

The relationship between maternal glomerular filtration rate and fetal growth: a systematic review of the evidence

Protocol Information

Authors: Hanna M Vesterinen, Paula I Johnson, Dylan S Atchley, Patrice Sutton, Juleen Lam, Marya G Zlatnik, Saunak Sen & Tracey J Woodruff

Contact Person: Hanna Vesterinen, UCSF Program on Reproductive Health and the Environment, 1330 Broadway St., Suite 1135, Oakland, CA 94612. E-mail vesterinenh@obgyn.ucsf.edu. Phone 510-350-1247.

What's New: Applying a systematic and transparent methodology in the field of environmental health science to ascertain the quality and strength of the scientific evidence.

History: This document is part of the demonstration of proof of concept, by the UCSF of *The Navigation Guide*.(Woodruff and Sutton, 2011).

Publication: Our intent is to publish the review in a peer-reviewed journal.

PROTOCOL

Background. Both human and animal studies have shown that impaired prenatal growth impacts negatively on postnatal development and later life disease and morbidity (Barker, 1995). It is therefore important to discern the causes of impaired prenatal growth, and one such factor may be maternal exposure to certain environmental chemicals. A difficulty faced by researchers investigating such exposures is that human studies are largely observational for ethical reasons. Consequently, discerning causality between the two variables is not trivial. Here we describe a specific example of this issue.

Our research group recently conducted a systematic review and meta-analysis on the impact of perfluooctanoic acid (PFOA) on fetal growth, concluding from consideration of both observational human and experimental animal studies that PFOA exposure prior to or during pregnancy is inversely associated with birth weight (Johnson et al., 2013, Koustas et al., 2013). However researchers have hypothesized that reduced fetal growth may lead to an increased measured load of maternal PFOA (termed “reverse causality” (Whitworth et al., 2012)). Pregnancy is associated with both plasma volume expansion (PVE) to accommodate the growing fetus and an increase in kidney function, including glomerular filtration rate (GFR; Chang and Streitman, 2012, Chesnutt, 2004); therefore central to the reverse causality hypothesis is that reduced fetal growth causes a reduction in GFR either directly or indirectly, through a lower plasma volume expansion, and therefore a lower filtered load of PFOA resulting in a higher PFOA concentration measured in maternal or cord blood (Figure 1B).

Systematic reviews are becoming increasingly common in environmental health research. One approach to systematic review in this field has been documented in The Navigation Guide which aims to bridge the gap between clinical and environmental health sciences (Woodruff et al., 2010). The method consists of transparent and reproducible steps to generate a statement on the toxic effects of a specific chemical on reproductive outcomes. In this review we will adapt and apply the first three steps of the Navigation Guide systematic review methodology (Woodruff et al., 2010): (1) specify the study question; (2) select the evidence; and (3) rate the quality and strength of the evidence. We will adapt the Navigation Guide’s ratings about the toxicity of an exposure to apply to a situation that is not necessarily toxic, i.e. the relationship between fetal growth and GFR.

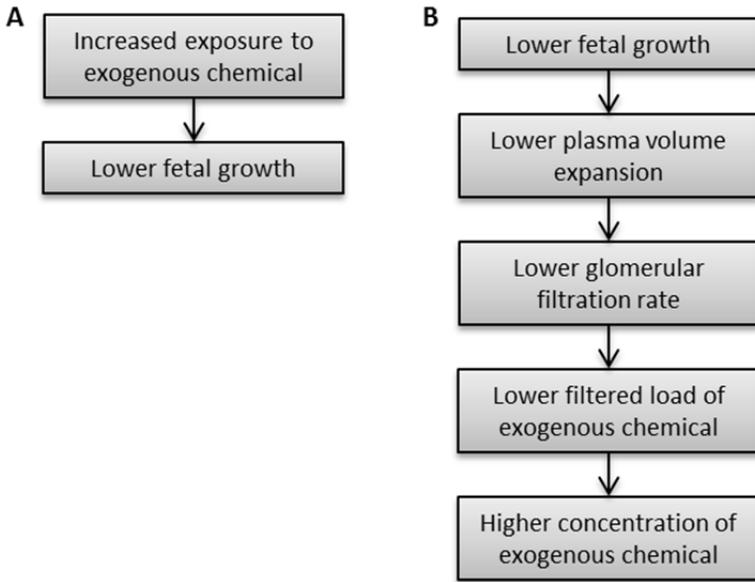


Figure 1. A flow diagram outlining two potential hypotheses for the relationship between exogenous chemicals and fetal growth. Increased exposure to exogenous chemicals may cause changes in fetal growth (A), or changes in fetal growth may affect the concentration of measurable chemical due to changes in the maternal plasma volume and subsequent changes in maternal glomerular filtration rate (B). The present review evaluates the evidence in support of hypothesis B.

Objectives

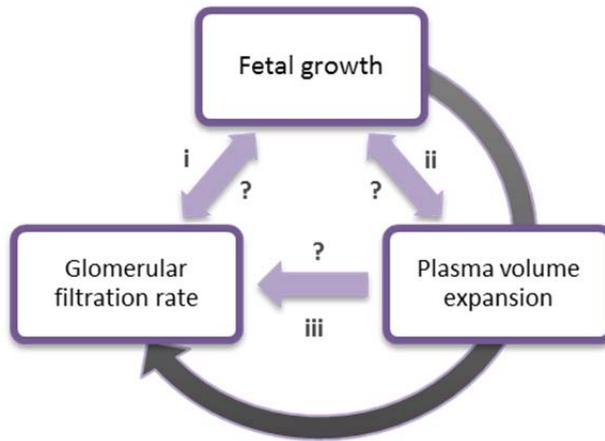
- Our objective is to answer the question: 1) “Is there an association between fetal growth and maternal glomerular filtration rate (GFR) in humans” Specifically we set out to assess the plausibility of the hypothesis that reduced fetal growth causes a reduction in GFR either directly or indirectly via a reduction in PVE. We included only evidence from observational or experimental studies on mammals (human and non-human) and these constituted our streams of evidence which we considered to be directly applicable to answer the study question.

