



HAL
open science

A proposed method of bias adjustment for meta-analyses of published observational studies

Simon G Thompson, Ulf Ekelund, S Jebb, Anna Karin Lindroos, Stephen
Sharp, R Turner, Desiree Wilks

► **To cite this version:**

Simon G Thompson, Ulf Ekelund, S Jebb, Anna Karin Lindroos, Stephen Sharp, et al.. A proposed method of bias adjustment for meta-analyses of published observational studies. *International Journal of Epidemiology*, 2010, 10.1093/ije/DYQ248 . hal-00654802

HAL Id: hal-00654802

<https://hal.science/hal-00654802>

Submitted on 23 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**A proposed method of bias adjustment for meta-analyses of
published observational studies**

Journal:	<i>International Journal of Epidemiology</i>
Manuscript ID:	IJE-2010-02-0122.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	16-Sep-2010
Complete List of Authors:	Thompson, Simon; MRC Biostatistics Unit Ekelund, Ulf; MRC, Institute of Metabolic Science Jebb, S; <none> Lindroos, Anna Karin; MRC Human Nutrition Research Sharp, Stephen; MRC Epidemiology Unit Turner, R; MRC Biostatistics Unit, Institute of Public Health Wilks, Desiree; MRC Human Nutrition Research
Key Words:	meta-analysis, study quality, bias adjustment, observational studies, physical activity, obesity

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A PROPOSED METHOD OF BIAS ADJUSTMENT FOR META-ANALYSES OF PUBLISHED OBSERVATIONAL STUDIES

Thompson SG¹, Ekelund U², Jebb SA³, Lindroos AK³, Mander AP¹,
Sharp SJ², Turner RM¹, Wilks DC³

¹MRC Biostatistics Unit, Cambridge, UK; ²MRC Epidemiology Unit, Cambridge, UK;

³MRC Human Nutrition Research, Cambridge, UK.

16 September 2010; revised for *Int J Epidemiol*

Running title: Bias adjustments in meta-analysis

Correspondence: Prof Simon G Thompson, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, UK. Tel: +44 1223 330366. Fax: +44 1223 330388. Email: simon.thompson@mrc-bsu.cam.ac.uk

Word count: 4632 (abstract and main text)

Abstract

Objective: Interpretation of meta-analyses of published observational studies is problematic because of numerous sources of bias. We develop bias assessment, elicitation and adjustment methods, and apply them to a systematic review of longitudinal observational studies of the relationship between objectively measured physical activity and subsequent change in adiposity in children.

Methods: We separated internal biases that reflect study quality from external biases that reflect generalisability to a target setting. Since published results were presented in different formats, these were all converted to correlation coefficients. Biases were considered as additive or proportional on the correlation scale. Opinions about the extent of each bias in each study, together with its uncertainty, were elicited in a formal process from quantitatively-trained assessors for the internal biases and subject-matter specialists for the external biases. Bias-adjusted results for each study were combined across assessors using median pooling, and results combined across studies by random-effects meta-analysis.

Results: Before adjusting for bias, the pooled correlation is difficult to interpret because the studies varied substantially in quality and design, and there was considerable heterogeneity. After adjusting for both the internal and external biases, the pooled correlation provides a meaningful quantitative summary of all available evidence, and the confidence interval incorporates the elicited uncertainties about the extent of the biases. In the adjusted meta-analysis, there was no apparent heterogeneity.

Conclusion: This approach provides a viable method of bias adjustment for meta-analyses of observational studies, allowing the quantitative synthesis of evidence from otherwise incompatible studies. From the meta-analysis of longitudinal observational studies, we conclude that there is no evidence that physical activity is associated with gain in body fat.

Keywords / MeSH headings: Meta-analysis, study quality, bias adjustment, observational studies, physical activity, obesity

1
2
3
4
5
6 **Key messages**
7

8
9 We present novel methods for undertaking a quantitative meta-analysis, when the
10 component studies are observational and thus prone to many biases.
11

12
13
14 We describe how the process can be broken down into small manageable steps, and how
15 to incorporate opinion elicited in a formal manner about the size and uncertainty of the
16 biases in each study.
17
18

19
20
21 Bias checklists, elicitation scales and computer code are made available so that others can
22 carry out similar analyses.
23
24

25
26
27 These methods, or ones similar to them, will increasingly need to be adopted when
28 formulating guidance on public health issues for which randomised trial evidence is not
29 available.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Introduction

Many issues of public health importance cannot be investigated in intervention studies or randomised trials, for either ethical or practical reasons^{1,2}. Observational studies then provide the only source, or a large component, of relevant evidence. Such studies are notoriously prone to biases, caused for example through selection of participants, confounding and loss to follow-up. Especially when only published information is available, the potential impact of biases on the reported results and their interpretation is often unclear³.

This issue comes to the fore when undertaking a systematic review⁴, for then the objective is to collate and synthesize all the available evidence in a rigorous way. Systematic reviews have in the main focused on intervention studies, and especially randomised trials⁵. In these situations, although potential biases still have to be considered, there is an appreciation of their major sources and potential impact⁶. Reviews of observational studies commonly reach rather qualitative conclusions, for example based on a tabulation of study-specific results together with a commentary on their idiosyncrasies and potential biases. An overall quantitative conclusion using meta-analysis is often avoided because of the intangible nature of some of the biases, the incompatibility of methods of presenting results in different papers⁷, and the fact that relevant information is often missing in publications. Alternatively, a rather arbitrary dichotomy is introduced to separate the 'better' from the 'poorer' quality studies, and a quantitative meta-analysis of the former presented. This simplistic approach essentially disregards any biases in the 'better' studies, and assumes that the 'worse' studies are totally non-informative. Similarly, simple scoring of studies according to some measure of quality does not directly address their biases⁸.

In the context of systematic reviews of intervention studies, both randomised and non-randomised, work has recently been developed to quantify the potential biases using subjective opinion elicited from experts so that meta-analysis can be undertaken⁹. Using elicited opinion is necessary, because there is rarely sufficient empirical evidence about

1
2
3 the potential size of particular biases relevant to an individual study¹⁰. The magnitude of
4 biases always of course remains uncertain, and quantifying this uncertainty is part of the
5 elicitation process. Here we extend this work on intervention studies to the more
6
7 problematic context of observational studies.
8
9

10 11 12 **Methods**

13
14
15
16 Our aim is to make a quantitative conclusion, on the basis of observational studies, about
17 a particular association of public health importance. As an example, we consider the
18 relationship between physical activity and subsequent change in adiposity in children.
19 Relevant studies were undertaken in different contexts (populations, methods, lengths of
20 follow-up), but we aim to make a conclusion relevant to a specific target setting. The
21 studies then suffer from two forms of bias: first internal bias (or lack of rigour) and
22 second external bias (or lack of relevance to the target setting). In the following
23 explanation of our proposed approach, the focus is on the methods; more details of the
24 example and its interpretation are provided elsewhere¹¹.
25
26
27
28
29
30
31
32

