

**Title:** A Risk of Bias instrument for Non-randomized Studies of Exposures: a users' guide to its application in the context of GRADE

## **Abstract**

The objective of this paper is to explain how to apply, interpret, and present the results of a new instrument to assess the risk of bias (RoB) in non-randomized studies (NRS) dealing with effects of environmental exposures on health outcomes. This instrument is modelled on the Risk Of Bias In Non-randomised Studies of Interventions (ROBINS-I) instrument. The RoB instrument for NRS of exposures assesses RoB along a standardized comparison to a randomized target experiment, instead of the study-design directed RoB approach. We provide specific guidance for the integral steps of developing a research question and target experiment, distinguishing issues of indirectness from RoB, making individual-study judgments, and performing and interpreting sensitivity analyses for RoB judgments across a body of evidence. Also, we present an approach for integrating the RoB assessments within the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework to assess the certainty of the evidence in the systematic review. Finally, we guide the reader through an overall assessment to support the rating of all domains that determine the certainty of a body of evidence using the GRADE approach.

1  
2  
3  
4 **Title:** A Risk of Bias instrument for Non-randomized Studies of Exposures: a users' guide to its  
5 application in the context of GRADE  
6  
7  
8  
9  
10

11  
12  
13 **Author list**  
14

15  
16 Rebecca L. Morgan <sup>a</sup>; Kristina A. Thayer <sup>b</sup>; Nancy Santesso <sup>a</sup>; Alison C. Holloway <sup>c</sup>; Robyn Blain <sup>d</sup>;  
17 Sorina E. Eftim <sup>d</sup>; Alexandra E. Goldstone <sup>d</sup>; Pam Ross <sup>d</sup>; Mohammed Ansari <sup>e</sup>; Elie Akl <sup>a, f</sup>; Tommaso  
18 Filippini <sup>g</sup>; Anna Hansell <sup>h, i, j</sup>; Joerg J. Meerpohl <sup>k</sup>; Reem A. Mustafa <sup>a, l</sup>; Jos Verbeek <sup>m</sup>; Marco Vinceti <sup>g</sup>;  
19  
20  
21  
22 <sup>n</sup>; Paul Whaley <sup>o</sup>; Holger J. Schünemann <sup>a, p</sup>, GRADE Working Group  
23  
24  
25  
26  
27

28 <sup>a</sup> Department of Health Research Methods, Evidence, and Impact, McMaster University, Health  
29 Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1 Canada  
30  
31 [morganrl@mcmaster.ca](mailto:morganrl@mcmaster.ca), [santesna@mcmaster.ca](mailto:santesna@mcmaster.ca), [ramustafa@gmail.com](mailto:ramustafa@gmail.com), [schuneh@mcmaster.ca](mailto:schuneh@mcmaster.ca)  
32  
33  
34

35 <sup>b</sup> Integrated Risk Information System (IRIS) Division, National Center for Environmental Assessment  
36 (NCEA), Office of Research and Development, US Environmental Protection Agency, Building B  
37 (Room 211i), Research Triangle Park, NC USA 27711. [thayer.kris@epa.gov](mailto:thayer.kris@epa.gov)  
38  
39  
40  
41

42 <sup>c</sup> Department of Obstetrics and Gynecology, McMaster University, Health Sciences Centre, Room  
43 3N52A, 1280 Main Street West, Hamilton, ON L8S 4K1 Canada. [hollow@mcmaster.ca](mailto:hollow@mcmaster.ca)  
44  
45  
46  
47

48 <sup>d</sup> ICF International Inc., 9300 Lee Highway, Fairfax, VA, USA. [Robyn.Blain@icfi.com](mailto:Robyn.Blain@icfi.com),  
49  
50 [Pam.Ross@icfi.com](mailto:Pam.Ross@icfi.com), [Ali.Goldstone@icfi.com](mailto:Ali.Goldstone@icfi.com), [Sorina.Eftim@icfi.com](mailto:Sorina.Eftim@icfi.com)  
51  
52  
53  
54  
55  
56

57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112

<sup>e</sup> School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, ON K1H 8M5  
Canada. [tosansari@gmail.com](mailto:tosansari@gmail.com)

<sup>f</sup> Department of Internal Medicine, Faculty of Health Sciences, American University of Beirut, P.O.  
Box: 11-0236, Riad-El-Solh Beirut 1107 2020 Lebanon. [ea32@aub.edu.lb](mailto:ea32@aub.edu.lb)

<sup>g</sup> Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia,  
Italy. [tommaso.filippini@unimore.it](mailto:tommaso.filippini@unimore.it), [mvinceti@bu.edu](mailto:mvinceti@bu.edu)

<sup>h</sup> MRC-PHE Centre for Environment and Health, Imperial College London, St Mary's Campus, Praed  
St, Paddington London W2 1PG, UK. [a.hansell@imperial.ac.uk](mailto:a.hansell@imperial.ac.uk)

<sup>i</sup> Public Health Directorate, Imperial College Healthcare NHS Trust, St Mary's Hospital, Paddington,  
London W2 1PG, UK. [a.hansell@imperial.ac.uk](mailto:a.hansell@imperial.ac.uk)

<sup>j</sup> Centre for Environmental Health and Sustainability, University of Leicester, George Davies Building,  
University Road, Leicester, LE1 7RH, UK. [a.hansell@imperial.ac.uk](mailto:a.hansell@imperial.ac.uk)

<sup>k</sup> Institute for Evidence in Medicine (for Cochrane Germany Foundation), Medical Center - University of  
Freiburg, Breisacher Strasse 153, 79110 Freiburg, Germany. [meerpohl@cochrane.de](mailto:meerpohl@cochrane.de)

<sup>l</sup> Division of Nephrology and Hypertension, Department of Medicine, University of Kansas Medical  
Center, 3901 Rainbow Boulevard, Kansas City, KS 66160, USA. [ramustafa@gmail.com](mailto:ramustafa@gmail.com)

<sup>m</sup> Finnish Institute of Occupational Health, Cochrane Work, Neulaniementie 4, 70701 Kuopio, Finland.  
[Jos.Verbeek@ttl.fi](mailto:Jos.Verbeek@ttl.fi)

113  
114  
115 <sup>n</sup> Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts,  
116 USA. [mvinceti@bu.edu](mailto:mvinceti@bu.edu)  
117

118  
119  
120 <sup>o</sup> Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK.  
121 [p.whaley@lancaster.ac.uk](mailto:p.whaley@lancaster.ac.uk)  
122

123  
124  
125 <sup>p</sup> Department of Medicine, McMaster University, Health Sciences Centre, Room 2C14, 1280 Main  
126 Street West, Hamilton, ON L8S 4K1 Canada; [schuneh@mcmaster.ca](mailto:schuneh@mcmaster.ca)  
127

128  
129  
130  
131  
132  
133  
134 **Corresponding author:** Holger J. Schünemann. Department of Health Research Methods, Evidence  
135 and Impact, Health Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1  
136  
137  
138 Canada; [schuneh@mcmaster.ca](mailto:schuneh@mcmaster.ca)  
139

140  
141  
142  
143  
144 **Conflict of interest**  
145

146  
147 The authors declare they have no competing financial interests with respect to this manuscript, or its  
148 content, or subject matter.  
149

150  
151  
152 The views expressed are those of the authors and do not necessarily represent the views or policies of  
153 the U.S. Environmental Protection Agency.  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168

169  
170  
171 **Abstract**  
172  
173

174 The objective of this paper is to explain how to apply, interpret, and present the results of a new  
175 instrument to assess the risk of bias (RoB) in non-randomized studies (NRS) dealing with effects of  
176 environmental exposures on health outcomes. This instrument is modelled on the Risk Of Bias In Non-  
177 randomised Studies of Interventions (ROBINS-I) instrument. The RoB instrument for NRS of exposures  
178 assesses RoB along a standardized comparison to a randomized target experiment, instead of the study-  
179 design directed RoB approach. We provide specific guidance for the integral steps of developing a  
180 research question and target experiment, distinguishing issues of indirectness from RoB, making  
181 individual-study judgments, and performing and interpreting sensitivity analyses for RoB judgments  
182 across a body of evidence. Also, we present an approach for integrating the RoB assessments within the  
183 Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework to  
184 assess the certainty of the evidence in the systematic review. Finally, we guide the reader through an  
185 overall assessment to support the rating of all domains that determine the certainty of a body of evidence  
186 using the GRADE approach.  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203

204 **Keywords (6):** Risk of bias; environmental health; GRADE; non-randomized studies; study limitations;  
205 ROBINS  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224

## 1. Introduction

The evidence on the impact of environmental or occupational exposures on human health outcomes typically comes from non-randomized studies (NRS). Objective and transparent evaluation of evidence of exposures requires the use of systematic reviews [1]. A highly credible systematic review should include a standardized, rigorous, and transparent assessment of the risk of bias (RoB) in each included study and across the body of evidence [2, 3]. This is applicable when referring to studies evaluating the impact of an environmental, occupational or other type of exposure.

A recent study evaluated five RoB methods used in environmental health hazard assessments [4]. While all five methods considered similar issues (or domains) in RoB assessment, their relative emphasis on these issues varied. The study suggested a need for the harmonization and improvement of these methods. We developed the RoB instrument for NRS of exposures based on the feedback from developers of existing instruments and methods to address limitations such as outlining the ideal study, labelling of study designs, and the use of signaling questions [4, 5]. The objective of this paper is to explain how to apply, interpret, and present the results of a new instrument to assess the RoB in NRS dealing with effects of environmental exposures on health outcomes.

## 2. Overview of the instrument

The RoB instrument for NRS of exposures is modeled after the Risk Of Bias In Non-randomised Studies of interventions (ROBINS-I) instrument [6]. In 1965, Cochran proposed evaluating NRS using the criteria for RCTs [7]. Hernan et al. recently suggested that causal inference from NRS represents an attempt to emulate the ideal randomized trial (the target trial) that would answer the question of interest [8]. In fact, ROBINS-I uses a hypothetical ideal target trial that would be free of bias as a reference

281  
282  
283 point. By using the target trial as the reference point, ROBINS-I moves away from a study-design  
284 directed approach. That is, the specific design of the NRS, e.g. a case-control design, does not *a priori*  
285 determine absence or presence of RoB [9]. RoB instrument for NRS of exposures emulates these  
286 features of ROBINS-I.  
287  
288  
289  
290  
291

292  
293 In brief, the application of the RoB instrument for NRS of exposures consists of three steps:  
294  
295

- 296 1. Step I: presents the review question, potential confounders, co-interventions, and exposure and  
297 outcome measurement accuracy information;
- 298 2. Step II: describes each eligible study as a hypothetical target experiment, including specific  
299 confounders and co-interventions from that study that will require consideration; and
- 300 3. Step III: assesses RoB across seven items about the strengths and limitations of studies of  
301 environmental exposure.  
302  
303  
304  
305  
306  
307  
308

309  
310 The seven RoB items are: 1) Bias due to confounding, 2) Bias in selection of participants into the study,  
311 3) Bias in classification of exposures, 4) Bias due to departures from intended exposures, 5) Bias due to  
312 missing data, 6) Bias in measurement of outcomes, and 7) Bias in selection of reported results.  
313

314  
315 Judgments for each RoB item can be: ‘Low RoB’, ‘Moderate RoB’, ‘Serious RoB’, or ‘Critical RoB’.