To assess this hypothesis we require:

- Evidence of an association between: reduced fetal growth and GFR (Figure 2A, comparison i) *or* between fetal growth and PVE (Figure 2B comparison ii) *and* between PVE and GFR (Figure 2B, comparison iii).

We will evaluate both human and non-human mammalian models of the relationships described in Figure 2. We (i) assess the breadth of the evidence; (ii) assess the quality of individual studies within each stream of evidence; and (iii) assess the strength of the human and non-human mammalian evidence separately according to one of the following four ratings: 1. sufficient; 2. limited; 3. inadequate; or 4. evidence of the lack of a relationship between fetal growth and GFR; and iv) integrated the strength of the human and non-human mammalian evidence to generate an overall conclusion on the strength of the association between fetal growth and GFR according to one of the following five statements: “known to be related,” “probably related,” “possibly related,” “not classifiable” or “probably not related.”

A



B

	Independent variable		Dependent Variable
	"Exposure"	Comparator	Outcome
i.	Higher value for fetal growth	Lower value for fetal growth	Δ GFR
ii.			Δ PVE
iii.	Higher value for PVE	Lower value for PVE	Δ GFR

Figure 2. A schematic of the three relationships under review. To assess the relationship between fetal growth and GFR we considered both the direct evidence (A, comparison i) and the indirect evidence via PVE, by assessing the relationship between fetal growth and PVE (B, comparison ii), and PVE and GFR (B, comparison iii). Additionally we always considered variables downstream on the circular grey arrow to be the dependent variable in post-hoc analyses of study data, as outlined in the table (C). We will adapt our method from The Navigation Guide (Woodruff et al., 2010) which describes steps to review the association between exposure to potentially toxic chemicals and adverse health outcomes; we therefore interchange the term “exposure” in the Navigation Guide with higher and lower values for fetal growth or PVE.

Methods

Criteria for Selecting Studies

Studies that are eligible for review will address the study question and the characteristics outlined in the following “PECO” aid.

PECO

“PECO” (Participants, Exposure, Comparator, Outcomes) is an aid used to formulate an answerable question in a systematic review.

Participants: humans or mammals studied during or till term pregnancy.

Exposure: for studies where change in fetal growth and any other variable is being assessed, we will take higher levels of fetal growth relative to the comparator group as our exposure group; for studies which report change in maternal PVE and maternal GFR we will take higher levels of PVE relative to the comparator group as our exposure group..

Comparator: for studies where change in fetal growth and any other variable is being assessed, we will take lower levels fetal growth relative to the exposure group as our comparator group; for studies which report change in maternal PVE and maternal GFR is being assessed we will take lower levels of PVE relative to the exposure group as our comparator group..

Outcome: for studies where change in fetal growth and maternal plasma volume expansion, we will take change in maternal plasma volume as the outcome; for studies which report change in fetal growth and maternal GFR, we will take change in maternal GFR as the outcome (see Table 1). In addition we will consider both direct and indirect measures of maternal GFR (e.g. inulin or creatinine clearance), fetal growth (e.g. birth weight, head circumference, ponderal index), and plasma volume expansion (e.g. haemodilution or blood volume expansion). We will also include absolute plasma volume or blood volume as surrogate markers for PVE if maternal size is either adjusted for (e.g. multivariate analysis) or sufficiently similar between the exposure and comparator group (e.g. statistical non-significance).

Search Methods

We will search three online databases (Biosis Previews, ISI Web of Science and Pubmed and Embase) using the search terms described below. Our search will not be limited by language or publication date.

Electronic Searches

Pubmed

To assist in the development of a list of terms relevant to our PubMed search strategy, we conducted an analysis of the Medical Subject Headings (MeSH), title, and abstract text terms in a non-random group of publications which address this question. Specifically we searched Pubmed for publications with “glomerular filtration rate” AND pregnancy in the title; this yielded 20 hits from which we extracted relevant terms relating to glomerular filtration rate from the titles, abstracts and MeSH headings. This analysis produced a list of common and unique terms from these papers that were incorporated into the following title and abstract search strategy:

<p>#1 <i>Glomerular filtration</i></p>	<p>Glomerular filtration rate* [tiab] OR glomerular ultrafiltration [tiab] or glomerulofiltration rate* [tiab] OR glomerulus filtration rate* [tiab] OR glomerular plasma flow rate* [tiab] OR glomerular ultrafiltrate [tiab] OR ((creatinine [tiab] OR inulin OR iohexol OR cystatin-c OR s-creatinine OR s-urate OR Omnipaque) AND (clearance [tiab] OR excretion [tiab] OR plasma [tiab] OR serum [tiab] OR level [tiab])) OR “MDRD formula” [tiab] OR Cockcroft-Gault [tiab] OR (inulin [tiab] AND infusion* [tiab]) OR “eGFR(cystc)” [tiab] OR “eGFR(MDRD)” [tiab] OR 51cr-EDTA [tiab] OR 100/SCr [tiab] OR glomerulotubular mechanisms [tiab] OR glomerular filtration rate [mh] OR glomerular filtration rate/physiology [mh]</p>
<p>#2 <i>Plasma volume expansion</i></p>	<p>(Plasma volume* [tiab] OR haemodilution [tiab] OR hemodilution [tiab] OR blood volume [tiab] OR “hypervolemia of pregnancy” [tiab] OR (physiologic [tiab] AND (anemia [tiab] OR anaemia [tiab] OR anemic [tiab] OR anaemic [tiab]))) AND (pregnancy [tiab] OR pregnant [tiab] OR pregnancy [mh])</p>
<p>#3 <i>Pregnancy/fetal growth</i></p>	<p>(fetal [tiab] OR fetus [tiab] OR embryo [tiab] OR embryonic [tiab] OR infant [tiab] OR baby [tiab] OR babies [tiab]) OR birth [tiab] OR intrauterine [tiab]) AND (weight [tiab] OR development [tiab] OR size [tiab] OR length [tiab] OR growth [tiab] OR circumference [tiab] OR restriction [tiab]) OR birthweight [tiab] OR IUGR [tiab] OR</p>

