



Original article

# Intervention effect estimates in cluster randomized versus individually randomized trials: a meta-epidemiological study

Clémence Leyrat <sup>1\*</sup>, Agnès Caille<sup>2,3</sup>, Sandra Eldridge<sup>4</sup>, Sally Kerry<sup>4</sup>, Agnès Dechartres<sup>5</sup> and Bruno Giraudeau<sup>2,3</sup>

<sup>1</sup>Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK, <sup>2</sup>Université de Tours, Université de Nantes, INSERM, SPHERE U1246, Tours, France, <sup>3</sup>INSERM CIC1415, CHRU de Tours, Tours, France, <sup>4</sup>Centre for Primary Care and Public Health, Queen Mary University of London, London, UK and <sup>5</sup>Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Département Biostatistique Santé Publique et Information Médicale, Hôpitaux Universitaires Pitié Salpêtrière – Charles Foix, AP-HP, Paris, France

\*Corresponding author. Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, WC1E 7HT London, UK. E-mail: clemence.leyrat@lshtm.ac.uk

Editorial decision 18 September 2018; Accepted 4 October 2018

## Abstract

**Background:** Cluster (CRTs) and individually randomized trials (IRTs) are often pooled together in meta-analyses (MAs) of randomized trials. However, the potential systematic differences in intervention effect estimates between these two trial types has never been investigated. Therefore, we conducted a meta-epidemiological study comparing intervention effect estimates between CRTs and IRTs.

**Methods:** All Cochrane MAs including at least one CRT and one IRT, published between 1 January 2010 and 31 December 2014, were included. For each MA, we estimated a ratio of odds ratios (ROR) for binary outcomes or a difference of standardized differences (DSMD) for continuous outcomes, where less than 1 (or 0, respectively) indicated a greater intervention effect estimate with CRTs.

**Results:** Among 1301 screened reviews, we selected 121 MAs, of which 76 had a binary outcome and 45 had a continuous outcome. For binary outcomes, intervention effect estimates did not differ between CRTs and IRTs [ROR 1.00, 95% confidence interval (0.93 to 1.08)]. Subgroup and adjusted analyses led to consistent results. For continuous outcomes, the DSMD was 0.13 (0.06 to 0.19). It was lower for MAs with a pharmacological intervention [-0.03, (-0.12 to 0.07)], an objective outcome [0.05, (-0.08 to 0.17)] or after adjusting for trial size [0.06, (-0.01 to 0.15)].

**Conclusion:** For binary outcomes, CRTs and IRTs can safely be pooled in MAs because of an absence of systematic differences between effect estimates. For continuous outcomes, the results were less clear although accounting for trial sample sizes led to a non-significant difference. More research is needed for continuous outcomes and, meanwhile, MAs should be completed with subgroup analyses (CRTs vs IRTs).

**Key words:** Cluster randomized trial, individually randomized trial, intervention effect estimate, meta-epidemiological study, systematic review

### Key Messages

- Cluster randomized trials are known to be more pragmatic than individually randomized trials but also more susceptible to bias.
- Cluster randomized and individually randomized trials are often pooled in MAs, but no study has investigated potential systematic differences in intervention effect estimates between these two trial types.
- In MAs of binary outcomes, intervention effect estimates for cluster and individually randomized trials did not differ.
- In MAs of continuous outcomes, intervention effect estimates were moderately more favourable for individually randomized trials. However, the difference in intervention effects was moderated by study size and characteristics of outcome and intervention. Therefore, the inconsistent results in subgroup analyses invite further studies.

## Introduction

Cluster randomized trials (CRTs) are defined as trials in which clusters of participants such as wards, practices, schools or villages are randomized rather than the participants themselves.<sup>1</sup> These trials are known to be more susceptible to bias than individually randomized trials (IRTs). For instance, recruitment bias may occur when participants are recruited after cluster randomization by a non-blinded recruiter.<sup>2–6</sup> This situation shares some similarities with the lack of allocation concealment, shown to be associated with an over-estimation of intervention effects.<sup>7</sup> In this situation, study groups may be unbalanced in regard to individual baseline characteristics, as the individual is not the unit of randomization.