33 34 *Physical activity and obesity example*

35
36
37 Obesity is a major global health issue¹², and the increase in obesity of children is of
38 particular concern¹³. It is proposed that increasing physical activity, which raises energy
39 expenditure, may protect against excess weight gain. But the evidence underpinning this
40 assertion is incomplete. Most cross-sectional studies of physical activity and body weight
41 indeed show an inverse association¹⁴. However their interpretation is problematic,
42 because the direction of any causal link is unclear (does physical activity lead to lower
43 weight, or does obesity lead to lower levels of physical activity?). In addition, studies not
44 using objective measures may be distorted by reporting biases for physical activity¹⁵.
45
46 Thus we focus on longitudinal observational studies in children, which relate objective
47 measures of baseline physical activity to objective measures of subsequent change in
48 adiposity, found in a thorough literature search from January 2000 up to September
49 2008^{11,16}.
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 Six studies fulfilling the eligibility criteria were found¹⁷⁻²². They are characterised by
6 heterogeneity in populations studied, age and sex groups recruited, follow-up times,
7 measures of physical activity level and body composition, and which confounders are
8 adjusted for. One of the studies is summarised in Table 1, and will provide a running
9 example in this paper. Most studies measured percentage of total weight as body fat
10 (%BF) at baseline and follow-up, and regressed change in %BF on baseline physical
11 activity level and confounders. Results from these regression analyses were presented in
12 various ways, for example either a partial regression or a partial correlation coefficient
13 with a p-value, and often without a direct measure of uncertainty such as a standard error
14 or confidence interval.
15
16
17
18
19
20
21
22
23

24 ***Target setting and categories of bias***

25
26
27
28 The overall approach to identify and quantify the biases of original studies in relation to a
29 target setting is depicted in Figure 1. For each original study undertaken, an idealised
30 version is described that is not subject to any internal biases. This separates the internal
31 biases from the external biases, which are themselves broken down into components so
32 that they can be more easily assimilated and opinions about their magnitude elicited.
33
34
35
36
37
38

39 The key components of a well-defined target setting in our example were considered to
40 be the population, the measure of physical activity, the measure of change in adiposity,
41 and the duration of follow-up. The specific target setting chosen is shown in Table 2, in
42 order to address the most relevant public health question in the UK. While some aspects
43 (for example the choice of change in %BF as the outcome measure) were well
44 represented within the studies undertaken, others were not (all the studies were conducted
45 in the USA but the target population was the UK). Also shown in Table 2 is the idealised
46 version of the example study from Table 1. The idealised study uses the same design,
47 population, measures and context as the original study, but is not subject to any internal
48 biases (for example no loss to follow-up, proper control of confounding). There is no
49 subjectivity involved in defining the idealised study; it does not have to be practicable but
50
51
52
53
54
55
56
57
58
59
60

1
2
3 is merely a mechanism to enable internal and external biases to be separated. The
4 differences between the original and idealised study represent internal biases, and
5 differences between the idealised study and the target setting (Table 2) represent potential
6 external biases.
7
8
9

10
11
12 The sources of internal bias were put into six categories (Figure 1): selection bias
13 (whether the sample recruited was representative of the intended population), control of
14 confounding (whether essential confounders have been adjusted for), exposure measure
15 (problems in assessing physical activity), attrition (loss to follow-up), outcome measure
16 (problems in measuring change in adiposity), and any other biases (for example when the
17 statistical analysis used was thought to have introduced bias). These six categories of bias
18 were generally mutually exclusive, so that each potential bias in each study could be
19 placed in one category, and considered to operate independently of each other. The
20 external biases were in four categories (population, exposure measure, outcome measure
21 and follow-up time) that relate to the definition of the target setting. To help itemise the
22 specific biases for each study, a checklist was developed (Figure 2) based on previous
23 work^{3,9,23} and this was completed for each study.
24
25
26
27
28
29
30
31
32
33
34

35 The choice of appropriate confounders to adjust for is a difficult issue. Rather than
36 attempt to say whether the choice of a particular set of confounders was 'correct', we
37 judged the bias from the adjustment presented in relation to using a standard set of
38 confounders (namely age, sex, ethnic group, sexual maturity, baseline fat mass and
39 baseline lean mass). Moreover we did not consider the effects of within-subject variation
40 over time in the assessment of physical activity. Thus the target parameter to be estimated
41 in the meta-analysis is that for the association between change in %BF and observed
42 baseline physical activity energy expenditure (PAEE) adjusted for a specific set of
43 confounders.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Extracting results

The principal quantitative result extracted from each study, which would form the basis for the meta-analysis, was chosen to be as close as possible to an estimate of the target parameter. Then the extent to which biases would have to be assessed was minimised. For example adjusted associations were chosen if available, and reported associations with PAEE were preferred over associations with total energy expenditure. Since the exposure and outcome variables were on different scales in different studies, and because results were presented in different formats, it was necessary to convert all extracted results to a common scale. Moreover standard errors were not always provided. Our solution was to transform all associations into correlation coefficients using, if nothing else were available, the sample size and the p-value to derive these.

We use the result that the Fisher-transformation of a correlation coefficient r , namely $z = 0.5 \ln[(1+r)/(1-r)]$, has an approximate normal distribution with standard error $\sqrt{1/(n-3)}$ where n is the sample size²⁴. Thus the relevant (2-sided) p-value reported in the paper is first converted into a standard normal score S taking due regard of the sign of the association in the paper, the Fisher-transformed correlation derived as

$z = S \times \sqrt{1/(n-3)}$, and the correlation as $r = (e^{2z} - 1)/(e^{2z} + 1)$. Where papers presented both a correlation coefficient and a p-value, our derived correlation agreed well with the published value.

Bias assessments

The process of eliciting biases was as follows, for each study in turn. The same subject-matter specialist and one statistician reviewed each study's publication, defined the idealised version of the study, and completed the checklist in Figure 2 by qualitatively describing each potential source of bias. The internal biases were then assessed by a group of six quantitatively-trained assessors (primarily statisticians) and the external biases by a group of five subject-matter specialists (primarily physical activity

1
2
3 epidemiologists). Having read the paper and checklist, the group agreed any
4 modifications to be made to the checklist, but avoided discussing the seriousness or
5 magnitude of potential biases. Each bias was classified by the group as operating either
6 additively or proportionally on a correlation scale. An additive bias could introduce a
7 correlation where none were in truth present; examples included inadequate control of
8 confounding or biases caused through missing data or loss to follow-up. A proportional
9 bias would change the magnitude but not the sign of the correlation, thus exaggerating or
10 attenuating a true effect; examples included differences between populations, and biases
11 caused by undertaking stepwise regression and retaining only statistically significant
12 predictors.
13
14
15
16
17
18
19
20
21
22