	“low birth weight” [tiab] OR (small [tiab] AND gestational age [tiab]) OR “intrauterine growth restriction” [tiab] OR “intrauterine growth retardation” [tiab] OR “fetal growth restriction” OR fetal growth retardation [mh] OR fetal development [mh] OR embryonic and fetal development [mh] OR infant, low birth weight [mh] OR birth weight [mh]
<i>Search: (#1 AND #2) OR (#1 AND #3) OR (#2 AND #3)</i>	

ISI Web of Science and Biosis Previews

To develop a Web of Science and Biosis Previews search filter, we modified the PubMed search filter. This consisted of removing all MeSH terms and formatting the search terms for the Web of Science and Biosis Previews database. Specifically, we will perform a topic search in Web of Science, which will search all keywords, titles, and abstracts for the following search terms:

<i>#1 Glomerular filtration</i>	TS=(Glomerular filtration rate* OR glomerular ultrafiltration OR glomerulofiltration rate* OR glomerulus filtration rate* OR glomerular plasma flow rate* OR glomerular ultrafiltrate OR MDRD fomula* OR “modification of diet in renal disease” OR ((creatinine OR inulin OR iohexol OR cystatin-c OR Omnipaque OR S-Creatinine OR S-urate) NEAR/5 (clearance OR excretion OR plasma OR serum OR level)) OR Cockcroft-Gault OR (inulin NEAR/2 infusion*) OR “eGFR(cystc)” OR “eGFR(MDRD)” OR 51cr-EDTA OR 100/SCr OR glomerulotubular mechanisms)
<i>#2 Plasma volume expansion</i>	TS=(((Plasma NEAR/2 volume*) OR plasma-volume* OR haemodilution OR hemodilution OR (blood NEAR/2 volume) OR “hypervolemia of pregnancy” OR (physiologic AND (anemia OR anaemia OR anemic OR anaemic))) AND (pregnancy OR pregnant))
<i>#3 Pregnancy/fetal growth</i>	TS=((fetal OR fetus OR embryo OR embryonic OR infant OR baby OR babies OR birth OR intrauterine) NEAR/4 (weight OR development OR size OR length OR growth OR circumference OR restriction OR small) OR birthweight OR IUGR OR low birth weight OR “intrauterine growth restriction” OR “intrauterine growth retardation” OR “fetal growth restriction”)
<i>Search: (#1 AND #2) OR (#1 AND #3) OR (#2 AND #3)</i>	

Embase

Embase (www.embase.com) is a database that is complementary to PubMed, particularly when looking for international coverage of biomedical sciences, and includes over 5 million records not indexed in PubMed. To develop an Embase search filter, we will modify the PubMed search filter. This will consist of removing all MeSH terms and formatting the search terms for the Embase database. Specifically, we will perform a topic search in Embase, which will search all titles and abstracts for search terms, with no mapping to preferred terminology and no explosion of terms.

<i>#1 Glomerular filtration</i>	(Glomerular filtration rate OR glomerular ultrafiltration OR glomerulus filtration rate OR glomerular plasma flow rate OR glomerular ultrafiltrate OR ((creatinine OR inulin OR iohexol OR cystatin-c OR Omnipaque OR S-Creatinine OR S-urate) adj5 (clearance OR excretion OR plasma OR serum OR level)) OR MDRD fomula OR Cockcroft-Gault OR (inulin adj2 infusion) OR 51cr-EDTA OR glomerulotubular mechanisms).ti,ab.
<i>#2 Plasma volume expansion</i>	((Plasma adj2 volume) OR plasma-volume OR haemodilution OR hemodilution OR (blood adj2 volume) OR hypervolemia OR (physiologic AND (anemia OR anaemia OR anemic OR anaemic))) AND (pregnancy OR pregnant)).ti,ab.
<i>#3 Pregnancy/fetal growth</i>	((fetal OR fetus OR embryo OR embryonic OR infant OR baby OR babies OR birth OR intrauterine) adj4 (weight OR development OR size OR length OR growth OR circumference OR restriction OR small) OR birthweight OR IUGR OR low birth weight OR intrauterine growth restriction OR intrauterine growth retardation OR fetal growth restriction).ti,ab.
<i>Search: (#1 AND #2) OR (#1 AND #3) OR (#2 AND #3)</i>	

Snowball Search

In addition to the original search strategy described above we will hand search the reference list of all articles which are included in the systematic review for additional relevant articles.

Study Selection Criteria

All search results will be imported or manually entered into EndNote (ThomsonReuters) reference management software. We will eliminate duplicate references, reviews and book chapters or editorials before we begin evaluating the eligibility of the studies identified.

Two reviewers (HV, DA) will independently conduct a title and abstract screen of the literature search results to determine whether a reference meets the criteria for inclusion. The same two individuals will then evaluate the entire text of the remaining articles to identify studies meeting the inclusion criteria. A third reviewer (PJ) will be brought in to resolve any discrepancies between the reviewers resulting from each step of the review process. The review of selection criteria for inclusion will be performed using a structured form designed and accessed in DistillerSR (Evidence Partners; available at: <http://www.systematic-review.net>), an online application designed specifically for the screening and data extraction phases of a systematic review.

Reports of human or mammalian studies in any language or date will be eligible for inclusion. All reports that evaluate two or more of the following: GFR or PVE, either directly or indirectly, during pregnancy or fetal growth either during pregnancy or at term will be eligible for inclusion.

Citation titles and abstracts will be screened and excluded if one or more of the following criteria are met:

1. Article does not contain original data or observations (e.g. a review article).
2. Study subjects were not humans or mammals.
3. Two or more of the following variables: fetal growth, PVE or GFR, were not measured directly or estimated for the study subjects.
4. Estimates for plasma volume expansion or glomerular filtration rate were not made during pregnancy and fetal weight was not measured at term.
5. Exclude for other reason (explanation required).

