Puljak L. Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks. *BMC Med Res Methodol*. 2019;19(1):203.
53. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations Evid Based Med Working Group. *JAMA*. 1995;274(22):1800–4.
54. Goossen K, Hess S, Lunny C, Pieper D. Database combinations to retrieve systematic reviews in overviews of reviews: a methodological study. *BMC Med Res Methodol*. 2020;20(1):138.
55. Salvador-Olivan JA, Marco-Cuenca G, Arquero-Aviles R. Development of an efficient search filter to retrieve systematic reviews from PubMed. *J Med Libr Assoc*. 2021;109(4):561–74.
56. Navarro-Ruan T, Haynes RB. Preliminary comparison of the performance of the National Library of Medicine's systematic review publication type and the sensitive clinical queries filter for systematic reviews in PubMed. *J Med Libr Assoc*. 2022;110(1):43–6.
57. Wright J, Walwyn R. Literature search methods for an overview of reviews ('umbrella' reviews or 'review of reviews'). Leeds Institute of Health Sciences: University of Leeds; 2016.
58. Koster J. PubMed PubReminer Academic Medical Center, University of Amsterdam; 2014. Available from: <https://hgserver2.amc.nl/cgi-bin/miner/miner2.cgi>.
59. Scells H, Zucco G. searchrefiner: a query visualisation and understanding tool for systematic reviews. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino: Association for Computing Machinery; 2018. p. 1939–42.
60. McGill Library. Grey literature and other supplementary search methods; 2022. Available from: <https://libraryguides.mcgill.ca/knowledge-syntheses>.
61. Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225–34.
62. Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, et al. *AHRQ Methods for Effective Health Care Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions*. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.
63. Wang Z, Taylor K, Allman-Farinelli M, Armstrong B, Askie L, Ghersi D, et al. A systematic review: tools for assessing methodological quality of human observational studies. 2019.

64. 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):1007–14.
65. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64(12):1311–6.
66. Office of Health Assessment and Translation (OHAT). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Division of the National Toxicology Program, National Institute of Environmental Health Sciences; 2019.
67. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol*. 2011;64(12):1303–10.
68. Suades-González E, Gascon M, Guxens M, Sunyer J. Air pollution and neuropsychological development: a review of the latest evidence. *Endocrinology*. 2015;156(10):3473–82.
69. Morales-Suárez-Varela M, Peraita-Costa I, Llopis-González A. Systematic review of the association between particulate matter exposure and autism spectrum disorders. *Environ Res*. 2017;153:150–60.
70. Tenero L, Piacentini G, Nosetti L, Gasperi E, Piazza M, Zaffanello M. Systematic review indoor/outdoor not-voluptuary-habit pollution and sleep-disordered breathing in children: a systematic review. *Transl Pediatr*. 2017;6(2):104–10.
71. King C, Kirkham J, Hawcutt D, Sinha I. The effect of outdoor air pollution on the risk of hospitalisation for bronchiolitis in infants: a systematic review. *PeerJ*. 2018;6:e5352.
72. Fu L, Chen Y, Yang X, Yang X, Liu S, Pei L, et al. The associations of air pollution exposure during pregnancy with fetal growth and anthropometric measurements at birth: a systematic review and meta-analysis. *Environ Sci Pollut Res Int*. 2019;26(20):20137–47.
73. Rappazzo KM, Nichols JL, Rice RB, Luben TJ. Ozone exposure during early pregnancy and preterm birth: a systematic review and meta-analysis. *Environ Res*. 2021;198:111317.
74. Ravindra K, Chanana N, Mor S. Exposure to air pollutants and risk of congenital anomalies: a systematic review and meta-analysis. *Sci Total Environ*. 2021;765:142772.
75. Stenson C, Wheeler AJ, Carver A, Donaire-Gonzalez D, Alvarado-Molina M, Nieuwenhuijsen M, et al. The impact of Traffic-Related air pollution on child and adolescent academic Performance: a systematic review. *Environ Int*. 2021;155:106696.
76. Uwak I, Olson N, Fuentes A, Moriarty M, Pulczinski J, Lam J, et al. Application of the navigation guide systematic review methodology to evaluate prenatal exposure to particulate matter air pollution and infant birth weight. *Environ Int*. 2021;148:106378.