316  
317 Similarly, an overall judgment about the bias at the study level is either ‘Low RoB’, ‘Moderate RoB’,  
318 ‘Serious RoB’, or ‘Critical RoB’. In order to reach a judgment for each RoB item, the rater first answers  
319 one or more signalling questions with ‘Yes’, ‘Probably yes’, ‘Probably no’, ‘or No’. The answer should  
320 be based on the information available in the publications/reports of the individual study and be justified  
321 in an accompanying free-text field.  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336

337  
338  
339 Previously published guidance for the ROBINS-I instrument proposes that the study-level RoB should  
340 be the most concerning level among the RoB items for that study, unless raters determine the study-level  
341 RoB to be more severe because of compounded risks of more than one individual RoB item [6].  
342  
343

344  
345 Identifying RoB per item and across items per study allows systematic-review authors to explore the  
346 possible influence of studies at less compared to more severe RoB on the pooled estimates of effect [10].  
347  
348

349  
350 As in the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach  
351 for the body of evidence, risk of bias is assessed by outcome in a study and study RoB could vary by  
352 outcome (e.g. subjective outcomes may have different levels of bias than objective outcomes) or group  
353 of outcomes, if pragmatic rationale supports the grouping of outcomes.  
354  
355  
356  
357

358  
359 Systematic-review authors can then use the RoB instrument as part of the assessment of the certainty of  
360 the body of evidence using the GRADE framework. Within the GRADE framework, RoB is one domain  
361 for assessing the certainty of evidence (CoE), the others being inconsistency, indirectness, imprecision,  
362 publication bias, magnitude of effect, dose-response gradient, and plausible opposing residual  
363 confounding [2]. As per the current GRADE guidance, evidence from NRS, appraised using existing  
364 design-specific RoB instruments, starts with a default initial certainty of “Low” due to concerns of  
365 confounding and selection bias when randomization is lacking. Raters then downgrade or upgrade the  
366 body of evidence according to specific GRADE domain assessments, including a more detailed  
367 evaluation for RoB other than confounding. However, since the RoB instrument for NRS of exposures  
368 takes into account lack of randomization, evidence will not be automatically rated down because  
369 judgments of risk of bias would have been made with reference to a hypothetical target experiment  
370 (ideal target trial). Bodies of evidence of any study design will undergo the same RoB evaluation  
371 without specific reference to the study design. In the context of using ROBINS-like instruments, all  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392



393  
394  
395 studies within the bodies of evidence will start at the same ‘High’ initial certainty within GRADE  
396  
397 regardless of study design. However, in general, NRS, due to potential for confounding and selection  
398  
399 bias when compared with RCTs will receive a rating of low or very low depending on the degree of  
400  
401 RoB. Raters must justify not rating down only in the presence of specific study design and execution or  
402  
403 result features [9].  
404  
405

406  
407 When conducting a systematic review, results from the study-level RoB instrument for NRS of health  
408  
409 effects of exposures inform judgments about overall RoB for the body of evidence across studies. So far,  
410  
411 no guidance on the use of the RoB instrument for NRS of effects of exposures for this purpose exists.  
412  
413 This article provides guidance for the application of the RoB instrument for NRS of exposures at the  
414  
415 study level and as part of a RoB judgment within the GRADE framework to determine the certainty  
416  
417 across a body of evidence [5]. Although the RoB instrument for NRS of exposures is still being refined  
418  
419 in consultation with a diverse group of subject matter experts, we highlight a number of important  
420  
421 procedural questions. Thus, describing our experience in implementing the RoB instrument for NRS of  
422  
423 effects of exposures will facilitate future testing and clarification of the use of the instrument by  
424  
425 systematic review authors and guideline developers.  
426  
427  
428

### 429 **3. Approach when conducting systematic reviews for studies of exposure**

430  
431 We previously described the development of the RoB instrument for NRS of exposures [5]. In addition  
432  
433 to this effort, we have solicited broader input on this instrument at workshops held at GRADE Working  
434  
435 Group meetings in March 2015, October 2015, and April 2016; during a meeting to develop ROBINS of  
436  
437 Exposures (ROBINS-E; an instrument based on the RoB instrument for NRS of exposures and  
438  
439 ROBINS-I) in January 2017; and at the Global Evidence Summit in September 2017. Findings from  
440  
441  
442  
443  
444  
445  
446  
447  
448

449  
450  
451 these workshops, through this iterative process, have led to further refinement and pilot-testing of the  
452  
453 RoB instrument for NRS of exposures.  
454

455  
456 Figure 1 presents a schematic of how the RoB instrument for NRS of exposures fits into the systematic  
457  
458 review process. It illustrates steps for evaluating the RoB of individual studies in a systematic review  
459  
460 and integrating the results across a body of evidence into the GRADE evidence-assessment framework.  
461  
462 For each outcome in the review, authors of systematic reviews would go through Steps II and III, and  
463  
464 GRADE.  
465  
466

### 467 468 **3.1. Complete Step I of the RoB instrument for NRS of exposures** 469

#### 470 471 **3.1.1. Define the research question**

472  
473 This process begins with the definition of the research question. For questions about exposures (i.e.  
474  
475 unintentional interventions), namely the environmental and occupational type, the research question is  
476  
477 formatted as a PECO (population, exposure, comparator(s), and outcomes) question [5, 11]. For  
478  
479 example, we may ask the following research question “In production workers exposed to steady state  
480  
481 noise for ten years (population), what is the effect of exposure to a noise level of 80 dB(A) measured as  
482  
483 LAeq,8h or greater (exposure) compared to less than 80 dB(A) also measured as LAeq,8h (comparison)  
484  
485 in the same population on hearing level?” To understand the relation between noise and hearing loss, we  
486  
487 may also ask the following PECO: “In production workers exposed to steady state noise louder than 80  
488  
489 dB(A) during ten years measured as L Aeq,8h, what is the effect of an increase of 5 dB(A) on hearing  
490  
491 level compared to the level from where the increase started, over the whole range of exposure, assuming  
492  
493 an exponential relationship between exposure and hearing level?”  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504

505  
506  
507 Since the RoB instrument for NRS of exposures is set up as a comparison between groups that can be  
508 exposed or not, or exposed to different levels, it is necessary to clearly identify what is the exposure  
509 level of interest and what is the comparison. In some situations, little or nothing may be known about the  
510 relationship between an exposure and outcome to inform the PECO. There are at least five approaches to  
511 facilitate formulating and defining the levels of exposure within the PECO (Table 1) [12]. Researchers  
512 should be transparent about which of these approaches they are using for definition of their PECO and  
513 ensure that the exposure and comparator(s) are explicitly defined.  
514  
515

### 523 3.1.2. Identify confounders, co-interventions, and measures of exposures and 524 outcomes 525

526 In Step I, systematic-review authors list confounders and co-interventions that are associated with both  
527 the exposure and outcome. In addition, review authors assess the accuracy of the exposure and outcome  
528 measurements. These sections must be populated by knowledgeable members of the review team. While  
529 working through these sections, raters respond to signaling questions in the confounding, participant  
530 selection, and exposure measurement RoB items. Consideration of these issues may lead to the  
531 identification of different sources of indirectness [5]. For example, the review team may identify obesity  
532 as one of their important outcomes; however, studies may measure waist circumference (and measure it  
533 accurately within the study) to inform the outcome of obesity. The review team may label waist  
534 circumference as an indirect measure of obesity.  
535  
536

546 We present the text used in the review-level protocol for an example on bisphenol A (BPA), comparing  
547 the highest exposure stratum and lowest exposure stratum of BPA in each eligible study (Appendix A).  
548 The PECO being: “What is the effect of highest levels compared with lowest levels of BPA exposure on  
549 body weight?” We reviewed published literature, as well as consulted with topic-specific experts, to  
550  
551  
552  
553  
554

561  
562  
563 determine the final set of responses to the Step I fields. For some exposures, a public database of  
564 confounders for measures of environmental exposures and health outcomes (i.e., PhenX Toolkit;  
565 <https://www.phenxtoolkit.org/>) may provide additional information.  
566  
567  
568  
569  
570

### 571 **3.2. Complete Step II of the RoB instrument for NRS of exposures for eligible studies**

#### 572 **3.2.1. Construct the target experiment**

573  
574 At this point, the studies that meet the eligibility criteria of the review should have been identified. The  
575 reviewers should complete separate forms for each relevant outcome (group) within each study. At the  
576 start of Step II, reviewers construct a study-specific target experiment informed by the PECO question,  
577 the exposure and comparator exposure thresholds, outcome specific confounders, and health outcome  
578 measurements. As explained in previous GRADE guidance for the use of ROBINS-I, the target  
579 experiment provides a structured comparison with a reference experiment that is considered to be at low  
580 RoB [9]. The target experiment need not be realistic, as it should reflect a study design that reduces  
581 known and unknown imbalance in prognostic factors and confounding [5]. It then allows RoB  
582 assessment of individual studies and across studies at a later stage against the lowest possible bias that  
583 research could yield for the question at hand. Also, in Step II, the reviewer records how the individual  
584 studies measured the exposure and health outcome. The information recorded in Step II informs the RoB  
585 judgments made in Step III.  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600

601 For example, let's consider our review on BPA and weight. The PECO of the review is comparing the  
602 highest to the lowest level of BPA exposure. In Step II, we determine the target experiment for the  
603 included study (Appendix B). Based on the quantities identified in the study by Carwile & Michels [13],  
604 the target hypothetical experiment would be framed as an experiment in which the general adult  
605 population is randomly allocated to a high level of BPA exposure ( $\geq 4.7$  ng/mL) or a low level of BPA  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616

617  
618  
619 exposure ( $\leq 1.1$  ng/mL) and body weight measured. In this situation, we compared two exposure cut-  
620  
621 offs to determine the effect on obesity.  
622

623  
624 Confounders must be explored in each eligible study, as studies and outcomes may be affected by  
625 different confounders. For example, the review question may be about the general population, but the  
626 study includes only industrial workers, which may introduce additional confounders, such as exposure to  
627 other chemicals. Note that it may have impact on judging indirectness or selection bias, too. Also, in  
628 Step II, the reviewer makes a judgment of the potential magnitude and direction of the impact of the  
629 confounding factor on the effect estimate. For example, when examining the effect of BPA on body  
630 weight, consumption of processed foods is considered a confounder as it both increases the participants'  
631 exposure to BPA through food packaging and increases overall caloric intake [14]. We present the  
632 completed Step II sections for two studies from our BPA and obesity example: Carwile & Michels, 2011  
633 and Harley et al., 2013 (Appendices B & C) [13, 15].  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645

### 646 3.2.2. Identifying sources of indirectness to integrate within GRADE and their 647 relation to risk of bias 648

649 While establishing the target experiment in Step II, individuals may identify studies that present  
650 evidence different from the PECO question (i.e., a restricted version of any concept such as only part of  
651 the population of interest or a section of the range of interest for high exposure) [16]. For example,  
652 consider again the review of hearing loss due to noise exposure. Studies with only shift workers may be  
653 considered indirect evidence for effects in the general population. Studies reporting on waist  
654 circumference may be considered indirect evidence for the measure of the outcome of obesity. Sources  
655 of indirectness may also come from studies that do not have a direct comparison (and therefore results  
656 would be compared to results from an external control or comparator group) or when using surrogate  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672

673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728

measures. While the review team may decide to include this study in the review, when evaluating the evidence within GRADE, differentiation between the domain of risk of bias and indirectness may be rather nuanced. Consider the following: the target experiment serves as the anchor point. If the study at hand tries to emulate the exposure specified in the target experiment but does not achieve what it sets out to do, it is subject to bias. If it acknowledges difficulty in mimicking and defines a proxy experiment, which the study appropriately implements, then it could be considered indirectness in relation to the question of interest.

Subsequent considerations for RoB when using indirect evidence in a review require critical evaluation to identify potential for misclassification of the exposure. While it is important to recognize the potential for more serious bias in classification of exposures when using an indirect comparison, there are situations in which they may present less risk because of clearly delineated exposure and comparison groups (e.g. there is little to no concern that the exposure groups are overlapping)..

Similarly, studies identified for the review may use exposure measures that are indirect to those identified in the PECO, i.e., proxy or intermediate markers of measures. Within the BPA example, , the measurement of exposure level based on a participant’s job title (e.g. cashier) would be indirect [17]. Extrapolating BPA exposure levels based on a participant’s job title may also introduce a risk to bias based on specific prognostic factors or the ability to differentiate between the levels of exposure.