If an abstract for a citation is not within the database and the citation cannot be excluded based on title alone, or if it was not clear if the study was relevant, it will be tagged as “unclear” and included for full-text review.

Articles which progress from the title and abstract screening level will have their full-text screened and excluded if any of the criteria from the title/abstract screening level are met, as well as any of the following additional exclusion criteria:

1. Pregnancies are complicated by other diseases such as pre-eclampsia, or births are spontaneously aborted. and where data from a normal-pregnancy comparison group cannot be used to assess the relationship between two or more variables (e.g. individual participant data, or measurements made at multiple time points during pregnancy).
2. Exogenous drugs are taken during or prior to pregnancy and data from a normal pregnancy comparison group cannot be used to assess the relationship between two or more variables (e.g. individual participant data, or measurements made at multiple time points during pregnancy).
3. Pregnancies are complicated by diet such as under-nutrition or overfeeding and where data from a normal-pregnancy comparison group cannot be used to assess the relationship between two or more variables (e.g. individual participant data, or measurements made at multiple time points during pregnancy).
4. PVE or GFR measurements are not from the mother or not made during pregnancy or at term.

For articles (including non-English articles) which were not available in the database we attempted to obtain these from a broad internet search or through inter-library loans. Articles which were in a foreign language were translated to determine eligibility.

Data Collection

One reviewer (HV) will independently extract data pertaining to study characteristics, exposure assessment and outcome measurements from all of the included articles; one additional reviewer (DA) will check all the outcome measurements data for accuracy and a third reviewer (JL) will independently extract and analyze outcome measurements data from 10% of the included studies. Two reviewers (HV and DA) will independently extract data pertaining to reporting quality and the risk of bias. The data extraction will be performed using a custom form in DistillerSR. This form was created by combining aspects of a variety of checklists and criteria (Hooijmans et al., 2010, Kilkenny et al., 2010, Guyatt et al., 2011, Higgins).

Risk of Bias Determination

Risk of bias will be assessed using domains similar to the Cochrane Collaboration's 'Risk of Bias' tool that address selection bias, performance bias, attrition bias, detection bias, and reporting bias (Higgins et al., 2008). We have modified terminology and concepts in these domains to make appropriate for human observational studies and experimental or observational mammalian studies and included concepts from

the AHRQ's Methods Guide for Comparative Effectiveness Reviews (Viswanathan et al., 2008) (see Appendix I for detailed instructions on making a risk of bias determination for review authors).

Informed by empirical data from meta-analyses conducted on pharmacological treatments (Roseman et al., 2011), we will also assess funding source and declared conflicts of interest as sources of bias. Two review authors (HV, DA) will independently make risk of bias determinations for each study across all domains and then compare their results. Any discrepant results that cannot be resolved between these two authors will then be reviewed by 2 other co-authors. If, upon further discussion the 4 co-authors cannot reach agreement on an appropriate risk of bias determination for a particular domain, the more conservative judgment will be selected (e.g. if one reviewer makes a judgment of 'yes' and the other makes a judgment of 'probably yes', the 'probably yes' judgment will be used). Relevant domains and the streams of evidence they pertain to will include:

- Sequence generation (mammalian studies)
- Allocation concealment (mammalian studies) Blinding of outcome assessment (personnel and outcome assessors; mammalian studies and human studies)
- Blinded assessment
- Participant selection (human studies)
- Confounding (human studies)
- Incomplete outcome data (mammalian and human studies)
- Selective outcome reporting (mammalian and human studies)
- Other potential threats to validity (mammalian and human studies)
- Financial conflict of interest (mammalian and human studies)

Publication bias will be addressed by: (1) implementing a comprehensive search of the literature using multiple sources and methods in order to identify all published and unpublished studies that meet the eligibility criteria; and (2) as possible, using funnel plot analysis and/or other statistical analyses of the studies included in the systematic review.

Analysis

Characteristics from each study will be compiled and reviewed to establish comparability between studies or to identify data transformations necessary to ensure such comparability. Several characteristics will be evaluated across all eligible studies. Examples of these include:

- Study design (e.g. cross-sectional, retrospective cohort)
- Details on how participants were classified into exposure groups, if any (e.g. quartiles of birth weight)
- Outcome reported (e.g. birth weight, PVE or GFR, or other measurement of these)
- Type of outcome data (e.g. continuous or dichotomous)

These characteristics will be assessed by two reviewers to determine comparability between studies and to identify any major heterogeneity concerns that render the studies unable to combine in a meta-analysis. A determination will then be made whether the inclusion of studies in a meta-analysis of the data is appropriate. Situations in which it may not be appropriate to include a study are: data on exposure or outcome are too different to be combined; there are concerns about high risk of bias; or other circumstances which may indicate that averaging study results would not produce meaningful results. Although certain studies may be excluded from a meta-analysis based on these concerns, sensitivity analyses can and should be conducted that include studies when reasons for exclusion are in question. For example, if a study is excluded from the meta-analysis because of differing methods of exposure measurement, the effect of including the differing study on the meta-analysis result can be examined, and the heterogeneity statistics may help to support the exclusion of the study in question. Additionally, all eligible studies (not only those combined in a meta-analysis) will be reviewed and included for evaluating and rating the quality/strength of the human and mammalian evidence.

Data extracted from eligible studies will be imported to statistical software for analysis. A “digital ruler” (Universal Desktop Ruler) will be used when necessary to estimate data points only presented in graphs. Where possible the following fields from the data extraction will be used in the meta-analysis:

- Mean estimates of fetal growth, PVE or GFR
- Precision estimate for outcome measurements for each exposure group

If the type of exposure data differs among studies (e.g. the measure of birth weight), the data will be normalized when possible to the same metric. If there is a mixture of outcome measurements such that some data are expressed as an empirical or percent change in outcome measurement while other data are

expressed as a prevalence of the outcome (such as prevalence of low birth weight), then the possibility of including both types of data will be explored. The results from subgroup, combined and any sensitivity analyses will be compared.