Some interventions have been assessed with both CRTs and IRTs. In two reviews of hip protectors, large positive effects were seen in CRTs whereas effects in IRTs were more equivocal.<sup>8,9</sup> Although the intervention assessed may appear simple, Hahn *et al.*<sup>9</sup> explained that it actually may differ in two ways between CRTs and IRTs: (i) CRTs may benefit from a ‘herd effect’ with higher compliance; and (ii) IRTs may suffer from inter-group contamination which is often a reason for adopting cluster randomization. Both elements could lead to larger intervention effect estimates in CRTs than in IRTs. Conversely, Gilbody *et al.*<sup>10</sup> found similar results between CRTs and IRTs when investigating collaborative care for depression, and Selvaraj and Prasad<sup>11</sup> found similar proportions of positive results between CRTs and IRTs

These examples remain anecdotal, and to date we lack general findings as to whether intervention effect estimates are, on average, larger in CRTs than in IRTs. It then remains unclear whether these trial types can be pooled in meta-analyses (MAs). The Cochrane Handbook<sup>12</sup> considers

the unit-of-analysis error for CRTs, but nothing is said regarding a potential systematic difference in intervention effect estimates between these two trial types. Knowing if such a difference exists is, however, crucial for different reasons. First, CRTs and IRTs are often meta-analysed together, but this relies on the assumption that they estimate the same quantity of interest. Second, if there is a systematic difference between the estimates from the two types of trials, it might suggest that CRTs and IRTs lead to different estimands and therefore the interpretations of the results are different; CRTs keep existing ‘social units’ in which participants can interact. Therefore, CRTs may lead to real-world evidence and estimation of the effectiveness, as opposed to the ‘ideal-world’ estimation of efficacy obtained from IRTs. Third, the presence of systematic differences would imply that the intervention effect estimates from an IRT could not be used (at least as it is) to inform the sample size of a future CRT and vice versa.

For these reasons, we performed a meta-epidemiological study to assess whether intervention effect estimates are larger in CRTs than in IRTs. With this approach, we aim to understand whether the specificities of the two types of trials lead to systematic differences in intervention effect estimates. Indeed, CRTs and IRTs not only differ in the randomization procedure but also in ways participants are recruited, the intervention delivered, etc. Hence, we want to quantify the overall impact of these differences on the intervention effect estimates. To do so, we compared intervention effect estimates for the same intervention on the same outcome in studies using cluster randomization and studies using individual randomization. In order to ensure a comparability of the intervention and outcomes, we used trials that have been meta-analysed together in systematic reviews, adopting a quantitative approach called a meta-epidemiological study.

## Methods

Meta-epidemiological studies are used to determine which trial characteristics are associated with treatment effect estimates.<sup>13</sup> In this study, the characteristic of interest is the design (cluster vs individual randomization). A meta-epidemiological study is generally conducted with a two-step approach using a collection of MAs. First, for each selected MA and considering the trial as the unit of analysis, the difference in intervention effect estimates between studies which have the characteristic of interest (i.e. which are cluster randomized, in the present study) and those which are not, is assessed. This is done by fitting one meta-regression for each selected MA. Second, results obtained after this first step are meta-analysed. In this second step, the units of analysis are the MAs which have initially been selected. Using this approach, our null hypothesis for our primary analysis was an absence of a systematic difference in intervention effect estimates between CRTs and IRTs.

## Data sources

On 10 March 2015, we searched for eligible MAs published between 1 January 2010 and 31 December 2014 in the Cochrane Database of Systematic Reviews, by using the following keywords: 'cluster randomized' OR 'group randomized' OR 'community randomized' in the full text.

## MA and trial selection

Identified systematic reviews were screened to select only those including both CRTs and IRTs. We selected eligible MAs with at least three randomized trials, including at least one CRT and one IRT. Where more than one MA was eligible within the same review, the MA corresponding to the primary outcome, if clearly stated, was selected. Otherwise, the MA with the largest number of trials was selected. We excluded MAs with safety or compliance outcomes and those for which a control group was not clearly identifiable.

Trials were classified as CRT or IRT according to what was reported in the systematic review. Because we were interested in comparing trials that used different randomization units, we discarded quasi-randomized trials and controlled before–after studies. We discarded duplicate trials within MAs and kept the duplicate with the largest sample size. We finally discarded duplicate trials between MAs, keeping the duplicate from the most recently published systematic review. All those steps were performed independently by two of us (C.L., B.G.), with

disagreements resolved by discussion, referring to a third opinion (A.C.) when necessary.