23 After the group discussion, each assessor individually considered biases in each category
24 (Figure 1). A first qualitative stage was to consider whether the bias was potentially large,
25 medium, small, or negligible, and in what direction. They then indicated their view about
26 the magnitude of an additive bias, and their uncertainty about this, on the upper scale in
27 Figure 3. This required marking an interval on the untransformed correlation scale such
28 that they believed there was a two-thirds chance that the bias lay inside this interval, and
29 a one-third chance that it lay outside. To help guide these judgements, Figure 4 shows the
30 impact of different biases on the magnitude of the confidence interval for the correlation
31 according to sample size. From this, a guideline was suggested that additive biases of
32 magnitude more than 0.2 were large, those between 0.1 and 0.2 were moderate, and those
33 less than 0.1 were small. If an assessor had no opinion about the direction of the bias,
34 then the interval would be placed symmetrically about zero. If an assessor thought that
35 the bias would tend to favour a negative correlation, the centre of the interval would be
36 on the left hand side of the upper scale in Figure 3, and vice-versa for a bias favouring a
37 positive correlation. If there was thought to be no or negligible bias, the 'interval' became
38 a point at zero on the scale. Biases considered proportional by the group were indicated
39 on the lower scale in Figure 3 in a similar way, indicating exaggeration or attenuation of
40 effect.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Meta-analysis

We performed meta-analysis of correlation coefficients on the Fisher-transformed scale because the distribution of z is more symmetric than that of r . We incorporated assessments of the biases elicited on the correlation scale but transformed onto the z scale. Since the range of z is from minus to plus infinity, this has the theoretical advantage that additive biases cannot produce impossible values of the underlying correlation. However, in our context where correlations are modest in magnitude, this is of limited practical importance because the values of z and r are numerically close in the range -0.3 to $+0.3$.

The calculations for including the bias assessments in the meta-analysis follow published methods⁹, and Stata code is available²⁵. In brief, a bias assessment interval marked on the scales in Figure 3 is considered to be an estimated bias \pm one standard deviation (SD), since this corresponds to a two-thirds (67%) interval for a normal probability distribution. For each study and assessor, the total internal additive bias is calculated by adding the individual bias estimates and summing their variances (squared SDs). The total internal proportional bias and estimated variance for each study and assessor are also calculated⁹. These two quantities are combined to give a total internal bias and variance. This total bias for each assessor and study is subtracted from the observed study result, and the total variance of the bias added to the study result variance, to give an internal bias-adjusted estimate and variance for each study and each assessor. The external biases are then incorporated using a similar procedure. These adjusted estimates for each study are then averaged across assessors by median pooling²⁶, taking the median of the bias-adjusted estimates and the median of the variances; this corresponds to a ‘typical’ assessor.

Random-effects meta-analysis across studies was undertaken on the Fisher-transformed correlation scale. The impact of heterogeneity was summarised by the I^2 statistic²⁷, which estimates the percentage of variation between study results explained by true heterogeneity rather than chance. Values of I^2 close to 0% represent little heterogeneity beyond that compatible with chance. Summary estimates and intervals were converted back to the correlation scale for presentation.

Results

To explain the process, we first consider the biases, elicitations and adjustments performed for the one example study¹⁸ summarised in Table 1. The study result extracted was based on a sample size of 39 and a reported p-value of 0.04 from a multiple regression for the association of baseline PAEE and other covariates with change in %BF, yielding a calculated (partial) correlation of -0.33 (95% CI -0.59 to -0.01).

The internal biases reflect differences between the study undertaken and the idealised version of the study (Table 2); the elicited internal biases are shown in Figure 5 (top). Since there was little information about recruitment, it is possible that the girls in the example study were not representative of the population intended. The resulting selection bias was considered an additive bias; the assessors generally had no opinion about the direction of the bias but some were more uncertain about its impact than others. The reported result included adjustment for age and baseline fat-free mass but not ethnic group or baseline fat mass; the assessors generally thought that the resulting bias was quite modest (compared to the standard specified set of confounders), with no strong opinion about its direction. There were no differences in implementation of the exposure and outcome measures between the actual study and the idealised study, so no biases were recorded for these items. Only 39 out of the original 47 study entrants had the requisite follow-up data, and there was no comment in the published study about whether the girls omitted were similar to those included in the analysis. The assessors again did not have an opinion about the direction of the resulting bias. Finally, the study reported results from a stepwise regression, where non-significant effects had been excluded. The assessors regarded this as a proportional bias, generally likely to exaggerate the size of the reported association between PAEE and change in %BF.

For the external biases, the idealised version of the study is compared with the target setting (Table 2); the elicited external biases, which were all considered as proportional, are shown in Figure 5 (bottom). Since %BF was the outcome in both the idealised study

1
2
3 and target setting, there is no bias for this component. The potential biases relate to
4 differences in population (age range, gender and country), PAEE being measured under
5 laboratory rather than free-living conditions, and a slight difference in follow-up interval.
6
7 The assessors generally thought that the PAEE measurement used in the study might
8 diminish the association as compared to the target setting, but the other biases were
9 generally thought to be small (proportional bias near 1).
10
11
12
13
14

15
16 The effect of adjusting for these biases, pooled over assessors, is shown in Table 3. The
17 anticipated direction of the internal biases overall brings the correlation slightly nearer
18 zero, and the confidence interval width increases to reflect the uncertainty in the biases.
19 The effect of the external biases is to further increase the confidence interval width, but
20 the correlation estimate remains almost the same. These results are also shown in Figure
21 6 (second study).
22
23
24
25
26

27
28 A similar exercise was undertaken for each of the six studies in our example, leading to
29 bias-adjusted results for each study and corresponding meta-analyses (Figure 6, Table 3).
30 The meta-analysis of unadjusted correlations gave a summary estimate of -0.04 (95% CI
31 -0.21 to 0.14), but with substantial heterogeneity ($I^2=78\%$). This heterogeneity reflects
32 both the different study designs and measures used, but also the effect of biases.
33
34

35
36 Adjusting for the internal biases reduced the heterogeneity ($I^2=15\%$). After also taking
37 into account the external biases, there was no apparent heterogeneity ($I^2=0\%$) and the
38 pooled correlation was -0.01 (95% CI -0.18 to 0.16). The overall conclusion from the
39 bias-adjusted meta-analysis, now consistently expressed in terms of the correlation
40 between PAEE and subsequent change in %BF, is that there is little or no association. To
41 help interpretation, the pooled correlation can be converted to a regression coefficient;
42 using published standard deviations of PAEE and change in %BF¹⁷, the estimated bias-
43 adjusted regression coefficient was -0.05 (95% CI -1.00 to 0.91) change in %BF per 1
44 MJ/d (239 kcal/d) increase in PAEE.
45
46
47
48
49
50
51
52

53
54
55 After adjustment for the biases, the relative weights the different studies receive in the
56 meta-analysis change. For example, the fourth study²⁰ in Figure 6 received 17% of the
57
58
59
60

1
2
3 weight in the unadjusted meta-analysis but only 2% in the fully adjusted meta-analysis.
4
5 This in part reflects the uncertainty in the external biases for this study, since
6
7 accelerometer counts were used as a measure of physical activity rather than a direct
8
9 measure of PAEE, and skinfold thickness as a measure of body composition rather
10
11 than %BF.