For studies which dichotomize the independent variable (see Table 1) and report a mean score and measure of precision for the dependent variable we will calculate effect sizes as difference in means. Where studies report raw data for the independent and dependent variable (individual participant data; IPD), we will perform a post-hoc regression analysis selecting the independent and dependent variables as described in Table 1 and using the beta-coefficient as our effect size estimate.

The effect estimates from individual studies (either difference in means or beta-coefficients) will be combined using a fixed effects model to account for potential heterogeneity across studies. The final quantitative result will be the combined estimate with an associated confidence interval. Consultation with a statistician (SS) will guide the determination of whether the chosen statistical approach is appropriate for the study data available and if further modifications are required.

Statistical heterogeneity

To test statistical heterogeneity across the study estimates, we will estimate the variance component corresponding to between-study variability, and use a likelihood ratio test for the null hypothesis that the between study variability is absent. A p-value of 0.05 or less will be considered statistically significant. To assess the impact of existing study heterogeneity on the meta-analysis, the I^2 test statistic will be calculated and evaluated, by considering the magnitude/direction of effect, the strength of evidence of heterogeneity (e.g., p-value from a chi squared test or a confidence intervals for I^2), and the Cochrane's guide to interpretation as follows:

- 0%-40%: might not be important;
- 30%-60%: may represent moderate heterogeneity;
- 50%-90%: may represent substantial heterogeneity;
- 75%-100%: considerable heterogeneity.

Sensitivity analysis

Sensitivity analyses may be conducted in which a study is added or removed from the meta-analysis to evaluate if the results are significantly affected by one particular study. Subgroup analyses based on any heterogeneous characteristics identified from the review for comparability across studies may also be conducted. Additionally, we will make a funnel plot, if possible, of the estimated effects to visually assess the possibility of publication bias.

Assessment of Body of Evidence

Upon completion of the data collection, risk of bias determination, and data analysis, co-authors will assess the quality of evidence. **The instructions to co-authors are provided in a separate document, *Navigation Guide Protocol for Rating the Quality and Strength of the Human and Non-Human Evidence*.**

The initial quality level of human observational data will be considered “moderate,” in contrast to GRADE guidelines, developed for clinical interventions, which assign observational studies an initial rating of “low” quality.(Balslem et al., 2011) In environmental health, human observational data are the “best” data available for decision-making, and in this regard they are comparable to human randomized controlled trials (RCTs) in the clinical sciences. Because ethics virtually precludes human RCTs in environmental health, beginning human observational studies at “moderate” quality captures the value of these data relative to what data are available.

Factors that may decrease the quality level of the body of evidence include:

1. Risk of bias: Study limitations – a substantial risk of bias across body of evidence.
2. Indirectness: Evidence was not directly comparable to the question of interest (i.e., population, exposure, comparator, outcome).
3. Inconsistency: Widely different estimates of effect (heterogeneity or variability in results).
4. Imprecision: Studies had few participants and few events (wide confidence intervals).
5. Publication Bias: Studies missing from body of evidence, resulting in an over or underestimate of true effects from exposure.

Factors that may increase the quality level of the body of evidence include:

1. Large magnitude of effect: Upgraded if modeling suggested confounding alone unlikely to explain associations with relative risk greater than 2 or very unlikely to explain relative risk greater than 5.
2. Dose response: Upgraded if consistent dose response gradient in one or multiple studies, and/or dose response across studies.
3. Confounding minimizes effect: Upgraded if consideration of all plausible residual confounders or biases would underestimate the effect or suggest a spurious effect when results show no effect.

Possible ratings for each criteria are 0 (no change from “moderate” quality), -1 (1 level downgrade) or - 2 (2 level downgrade); +1 (1 level upgrade) or + 2 (2 level upgrade). It is important to note, the ratings of the 8 criteria are not added together to create a score. Authors who decide to rate quality down or up are required to specify the criteria most responsible for their decision and document all factors that contributed to their final quality rating. Coauthors will independently evaluate the quality of the evidence and then compare their evaluations. Discrepancies between the co-authors’ decisions will be resolved through discussion until consensus is reached.

Subsequent to consensus on the quality of the evidence, the co-authors will rate the strength of evidence. The overall strength of the body of human evidence is based on a combination of four criteria: (1) Quality of body of evidence (i.e., the rating from the previous step); (2) Direction of effect; (3) Confidence in effect; and (4) Other compelling attributes of the data that may influence certainty. Review authors will consider separately the strength of each relationship in the model (Figure 2), and base the overall strength rating on evidence for the whole model; that is, we require evidence of a relationship between fetal growth and GFR directly, or evidence of a relationship between fetal growth and PVE *and* a relationship between PVE and GFR. We will assess the strength of the streams of evidence separately against criteria adapted from the Navigation Guide (Woodruff et al., 2010) to determine an overall rating of: “sufficient evidence of an association”; “limited evidence of an association”; “inadequate evidence of an association”; or “evidence of lack of association”. Review authors independently evaluated the strength of the evidence according to the same 4 criteria and compared their evaluations, resolved discrepancies through discussion, and recorded the rationale for every decision. Any discrepancies between the reviewers’ decisions will be resolved through discussion. The senior author (TW) will be the ultimate arbiter of the discrepancies that cannot be resolved through consensus among the co-authors.

Integrating the evidence

The Navigation Guide systematic review methodology describes the process of integrating the strength of human and non-human evidence to generate one of five ratings on the toxicity of the chemical under review (Figure 3, modified from Woodruff et al., 2010); here we will modify the names of these ratings to be relevant to the strength of the evidence of an association between fetal growth and GFR: “known to be associated,” “probably associated,” “possibly associated,” “not classifiable” or “probably not associated.” (Figure 3).