## Data extraction and coding

We extracted data related to MAs and trials by using two standardized and pilot-tested spreadsheets. The items we extracted are presented in [Supplementary Table 1](#), available as [Supplementary data](#) at *IJE* online. Data were collected from the systematic reviews, except when otherwise specified. Data extraction was performed independently by two of us (C.L., B.G.) and any discrepancy was adjudicated, referring to a third opinion (A.C.) when necessary.

If the number of patients per group, number of events, means and standard deviations were not reported in the systematic review, we collected them from the trial reports, or in case of doubt, we contacted the authors of the systematic reviews (which happened for nine MAs).

## Statistical analysis

### Accounting for clustering: calculating effective sample size

For CRTs which did not adjust for clustering during analysis, we applied the method described in the Cochrane Handbook<sup>12</sup>; the sample size of the trial was reduced to an effective sample size by dividing the original sample size by the design effect. The design effect is defined as  $[1 + (M - 1)\rho]$ , where  $M$  is the average cluster size and  $\rho$  the intra-class correlation coefficient (ICC), the parameter classically used to quantify the clustering effect. We collected the ICC from the trial report and, if not reported, a value of 0.03 was chosen, corresponding to the median ICC value for outcome variables observed in the Campbell *et al.*<sup>14</sup> review. A sensitivity analysis doubled this generic value to 0.06 and also considered two extreme situations of no correlation ( $ICC = 0$ ) and a very strong correlation ( $ICC = 0.50$ ). If clustering was accounted for, we collected the effective sample size reported in the systematic review.

### Estimation of intervention effects within each MA

For binary outcomes, intervention effects estimates were expressed as odds ratios. For all outcomes, an odds ratio of less than 1 indicated a beneficial effect of the experimental intervention. For continuous outcomes, intervention effects estimates were expressed as standardized mean differences

using the Hedges and Olkin unbiased estimator of effect size.<sup>15</sup>

$$g = \left(1 - \frac{3}{4(n_1 + n_2 - 2) - 1}\right)d,$$

where  $n_1$  and  $n_2$  are the size of the control and experimental group, respectively, and  $d$  is the traditional Cohen's standardized difference:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}}.$$

$\bar{X}_1$  and  $\bar{X}_2$  are the observed means in the control and experimental group, respectively, and  $s_1^2$  and  $s_2^2$  are the two sample variance estimates. An effect of less than 0 always indicated a beneficial effect of the experimental intervention. For each MA, the intervention effect was estimated by using a random effect MA.

### Meta-epidemiologic analyses

We analysed binary and continuous outcomes separately, using the two-step approach proposed by Sterne *et al.*<sup>16</sup> First, for each MA, we performed an inverse-variance weighted random effects meta-regression, thus accounting for between-trial heterogeneity. The only covariate was the type of trial (cluster or individual randomization), with individual randomization as the reference category. For binary outcomes, we estimated the ratio of odds ratios (ROR), where a ratio of odds ratios less than 1 indicated more favourable intervention effect estimates in cluster trials, meaning that either the intervention was more beneficial or less detrimental in CRTs than in IRTs. For continuous outcomes, we estimated the difference in standardized mean differences (DSMD), where less than 0 indicated more favourable intervention effect estimates in cluster trials. In the second stage, the ratio or difference in intervention effects was combined across MAs using random effects MAs. The heterogeneity between MAs was quantified with the  $I^2$ , Cochran Q chi-squared test, and between MAs variance  $\tau^2$ <sup>17</sup> using a REML estimation.<sup>18</sup> Analyses involved use of SAS 9.4 and R 3.2.0 with the package *metafor*. All the statistical tests were done at a 5% significance level.

### Subgroup and adjusted analyses

The type of outcome (objective vs subjective) was a pre-specified subgroup analysis motivated by the fact that Savović *et al.*<sup>7</sup> observed differences in their meta-meta-epidemiological study according to whether the outcome

was an objective one or not, especially when looking at blinding. The type of intervention (pharmacological vs non-pharmacological) and control intervention (active vs inactive) were *post hoc* subgroup analyses. For these subgroup analyses, interaction *P*-values were obtained fitting a random effects meta-regression model with MAs as the unit of analysis and including the variable defining the subgroup. Then, the ROR or DSMD was estimated separately in each subgroup. Planned sensitivity analyses involved adjusting the meta-regression models on each domain of the Risk of Bias tool.<sup>19</sup> We adjusted the analysis using each item one at a time, considering low vs high or unclear risk. Further *post hoc* sensitivity analyses were also conducted adjusting on trial sample size. Adjusted analyses were conducted excluding MAs with missing data.