### **3.3. Complete Step III of the RoB instrument for NRS of exposures assessment for eligible studies**

Raters evaluate eligible studies and determine RoB by responding to signaling questions for each of the seven RoB items listed previously. Appendices D & E present summaries from two studies addressing BPA and body weight (as measured by prevalent overweight and prevalent obesity). We present

729  
730  
731 judgments across assessments of the RoB instrument items for NRS of exposures in a RoB matrix for all  
732  
733 eligible studies in Table 2.  
734

735  
736 Due to the lack of randomization and allocation concealment, studies will typically be judged as  
737  
738 ‘Serious’ RoB within the item of bias due to confounding and, also, may be judged as ‘Serious’ due to  
739  
740 selection of participants. While RoB items 4-to-7 are similar to those used to evaluate RCTs [6, 18], bias  
741  
742 due to confounding, selection of participants, and classification of the exposure present considerations  
743  
744 unique to studies of exposures [5]. Below, we highlight some of these nuances and how raters can  
745  
746 address them in their item- and study-level RoB judgments.  
747  
748  
749

### 750 Bias due to Confounding

751  
752 Three situations require particular attention when evaluating bias due to confounding for exposures: 1)  
753  
754 the evaluation of cross-sectional studies; 2) considerations of large effects; and 3) opposing residual  
755  
756 plausible confounding.  
757

758  
759 Cross-sectional studies can impact the judgment on the item-level RoB due to confounding (e.g. time-  
760  
761 varying confounding). This is because we might be unable to evaluate time-varying confounding and it  
762  
763 makes the measurement of the effect of known confounders more challenging. We present two examples  
764  
765 from the BPA and body weight review. While Carwile & Michels adjusted for all critical confounders,  
766  
767 the measurement of exposure and outcome at one time point lowers our certainty that temporal  
768  
769 confounders (e.g. dietary preference for canned food) are not responsible for any observed long-term  
770  
771 association (Appendix D) [13]. In this specific study, the data collection point is part of the National  
772  
773 Health and Nutrition Examination Survey (NHANES), a nationally-representative dataset with years of  
774  
775 prior data collection, therefore providing supplemental information about the adjustment of confounders.  
776  
777  
778  
779  
780  
781  
782  
783  
784

785  
786  
787 In contrast, within that review, neither Li nor Wang provide that same level of information about the  
788 data collection, therefore presenting “Critical” bias due to confounding (Table 1) [19, 20].  
789

790  
791  
792 Studies judged as biased due to confounding with evidence of a large effect or opposing residual  
793 confounding (i.e. when residual confounders would result in the underestimating of an apparent  
794 exposure effect) may not require severe RoB item-level judgment [21]. This is due to the magnitude of  
795 the effect outweighing the size of the bias that might exist in the study or that all plausible biases go in a  
796 direction that would have reduced the observed effect or increased the observed lack of effect. These  
797 latter two domains contribute to increasing the CoE in a body of evidence of NRS in GRADE; however,  
798 within the RoB instrument for NRS of exposures they may also influence the study-level judgments  
799 [21]. To demonstrate this situation, we present an example on smoking and lung cancer-related mortality  
800 [22, 23]. A prospective cohort study compared lung cancer-related mortality rates among smokers and  
801 non-smokers [23]. Although there are some concerns due to residual and unmeasured confounders, such  
802 as occupational or air pollution exposures, the large magnitude of effect (30 times greater mortality rate  
803 due to lung cancer among persons smoking 25 or more cigarettes vs. non-smokers) warrants a less  
804 severe RoB item-level judgment of ‘Low’ or ‘Moderate’, instead of ‘Serious’ for the RoB item of  
805 confounding [23]. In this example, the large magnitude of effect reduces our concern that bias alone  
806 creates a spurious effect [24].  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824

825 In addition, exploratory research conducted has suggested there is no relation between the 10 most  
826 common occupational exposures (i.e., sulfur dioxide, welding fumes, engine emissions, gasoline,  
827 lubricating oil, solvents, paints/varnishes, adhesives, excavation dust, and wood dust) and smoking  
828 history [25]. This exploration into the relationship between exposures and the outcome of interest  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840



841  
842  
843 reduces our concern for potential residual plausible confounding due to other occupational or air  
844  
845 pollution exposures even more.  
846  
847

## 848 Bias due to Misclassification of Exposure

849  
850 In NRS of exposure, there is a particular concern with distinguishing between the exposed and reference  
851  
852 groups, as measuring exposure is difficult and the reference groups are often assumed to be non-  
853  
854 exposed. Bias relating to exposure assessment is a major source of systematic error in studies of  
855  
856 environmental exposures [26]. This is dealt with explicitly in a separate paper [27]. It is crucial to  
857  
858 identify the source and type of exposure misclassification. If non-differential, the exposure  
859  
860 misclassification will usually bias associations to the null, although the final impact on the observed  
861  
862 relative risk is also dependent on other factors [28].  
863  
864

865  
866 Systematic reviewers may be faced with different approaches to exposure assessment. In the example of  
867  
868 noise exposure, this may be assessed by (in order of most severe to least severe exposure  
869  
870 misclassification bias) [29]:  
871  
872

- 873  
874 • Self-report questionnaire: Do you have to raise your voice to carry out a normal conversation  
875  
876 with a colleague when approximately two metres apart for at least part of the working day (may  
877  
878 indicate noise levels >80dB);
- 879  
880 • Modelling: in the occupational setting, a job-exposure matrix would be an example, whereby an  
881  
882 occupational hygienist classifies likely exposure ranges based on job title;
- 883  
884 • Environmental monitoring: using a noise monitor to measure noise in the workplace  
885  
886 environment will give a continuous measurement but sensor measurement error likely to be  
887  
888 optimised for certain exposure ranges;  
889  
890

- Personal monitoring: using a personal noise monitor to measure exposure but sensor measurement error likely to be optimised for certain exposure ranges;
- Individual dose: personal monitoring, additionally taking account of use of ear defenders, hearing acuity, etc.

In our example of BPA and body weight, the review team and topic-specific experts note the accuracy of the measurement of exposure requires multiple measurements (cited here from five-to-13 repeated measurements) at different time points, due to the non-persistent nature of BPA in the body [30]. If an individual study uses fewer than the recommended number of samples, or since diagnostic accuracy of BPA with the collection of between five and 13 samples only yields  $\geq 0.80$  sensitivity and specificity depending on level of exposure (small, moderate, high), there are concerns for non-differential misclassification (i.e. random error) potentially conflating participants in the exposure and comparator groups, likely leading to little difference in the outcomes (i.e. bias toward the null). When the exposure is non-persistent, we have more confidence when studies use multiple timepoints to measure the exposure level. The number of collected samples increases our certainty in the correct classification of the higher exposed and lower exposed groups. In this situation we may consider the exposure domain for Harley to be of less potential risk of bias for misclassification of the exposure. Although repeated measures in urine is acceptable, there is still some scientific uncertainty about the most direct measure of BPA exposure (i.e. urine vs blood)[31, 32]. In Carwile & Michels, participants provided only one sample; therefore we may have critical concerns about bias due to misclassification of the exposure (Appendix D) [13].

953  
954  
955 The single sampling method used in Carwile & Michels decreases our certainty that the higher exposed  
956 and lower exposed participants can be accurately distinguished. Returning to Figure 1, in their protocol,  
957 review authors could have specified to exclude such studies a priori or identified this risk of bias item as  
958 a reason to conduct a sensitivity analysis (see below).  
959  
960  
961  
962  
963

### 964 965 **3.4. RoB judgments for an individual study for an outcome** 966

967 According to ROBINS-I guidance, raters should assign the study-level RoB according to the most  
968 severe of the RoB item-level judgments unless they determine the study to have more severe RoB based  
969 on a combination of RoB judgments across items [6]. We demonstrate this in our example of BPA and  
970 weight in Table 3. This approach relies on individuals critically evaluating the rationale and direction of  
971 the bias. For example, if more than one RoB item within a study were rated as serious RoB but no RoB  
972 items were of critical RoB, then the study-level RoB could either be serious or could be critical if the  
973 consideration of all serious ratings leads to greater concern than would be expressed by a rating of  
974 serious on the study level.  
975  
976  
977  
978  
979  
980  
981  
982  
983

### 984 985 **3.5. Sensitivity analyses and overall RoB across studies** 986

987 Sensitivity analyses allow for exploration across a body of evidence to determine whether the pooled  
988 results are robust with including, versus excluding, studies with certain RoB [33]. The variability in RoB  
989 judgments across individual studies may inform whether a selection of studies, rather than the whole  
990 body of evidence, best informs the research question. The approach to conducting sensitivity analyses  
991 (not to be confused with the sensitivity of a study) should be specified at the protocol step of the  
992 systematic review; however, may be identified after the preliminary analysis. For example, studies may  
993 be deemed critical in the domain of bias due to confounding resulting from unadjusted analyses of  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008

1009  
1010  
1011  
1012 covariates. If a body of evidence includes studies with adjusted and unadjusted analyses, a sensitivity  
1013  
1014 analysis could compare the estimates of effect for the adjusted (removing those studies not adjusting for  
1015  
1016 covariates) and the total pooled estimate. If the effect estimates are not robust and differ between  
1017  
1018 analyses (i.e. confounding may have an influence on the results), then review authors might consider  
1019  
1020 whether to exclude the studies with unadjusted analyses; however, if the effect estimates do not differ  
1021  
1022 (e.g. confounding apparently has no influence on the results), then the review authors may keep the  
1023  
1024 unadjusted studies in the analysis because the suspicion of confounding apparently does not have a big  
1025  
1026 impact. In these instances when the effect estimate is similar across studies then authors could consider  
1027  
1028 updating the individual study level ratings to indicate a less severe RoB for the item and include the  
1029  
1030 rationale that the sensitivity analysis showed no effect of RoB on the results.  
1031  
1032

1033  
1034 Using BPA as an example, we compared studies for the body weight outcomes of prevalent overweight  
1035  
1036 and prevalent obesity at higher and lower RoB in sensitivity analyses specifically across the domain of  
1037  
1038 confounding (Tables 4 & 5; Appendices F & G). We conducted these sensitivity analyses to explore the  
1039  
1040 potential for bias introduced by studies that did not adjust for all critical confounders. The sensitivity  
1041  
1042 analysis for the outcome of prevalent overweight resulted in a difference between the effect estimates,  
1043  
1044 demonstrating that bias due to confounding impacted the pooled estimate; therefore, the judgment would  
1045  
1046 be reflective of the more severe RoB (Table 4). An additional option would be to only show results from  
1047  
1048 Harley, Eng, and Carwile in the GRADE evidence assessment. In contrast, the sensitivity analysis of  
1049  
1050 studies reporting on prevalent obesity demonstrated similar effect estimates (Appendix G). In this  
1051  
1052 situation, all studies reflect the less serious RoB judgment (Table 5).  
1053  
1054

### 1055 1056 **3.6. Integration of RoB judgments across a body of evidence into GRADE assessment** 1057 1058 1059 1060 1061 1062 1063 1064

1065  
1066  
1067  
1068 The overall rating of RoB across the body of evidence for an outcome is integrated into the GRADE  
1069  
1070 assessment similar to what has been previously described in the literature for the result of RCTs and  
1071  
1072 observational studies [10]. It is also during this process where indirectness, if identified during Steps I or  
1073  
1074 II within the RoB instrument for NRS of exposures, would be integrated in the overall assessment of the  
1075  
1076 evidence. When evaluating RoB using ROBINS-I and the RoB instrument for NRS of exposures, the  
1077  
1078 body of evidence starts at ‘High’ initial CoE within GRADE. For the example of BPA and its effect on  
1079  
1080 body weight, we present the outcomes of prevalent overweight (i.e., BMI  $\geq$  85th percentile for age/sex  
1081  
1082 in children;  $25 \leq$  BMI  $<$  30 kg/m<sup>2</sup> in adults) and prevalent obesity (BMI  $\geq$  95th percentile for age/sex in  
1083  
1084 children; BMI  $\geq$  30 kg/m<sup>2</sup> in adults) in a GRADE evidence profile (Table 6). It is across this body of  
1085  
1086 evidence that we look for evidence of the three factors (magnitude of effect, dose-response gradient, and  
1087  
1088 opposing residual confounding) considered in the past as mechanisms to upgrade the quality of the  
1089  
1090 evidence for NRS within GRADE [21]. The BPA example does not demonstrate any situation, based on  
1091  
1092 these three factors, which may lead to a less severe RoB judgment. Across the body of evidence for  
1093  
1094 prevalent overweight, our RoB based on the RoB instrument for NRS of exposures and sensitivity  
1095  
1096 analysis of the item of confounding is ‘Critical’, resulting in a rating down of three levels for RoB. In  
1097  
1098 addition, we rate down for imprecision because the effect estimate crosses the null. Our final CoE would  
1099  
1100 be ‘Very low’. Across the body of evidence for prevalent obesity, our RoB is ‘Serious’; therefore, we  
1101  
1102 rate down two levels for RoB. There are no other GRADE domains that we would rate down for. Our  
1103  
1104 final CoE would be ‘Low’.  
1105  
1106  
1107  
1108

## 1109 **4. Discussion**

1110  
1111 The RoB instrument for NRS of exposures presents a novel instrument for conducting the RoB  
1112  
1113 assessment of individual studies included in a systematic review of the health effects of exposure. In this  
1114  
1115

1121  
1122  
1123 users' guide, we suggest that the RoB instrument for NRS of exposures provides a standardized  
1124  
1125 instrument for the transparent evaluation of RoB for NRS of exposures. We present an overview of the  
1126  
1127 process, using examples to demonstrate specific issues encountered when formulating the PECO for the  
1128  
1129 review, outlining a target experiment for an individual study, evaluating bias in individual studies, and  
1130  
1131 summarizing judgments across the body of evidence. We highlight the need for critical consideration of  
1132  
1133 the RoB judgments, including situations within individual studies and across a body of evidence when  
1134  
1135 the judgments may be less severe. In addition, we present sources of indirectness identified in eligible  
1136  
1137 studies that would inform the GRADE evidence assessment. We also present the steps for integrating the  
1138  
1139 RoB across a body of evidence into a GRADE evidence profile.  
1140  
1141  
1142  
1143