12
13
14 To investigate the consistency across assessors, we repeated the bias adjustments for each
15
16 of the internal bias assessors separately, and then for each of the five external bias
17
18 assessors. The results from the meta-analysis (Figure 7) show consistency across
19
20 assessors, and do not change the overall conclusion based on a 'typical' assessor (Figure
21
22 6, Table 3).

23 24 25 **Discussion**

26
27
28 We have presented a method of obtaining an overall quantitative summary in a
29
30 systematic review of observational studies, where numerous potential biases may operate.
31
32 This is in contrast to the common approach where a vague and non-committal qualitative
33
34 conclusion is drawn, because of the poor quality, reporting or relevance of the component
35
36 studies. This apparently daunting task has been achieved by breaking it down into small
37
38 manageable steps as follows: (i) define a target question, (ii) describe an idealised version
39
40 of each study, (iii) separate internal from external biases, (iv) separate categories of these
41
42 biases, (v) compile a checklist of the possible biases in each study, (vi) agree this
43
44 checklist within a group of assessors, (vii) elicit the biases and their uncertainty from
45
46 assessors independently for each category of bias for each study, and (viii) perform a
47
48 bias-adjusted meta-analysis. Although this is a time-consuming process, there are no
49
50 obvious alternatives, since empirical evidence on the size and uncertainty of all the biases
51
52 is not available.

53
54 Other methods of adjusting for biases in meta-analysis have previously been proposed.
55
56 Some have adjusted for specific biases by specifying a model with parameters that
57
58 together determine the bias in the target effect²⁸. These methods have been developed, for
59
60

1
2
3 example employing external empirical data, to address misclassification of exposure or
4 outcome²⁹ and uncontrolled confounding³⁰, using a full or approximate likelihood
5 approach. Others have used distributions to represent directly the overall internal and
6 external biases in the effect of interest in each study³¹. Like the former more complex
7 methods²⁸⁻³⁰, we model biases due to individual sources, but like the latter simpler
8 method³¹ we assume a direct form for the bias in the target effect. Our aim has been to
9 present generic methods that can be used in a routine setting. Specifically, we have
10 extended a previous approach for intervention studies where the outcome scale was
11 relative risk⁹, tailored to the context of observational studies where the outcome scale is
12 correlation. In contrast, simple methods based on weighting by quality scores are known
13 to be inadequate³².

14
15
16
17
18
19
20
21
22
23
24 For the example considered, we conclude that there is little or no relationship between
25 physical activity and subsequent change in percent body fat in children, since the
26 estimated pooled correlation is almost zero with tight confidence limits. While physical
27 activity is no doubt important for various aspects of health, a policy focusing on
28 increasing physical activity alone, without changing dietary habits, is unlikely to be
29 effective in reducing obesity in children¹¹. Before biases are considered, the results of the
30 different studies were severely heterogeneous, which makes a pooled result very difficult
31 to interpret. After adjusting for internal biases, the results are less heterogeneous across
32 studies, but the pooled result still refers to the associations between the measures of
33 physical activity and change in adiposity used in the different studies. After also adjusting
34 for external biases, the correlation refers to that between PAEE and change in percent
35 body fat, as in the target setting, and so is directly interpretable. The lack of heterogeneity
36 between studies at this stage is what one would expect if the bias adjustment process was
37 working as intended. The confidence interval for the pooled correlation now incorporates
38 the uncertainty about the magnitudes of the biases, rather than the heterogeneity between
39 studies as in the unadjusted analysis.

40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55 In the example presented, the pooled estimate and confidence interval are quite similar
56 between the unadjusted and bias-adjusted meta-analyses. In other examples we have
57
58
59
60

1
2
3 undertaken the bias-adjusted pooled estimate or its precision were rather different from
4 the unadjusted values. In a meta-analysis of intervention studies of the effect of routine
5 antenatal anti-D prophylaxis on maternal sensitization, bias adjustments led to a similar
6 overall odds ratio but a substantially wider confidence interval⁹. In a meta-analysis of
7 observational studies of the relationship between dietary energy density and subsequent
8 changes in adiposity in children, bias adjustments made the correlation both more
9 positive and more imprecise, suggesting that the near-null rather precise unadjusted
10 association might be misleading³³.
11
12
13
14
15
16
17

18
19 There are of course limitations to the approach we have adopted which add uncertainty
20 around the final conclusions. First, the elicited biases are subjective. Assessors may not
21 agree with each other, and different assessors might have reached different judgements,
22 including for example whether a particular bias is best represented as additive or
23 proportional. Assessors might also not be consistent in how they judge the same bias on
24 different occasions. We have minimised these problems by involving assessors who are
25 experienced in the biases being judged (either methodological or subject-matter
26 specialists), by using independent judgements from a group of assessors, and basing
27 results on median pooling (which corresponds to a 'typical' assessor and eliminates
28 extreme judgements). Moreover, in general, the judgements of the different assessors
29 were quite similar (Figures 5 and 7), and using more assessors would not have reduced
30 the uncertainty about the views of a typical assessor. The method would be improved if it
31 were better informed by empirical evidence, for example from meta-epidemiological
32 studies¹⁰, or if authors themselves investigated the potential for bias in their studies³⁴.
33
34 Analyses of individual participant data, when these are available for at least one of the
35 contributing studies, can help in the assessment of biases, for example in investigating the
36 potential impact of missing data, of adjustment for different confounders, or of
37 categorising a continuous exposure³³.
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52
53 A second issue relates to the limits necessarily placed on the process. We consider results
54 in terms of correlations, since these can always be derived from just the sample size and
55 reported p-value. Any approximations in extracting results from a published paper (for
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

example rounded p-values, uncertain sample sizes, or unclear analytical methods) can be considered as an additional internal bias. Although meta-analysis of correlations or regression coefficients is an established method^{35,36} and has been used before in the field of nutrition and energy expenditure³⁷, it is conceptually a somewhat difficult scale on which to elicit biases. Hence we provided some guidance, derived from Figure 4, on what might be considered small or large additive biases. Our process assesses confounding bias relative to a pre-specified set of confounders, and considers only the published studies available and so does not address publication or dissemination biases³⁸. It also does not adjust for biases resulting from within-person variability over time, in either the exposure or confounders, since these “multivariate measurement error” effects are very hard to judge. This is an example where parametric modelling of individual biases using empirical evidence^{29,30} would be more reliable. For these reasons, one needs to be somewhat cautious in making a causal interpretation from the summarised results.