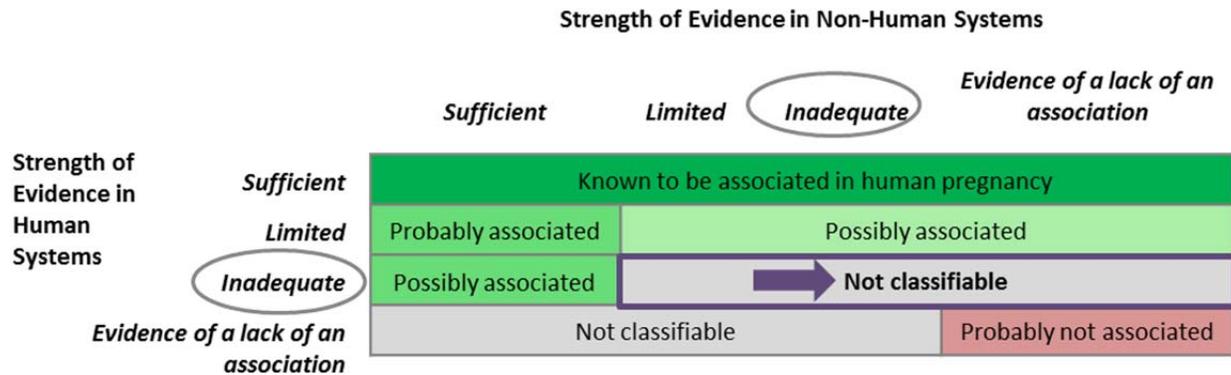


Figure 3. A schematic of the integration of evidence from human and non-human systems to generate one of five possible ratings on the association between fetal growth and GFR.

References

- BALSHEM, H., HELFAND, M., SCHUNEMANN, H. J., OXMAN, A. D., KUNZ, R., BROZEK, J., VIST, G. E., FALCK-YTTER, Y., MEERPOHL, J., NORRIS, S. & GUYATT, G. H. 2011. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*, 64, 401-6.
- BARKER, D. J. P. 1995. Fetal origins of coronary heart disease. *BMJ*, 311, 171-174.
- CHANG, J. & STREITMAN, D. 2012. Physiologic Adaptations to Pregnancy. *Neurologic Clinics*, 30, 781-789.

- CHESNUTT, A. N. 2004. Physiology of normal pregnancy. *Critical Care Clinics*, 20, 609-+.
- GUYATT, G. H., OXMAN, A. D., VIST, G., KUNZ, R., BROZEK, J., ALONSO-COELLO, P., MONTORI, V., AKL, E. A., DJULBEGOVIC, B., FALCK-YTTER, Y., NORRIS, S. L., WILLIAMS, J. W., JR., ATKINS, D., MEERPOHL, J. & SCHUNEMANN, H. J. 2011. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol*, 64, 407-15.
- HIGGINS, J., DEEKS J. (EDITORS). CHAPTER 7: SELECTING STUDIES AND COLLECTING DATA. IN: HIGGINS J., GREEN S. (EDITORS) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 (updated March 2011).The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
- HIGGINS, J. P. T., GREEN, S. & COCHRANE COLLABORATION. 2008. *Cochrane handbook for systematic reviews of interventions*, Chichester, England ; Hoboken, NJ, Wiley-Blackwell.
- HOOIJMANS, C. R., LEENAARS, M. & RITSKES-HOITINGA, M. 2010. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. *Altern Lab Anim*, 38, 167-82.
- JOHNSON, P. I., SUTTON, P., ATCHLEY, D. S., KOUSTAS, E., LAM, J., SEN, S., ROBINSON, K. A., AXELRAD, D. A. & WOODRUFF, T. J. 2013. The Navigation Guide - evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. *Under review*.
- KILKENNY, C., BROWNE, W., CUTHILL, I. C., EMERSON, M. & ALTMAN, D. G. 2010. Animal research: reporting in vivo experiments: the ARRIVE guidelines. *Br J Pharmacol*, 160, 1577-9.
- KOUSTAS, E., LAM, J., SUTTON, P., JOHNSON, P. I., ATCHLEY, D. S., SEN, S., ROBINSON, K. A., AXELRAD, D. A. & WOODRUFF, T. J. 2013. The Navigation Guide - evidence-based medicine meets environmental health: systematic review of non-human evidence for PFOA effects on fetal growth. *Under review*.
- ROSEMAN, M., MILETTE, K., BERO, L. A., COYNE, J. C., LEXCHIN, J., TURNER, E. H. & THOMBS, B. D. 2011. Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments. *JAMA*, 305, 1008-17.
- VISWANATHAN, M., ANSARI, M. T., BERKMAN, N. D., CHANG, S., HARTLING, L., MCPHEETERS, M., SANTAGUIDA, P. L., SHAMLIYAN, T., SINGH, K., TSERTSVADZE, A. & TREADWELL, J. R. 2008. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions.

WHITWORTH, K. W., HAUG, L. S., BAIRD, D. D., BECHER, G., HOPPIN, J. A., SKJAERVEN, R., THOMSEN, C., EGGESBO, M., TRAVLOS, G., WILSON, R., CUPUL-UICAB, L. A., BRANTSÆTER, A. L. & LONGNECKER, M. P. 2012. Perfluorinated compounds in relation to birth weight in the Norwegian Mother and Child Cohort Study. *Am J Epidemiol*, 175, 1209-16.

WOODRUFF, T. J. & SUTTON, P. 2011. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. *Health Aff (Millwood)*, 30, 931-7.

WOODRUFF, T. J., SUTTON, P. & GROUP, T. N. G. W. 2010. An Evidence-Based Medicine Methodology To Bridge The Gap Between Clinical And Environmental Health Sciences. *Health Affairs*, 30, 931-937.

Appendix I: Instructions for Making Risk of Bias Determinations

1. SEQUENCE GENERATION (Animal studies)

Was the allocation sequence adequately generated?

Criteria for a judgment of **'YES'** (i.e. low risk of bias):

The investigators describe a random component in the sequence generation process such as:

- Referring to a random number table;
- Using a computer random number generator;
- Coin tossing;
- Shuffling cards or envelopes;
- Throwing dice;
- Drawing of lots.