### 1144 **3.1. Advantages and disadvantages of using the RoB instrument for NRS of exposures approach**

1145  
1146 Some challenges remain, specifically when defining the target experiment and making judgments at the  
1147  
1148 study and review level. The major challenge when identifying a hypothetical target randomized  
1149  
1150 experiment is that much of the research on environmental health exposures focuses on a potential link  
1151  
1152 with a human health hazard. Defining a specific comparison to an exposure presents a challenge, as  
1153  
1154 there may be a paucity of evidence to support the distinct exposure and comparator; however, in this  
1155  
1156 paper we present five scenarios to facilitate the identification of an exposure and comparator [12]. In  
1157  
1158 addition, the best available studies to inform a review may only present data on one exposure category.  
1159  
1160 In this situation, we recommend other sources of comparative exposure data, such as historical controls  
1161  
1162 (i.e. source of data presents levels of exposure before and after introduction to a known source of  
1163  
1164 exposure).  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176

1177  
1178  
1179 Inter-rater reliability of the RoB instrument for NRS of exposures has not yet been measured; however,  
1180  
1181 the purpose of the RoB instrument for NRS of exposures is not necessarily to have different experts  
1182  
1183 reach the same judgment per study and across studies, but instead to justify the judgements and make the  
1184  
1185 judgements transparent. We present several examples when using the RoB instrument for NRS of  
1186  
1187 exposures. More examples are needed to highlight nuances of this instrument when applied on an  
1188  
1189 individual-study and across-study basis.  
1190  
1191  
1192

1193 Based on concerns from systematic-review authors and guideline developers in the environmental health  
1194  
1195 field, the RoB instrument for NRS of exposures evaluates bias using a standardized comparison to a  
1196  
1197 hypothetical target experiment. This allows the body of evidence to start at ‘High’ initial CoE within the  
1198  
1199 GRADE framework, potentially improving acceptability of this instrument and the use of GRADE for  
1200  
1201 environmental decision-making assessments. Of note is that randomized controlled exposure trials in  
1202  
1203 animals would be evaluated with the framework for randomized trials and not the herein described  
1204  
1205 instrument.  
1206  
1207  
1208

#### 1209 **4.2. Relation to other studies**

1210

1211 This is the first article describing examples from systematic reviews using the RoB instrument for NRS  
1212  
1213 of exposures to evaluate the RoB across a body of evidence for a specific outcome. We present one  
1214  
1215 option of a RoB matrix displaying the RoB study- and item-level judgments. In addition, we present  
1216  
1217 examples of when an individual and a body of evidence RoB judgment may be improved (determined to  
1218  
1219 be a less severe RoB) based on further exploration of residual and unmeasured confounding. We  
1220  
1221 highlight the value added by performing sensitivity analyses with the body of evidence to explore  
1222  
1223 sources of bias.  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232

1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288

The application of ROBINS-I for RoB assessment across a body of evidence is undergoing further development, as are the procedures for interpreting RoB within the GRADE approach when NRS are compared to RCTs as in the RoB instrument for NRS of exposures or ROBINS-I [9]. Collaboration between the developers of the RoB instrument for NRS of exposures and these projects allows for an iterative approach to methods advancements. We expect that this approach would be applicable to broader research of exposures conducted in the fields of public health and nutrition, not limited to environmental exposures.

#### **4.3. Implications for stakeholders using the RoB instrument for NRS of exposures**

Evaluating the RoB across the body of evidence for an outcome informs one domain within the GRADE framework's evidence assessment contributing to the understanding about the overall CoE. Using this instrument should not result in a final certainty distinct from the prior approach of starting NRS at 'Low' initial CoE within GRADE because the conceptual underpinnings are the same. However, the approach is fairer and more transparent. Indeed, users may prefer investigating the relationship between rating down for imbalances due to confounders, selection bias, or misclassification of the exposure instead of starting at 'Low' initial CoE as a general judgment about these items. The process and examples outlined in this manuscript provide guidance for researchers and guideline developers using evidence about exposures to inform their systematic reviews and decision making.

#### **4.4. Unanswered questions and future research**

This research provides many opportunities for further application and assessment of the RoB instrument for NRS of exposures and integration into GRADE. Specific areas of interest based on our research may include 1) how to apply the RoB instrument for NRS of exposures to primary studies that use different



1289 exposure measurement strategies; 2) the process for making a judgment about the body of evidence  
1290  
1291 when using different techniques to synthesize evidence of the effects; and 3) the role of dose-response  
1292  
1293 within RoB and GRADE assessments.  
1294  
1295  
1296  
1297  
1298

1299 We present several measurement strategies that may be used when direct measures of the exposure are  
1300  
1301 unfeasible or not available, such as modelling, or environmental or personal monitoring. Each method  
1302  
1303 may be associated with greater or lesser specificity and/or potential for exposure misclassification.  
1304  
1305 Application of the RoB instrument for NRS of exposures to topics using these measures is needed.  
1306  
1307

1308 In addition, we present the process for when the RoB across a body of evidence can be further explored  
1309  
1310 and assessed by using meta-analytic approaches; however, systematic reviews of exposures may use  
1311  
1312 other approaches to summarize evidence, such as a qualitative analysis or narrative summary. Further  
1313  
1314 exploration of how these methods may translate to different summary approaches is needed.  
1315  
1316  
1317

1318 Lastly, while we present situations of where magnitude of effect and opposing residual confounding  
1319  
1320 may decrease our concerns about bias within both individual assessments and across the body of  
1321  
1322 evidence, more exploration of the role of dose-response is needed. Future research should provide  
1323  
1324 examples of how to incorporate dose response into an assessment using the RoB instrument for NRS of  
1325  
1326 exposures.  
1327  
1328

## 1329 **5. Conclusions**

1330

1331 The RoB instrument for NRS of exposures provides a novel approach for evaluating RoB of exposures.  
1332  
1333 Determining the RoB across a body of evidence is critical to inform decision making about health  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344

1345  
1346  
1347 exposures. We present guidance and examples for systematic-review authors and guideline developers  
1348  
1349 to follow when using this instrument.  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400

## Acknowledgments

We are grateful to all systematic review and GRADE members who have collaborated with their feedback and suggestions for this work. We thank Jani Ruotsalainen, Finnish Institute of Occupational Health, Cochrane Work Review Group, for his help in formulating the PECO questions on noise exposure.

1457  
1458  
1459  
1460 **Funding Sources**  
1461

1462 This research was supported by the Intramural Research Program of the National Institute of  
1463  
1464 Environmental Health Sciences and the MacGRADE Centre at the McMaster University.  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512

## Authors' Contributions

RLM, KAT, and HJS designed and conceived of the study. RB, SEE, AEG, and PR conducted the risk of bias evaluations and provided feedback on its use. RLM and HJS developed the schematic. RLM, KAT, NS, ACH, and HJS reviewed suggestions for operationalization and integration of the instrument. RLM drafted the manuscript. KAT, NS, ACH, MA, EA, TF, AH, JM, RAM, JV, MV, PW and HJS reviewed the manuscript and provided major revisions. All authors read and approved the final manuscript.

## References

1. Woodruff TJ, Sutton P: **The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes.** *Environ Health Perspect* 2014, **122**(10):1007-1014.
2. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S *et al*: **GRADE guidelines: 3. Rating the quality of evidence.** *J Clin Epidemiol* 2011, **64**(4):401-406.
3. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D: **The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration.** *PLoS medicine* 2009, **6**(7):e1000100.
4. Rooney AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Boyles AL, Ratcliffe JM, Kraft AD, Schünemann HJ, Schwingl P: **How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards.** *Environment international* 2016, **92**:617-629.
5. Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, Goldstone AE, Ross P, Guyatt G, Schünemann HJ: **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.** *Environment international* 2018, **120**:382-387.
6. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I *et al*: **ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions.** *BMJ* 2016, **355**:i4919.
7. Cochran WG, Chambers SP: **The planning of observational studies of human populations.** *Journal of the Royal Statistical Society Series A (General)* 1965, **128**(2):234-266.
8. Hernán MA, Robins JM: **Using big data to emulate a target trial when a randomized trial is not available.** *American journal of epidemiology* 2016, **183**(8):758-764.
9. Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, Morgan RL, Gartlehner G, Kunz R, Katikireddi SV: **GRADE Guidelines: 18. How ROBINS-I and other tools to assess risk of bias in non-randomized studies should be used to rate the certainty of a body of evidence.** *Journal of clinical epidemiology* 2018.
10. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y *et al*: **GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias).** *J Clin Epidemiol* 2011, **64**(4):407-415.

- 1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680
11. Morgan RL, Thayer KA, Bero L, Bruce N, Falck-Ytter Y, Ghera D, Guyatt G, Hooijmans C, Langendam M, Mandrioli D *et al*: **GRADE: Assessing the quality of evidence in environmental and occupational health.** *Environ Int* 2016, **92-93**:611-616.
  12. Morgan RL, Whaley P, Thayer KA, Schunemann HJ: **Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes.** *Environ Int* 2018.
  13. Carwile JL, Michels KB: **Urinary bisphenol A and obesity: NHANES 2003-2006.** *Environ Res* 2011, **111**(6):825-830.
  14. Ranciere F, Lyons JG, Loh VH, Botton J, Galloway T, Wang T, Shaw JE, Magliano DJ: **Bisphenol A and the risk of cardiometabolic disorders: a systematic review with meta-analysis of the epidemiological evidence.** *Environ Health* 2015, **14**(1):46.
  15. Harley KG, Schall RA, Chevrier J, Tyler K, Aguirre H, Bradman A, Holland NT, Lustig RH, Calafat AM, Eskenazi B: **Prenatal and postnatal bisphenol A exposure and body mass index in childhood in the CHAMACOS cohort.** *Environmental health perspectives* 2013, **121**(4):514.
  16. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G *et al*: **GRADE guidelines: 8. Rating the quality of evidence--indirectness.** *Journal of clinical epidemiology* 2011, **64**(12):1303-1310.
  17. Thayer K, Rooney A, Boyles A, Holmgren S, Walker V, Kissling G, U.S. Department of Health and Human Services: **Draft protocol for systematic review to evaluate the evidence for an association between bisphenol A (BPA) exposure and obesity.** *National Toxicology Program* 2013.
  18. Higgins J, Sterne J, Savovic J, Page M, Hrobjartsson A, Boutron I, Reeves B, Eldridge S: **A revised tool for assessing risk of bias in randomized trials** In: *Cochrane Methods*. Edited by Chandler J, McKenzie J, Boutron I, Welch V. <http://www.cochranelibrary.com/dotAsset/ecafc5c7-0b9b-4cd1-a4c1-8b0013aea046.pdf>; 2016.
  19. Li D-K, Miao M, Zhou Z, Wu C, Shi H, Liu X, Wang S, Yuan W: **Urine bisphenol-A level in relation to obesity and overweight in school-age children.** *PloS one* 2013, **8**(6):e65399.
  20. Wang H-x, Zhou Y, Tang C-x, Wu J-g, Chen Y, Jiang Q-w: **Association between bisphenol A exposure and body mass index in Chinese school children: a cross-sectional study.** *Environmental Health* 2012, **11**(1):79.
  21. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, Atkins D, Kunz R, Brozek J, Montori V *et al*: **GRADE guidelines: 9. Rating up the quality of evidence.** *J Clin Epidemiol* 2011, **64**(12):1311-1316.

- 1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736
22. Doll R, Hill AB: **Smoking and carcinoma of the lung.** *British medical journal* 1950, **2**(4682):739.
  23. Doll R, Hill AB: **Mortality in relation to smoking: ten years' observations of British doctors.** *British medical journal* 1964, **1**(5395):1399.
  24. Bross ID: **Spurious effects from an extraneous variable.** *Journal of chronic diseases* 1966, **19**(6):637-647.
  25. Blair A, Stewart P, Lubin JH, Forastiere F: **Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures.** *American journal of industrial medicine* 2007, **50**(3):199-207.
  26. Steenland K, Savitz DA: **Topics in environmental epidemiology:** Oxford University Press, USA; 1997.
  27. Kogevinas M: **Epidemiological approaches in the investigation of environmental causes of cancer: the case of dioxins and water disinfection by-products.** In: *Environmental Health: 2011:* BioMed Central; 2011: S3.
  28. Jurek AM, Greenland S, Maldonado G, Church TR: **Proper interpretation of non-differential misclassification effects: expectations vs observations.** *International journal of epidemiology* 2005, **34**(3):680-687.
  29. Nieuwenhuijsen MJ: **Exposure assessment in environmental epidemiology:** Oxford University Press, USA; 2015.
  30. Cox KJ, Porucznik CA, Anderson DJ, Brozek EM, Szczotka KM, Bailey NM, Wilkins DG, Stanford JB: **Exposure classification and temporal variability in urinary bisphenol A concentrations among couples in Utah—the HOPE study.** *Environmental health perspectives* 2016, **124**(4):498.
  31. Vandenberg LN, Hunt PA, Myers JP, vom Saal FS: **Human exposures to bisphenol A: mismatches between data and assumptions.** *Reviews on environmental health* 2013, **28**(1):37-58.
  32. Thayer KA, Doerge DR, Hunt D, Schurman SH, Twaddle NC, Churchwell MI, Garantziotis S, Kissling GE, Easterling MR, Bucher JR: **Pharmacokinetics of bisphenol A in humans following a single oral administration.** *Environment international* 2015, **83**:107-115.
  33. Higgins J, Green S: **Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 (updated March 2011).** <http://handbook.cochrane.org/> [accessed 3 February 2013]. 2011.