### Sample size calculation

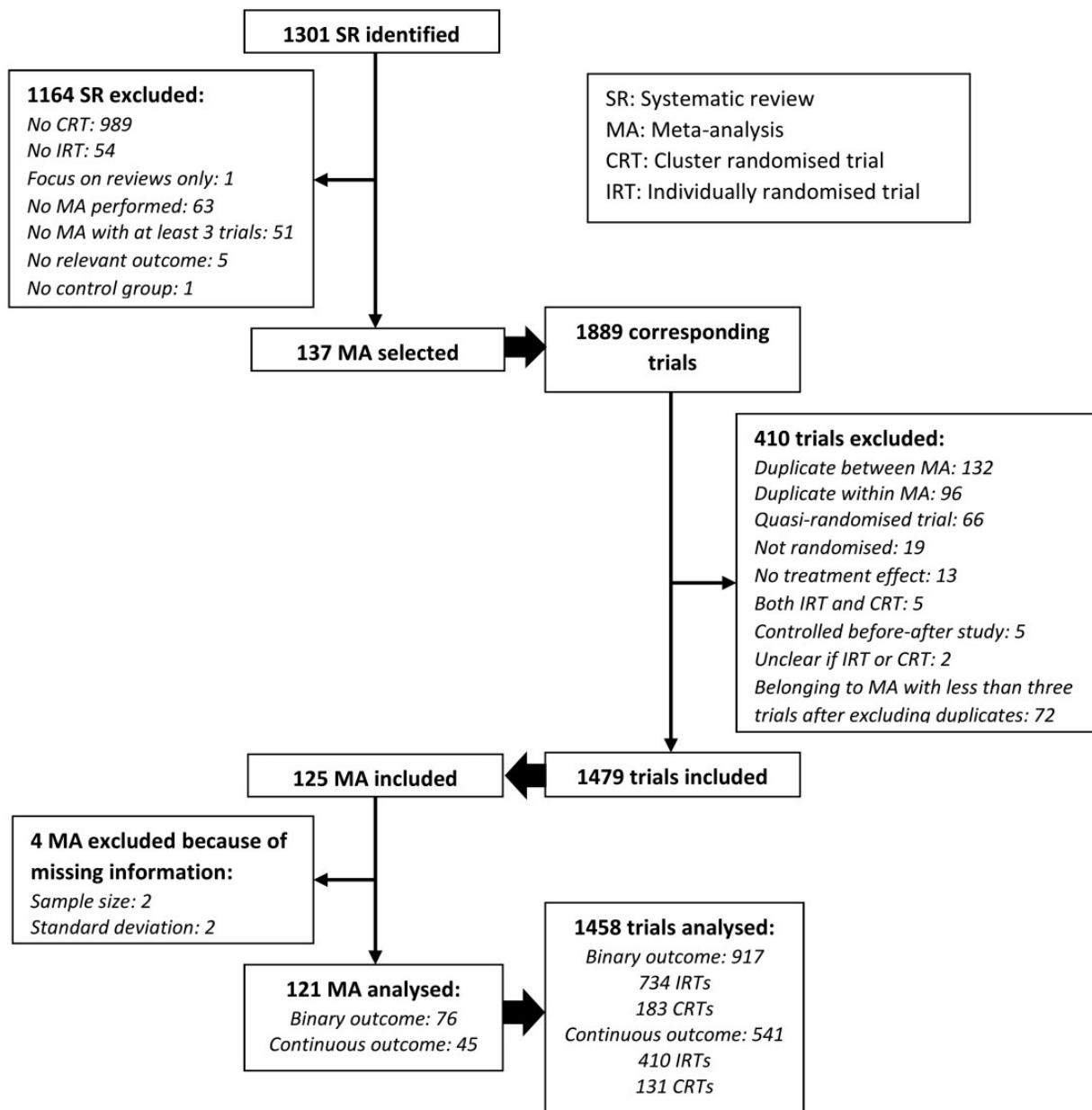
In order to detect a ratio of odds ratio of 0.85, we required 57 MAs to achieve 80% power using a two-sided 5% significance level,<sup>20</sup> assuming a mean number of eight trials per MA<sup>21</sup> with an average of three being CRTs, and the following variances: 0.25 for the within-trial variance of the intervention effect estimate; 0.08 for the between-trial within-meta-analysis variance of the intervention effect estimate; 0.0256 for the between-trial variance of the trial-specific impact of the cluster vs individual randomization; and 0.0016 for the between-meta-analysis variance of the trial-specific impact of cluster randomization. These assumptions were based on the Turner *et al.*<sup>22</sup> large epidemiological study of Cochrane MAs. Such a sample size calculation supposes a binary outcome. We decided to perform two separate analyses, according to whether the outcome was binary or continuous, and then aimed at identifying at least 57 MAs with a binary outcome.