Criteria for the judgment of **'PROBABLY YES'** (i.e. probably low risk of bias):

There is insufficient information about the sequence generation process to permit a judgment of 'YES', but there is indirect evidence that suggests the sequence generation process was random, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of **'PROBABLY NO'** (i.e. probably high risk of bias):

There is insufficient information about the sequence generation process to permit a judgment of 'NO', but there is indirect evidence that suggests a non-random component in the sequence generation process, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of **'NO'** (i.e. high risk of bias):

The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic, non-random approach, for example:

- Sequence generated by date of birth;
- Sequence generated by some rule based on date (or day) of arrival at facility;
- Sequence generated by some rule based on record number.

Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgment or some method of non-random categorization of animals, for example:

- Allocation by judgment of the investigator;
- Allocation by availability of the intervention.

Criteria for the judgment of **'NOT APPLICABLE'** (risk of bias domain is not applicable to study):

There is evidence that sequence generation is not an element of study design capable of introducing risk of bias in the study.

2. ALLOCATION CONCEALMENT (Animal studies)

Was allocation adequately concealed?

Criteria for a judgment of **'YES'** (i.e. low risk of bias):

Investigators could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:

- Sequentially numbered treatment containers of identical appearance to control; or
- Sequentially numbered prepared route of administration (e.g., pre-prepared water dosed with chemical) of identical appearance; or
- Sequentially numbered, opaque, sealed envelopes.

Criteria for the judgment of **'PROBABLY YES'** (i.e. probably low risk of bias):

There is insufficient information about allocation concealment to permit a judgment of **'YES'**, but there is indirect evidence that suggests the allocation was adequately concealed, as described by the criteria for a judgment of **'YES'**.

Criteria for the judgment of **'PROBABLY NO'** (i.e. probably high risk of bias):

There is insufficient information about allocation concealment to permit a judgment of **'NO'**, but there is indirect evidence that suggests the allocation was not adequately concealed, as described by the criteria for a judgment of **'NO'**.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

Investigators handling experimental animals could possibly foresee assignments and thus introduce selection bias, such as allocation based on:

- Using an open random allocation schedule (e.g. a list of random numbers); or
- Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered); or
- Alternation or rotation; or
- Non-random and known criteria, such as date of birth; or
- Record number; or
- Any other explicitly unconcealed procedure.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that allocation concealment is not an element of study design capable of introducing risk of bias in the study.

3. BLINDING OF PERSONNEL AND OUTCOME ASSESSORS (Animal and human studies)

Was knowledge of the allocated interventions adequately prevented during the study?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

Any one of the following:

- No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding; or
- Blinding of key study personnel ensured, and unlikely that the blinding could have been broken; or
- Some key study personnel were not blinded, but outcome assessment was blinded and the non-blinding of others unlikely to introduce bias.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about blinding to permit a judgment of 'YES', but there is indirect evidence that suggests the study was adequately blinded, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about blinding to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not adequately blinded, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of '**NO**' (i.e. high risk of bias):

Any one of the following:

- No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding; or
- Blinding of key study personnel attempted, but likely that the blinding could have been broken; or
- Some key study personnel were not blinded, and the non-blinding of others likely to introduce bias.

Criteria for the judgment of '**NOT APPLICABLE**' (risk of bias domain is not applicable to study):

There is evidence that blinding is not an element of study design capable of introducing risk of bias in the study.

4. CONSISTENT RECRUITMENT (Human studies)

Was the strategy for recruiting participants consistent across study groups?

Criteria for a judgment of '**YES**' (i.e. low risk of bias):

Protocols for recruitment and inclusion/exclusion criteria were applied similarly across study groups, and any one of the following:

- Study participants were recruited from the same population at the same time frame; or
- Study participants were not all recruited from the same population, but proportions of participants from each population in each study group are uniform

Criteria for the judgment of '**PROBABLY YES**' (i.e. probably low risk of bias):

There is insufficient information about participant selection to permit a judgment of 'YES', but there is indirect evidence that suggests that participant recruitment and inclusion/exclusion criteria was consistent, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of '**PROBABLY NO**' (i.e. probably high risk of bias):

There is insufficient information about participant selection to permit a judgment of 'NO', but there is indirect evidence that suggests that participant recruitment or inclusion/exclusion criteria was inconsistent, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of '**NO**' (i.e. high risk of bias):

Any one of the following:

- Protocols for recruitment or inclusion/exclusion criteria were applied differently across study groups; or
- Study participants were recruited at different time frames; or
- Study participants were recruited from different populations and proportions of participants from each population in each study group are not uniform

Criteria for the judgment of '**NOT APPLICABLE**' (risk of bias domain is not applicable to study):

There is evidence that participant selection is not an element of study design capable of introducing risk of bias in the study.

5. CONFOUNDING (Human studies)

Was confounding adequately addressed?

We considered the following factors when assessing the risk of confounding:

- Gestational age (where fetal growth was being assessed)
- Maternal size (pre-pregnancy BMI, weight, height or surface area)
- Maternal age

Criteria for a judgment of '**YES**' (i.e. low risk of bias):

The study accounted for (i.e., matched, stratified, multivariate analysis or otherwise statistically controlled for) important potential confounders, or reported that potential confounders were evaluated and omitted because inclusion did not substantially affect the results. The determination of specific confounders may be informed by the data, including the studies included in the review.

Criteria for the judgment of '**PROBABLY YES**' (i.e. probably low risk of bias):

The study accounted for most but not all of the important potential confounders
AND this lack of accounting is not expected to introduce substantial bias.

Criteria for the judgment of '**PROBABLY NO**' (i.e. probably high risk of bias):

The study accounted for some but not all of the important potential confounders
AND this lack of accounting may have introduced substantial bias.

Criteria for the judgment of '**NO**' (i.e. high risk of bias):

The study did not account for or evaluate important potential confounders.