77. Gong C, Wang J, Bai Z, Rich DQ, Zhang Y. Maternal exposure to ambient PM(2.5) and term birth weight: a systematic review and meta-analysis of effect estimates. *Sci Total Environ*. 2022;807(Pt 1):150744.
78. Lin LZ, Zhan XL, Jin CY, Liang JH, Jing J, Dong GH. The epidemiological evidence linking exposure to ambient particulate matter with neurodevelopmental disorders: a systematic review and meta-analysis. *Environ Res*. 2022;209:112876.
79. Yu Z, Zhang X, Zhang J, Feng Y, Zhang H, Wan Z, et al. Gestational exposure to ambient particulate matter and preterm birth: an updated systematic review and meta-analysis. *Environ Res*. 2022;212(Pt C):113381.
80. Zhu W, Zheng H, Liu J, Cai J, Wang G, Li Y, et al. The correlation between chronic exposure to particulate matter and spontaneous abortion: a meta-analysis. *Chemosphere*. 2022;286(Pt 2):131802.
81. Ziou M, Tham R, Wheeler AJ, Zosky GR, Stephens N, Johnston FH. Outdoor particulate matter exposure and upper respiratory tract infections in children and adolescents: a systematic review and meta-analysis. *Environ Res*. 2022;210:112969.
82. Blanc N, Liao J, Gilliland F, Zhang JJ, Berhane K, Huang G, et al. A systematic review of evidence for maternal preconception exposure to outdoor air pollution on Children's health. *Environ Pollut*. 2023;318:120850.
83. Liang W, Zhu H, Xu J, Zhao Z, Zhou L, Zhu Q, et al. Ambient air pollution and gestational diabetes mellitus: an updated systematic review and meta-analysis. *Ecotoxicol Environ Saf*. 2023;255:114802.
84. Tandon S, Grande AJ, Karamanos A, Cruickshank JK, Roever L, Mudway IS, et al. Association of ambient air pollution with blood pressure in adolescence: a systematic-review and meta-analysis. *Curr Probl Cardiol*. 2023;48(2):101460.
85. McGuinness LA, Higgins JPT. Risk-of-bias Visualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Res Synth Methods*. 2021;12(1):55–61.
86. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. Newcastle-Ottawa quality assessment scales. Available from: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
87. Mustafić H, Jabre P, Caussin C, Murad MH, Escolano S, Tafflet M, et al. Main air pollutants and myocardial infarction: a systematic review and meta-analysis. *JAMA*. 2012;307(7):713–21.
88. Herzog R, Álvarez-Pasquin MJ, Díaz C, Del Barrio JL, Estrada JM, Gil Á. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? a systematic review. *BMC Public Health*. 2013;13(1):154.
89. Modesti PA, Reboldi G, Cappuccio FP, Agemang C, Remuzzi G, Rapi S, et al. Panethnic differences in blood pressure in Europe: a systematic review and meta-analysis. *PLoS One*. 2016;11(1):e0147601.
90. National Institute for Environmental Health Sciences. OHAT risk of bias rating tool for human and animal studies. 2015.
91. Sterne J, Higgins J, Reeves B, on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for NonRandomized Studies of Interventions (ACROBAT-NRSI); 2014.
92. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
93. Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. *J Clin Epidemiol*. 2012;65(9):934–9.
94. Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, et al. Evaluation of the risk of bias in non-randomized studies of interventions (ROBINS-I) and the 'target experiment' concept in studies of exposures: rationale and preliminary instrument development. *Environ Int*. 2018;120:382–7.
95. Oxford Centre for Evidence-Based Medicine. Levels of Evidence 2009. Available from: <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009>.
96. OCEBM Levels of Evidence Working Group. The Oxford 2011 Levels of Evidence. Oxford Centre for Evidence-Based Medicine; 2011.
97. Scottish Intercollegiate Guidelines Network. SIGN 50. In: Healthcare improvement Scotland, editor. Edinburgh: Elliott House; 2011.
98. WHO Global Air Quality Guidelines Working Group on Certainty of Evidence Assessment. Approach to assessing the certainty of evidence from systematic reviews informing WHO global air quality guidelines. 2020. Available from: <https://ars.els-cdn.com/content/image/1-s2.0-S0160412020318316-mmc4.pdf>.