The work we have presented could be further developed, and it would be beneficial if our methods were applied to other examples by independent investigators in the future. Web-based software could be developed to aid the elicitation process and subsequent analysis. Ideally our approach needs validation, either against empirical evidence or for example in the context of a systematic review that pre-dates a planned large definitive study, where the design of the latter provides the relevant target setting. Most fundamentally, experience with the method needs to be gained in terms of real policy decision making, for example in national public health intervention assessments³⁹. It is exactly in this kind of context where quantitative summaries, acknowledging the uncertainties from methodologically limited studies, are required.

Acknowledgement

This work was supported by the UK Medical Research Council Population Health Sciences Research Network (grant number PHSRN27). We thank two referees for useful comments. SGT is guarantor for the paper.

References

1. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-1218.
2. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Quality and Safety in Healthcare* 2003;12:47-52.
3. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344-349.
4. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998;316:140-144.
5. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2*. The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org
6. Lundh A, Gøtzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Med Res Methodol* 2008;8:22.
7. Chene G, Thompson SG. Methods for summarizing the risk associations of quantitative variables in epidemiologic studies in a consistent form. *Am J Epidemiol* 1996;144:610-621.
8. Ioannidis JP, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 2008;336:1413-1415.
9. Turner R, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. *J Roy Statist Soc A* 2009;172:21-47.
10. Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirically based priors. *J Roy Statist Soc A* 2009;172:119-136.
11. Wilks DC, Sharp SJ, Ekelund U, Thompson SG, Mander AP, Turner RM, Jebb SA, Lindroos AK. Objectively measured physical activity and fat mass in children: a bias-adjusted meta-analysis of prospective studies. Submitted to *PLoS ONE*, 2010.
12. World Health Organisation. *Global health risks: Mortality and burden of disease attributable to selected major risks*. Geneva: World Health Organisation, 2009.
13. Deckelbaum RJ, Williams CL. Childhood obesity: the health issue. *Obes Res* 2001;9 Suppl 4:239S-243S.
14. Jimenez-Pavon D, Kelly J, Reilly JJ. Associations between objectively measured habitual physical activity and adiposity in children and adolescents: Systematic review. *Int J Pediatr Obes* 2010;5:3-18.
15. Westerterp KR. Assessment of physical activity level in relation to obesity: current evidence and research issues. *Med Sci Sports Exerc, Suppl* 1999;31:522-525.

16. Wilks DC, Besson H, Lindroos AK, Ekelund U. Objectively measured physical activity and obesity prevention in children, adolescents and adults: a systematic review of prospective studies. *Obes Rev* 2010, Jun 29 [Epub ahead of print].
17. DeLany JP, Bray GA, Harsha DW, Volaufova J. Energy expenditure and substrate oxidation predict changes in body fat in children. *Am J Clin Nutr* 2006;84:862-870.
18. Figueroa-Colon R, Arani RB, Goran MI, Weinsier RL. Paternal body fat is a longitudinal predictor of changes in body fat in premenarcheal girls. *Am J Clin Nutr* 2000;71:829-834.
19. Johnson MS, Figueroa-Colon R, Herd SL, et al. Aerobic fitness, not energy expenditure, influences subsequent increase in adiposity in black and white children. *Pediatrics* 2000;106:E50.
20. Moore LL, Gao D, Bradlee ML, et al. Does early physical activity predict body fat change throughout childhood? *Prev Med* 2003;37:10-17.
21. Salbe AD, Weyer C, Harper I, Lindsay RS, Ravussin E, Tataranni PA. Assessing risk factors for obesity between childhood and adolescence: II. Energy metabolism and physical activity. *Pediatrics* 2002;110:307-314.
22. Treuth MS, Butte NF, Sorkin JD. Predictors of body fat gain in nonobese girls with a familial predisposition to obesity. *Am J Clin Nutr* 2003;78:1212-1218.
23. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-384.
24. Armitage P, Matthews JNS, Berry G. *Statistical Methods in Medical Research*. Oxford: Blackwell Science, 2002.
25. Sharp SJ. Bias adjustment methods for meta-analyses of published observational studies: Stata code. MRC Epidemiology Unit, 2010. Available from www.mrc-epid.cam.ac.uk/Unit/How/FunctionalTeams/Statistics.html
26. Clemen R, Winkler R. Combining probability distributions from experts in risk analysis. *Risk Analysis* 1999;19:187-203.
27. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-560.
28. Eddy DM, Hasselblad V, Schachter R. *Meta-analysis by the confidence profile method: the statistical synthesis of evidence*. San Diego: Academic Press, 1992.
29. Wolpert RL, Mengerson KL. Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: effects of environmental tobacco smoke. *Statistical Science* 2004;19:450-471.
30. Greenland S. Multiple-bias modelling for analysis of observational data (with Discussion). *J Roy Statist Soc A* 2005;168:267-306.
31. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statist Med* 2003;22:3687-3709.

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
32. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics* 2001;2:463-471.
33. Wilks DC, Mander AP, Jebb SA, Thompson SG, Sharp SJ, Turner RM, Lindroos AK. Dietary energy density and adiposity: a bias-adjusted meta-analysis of prospective studies. Submitted to *BMC Public Health*, 2010.
34. Fox MP. Creating a demand for bias analysis in epidemiological research. *J Epidemiol Community Health* 2009;63:91.
35. Peterson RA, Brown SP. On the use of beta coefficients in meta-analysis. *J Appl Psychol* 2005;90:175-181.
36. Prevost AT, Mason D, Griffin S, Kinmonth AL, Sutton S, Spiegelhalter D. Allowing for correlations between correlations in random-effects meta-analysis of correlation matrices. *Psychol Methods* 2007;12:434-450.
37. Carpenter WH, Poehlman ET, O'Connell M, Goran MI. Influence of body composition and resting metabolic rate on variation in total energy expenditure: a meta-analysis. *Am J Clin Nutr* 1995;61:4-10.
38. Song F, Parekh-Bhurke S, Hooper L, Loke YK, Ryder JJ, Sutton AJ, Hing CB, Harvey I. Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies. *BMC Med Res Methodol* 2009;9:79.
39. National Institute of Health and Clinical Excellence. *Developing NICE public health guidance*. London: NICE, 2006. Available from www.nice.org.uk/guidance

Table 1: Characteristics of one example longitudinal study¹⁸ of physical activity level and subsequent change in adiposity, and data extracted

Sample	47 normal weight girls aged 5-9 years from Alabama, USA
Exposure	PAEE during 24 hours in a calorimetric chamber
Outcome	%BF by dual-energy X-ray absorptiometry
Time period	Baseline and after an average of 1.6 years (SD 0.4 years)
Analysis	Stepwise regression of change in %BF on predictors including PAEE
Sample size for longitudinal analysis n	39
Reported p-value	0.04
Fisher-transformed correlation z (SE)	-0.34 (0.17)
Correlation r calculated from z (95% CI)	-0.33 (-0.59 to -0.01)