## Appendices

### Appendix A. Step I of the RoB instrument for NRS of exposures for the PECO: “What is the effect of highest levels vs. lowest levels of BPA exposure on weight?”

Step I Items	Response
Confounding for BPA and obesity	<ul style="list-style-type: none"> <li>• Body composition (age, ethnicity, gender, height, race);</li> <li>• Weight (age, gender);</li> <li>• Waist circumference (age, gender);</li> <li>• Body mass index (age, ethnicity, gender, race);</li> <li>• In addition, consumption of canned or packaged food and drink (“processed” food) that is also energy dense and low-nutrient (e.g., soda) is a significant confounder because food packaging is a main source of exposure to BPA.</li> <li>• Co-exposures: There may be some concern for co-exposure to certain phthalates used in food packaging that have also been linked to obesity. However, phthalates are used in different types of food packaging than BPA (plastic wraps versus canned lining and polycarbonate materials). No other <i>a priori</i> co-exposures of particular concern are identified for general population studies. There may be some co-exposures that need to be considered in occupational studies and these should be assessed on a case by case basis if discovered.</li> </ul>
Co-interventions	<ul style="list-style-type: none"> <li>• None identified</li> </ul>
Accuracy of the measurement of exposure to BPA (CAS# 80-05-7)	<ul style="list-style-type: none"> <li>• BPA is a non-persistent compound (near 100% elimination within 24 hours after oral exposure, possible longer elimination time from non-oral exposure but on order of days), so blood and urine measures only assess recent exposure. This means current exposure levels may NOT be indicative of past exposures. This is problematic for assessment of BPA as a risk factor for health outcomes that are not acute and take time to develop like obesity.</li> <li>• BPA measures are variable over time in the same person (even during the same day) so methods that utilize repeated measures of exposure are preferred. Some experts on BPA exposure assessment express less concern for lack of repeated measures for NHANES data because it is a large sample survey of the general population.</li> <li>• Standard analytical measures: Measurement of urine or blood by quantitative techniques such as liquid chromatography-triple quadrupole mass spectrometry (LC-MS/MS) and high-pressure liquid chromatography with tandem mass spectrometry (HPLC/MS) are preferred. Measurements made at CDC are considered high-quality.</li> <li>• Measures to minimize sample contamination with BPA should be taken (e.g., glass pipettes, polypropylene plastic lab ware and sample collection materials, water blanks).</li> <li>• Measures of unconjugated BPA in blood need to be very carefully considered based on extent to which investigators controlled for background exposures.</li> <li>• Questionnaire or self-reported measures of BPA exposure are more problematic due to the ubiquity of exposure and lack of</li> </ul>

1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820

	knowledge on all possible routes of exposure, e.g., thermal paper, certain pharmaceuticals. However, there is some support for an association between higher urine/blood levels of BPA and higher reported use of BPA-containing food packaging (e.g., canned food consumption) or handling of BPA-containing thermal paper (cashiers) so questionnaire data that assess these types of exposure sources may have some utility in assessing longer-term time trends in exposure.
Accuracy of the measurement of outcome of obesity	<ul style="list-style-type: none"><li>• Body Composition: Dual-energy X-Ray absorptiometry, triceps skinfold thickness, subscapular skinfold thickness, suprailiac skinfold thickness</li><li>• Measured waist circumference</li><li>• Body mass index</li><li>• Measured weight</li></ul> *Obesity typically develops relatively slowly over time so preferred follow-up times after start of exposure would be on the order of several months to years.

1821  
1822  
1823 **Appendix B. Step II of the RoB instrument for NRS of exposures for Carwile & Michels, 2011**  
1824  
1825

1826 **Specify a target randomized trial specific to the study**

1827 Design	Individual randomized controlled trial
1828 Participants	Adults of all ages, predominantly 18-35 years (8.2% <18 years and 7.9% > 35 years). Civilian, non-institutionalized, United States population. Analyses restricted to participants 18–74 years of age, who were included in the random subsample of participants, who supplied a spot urine sample analyzed for BPA.
1829 Experimental intervention	BPA highest levels (quartile 4: $\geq 4.7$ ng/mL)
1830 Comparator	BPA lowest levels (quartile 1: $\leq 1.1$ ng/mL)

1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838 **Specify the outcome**  
1839

1840 Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table).  
1841 Specify whether this is a proposed benefit or harm of intervention.

1842  
1843 Prevalent overweight (Overweight:  $25 \leq \text{BMI} < 30$  kg/m<sup>2</sup> [reference: BMI <25 kg/m<sup>2</sup>])  
1844

1845  
1846 **Specify the numerical result being assessed**  
1847

1848 In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a  
1849 reference (e.g., to a table, figure or paragraph) that uniquely defines the result being assessed.

1850  
1851 Participants in the upper BPA quartile 4 vs. participants in the lowest BPA quartile 1: OR: 1.76, 95% CI: 1.06–2.94)  
1852  
1853  
1854  
1855  
1856  
1857

1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904

<b>(i) Confounding domains listed in Step I</b>				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favor the experimental intervention or the comparator?
			Yes / No / No information	Favor experimental / Favor comparator / No information
Age, gender	Weight	No	Yes	Favor experimental
Consumption of canned or packaged food and drink ("processed" food) that is also energy dense and low-nutrient (e.g., soda)	Daily caloric intake	No	No	Favor experimental because obese individuals (potentially caused by higher consumption of canned foods and drinks) have higher urinary BPA levels relative to those with normal weight.

<b>(ii) Additional confounding domains relevant to the setting of this particular study, or which the study authors identified as important</b>				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favor the experimental intervention or the comparator?
			Yes / No / No information	Favor experimental / Favor comparator / No information
Alcohol drinking, fish intake, protein, fat, carbohydrate, and energy intake	none	no	no	

Carwile JL, Michels KB: **Urinary bisphenol A and obesity: NHANES 2003–2006.** *Environmental research* 2011, **111**(6):825-830.

1905  
1906  
1907  
1908 **Appendix C. Step II of the RoB instrument for NRS of exposures for Harley et al., 2013**  
1909

1910 **Specify a target randomized trial specific to the study**

1911 Design	Individual randomized controlled trial
1912 Participants	Children at 5 and 9 years of age born to eligible pregnant women were at least 18 years of age, spoke English or Spanish, qualified for low-income health insurance, were at < 20 weeks gestation, and were planning to deliver at the county hospital. Must have had a singleton, live birth.
1913 Experimental intervention	BPA highest levels (tertile 3: 4.6–349.8 µg/g)
1914 Comparator	BPA lowest levels (tertile 1: <LOD-2.4 µg/g)

1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922 **Specify the outcome**

1923 Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table).  
1924 Specify whether this is a proposed benefit or harm of intervention.

1925  
1926  
1927 Prevalent overweight (Overweight: BMI  $\geq$  85<sup>th</sup> percentile at 5 and 9 years of age)  
1928

1929  
1930 **Specify the numerical result being assessed**

1931 In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a  
1932 reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.  
1933

1934  
1935 Participants in the upper BPA tertile 3 vs. participants in the lowest BPA tertile 1: OR = 1.36 (0.75–2.47)  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946

1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988

<b>(i) Confounding domains listed in Step I</b>				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favor the experimental intervention or the comparator?
			Yes / No / No information	Favor experimental / Favor comparator / No information
Age, gender	Weight	No	Yes	Favor experimental
Consumption of canned or packaged food and drink (“processed” food) that is also energy dense and low-nutrient (e.g., soda)	Child consumption of soda, fast food, and sweets	No	Yes	Favor experimental because obese individuals (potentially caused by higher consumption of canned foods and drinks) have higher urinary BPA levels relative to those with normal weight.

<b>(ii) Additional confounding domains relevant to the setting of this particular study, or which the study authors identified as important</b>				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favor the experimental intervention or the comparator?
			Yes / No / No information	Favor experimental / Favor comparator / No information
Television watching	Average daily TV time	No	Yes	Favor experimental
Environmental tobacco smoke exposure	Self-reported mother’s smoking status	No	Yes	No information
Time spent playing outdoors	Unknown	No	No information	No information

Harley KG, Schall RA, Chevrier J, Tyler K, Aguirre H, Bradman A, Holland NT, Lustig RH, Calafat AM, Eskenazi B: **Prenatal and postnatal bisphenol A exposure and body mass index in childhood in the CHAMACOS cohort.** *Environmental health perspectives* 2013, **121**(4):514.

**Appendix D. Summary of Step III of the RoB instrument for NRS of exposures and the direction of bias and reaching the overall bias judgement for Carwile & Michels, 2013**

Bias items	Risk of bias	Direction of bias	Rationale
Bias due to confounding	Serious	Unknown	<p>NHANES data were used. Specific details were not provided in the study report, but NHANES co-variate data were obtained from either a standardized questionnaire or laboratory methods (e.g., creatinine). The reliability/validity of the questionnaire was not reported, but it is not expected to appreciably bias the results. Most of the critical confounders were considered statistically, but there is possibility of residual unmeasured (and unidentified) confounding. For the most part, although certain post-exposure variables are relevant to evaluating obesity (e.g., caloric intake), there is little information on the association of these variables to BPA exposure.</p> <p>No indication that time-varying confounding is a major concern given the cross-sectional nature of the study.</p> <p>Critical confounders (age, gender, and ethnicity) were accounted for in the analysis. Model 1 was adjusted for age, sex, and urinary creatinine. Model 2 was adjusted for race, education, and smoking in addition to Model 1 covariates.</p>
Bias in selection of participants into the study	Low	N/A	<p>Study is cross-sectional. Subjects were randomly selected from NHANES subjects with urinary BPA data available using the same criteria. Selection of subjects was unrelated to either exposure or outcome.</p> <p>While there is no information on start of exposure, everyone is exposed to BPA throughout their life, but the levels will change over time. Although BPA is ubiquitous, start of exposure and how exposure changes over time are not known. Timing of recruitment was similar (2003-2006) but given that the age ranged from 18 to 74 years, exposure could range by more than a decade.</p>

2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072

Bias in classification of exposures	Critical	Concerns of bias toward the null due to non-differential misclassification of the exposure.	<p>Urinary BPA concentration was measured in 1 spot sample from each participant. The lower limit of detection (LLOD) was 0.36ng/ml in 2003/04 and 0.4ng/ml in2005/06. For BPA concentrations below the LLOD (2003/04: n=110/1373 [8%]; 2005/06: n=114/1374 [8%]) NHANES assigned a value of the LLOD divided by the square root of two. BPA is a non-persistent compound and exposure measures were not repeated. Therefore, there is no confidence that the current exposure reflects exposure over the subject's life time or even over any duration of time. Because this population is obtained from NHANES some experts consider the lack of repeated measures to be less of a concern because it is a large survey of the general population (this cross-sectional study had a population of 2747 adults).</p> <p>Exposure was measured at same time as outcome, but participants were likely exposed throughout life due to BPA being a ubiquitous exposure. Therefore, it is unlikely that entry into the cohort started with the exposure.</p> <p>Cross-sectional analyses with both BPA exposure and weight, height, and waist circumference used to define obesity assessed simultaneously.</p> <p>Urine samples were obtained at the time that obesity measurements were obtained and analyzed later in a laboratory separate from where the data were collected. In addition, NHANES collected data on a variety of compounds and health effects without knowledge of the intent for this current study indicating that exposure status is not likely to be biased by knowledge of the outcome.</p> <p>The range/variability in exposure was likely sufficient with a 25th to 75th percentile range of 1.18 to 3.33 ng/mL urinary BPA ng/mL and quartiles ranging from &lt;1.1 ng/mL to &gt;4.7 ng/mL. However, we are not confident that the subjects were exposed to this concentration for a long period of time. Lacking information on the duration that subjects were exposed to these levels, the single BPA measurement obtained at the same time as outcome is not of sufficient to detect an effect of exposure.</p> <p>Urinary BPA samples were collected at the same time that height, weight, and waist</p>
-------------------------------------	----------	---	--



2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114

			<p>circumference were measured. Because BPA is not persistent, and obesity is not an acute effect, there is not adequate follow-up period to allow for the development of the outcome of interest.</p> <p>Total (free and conjugated) urinary BPA concentrations were measured at the Division of Environmental Health Laboratory Sciences (National Center for Environmental Health, CDC) using online solid-phase extraction coupled to isotope dilution high-performance liquid chromatography–tandem mass spectrometry. Quality control (QC) procedures included analysis of reagent blanks and samples of pooled human urine spiked with BPA at low-and high-concentrations. Coefficients of variation calculated for low-and high-concentration QC samples were 19% and 12% in 2003–2004 and 13% and 11% in 2005–2006. Additional information on laboratory methods is available online (CDC, 2004b, 2006b).</p>
Bias due to deviations from intended exposures	Low	N/A	<p>There is little concern that changes in exposure status occurred among participants. Although BPA levels may change overtime, the cross-sectional nature of the study and the intention-to-treat analyses this is of little concern because participants are analyzed based on the exposure group they are assigned from the single measurement. No critical co-exposures were identified and nothing about the subject characteristics suggests likelihood of differential exposure to other environmental contaminants at lower versus higher concentrations of BPA.</p>