99. Morgan RL, Thayer KA, Bero L, Bruce N, Falck-Ytter Y, Ghersi D, et al. GRADE: assessing the quality of evidence in environmental and occupational health. *Environ Int*. 2016;92–93:611–6.
100. World Health Organization International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans- Preamble. France: Lyon; 2006.
101. Bigos SJ, Richard Bowyer RO, Richard Braen G. Acute low back problems in adults, AHCPR guideline no. 14. *J Manual Manipulative Ther*. 1996;4(3):99–111.
102. Dijkers M. Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development. *KT Update*. 2013;1(5):1–9.
103. GRADE Working Group. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. In: Holger Schünemann, Jan Brozek, Gordon Guyatt, Oxman A. 2013.
104. University of California San Francisco Program on Reproductive Health and the Environment. Navigation Guide Protocol for Rating the Quality and Strength of Human and Non-Human Evidence 2012. Available from: <https://prhe.ucsf.edu/sites/g/files/tksr341f/Instructions%20to%20Authors%20for%20GRADING%20QUALITY%20OF%20EVIDENCE.pdf>.
105. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58(5):295–300.
106. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401–6.

107. The GRADE Working Group. GRADE Handbook Schünemann H BJ, Guyatt G, Oxman A. 2013.
108. Pérez Velasco R, Jarosińska D. Update of the WHO global air quality guidelines: systematic reviews - An introduction. *Environ Int.* 2022;170:107556.
109. Ioannidis JPA. Massive citations to misleading methods and research tools: Matthew effect, quotation error and citation copying. *Eur J Epidemiol.* 2018;33(11):1021–3.
110. Johnson PJ, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, et al. The Navigation Guide—Evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. *Environ Health Perspect.* 2014;122(10):1028–39.
111. Stang A, Jonas S, Poole C. Case study in major quotation errors: a critical commentary on the Newcastle-Ottawa scale. *Eur J Epidemiol.* 2018;33(11):1025–31.
112. Savitz DA, Wellenius GA, Trikalinos TA. The problem with mechanistic risk of bias assessments in evidence synthesis of observational studies and a practical alternative: assessing the impact of specific sources of potential bias. *Am J Epidemiol.* 2019;188(9):1581–5.
113. Dekkers OM, Vandenbroucke JP, Cevallos M, Renehan AG, Altman DG, Egger M. COSMOS-E: guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. *PLoS Med.* 2019;16(2):e1002742-e.
114. Schubauer-Berigan MK, Richardson DB, Fox MP, Fritschi L, Canu IG, Pearce N, et al. IARC-NCI workshop on an epidemiological toolkit to assess biases in human cancer studies for hazard identification: beyond the algorithm. *Occup Environ Med.* 2023;80(3):119–20.
115. Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, et al. A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE. *Environ Int.* 2019;122:168–84.
116. Bero L, Chartres N, Diong J, Fabbri A, Ghersi D, Lam J, et al. The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures. *Syst Rev.* 2018;7(1):242.
117. ROBINS-E Development Group. Risk Of Bias In Non-randomized Studies - of Exposure (ROBINS-E). Launch version 20 June 2023 2023. Available from: <https://www.riskofbias.info/welcome/robins-e-tool>.
118. WHO Global Air Quality Guidelines Working Group on Risk of Bias Assessment. Risk of bias assessment instrument for systematic reviews informing WHO global air quality guidelines 2020. Available from: <https://apps.who.int/iris/handle/10665/341717>.
119. Rooney AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Boyles AL, et al. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. *Environ Int.* 2016;92–93:617–29.
120. Taylor KW, Wang Z, Walker VR, Rooney AA, Bero LA. Using interactive data visualization to facilitate user selection and comparison of risk of bias tools for observational studies of exposures. *Environ Int.* 2020;142:105806.
121. Schünemann H, Hill S, Guyatt G, Akl EA, Ahmed F. The GRADE approach and Bradford Hill's criteria for causation. *J Epidemiol Community Health.* 2011;65(5):392–5.
122. Durrheim DN, Reingold A. Modifying the GRADE framework could benefit public health. *J Epidemiol Community Health.* 2010;64(5):387.
123. Montgomery P, Movsisyan A, Grant SP, Macdonald G, Rehfuess EA. Considerations of complexity in rating certainty of evidence in systematic reviews: a primer on using the GRADE approach in global health. *BMJ Glob Health.* 2019;4(Suppl 1):e000848.