%BF percent body fat, PAEE physical activity energy expenditure

Table 2: Target setting for meta-analysis, and the idealised version of one example study¹⁸

	Target setting	Idealised version of one example study ¹⁸
Population	General population of children aged 4-11 years in the UK	Normal weight girls aged 5-9 years from Alabama, USA
Exposure	Free-living PAEE objectively measured at baseline	PAEE measured by whole-room indirect calorimetry (laboratory conditions)
Outcome	Subsequent change in %BF, objectively measured at baseline and follow-up	Subsequent change in %BF measured at baseline and follow-up by dual-energy X-ray absorptiometry
Time interval	Outcome assessed over a 2-year period	Follow-up at 1.6 years

%BF percent body fat, PAEE physical activity energy expenditure

Table 3: Unadjusted and bias-adjusted correlations between baseline physical activity level and change in percent body fat for one example study¹⁸, and meta-analysis of unadjusted and bias-adjusted correlations (95% CI)

	Correlation for one example study ¹⁸	Meta-analysis of correlations in all six studies; I ² for heterogeneity
Unadjusted	-0.33 (-0.59 to -0.01)	-0.04 (-0.21 to 0.14); I ² =78%
Adjusted for internal biases (corresponds to idealised versions of each study)	-0.26 (-0.62 to 0.19)	0.00 (-0.18 to 0.19); I ² =15%
Adjusted for internal and external biases (corresponds to target setting, Table 2)	-0.27 (-0.68 to 0.26)	-0.01 (-0.18 to 0.16); I ² =0%

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1: Overview of bias adjustment method: separating internal and external biases

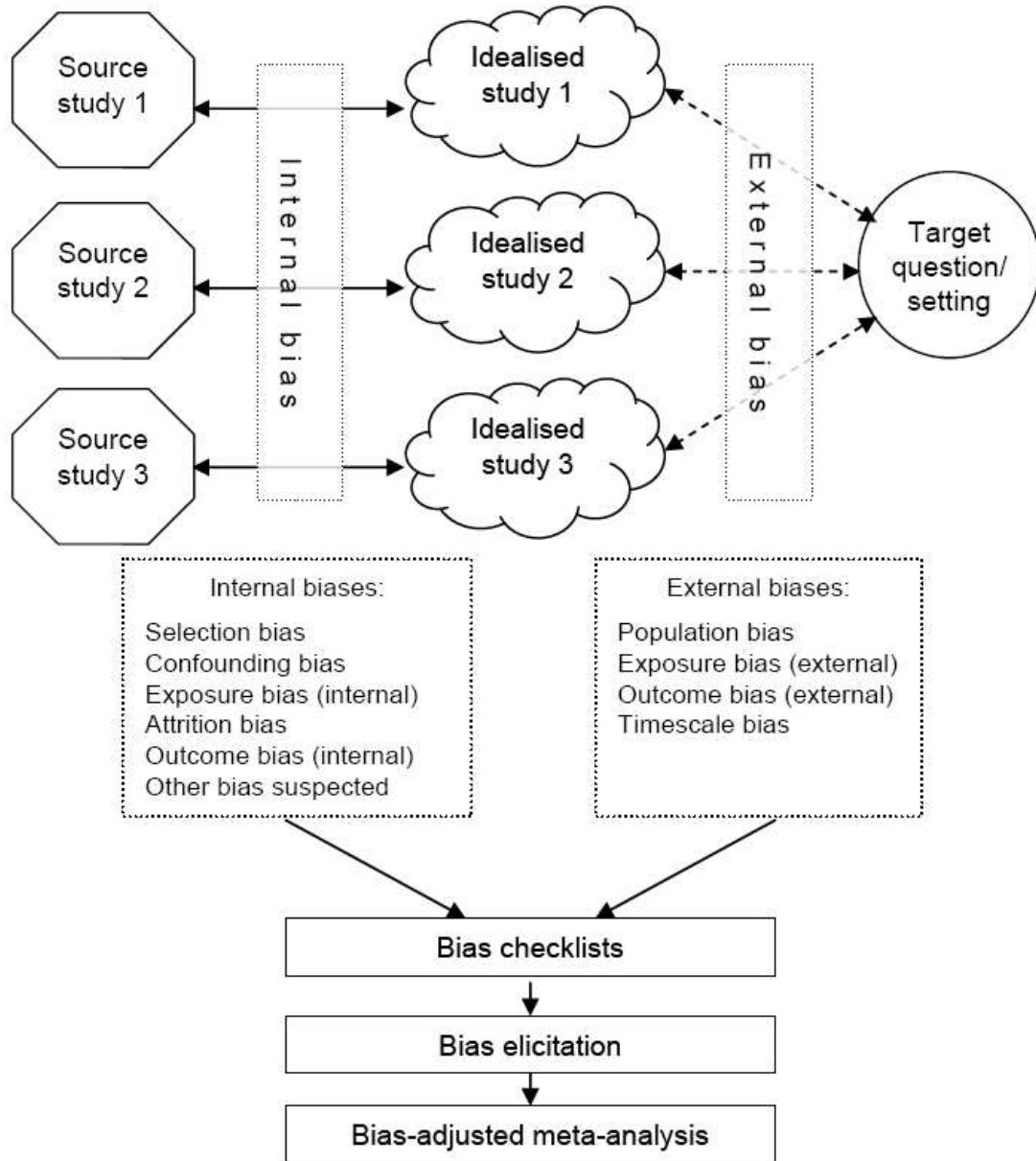


Figure 2: Checklists used for longitudinal studies of physical activity and obesity: internal and external biases

Checklist for sources of **internal** bias in longitudinal observational studies

	Yes/No/Unclear	Description
Selection bias		
Inclusion and exclusion criteria clear?		
Baseline measurements obtained for all participants recruited (i.e. no immediate drop-outs)?		
Confounding bias		
Appropriate choice of confounders (i.e. based on importance rather than convenience)?		
Adjustment made for all known important confounders? [1]		
Objective method of measuring confounders?		
Confounders measured accurately?		
Appropriate timing for measuring confounders?		
Exposure bias (internal)		
Was the exposure measure appropriate? [2]		
Objective method of measuring exposure?		
Exposure measured accurately?		
Appropriate timing for measuring exposure?		
Was the way that the exposure measure was used in the analysis appropriate?		
Attrition bias		
Are the results unlikely to be affected by losses to follow-up?		
Are the results unlikely to be affected by exclusions from analysis (e.g. because of extreme values or missing values of confounders)?		
Outcome bias (internal)		
Was the outcome measure appropriate? [3]		
Objective method of measuring outcome?		
Outcome measured accurately?		
Appropriate timing for measuring outcome?		
Was the way that the outcome measure was used in the analysis appropriate?		
Other bias suspected		
Was the statistical analysis appropriate?		