6. INCOMPLETE OUTCOME DATA (Animal and human studies)

Were incomplete outcome data adequately addressed?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

For human studies, participants were followed long enough to obtain outcome measurements and any one of the items in the following list; for animal studies the number of animals assessed for outcome of interest is reported and data is provided indicating adequate follow up of all treated animals:

- No missing outcome data; or
- Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); or
- Missing outcome data balanced in numbers across exposure groups, with similar reasons for missing data across groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a biologically relevant impact on the intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a biologically relevant impact on observed effect size; or
- Missing data have been imputed using appropriate methods

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about incomplete outcome data to permit a judgment of 'YES', but there is indirect evidence that suggests incomplete outcome data was adequately addressed, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about incomplete outcome data to permit a judgment of 'NO', but there is indirect evidence that suggests incomplete outcome data was not adequately addressed, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

For *human* studies, participants were not followed long enough to obtain outcome measurements OR any one of the following:

- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across exposure groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce biologically relevant bias in intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce biologically relevant bias in observed effect size; or
- Potentially inappropriate application of imputation.

For *animal* studies, any one of the following:

- The number of animals allocated not reported and no data is provided to indicate that there was adequate follow up of all treated animals. Additionally, any one of the following: Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce biologically relevant bias in intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce biologically relevant bias in observed effect size; or
- ‘As-treated’ analysis done with substantial departure of the intervention received from that assigned at randomization; or
- Potentially inappropriate application of simple imputation.

Criteria for the judgment of ‘**NOT APPLICABLE**’ (risk of bias domain is not applicable to study):

There is evidence that incomplete outcome data is not capable of introducing risk of bias in the study.

7. SELECTIVE OUTCOME REPORTING (Animal and human studies)

Are reports of the study free of suggestion of selective outcome reporting?

Criteria for a judgment of ‘**YES**’ (i.e. low risk of bias):

All of the study’s pre-specified (primary and secondary) outcomes outlined in the protocol, methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way.

Criteria for the judgment of ‘**PROBABLY YES**’ (i.e. probably low risk of bias):

There is insufficient information about selective outcome reporting to permit a judgment of 'YES', but there is indirect evidence that suggests the study was free of selective reporting, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of '**PROBABLY NO**' (i.e. probably high risk of bias):

There is insufficient information about selective outcome reporting to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not free of selective reporting, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of '**NO**' (i.e. high risk of bias):

Any one of the following:

- The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
- Not all of the study's pre-specified primary outcomes (as outlined in the protocol, methods, abstract, and/or introduction) have been reported; or
- One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; or
- One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected effect); or
- One or more outcomes of interest are reported incompletely
- The study report fails to include results for a key outcome that would be expected to have been reported for such a study.

Criteria for the judgment of '**NOT APPLICABLE**' (risk of bias domain is not applicable to study):

There is evidence that selective outcome reporting is not capable of introducing risk of bias in the study.

8. OTHER POTENTIAL THREATS TO VALIDITY (Animal and human studies)

Was the study apparently free of other problems that could put it at a risk of bias?

Criteria for a judgment of '**YES**' (i.e. low risk of bias):

The study appears to be free of other sources of bias.

Criteria for the judgment of '**PROBABLY YES**' (i.e. probably low risk of bias):

There is insufficient information to permit a judgment of 'YES', but there is indirect evidence that suggests the study was free of other threats to validity.

Criteria for the judgment of '**PROBABLY NO**' (i.e. probably high risk of bias):

There is insufficient information to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not free of other threats to validity, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of '**NO**' (i.e. high risk of bias):

There is at least one important risk of bias. For example, the study:

- Had a potential source of bias related to the specific study design used; or
- Stopped early due to some data-dependent process (including a formal-stopping rule); or
- Had extreme imbalance of characteristics among exposure groups; or
- Had differential surveillance for outcome between exposure groups or between exposed/unexposed groups
- The conduct of the study is affected by interim results (e.g. recruiting additional participants from a subgroup showing greater or lesser effect); or
- An insensitive instrument is used to measure outcomes (which can lead to under-estimation of both beneficial and harmful effects); or
- Selective reporting of subgroups; or
- Has been claimed to have been fraudulent; or
- Had some other problem

Criteria for the judgment of '**NOT APPLICABLE**' (risk of bias domain is not applicable to study):

There is evidence that other potential threats to validity are not capable of introducing risk of bias in the study.

9. **CONFLICT OF INTEREST (Animal and human studies)**

Was the study free of support from a company, study author, or other entity having a financial interest in any of the exposures studied?

Criteria for a judgment of '**YES**' (i.e. low risk of bias):

The study did not receive support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples include the following:

- Funding source is limited to government, non-profit organizations, or academic grants funded by government, foundations and/or non-profit organizations;
- Chemicals or other treatment used in study were purchased from a supplier;

- Company affiliated staff are not mentioned in the acknowledgements section;
- Authors were not employees of a company with a financial interest in the outcome of the study;
- Company with a financial interest in the outcome of the study was not involved in the design, conduct, analysis, or reporting of the study and authors had complete access to the data;
- Study authors make a claim denying conflicts of interest;
- Study authors are unaffiliated with companies with financial interest, and there is no reason to believe a conflict of interest exists;
- All study authors are affiliated with a government agency (are prohibited from involvement in projects for which there is a conflict of interest or an appearance of conflict of interest).

Criteria for the judgment of **'PROBABLY YES'** (i.e. probably low risk of bias):

There is insufficient information to permit a judgment of 'YES', but there is indirect evidence that suggests the study was free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of **'PROBABLY NO'** (i.e. probably high risk of bias):

There is insufficient information to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of **'NO'** (i.e. high risk of bias):

The study received support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples of support include:

- Research funds;
- Chemicals provided at no cost;
- Writing services;
- Author/staff from study was employee or otherwise affiliated with company with financial interest;
- Company limited author access to the data;
- Company was involved in the design, conduct, analysis, or reporting of the study;
- Study authors claim a conflict of interest

Criteria for the judgment of **'NOT APPLICABLE'** (risk of bias domain is not applicable to study):

- There is evidence that conflicts of interest are not capable of introducing risk of bias in the study.