2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156

Bias due to missing data	Low	N/A	<p>There is no information on the missing data by exposure level, but it is unlikely to be related to exposure level.</p> <p>The missing indicator method was used for covariates with missing data for <math>\geq 10\%</math> of observations, otherwise observations with missing covariate data were excluded. Data excluded from analysis did not exceed 4% and is considered relatively complete. 32 or 87 observations were stated excluded from analysis due to missing BMI data depending on the analysis conducted. 47 participants were excluded based on missing urinary BPA measurements. There were observations excluded based on missing covariate data. The number varied with the analysis but was only excluded if it was <math>&lt; 10\%</math>.</p>
--------------------------	-----	-----	---

2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  
2171  
2172  
2173  
2174  
2175  
2176  
2177  
2178  
2179  
2180  
2181  
2182  
2183  
2184  
2185  
2186  
2187  
2188  
2189  
2190  
2191  
2192  
2193  
2194  
2195  
2196  
2197  
2198

Bias in measurement of the outcome	Low	N/A	<p>It is unlikely that the outcome could be affected by knowledge of exposure. Height, weight, and waist circumference were measured using standard NHANES protocols (not described in the publication, but available on NHANES website). Body mass index was calculated (weight (kg)/height (m)<sup>2</sup>). The specific measurements would not be affected by knowledge of exposure, and it is unlikely that the calculation or assignment into obesity category would be affected by knowledge of exposure.</p> <p>Specific methods were not reported in the study report but are provided on NHANES website. Height and weight are likely sensitive measurements with waist circumference likely slightly less sensitive. Height, weight, and waist circumference were measured by trained technicians using a standardized protocol. Method details, including QA/QC procedures, are available on the NHANES website. BMI was calculated as weight in kilograms divided by height in meters squared and used to define overweight [25.0 &lt; BMI &lt; 29.9] and obesity [BMI &gt; 30.0].</p> <p>It is unlikely that any systematic error in measuring height, weight, or waist circumference (or in calculating the BMI or assigning obesity category) would have been related to exposure. NHANES has a standard protocol for measuring height, weight, and waist circumference that would have been used for all subjects. Outcome was assessed at the time of sample collection for exposure. Therefore, exposure was unknown at time of outcome assessment.</p>
Bias in selection of the reported result	Low	N/A	<p>Reporting of the results is consistent with an a priori plan and data were readily available from NHANES that provides all protocols for obtaining the data online. Results were provided for two measurements of obesity, which were reported in the methods making it unlikely that there is selective reporting based on outcome. Statistical methods reported in the methods section were used and presented in the results. Associations between urinary BPA and obesity were assessed for effect modification by gender, which were provided in the supplemental material.</p>

2199  
2200  
2201  
2202  
2203  
2204  
2205  
2206  
2207  
2208  
2209  
2210  
2211  
2212  
2213  
2214  
2215  
2216  
2217  
2218  
2219  
2220  
2221  
2222  
2223  
2224  
2225  
2226  
2227  
2228  
2229  
2230  
2231  
2232  
2233  
2234  
2235  
2236  
2237  
2238  
2239  
2240

<b>Overall bias</b>	<b>Serious</b>	<b>Possibly toward the null</b>	Overall bias was judged as Serious due to concerns of potential unknown confounders, unmeasured confounding due to the single time-point data collection, and concerns of non-differential misclassification of the exposure.
---------------------	----------------	---------------------------------	---

Carwile JL, Michels KB: **Urinary bisphenol A and obesity: NHANES 2003–2006.** *Environmental research* 2011, **111**(6):825-830.

2241  
2242  
2243  
2244  
2245  
2246  
2247  
2248  
2249  
2250  
2251  
2252  
2253  
2254  
2255  
2256  
2257  
2258  
2259  
2260  
2261  
2262  
2263  
2264  
2265  
2266  
2267  
2268  
2269  
2270  
2271  
2272  
2273  
2274  
2275  
2276  
2277  
2278  
2279  
2280  
2281  
2282

**Appendix E. Summary of Step III of the RoB instrument for NRS of exposures and the direction of bias and reaching the overall bias judgement for Harley et al., 2013**

<b>Bias items</b>	<b>Risk of bias</b>	<b>Direction of bias</b>	<b>Rationale</b>
-------------------	---------------------	--------------------------	------------------

2283  
2284  
2285  
2286  
2287  
2288  
2289  
2290  
2291  
2292  
2293  
2294  
2295  
2296  
2297  
2298  
2299  
2300  
2301  
2302  
2303  
2304  
2305  
2306  
2307  
2308  
2309  
2310  
2311  
2312  
2313  
2314  
2315  
2316  
2317  
2318  
2319  
2320  
2321  
2322  
2323  
2324

Bias due to confounding	Serious	Unknown	<p>Most of the critical confounders were considered statistically, but there is possibility of residual unmeasured (e.g., diet, pesticide exposure) confounding.</p> <p>The study evaluated the child's BPA exposure throughout several points in their life. And used each one separately in the evaluation.</p> <p>Changes in BPA exposure could be related to changes in food consumption over time as BPA exposure is mainly through canned or processed food including soda, which could also be related to obesity. Since Harley follows participants over time, there is some concern for time-varying confounding as they may have changed their diet while pregnant.</p> <p>Potential confounders were identified a priori using directed acyclic graphs. Potential confounders included maternal pre-pregnancy BMI, age, education, years of residence in the United States, smoking during pregnancy, soda consumption during pregnancy, and family income. Time-varying covariates considered were child consumption of soda, fast food, and sweets, television watching, environmental tobacco smoke exposure, and time spent playing outdoors, assessed at multiple times during childhood. Covariates were included in the final models if they were associated with both exposure and any of the growth outcomes at p-value &lt; 0.2 or if removing them changed the coefficient for the main BPA exposure variable by &gt; 10%. Maternal age and pre-pregnancy BMI were analyzed as continuous variables. Other variables were categorical. Mothers were interviewed twice during pregnancy, after delivery, and when their children were 2, 3.5, 5, 7, and 9 years of age to obtain information about demographic characteristics, diet, and behaviors. All interviews were conducted in English or Spanish using structured questionnaires, but no information was provided on reliability/validity. At the baseline interview, we asked mothers about their race/ethnicity, education, income, marital status, and number of years they had lived in the United States, as well as information about soda consumption, smoking, and alcohol and drug use during pregnancy. We calculated pre-pregnancy BMI from self-reported pre-pregnancy weight and measured height. If self-reported pre-pregnancy weight was unavailable or invalid, we used measured weight at first prenatal visit (n = 23) if the first prenatal visit occurred at or before 13 weeks gestation or used regression models to impute pre-pregnancy weight based on weight at all prenatal visits if the first prenatal visit occurred after 13 weeks (n = 16).</p>
-------------------------	---------	---------	---

2325  
2326  
2327  
2328  
2329  
2330  
2331  
2332  
2333  
2334  
2335  
2336  
2337  
2338  
2339  
2340  
2341  
2342  
2343  
2344  
2345  
2346  
2347  
2348  
2349  
2350  
2351  
2352  
2353  
2354  
2355  
2356  
2357  
2358  
2359  
2360  
2361  
2362  
2363  
2364  
2365  
2366

Bias in selection of participants into the study	Low	N/A	<p>Selection of subjects was unrelated to either exposure or outcome. The study sample consisted of participants in the Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS), a longitudinal cohort study of environmental factors and children’s growth and development. Pregnant mothers were enrolled Selection of subjects was unrelated to either exposure or outcome in 1999 and 2000 from prenatal clinics serving the farmworker population in the Salinas Valley, California. Eligible women were at least 18 years of age, spoke English or Spanish, qualified for low-income health insurance, were at &lt; 20 weeks gestation, and were planning to deliver at the county hospital. Mothers provided written informed consent for themselves and their children to participate in the study.</p> <p>Start of exposure occurred in the first trimester and all subjects were followed through 9 years of age.</p>
--	-----	-----	--

2367  
2368  
2369  
2370  
2371  
2372  
2373  
2374  
2375  
2376  
2377  
2378  
2379  
2380  
2381  
2382  
2383  
2384  
2385  
2386  
2387  
2388  
2389  
2390  
2391  
2392  
2393  
2394  
2395  
2396  
2397  
2398  
2399  
2400  
2401  
2402  
2403  
2404  
2405  
2406  
2407  
2408

Bias in classification of exposures	Moderate	Some concern of bias toward the null due to non-differential misclassification of the exposure.	<p>Urinary BPA concentration was measured in 4 spot samples, 2 during pregnancy and 2 from the child. LOD was 0.4 ng/mL. Concentrations &lt; LOD for which a signal was detected were reported as measured. Concentrations &lt; LOD with no signal detected were randomly imputed based on a log-normal probability distribution using maximum likelihood estimation. The number of collected samples increases our certainty in the correct classification of the higher exposed and lower exposed groups.</p> <p>Initial exposure was measured during the first trimester of pregnancy. While this may not be the exact date of start of exposure it would be very close for the children.</p> <p>Prenatal and five-year-old exposure measurements were taken prior to the assessment of BMI at 9 years.</p> <p>Exposure was assessed prior to the outcome at three different time points. Only one exposure measurement was obtained at the same time as the outcome; thus, it was not possible for classification of exposure to have been affected by the knowledge of the outcome.</p> <p>The range/variability in exposure was sufficient (range during pregnancy 0.5 to 4.6 ng/mL and during childhood 0.9 to 16.3 ng/mL). Although BPA levels change over time and we are not confident that the subjects were exposed to this concentration for a long period of time, the fact that there were 4 measurements per subject make us more confident in the exposure being represented of changes over time. In addition, since the child's exposure was first measured based on mother's levels when pregnant, then again when the children were 5 (4 years prior to measuring outcome) the duration of exposure would have been sufficient even if the level of this exposure was not consistent. BPA levels were also measured in the child at 9 years. However, data were not provided for the individual subjects to know how the BPA levels may have varied per subject.</p> <p>Children were followed up for 9 years, which would have been sufficient time for the outcome to develop.</p>
-------------------------------------	----------	---	---



2409  
 2410  
 2411  
 2412  
 2413  
 2414  
 2415  
 2416  
 2417  
 2418  
 2419  
 2420  
 2421  
 2422  
 2423  
 2424  
 2425  
 2426  
 2427  
 2428  
 2429  
 2430  
 2431  
 2432  
 2433  
 2434  
 2435  
 2436  
 2437  
 2438  
 2439  
 2440  
 2441  
 2442  
 2443  
 2444  
 2445  
 2446  
 2447  
 2448  
 2449  
 2450

			<p>Spot urine samples were collected from mothers at two time points during pregnancy: near the end of the first (mean ± SD, 13.8 ± 5.0 weeks gestation) and second (mean ± SD, 26.4 ± 2.4 weeks gestation) trimester of pregnancy and from the children when they were 5 (mean ± SD, 5.1 ± 0.2 years) and 9 (mean ± SD, 9.4 ± 0.4 years) years of age. Urine samples were collected in polypropylene urine cups, aliquoted into glass vials, and frozen at -80°C until shipment to the CDC for analysis. Analysis of field blanks showed no detectable contamination by BPA using this collection protocol. Solid-phase extraction coupled to high performance liquid chromatography–isotope dilution tandem mass spectrometry to measure total urinary BPA concentration (conjugated plus unconjugated). Concentrations &lt; LOD for which a signal was detected were reported as measured. Concentrations &lt; LOD with no signal detected were randomly imputed based on a log-normal probability distribution using maximum likelihood estimation. Specific gravity was measured with a refractometer (National Instrument Company Inc., Baltimore, MD) for the maternal urine samples, but was unavailable for the children’s samples. Thus, maternal concentrations were normalized for urinary dilution using urine specific gravity, and child BPA concentrations were normalized by dividing by urinary creatinine concentration.</p>
<p>Bias due to deviations from intended exposures</p>	<p>Low</p>	<p>N/A</p>	<p>There is little concern that changes in exposure status occurred among participants. Although BPA levels may change overtime, several measurements were obtained and evaluate separately by exposure they were assigned. Because each exposure was evaluated as an intent to treat, there is little concern about the potential changes in exposure. The study authors reanalyzed the models controlling separately for three important prenatal exposures in this population: organochlorine pesticides [using prenatal serum concentrations of dichlorodiphenyldichloroethylene (DDE)], organophosphate pesticides (using prenatal urinary metabolites of organophosphate pesticides), and brominated flame retardants [using prenatal serum concentrations of polybrominated diphenyl ethers (PBDEs)].</p>