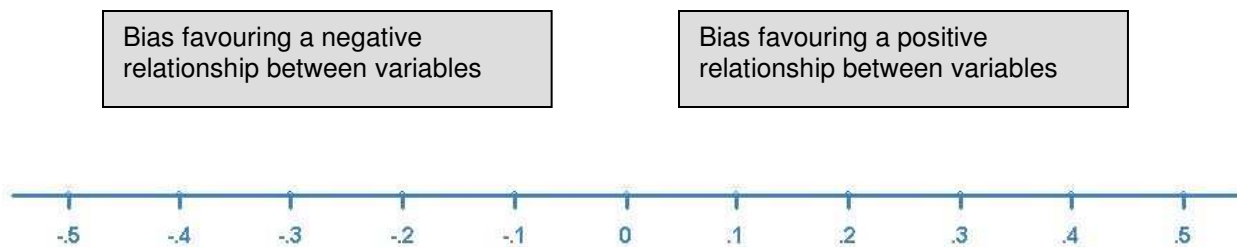
[1] Known important confounders could be listed here. [2] Appropriate measures of exposure could be listed here. [3] Appropriate outcome measures could be listed here.

Figure 2 (ctd)Checklist for sources of **external** bias in longitudinal observational studies

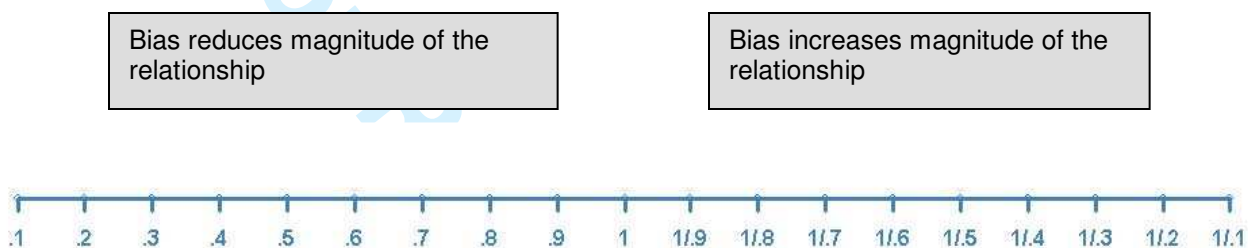
	Yes/No/Unclear	Description
Population bias		
Study subjects in idealised study drawn from population identical to target population, with respect to age, sex, health status etc.?		
Exposure bias (external)		
Exposure in idealised study identical to target exposure?		
Outcome bias (external)		
Outcome in idealised study identical to target outcome?		
Timescale bias		
Follow-up time in idealised study identical to target follow-up time?		

Figure 3: Elicitation scales for additive and proportional biases

Correlation scale for additive biases



Multiplicative scale for proportional biases

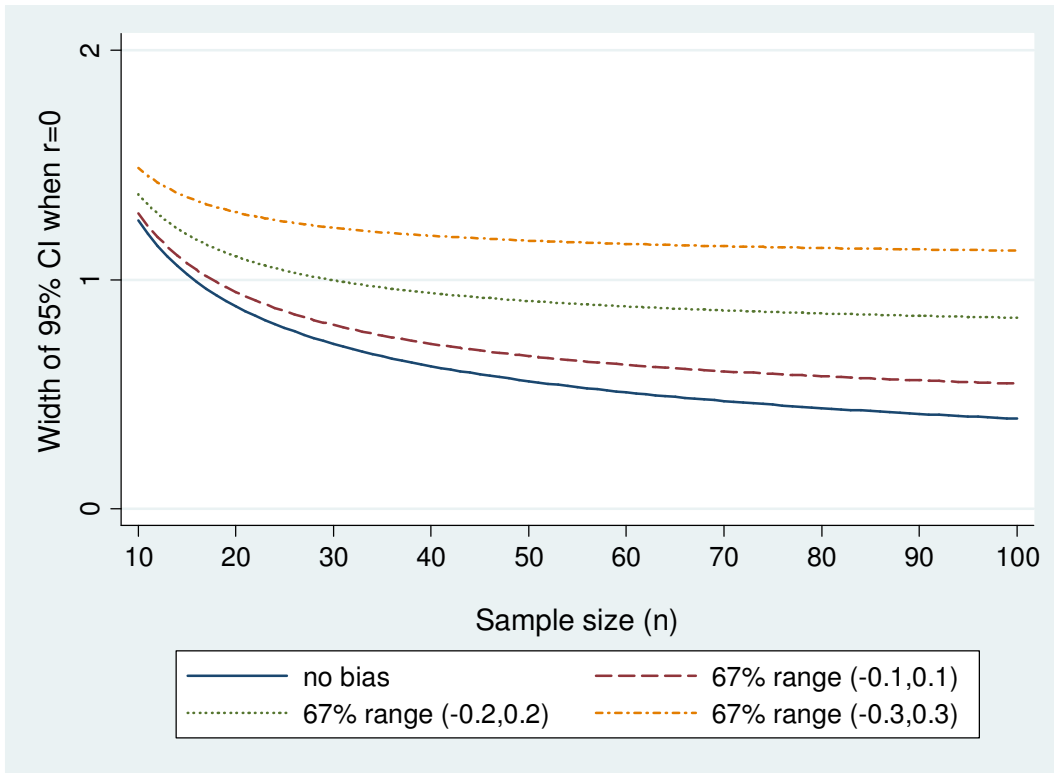


ew Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 4: Effect of ranges for an additive bias on the width of the 95% confidence interval for the bias-adjusted correlation coefficient



ew Only

Figure 5: Bias elicitations (67% intervals) for one study¹⁸ by six internal bias assessors and five external bias assessors; correlation scale. Blank sub-figures indicate the absence of that bias.