124. Rehfuess EA, Akl EA. Current experience with applying the GRADE approach to public health interventions: an empirical study. *BMC Public Health.* 2013;13(1):9.
125. Hilton Boon M, Thomson H, Shaw B, Akl EA, Lhachimi SK, López-Alcalde J, et al. Challenges in applying the GRADE approach in public health guidelines and systematic reviews: a concept article from the GRADE Public Health Group. *J Clin Epidemiol.* 2021;135:42–53.
126. Zähringer J, Schwingshackl L, Movsisyan A, Stratil JM, Capacci S, Steinacker JM, et al. Use of the GRADE approach in health policy-making and evaluation: a scoping review of nutrition and physical activity policies. *Implement Sci.* 2020;15:1–18.
127. Norris SL, Bero L. GRADE methods for guideline development: time to evolve? *Ann Intern Med.* 2016;165(11):810–1.
128. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* 2016;355:i4919.
129. Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *J Clin Epidemiol.* 2019;111:105–14.
130. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ.* 2015;350:h870.
131. Whaley P, Piggott T, Morgan RL, Hoffmann S, Tsaion K, Schwingshackl L, et al. Biological plausibility in environmental health systematic reviews: a GRADE concept paper. *Environ Int.* 2022;162:107109.
132. Morgan RL, Beverly B, Ghersi D, Schünemann HJ, Rooney AA, Whaley P, et al. GRADE guidelines for environmental and occupational health: a new series of articles in *Environment International*. *Environ Int.* 2019;128:11–2.
133. Moher D, Stewart L, Shekelle P. Establishing a new journal for systematic review products. *Syst Rev.* 2012;1(1):1.
134. U.S. Environmental Protection Agency (EPA). A framework for assessing health risks of environmental exposures to children. Washington, DC: National Center for Environmental Assessment; 2006.
135. Whaley P, Aiassa E, Beausoleil C, Beronius A, Bilotta G, Boobis A, et al. Recommendations for the conduct of systematic reviews in toxicology and environmental health research (COSTER). *Environ Int.* 2020;143:105926.
136. Vandenberg LN, Ågerstrand M, Beronius A, Beausoleil C, Bergman Å, Bero LA, et al. A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environ Health.* 2016;15(1):74.
137. Farid-Kapadia M, Joachim KC, Balasingham C, Clyburne-Sherin A, Offringa M. Are child-centric aspects in newborn and child health systematic review and meta-analysis protocols and reports adequately reported?—two systematic reviews. *Syst Rev.* 2017;6(1):31.
138. Cramer K, Wiebe N, Moyer V, Hartling L, Williams K, Swingler G, et al. Children in reviews: methodological issues in child-relevant evidence syntheses. *BMC Pediatr.* 2005;5:38.
139. Kapadia MZ, Askie L, Hartling L, Contopoulos-Ioannidis D, Bhutta ZA, Soll R, et al. PRISMA-Children (C) and PRISMA-Protocol for Children (P-C) Extensions: a study protocol for the development of guidelines for the conduct and reporting of systematic reviews and meta-analyses of newborn and child health research. *BMJ Open.* 2016;6(4):e010270.
140. Lawlor DA, Tilling K, Davey SG. Triangulation in aetiological epidemiology. *Int J Epidemiol.* 2017;46(6):1866–86.
141. National Academies of Sciences E, and Medicine. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press; 2021.
142. United States Environmental Protection Agency. Draft Protocol for Systematic Review in TSCA Risk Evaluations 2022. Available from: <https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/draft-protocol-systematic-review-tsca-risk-evaluations>.
143. Goldman GT, Dominici F. Don't abandon evidence and process on air pollution policy. *Science.* 2019;363(6434):1398–400.
144. Richmond-Bryant J. In defense of the weight-of-evidence approach to literature review in the integrated science assessment. *Epidemiology.* 2020;31(6):755–7.
145. McPartland J. Finally—EPA takes steps to unify its approach to the evaluation of chemicals for cancer and non-cancer endpoints. *Environ Defense Fund.* 2021. <https://blogs.edf.org/health/2021/07/13/final-ly-epa-takes-steps-to-unify-its-approach-to-the-evaluation-of-chemicals-for-cancer-and-non-cancer-endpoints/>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.