2451  
2452  
2453  
2454  
2455  
2456  
2457  
2458  
2459  
2460  
2461  
2462  
2463  
2464  
2465  
2466  
2467  
2468  
2469  
2470  
2471  
2472  
2473  
2474  
2475  
2476  
2477  
2478  
2479  
2480  
2481  
2482  
2483  
2484  
2485  
2486  
2487  
2488  
2489  
2490  
2491  
2492

Bias due to missing data	Low	N/A	<p>Reasons for exclusion were documented and unlikely to differ across exposures threshold. Although some subjects were lost to follow-up and the missing data were not described by exposure status, the study authors conducted analyses that addressed loss to follow-up and are likely to have removed any risk of bias thus judged low risk of bias. There is no statement that participants with missing covariate data were excluded from analyses. There is no information on the missing data by exposure level. Although it is unlikely to be related to exposure level, they had the data in order to compare those lost to follow-up with those included in the analysis, but no information was provided.</p> <p>Of the 527 mothers meeting the inclusion criteria, 402 had at least one urine measurement available. There were 325 measurements in children at 5 years and 304 available at 9 years. Of the 402 children included in the analysis, anthropometric measurements were available for 319 children at 5 years and 311 children at 9 years.</p>
--------------------------	-----	-----	---

2493  
2494  
2495  
2496  
2497  
2498  
2499  
2500  
2501  
2502  
2503  
2504  
2505  
2506  
2507  
2508  
2509  
2510  
2511  
2512  
2513  
2514  
2515  
2516  
2517  
2518  
2519  
2520  
2521  
2522  
2523  
2524  
2525  
2526  
2527  
2528  
2529  
2530  
2531  
2532  
2533  
2534

Bias in measurement of the outcome	Low	N/A	<p>It is unlikely that the outcome could be affected by knowledge of exposure. It was not noted that outcome assessors were blind to the exposure level, but it was likely given that separate individuals were used to measure the outcome parameters than conducted the exposure analysis (i.e., CDC).</p> <p>The same methods were used for all participants at all times measured. It is unlikely that any systematic error in anthropometric measurements (or calculating the BMI or assigning obesity category) would have been related to exposure. Children were weighed and measured without jackets or shoes by trained study staff. Weight was measured using a digital scale and rounded to the nearest 0.1 kg. Height was measured using a stadiometer and rounded to the nearest 0.1 cm. Starting at 5 years of age, waist circumference was measured at each visit by placing a measuring tape around the abdomen at the level of the iliac crest, parallel to the floor. Height and waist circumference measurements were conducted in triplicate and averaged for analysis. When the children were 9 years of age, fat percentage was measured using “foot-to-foot” bio-impedance technology with a Tanita TBF-300A body composition analyzer (Tanita Corp.). BMI was calculated as weight (kilograms) divided by height squared (square meters) and compared with the sex-specific BMI-for-age percentile data issued by CDC in 2000 (National Center for Health Statistics 2005). Children who were <math>\geq</math> 85th but <math>&lt;</math> 95th percentile for their age and sex were classified as overweight. Age- and sex-standardized BMI z-scores were also generated using the CDC norms. These methods are considered sensitive.</p>
------------------------------------	-----	-----	--

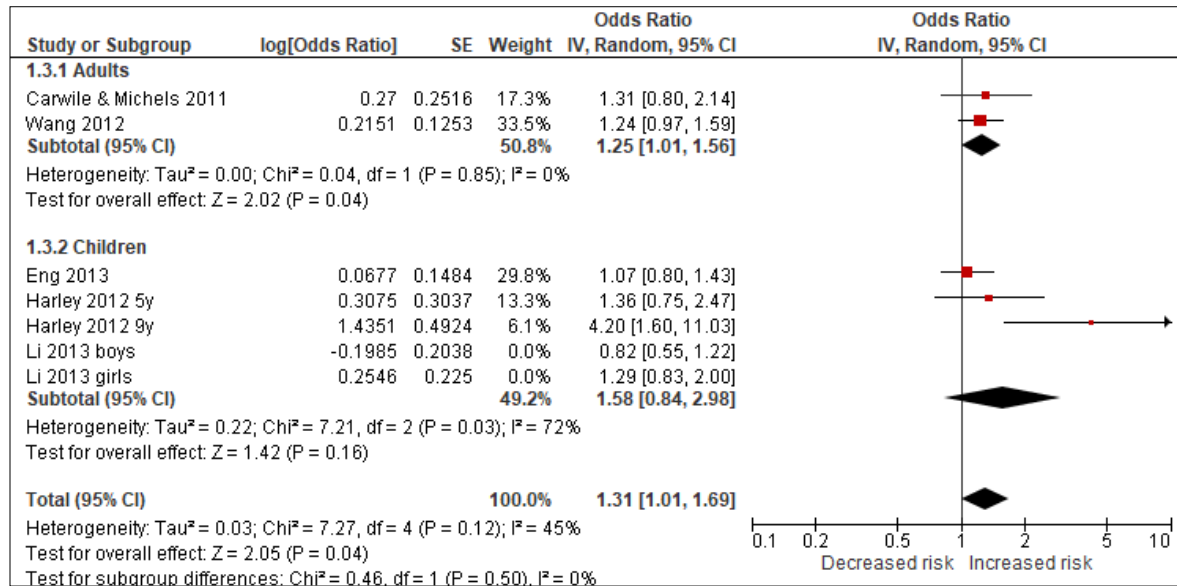
2535  
2536  
2537  
2538  
2539  
2540  
2541  
2542  
2543  
2544  
2545  
2546  
2547  
2548  
2549  
2550  
2551  
2552  
2553  
2554  
2555  
2556  
2557  
2558  
2559  
2560  
2561  
2562  
2563  
2564  
2565  
2566  
2567  
2568  
2569  
2570  
2571  
2572  
2573  
2574  
2575  
2576

Bias in selection of the reported result	Moderate	Potential for bias away from the null.	<p>Reported results are consistent with an a priori plan; however, as no protocol was published prior to the study there is potential for reporting bias to inflate results for publication success.</p> <p>Several measurements of obesity were evaluated and reported. These were also assessed at several different time periods in the children. Although the publication only shows a few of the results (both positive and negative), the BMI-z-scores for all ages are presented in the supplemental data indicating that it is unlikely that there was bias from selective reporting of outcome. Gender and age were evaluated as separate subgroups as described in the report.</p> <p>Statistical methods reported in the methods section were used and presented in the results or discussion. BPA was analyzed as categorical and continuous variable.</p>
<b>Overall bias</b>	<b>Moderate</b>	<b>Unknown</b>	Overall bias was judged as Moderate due to concerns of potential unknown confounders, some concerns of non-differential misclassification of the exposure, and some concerns with bias in reported results.

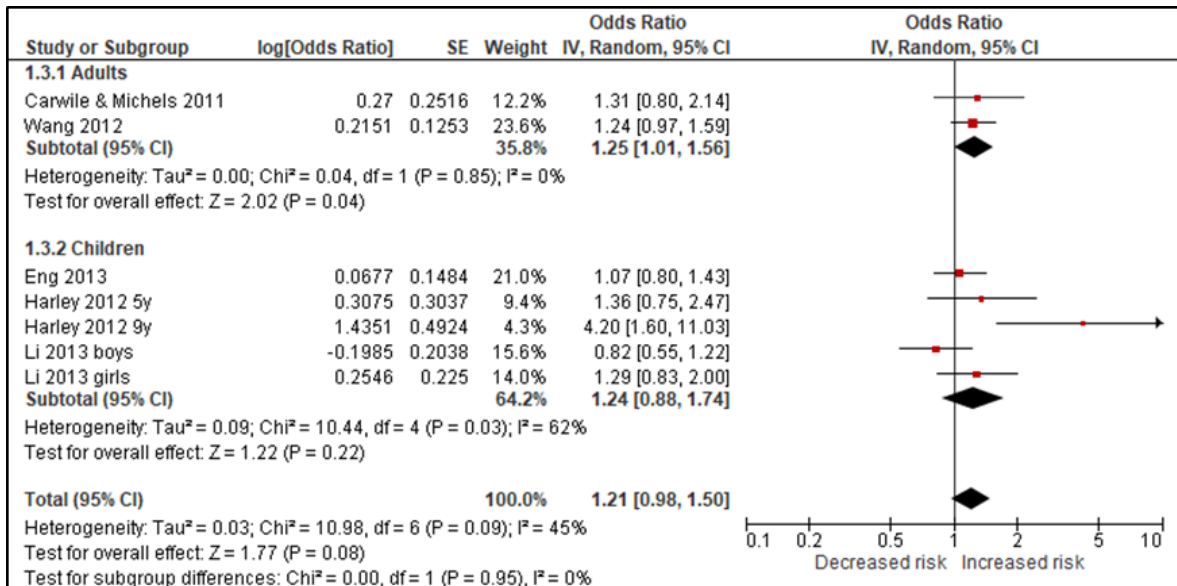
Harley KG, Schall RA, Chevrier J, Tyler K, Aguirre H, Bradman A, Holland NT, Lustig RH, Calafat AM, Eskenazi B: **Prenatal and postnatal bisphenol A exposure and body mass index in childhood in the CHAMACOS cohort.** *Environmental health perspectives* 2013, **121**(4):514.

## Appendix F. Sensitivity analysis for the outcome of prevalent overweight

Figure F.1. Sensitivity analysis of studies with ‘Serious’ bias due to confounding