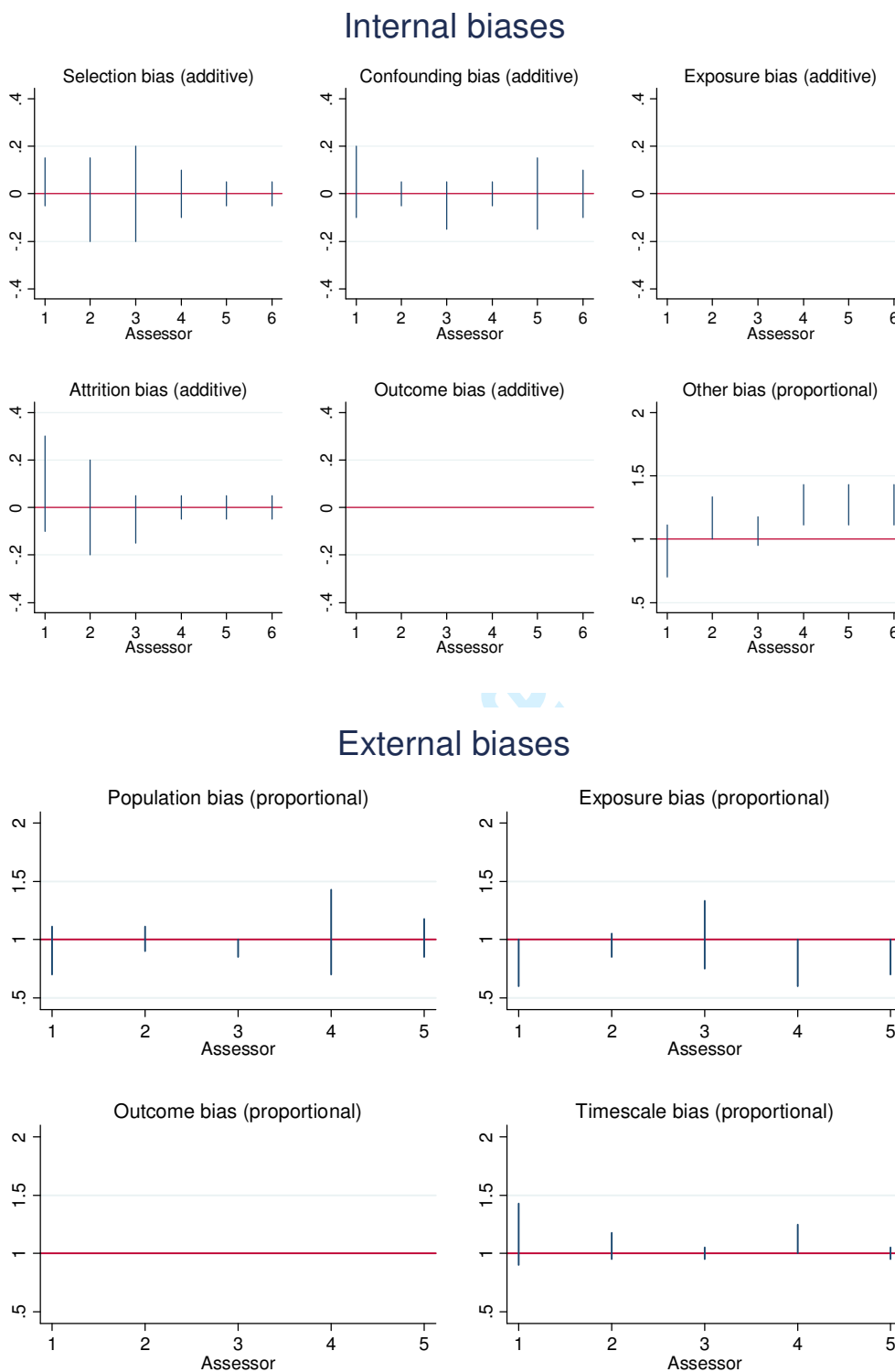


Figure 6: Meta-analysis of six studies¹⁷⁻²² for the association between physical activity and subsequent change in adiposity on the correlation scale. Results are shown unadjusted for any biases, adjusted for internal biases, and adjusted for both internal and external biases.

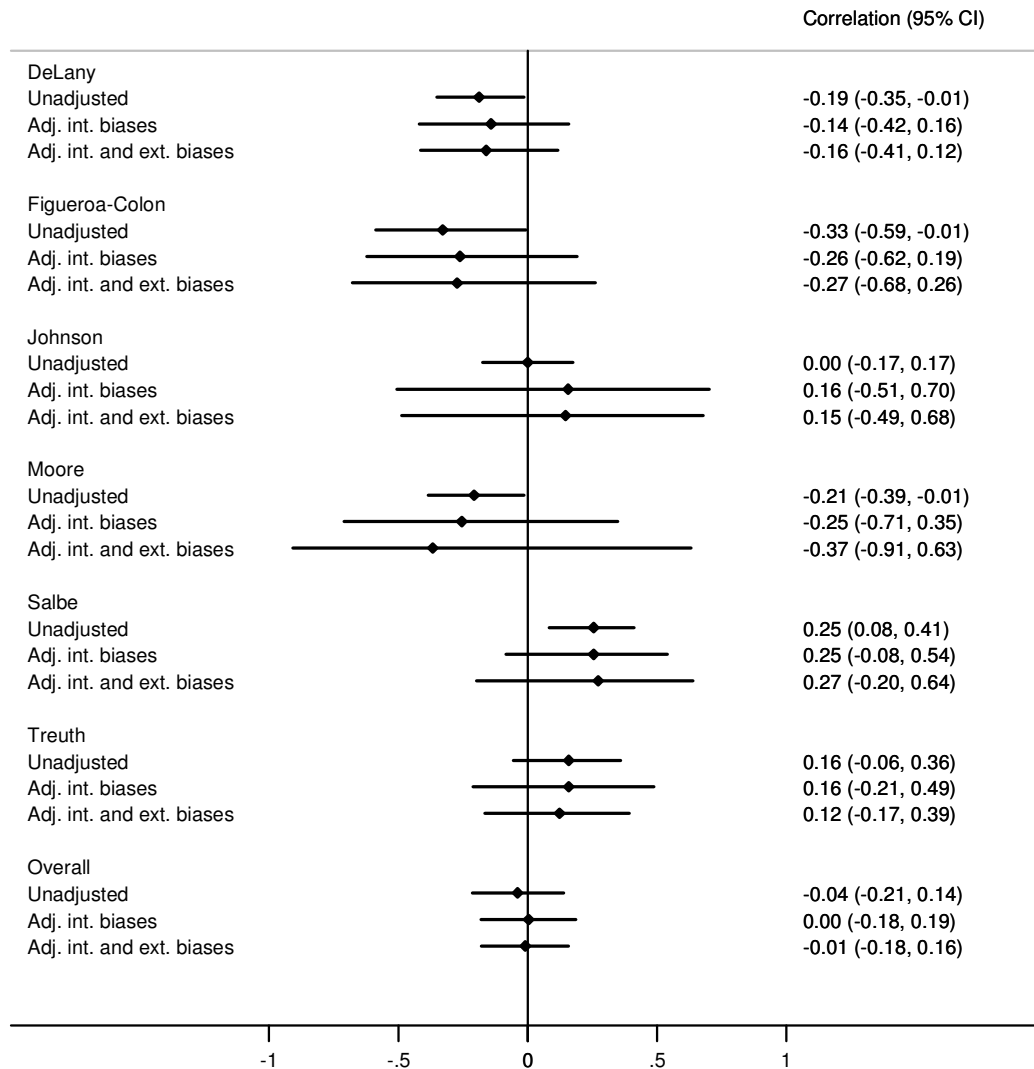


Figure 7: Meta-analysis of six studies¹⁷⁻²² for the association between physical activity and subsequent change in adiposity on the correlation scale. Results are shown using the internal bias adjustments from each of 6 internal bias assessors (IA1-IA6) separately, the overall internal bias-adjusted result, and adjusted for internal and external biases using the external biases from each of 5 external bias assessors (EA1-EA5) separately, and the overall result adjusted for both internal and external bias.