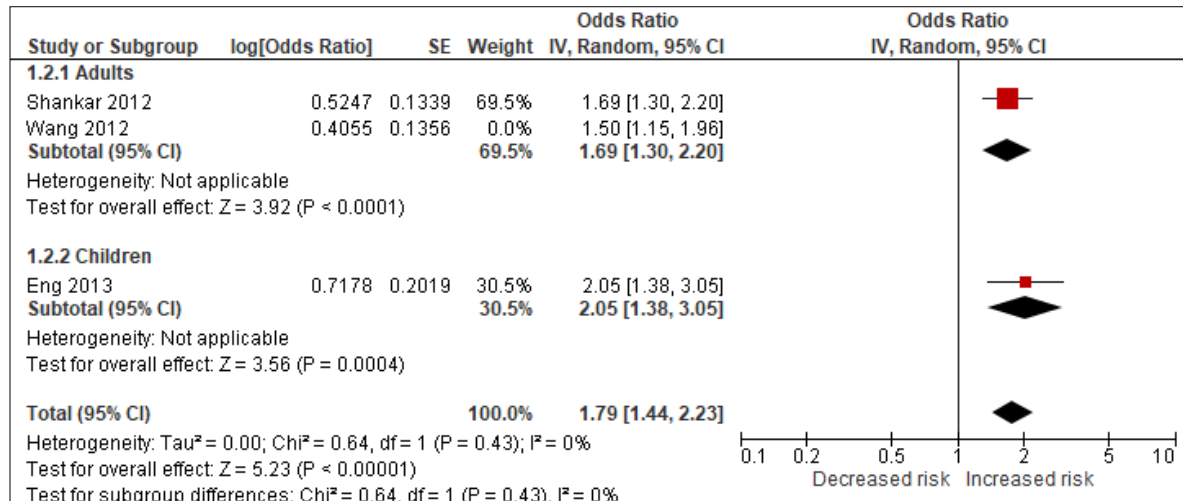


F.2. Sensitivity analysis of all studies

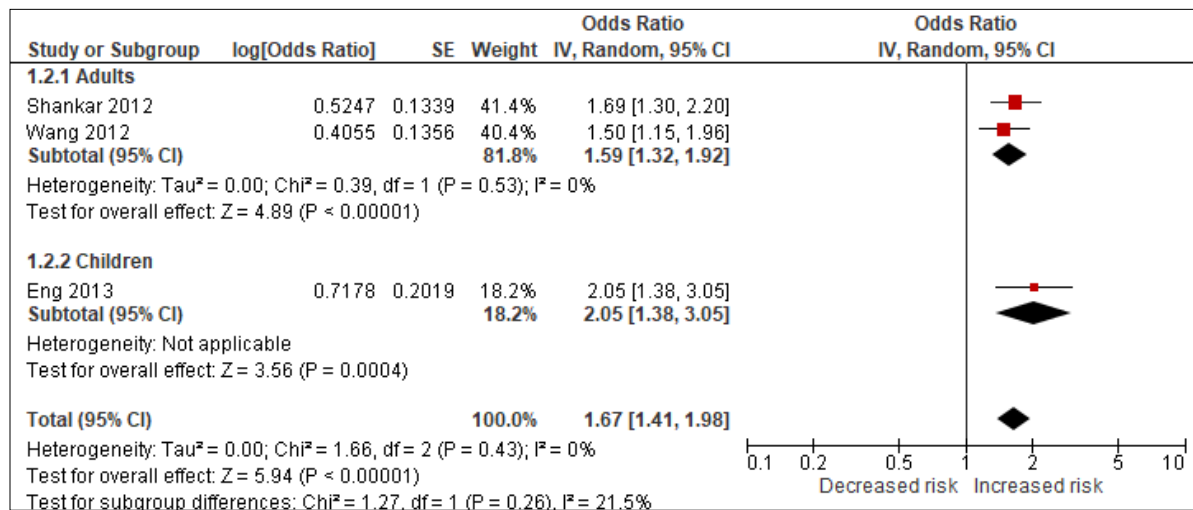


## Appendix G. Sensitivity analysis for the outcome of prevalent obesity

### G.1. Sensitivity analysis of studies with 'Serious' bias due to confounding

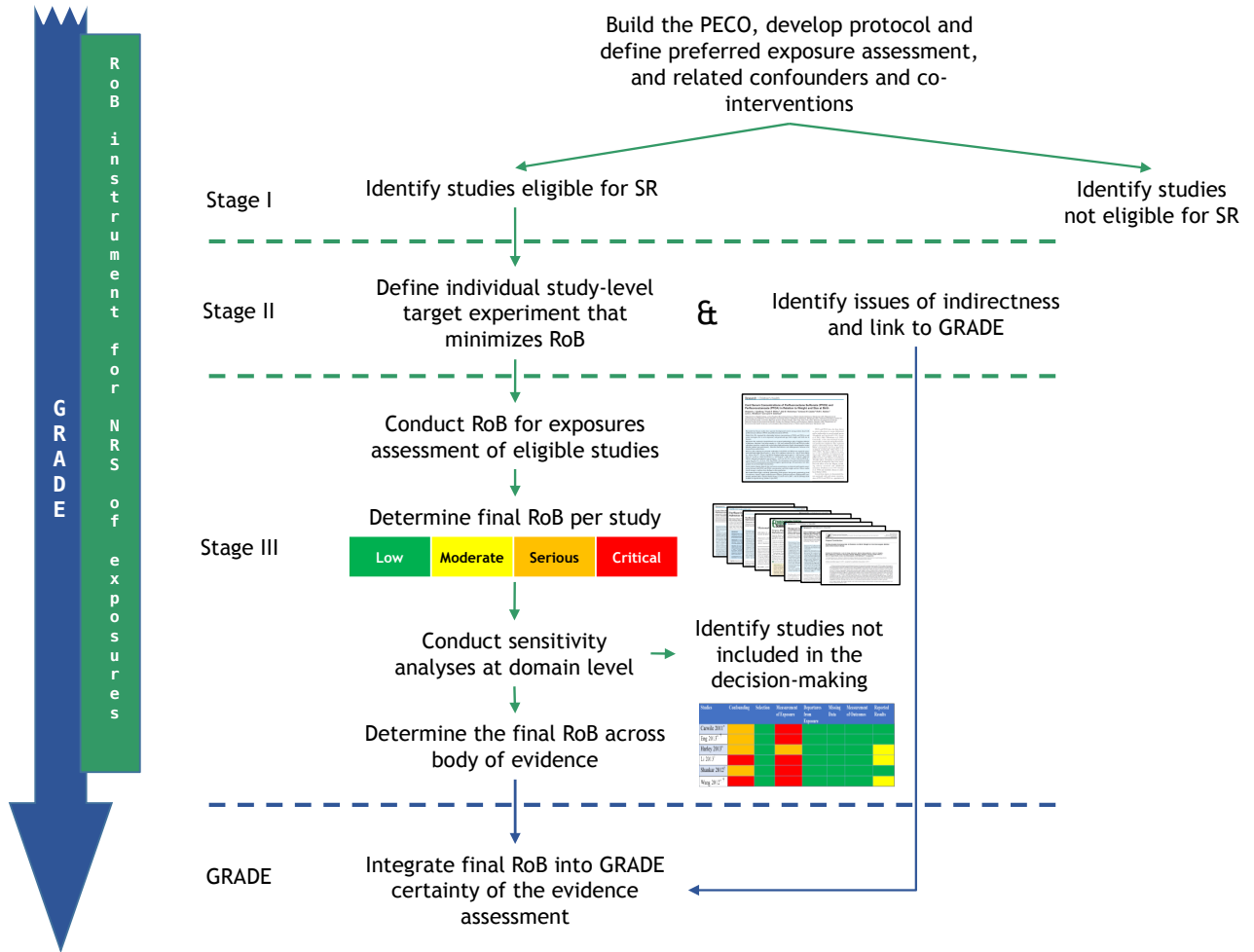


### G.2. Sensitivity analysis of all studies



# Figures.

Figure 1. Approach for conducting an assessment using the RoB instrument for NRS of exposures and the integration into GRADE when conducting systematic reviews of exposure.



GRADE: Grading of Recommendations Assessment, Development and Evaluation; PECO: population, exposure, comparator, outcome; RoB: risk of bias; SR: systematic review.

## Tables

Table 1. Five paradigmatic approaches and examples for identifying the exposure and comparator in systematic review and decision-making questions (from Morgan RL, Whaley P, Thayer KA, Schünemann HJ: Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. Environment international 2018.)

Potential systematic-review or research context	Approach	PECO example
1. Calculate the health effect from an exposure; describing the dose-effect relationship between an exposure and an outcome for risk characterisation.	Explore the shape and distribution of the relationship between the exposure and the outcome in the systematic review.	Among newborns, what is the incremental effect of 10 dB increase during gestation on postnatal hearing impairment?
2. Evaluate the effect of an exposure cut-off on health outcomes, when the cut-off can be informed iteratively by the results of the systematic review.	Use cut-offs defined based on distribution in the studies identified in the systematic review.	Among newborns, what is the effect of the highest dB exposure compared to the lowest dB exposure (e.g. identified tertiles, quartiles, or quintiles) during pregnancy on postnatal hearing impairment?
3. Evaluate the association between an exposure cut-off and a comparison cut-off, when the cut-offs can be identified or are known from other populations.	Use mean cut-offs from external or other populations (may come from other research).	Among commercial pilots, what is the effect of noise corresponding to occupational exposure compared to noise exposure experienced in other occupations on hearing impairment?
4. Identify an exposure cut-off that ameliorates the effects on health outcomes.	Use existing exposure cut-offs associated with known health outcomes of interest.	Among industrial workers, what is the effect of exposure to < 80 dB compared to ≥ 80 dB on hearing impairment?
5. Evaluate the potential effect of a cut-off* that can be achieved through an intervention to ameliorate the effects of exposure on health outcomes.	Select the comparator based on what exposure cut-offs can be achieved through an intervention.	Among the general population, what is the effect of an intervention that reduces noise levels by 20 dB compared to no intervention on hearing impairment?



Table 2. Risk of bias matrix presenting judgments for highest BPA exposure vs. lowest BPA exposure on the outcome of body weight, for the 7 RoB items, for 6 included studies.

Studies	Confounding	Selection	Measurement of Exposure	Departures from Exposure	Missing Data	Measurement of Outcomes	Reported Results
Carwile 2011*	Moderate	Low	Critical	Low	Low	Low	Low
Eng 2013*, †	Moderate	Low	Critical	Low	Low	Low	Low
Harley 2013*	Moderate	Low	Moderate	Low	Low	Low	Moderate
Li 2013*	Critical	Low	Critical	Low	Low	Low	Moderate
Shankar 2012†	Moderate	Low	Critical	Low	Low	Low	Low
Wang 2012*, †	Critical	Low	Critical	Low	Low	Low	Moderate

\* Prevalent overweight

† Prevalent obesity

Low	Moderate	Serious	Critical
-----	----------	---------	----------

Tables 3, 4, & 5. Risk of bias matrix presenting study-level and item-level judgments for exposure to highest BPA vs. exposure to lowest BPA on the outcomes of prevalent overweight and obesity.

Table 3. Study-level judgments for prevalent overweight and prevalent obesity

Studies	Confounding	Selection	Measurement of Exposure	Departures from Exposure	Missing Data	Measurement of Outcomes	Reported Results	Study-level RoB Judgment
Carwile 2011*	Moderate	Low	Critical	Low	Low	Low	Low	Critical
Eng 2013*, †	Moderate	Low	Critical	Low	Low	Low	Low	Critical
Harley 2013*	Moderate	Low	Moderate	Low	Low	Low	Moderate	Moderate
Li 2013*	Critical	Low	Critical	Low	Low	Low	Low	Critical
Shankar 2012†	Moderate	Low	Critical	Low	Low	Low	Low	Critical
Wang 2012*, †	Moderate	Low	Critical	Low	Low	Low	Moderate	Critical

\* Prevalent overweight

† Prevalent obesity

Low	Moderate	Serious	Critical
-----	----------	---------	----------

Table 4. Item-level judgments for prevalent overweight

Studies	Confounding	Selection	Measurement of Exposure	Departures from Exposure	Missing Data	Measurement of Outcomes	Reported Results
Carwile 2011	Moderate	Low	Critical	Low	Low	Low	Low
Eng 2013	Moderate	Low	Critical	Low	Low	Low	Low
Harley 2013	Moderate	Low	Moderate	Low	Low	Low	Moderate
Li 2013	Critical	Low	Critical	Low	Low	Low	Moderate
Wang 2012	Critical	Low	Critical	Low	Low	Low	Moderate
<b>Item-level judgment</b>	Critical	Low	Critical	Low	Low	Low	Moderate

Low	Moderate	Serious	Critical
-----	----------	---------	----------

Table 5. Item-level judgments for prevalent obesity

Studies	Confounding	Selection	Measurement of Exposure	Departures from Exposure	Missing Data	Measurement of Outcomes	Reported Results
Eng 2013	↓	↓	↓	↓	↓	↓	↓
Shankar 2012	↓	↓	↓	↓	↓	↓	↓
Wang 2012	↓	↓	↓	↓	↓	↓	↓
<b>Item-level judgment</b>							

Low	Moderate	Serious	Critical
-----	----------	---------	----------

Table 6. Exposure to BPA on the outcome of birthweight GRADE evidence assessment

<b>Question:</b> Exposure to highest levels of BPA (CAS# 80-05-7) compared to exposure to lowest levels of BPA in general population <b>Setting:</b> Community <b>Bibliography:</b> Rancière, F., Lyons, J. G., Loh, V. H., Botton, J., Galloway, T., Wang, T., ... & Magliano, D. J. (2015). Bisphenol A and the risk of cardiometabolic disorders: a systematic review with meta-analysis of the epidemiological evidence. <i>Environmental Health</i> , 14(1), 46.												
Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exposure to highest BPA levels	exposure to lowest BPA levels	Relative (95% CI)	Absolute (95% CI)		
Prevalent overweight (assessed with: BMI ≥85th percentile for age/gender in children; BMI 18.5-25/30 kg/m <sup>2</sup> )												
5	studies	very, very serious <sup>a</sup>	not serious <sup>b</sup>	not serious <sup>c</sup>	serious <sup>d</sup>	none	1774/5403 (32.8%)	1584/5657 (28.0%)	OR 1.21 (0.98 to 1.56)	40 more per 1,000 (from 4 fewer to 98 more)	⊕○○○ VERY LOW	CRITICAL
Prevalent obesity (assessed with: BMI ≥95th percentile for age/gender in children; BMI ≥25-30 kg/m <sup>2</sup> )												
3	studies	very serious <sup>a</sup>	not serious	not serious <sup>c</sup>	not serious	none	1425/5178 (27.5%)	1204/5342 (22.5%)	OR 1.67 (1.32 to 1.93)	102 more per 1,000 (from 52 more to 134 more)	⊕⊕○○ LOW	CRITICAL

CI: Confidence interval; OR: Odds ratio

### Explanations

- Most studies adjusted for known confounders of weight (age and gender) and diet; however, two studies did not account for caloric intake or diet which is relevant for evaluating weight-related outcomes, there is some risk of unmeasured confounding; BPA measurement present potential for bias as the chemical is non-persistent with a short half-life and exposure measurements were not repeated (except in one study), one study measures BPA three months post-BMI measurement, remaining studies measure BPA and BMI at the same time; however, the effect estimates may underestimate the true effect reducing our concern of non-differential misclassification; potential risk of reporting bias because three studies did not report prior publication of a protocol; however, all studies present outcome measures and analyses consistent with a priori plan outlined in the manuscript.
- The I<sup>2</sup> value = 45% and exploration of the forest plot suggests some inconsistency introduced by one outlying study contributing 4.3% of the weight to the analysis of children.
- Studies measured BPA concentration through urinary output. uBPA (BPA in urine) is considered a reliable and direct measure of BPA consumption and was not downgraded for indirectness.
- Imprecision is present because the width of the confidence interval is consistent with no association.