## Results

### Characteristics of selected MAs

Considering Cochrane reviews published over a period of five full years, we identified 1301 systematic reviews by the electronic search. We selected 121 MAs (full references in in the [Supplementary Material 1a, 1b and 2](#), available as [Supplementary data](#) at *IJE* online), corresponding to 1458 trials ([Figure 1](#)): 76 MAs (917 trials) had binary outcomes and 45 (541 trials) had continuous outcomes. MAs concerned very different medical and educational fields and interventions ([Supplementary Material 1a and 1b](#), available as [Supplementary data](#) at *IJE* online).

[Table 1](#) shows that pharmacological interventions were investigated in 25 (32.5%) MAs with a binary



**Figure 1.** Flow diagram of the selection of MAs and randomized trials.

outcome but in only six (13.3%) of those with a continuous outcome. Less than one-third of the MAs had active controls, both for binary and continuous outcomes. Assessed outcomes were objective in one-third of MAs with a binary outcome and in one-quarter with a continuous outcome. The median number of trials (interquartile range: IQR) included was 8 (5 to 15) for MAs with a binary outcome and 10 (5 to 14) for those with a continuous outcome. Finally, for MAs with a continuous outcome, more than half showed substantial heterogeneity, as defined in the Cochrane Handbook,<sup>12</sup> with median  $I^2$  of 60.4% (IQR 22.8%; 81.7%), whereas for MAs with a binary outcome, the median  $I^2$  was 26.5% (IQR 0.0%; 53.5%).

### Characteristics of selected trials

Among the 917 trials with a binary outcome, 183 (20.0%) were CRTs and 734 (80.0%) were IRTs (Table 1). The median sample size was 570 for CRTs (213 to 1764) and 208 for IRTs (83 to 527). The median number of randomized clusters was 31 (12 to 76) and in half, randomized clusters correspond to clinical settings. For 64 of them we used the 0.03 generic value for the ICC to correct the sample size.

Among the 541 trials with a continuous outcome, 131 (24.2%) were CRTs and 410 (75.8%) were IRTs. The median sample size was 139 for CRTs (55 to 291) and 113 for IRTs (56 to 211). Among the 131 CRTs, the median number of randomized clusters was 20 (10 to 40)

**Table 1.** Characteristics of MAs and trials included

MA characteristics	MAs with a binary outcome		MAs with a continuous outcome	
	<i>(n</i> = 76)		<i>(n</i> = 45)	
Intervention, <i>n</i> (%)				
Pharmacological	25 (32.9)		6 (13.3)	
Non-pharmacological	51 (67.1)		39 (86.7)	
Intervention in control group, <i>n</i> (%)				
Inactive	52 (68.4)		34 (75.6)	
Active	24 (31.6)		11 (24.4)	
Outcome objectivity, <i>n</i> (%)				
All-cause mortality	14 (18.4)		-	
Objectively assessed	11 (14.5)		11 (24.4)	
Objectively assessed but influenced by clinician or patient	30 (39.5)		5 (11.1)	
Subjectively assessed	21 (27.6)		29 (64.4)	
Number of trials, median (first and third quartiles) (range) <sup>a</sup>				
Total	8 (5; 15) (3 to 46)		10 (5; 14) (3 to 44)	
Cluster randomized trial	2 (1; 3) (1 to 9)		1 (1; 2) (1 to 24)	
Individually randomized trial	6 (3; 14) (1 to 45)		7 (3; 10) (1 to 38)	
<i>I</i> <sup>2</sup> , median (first and third quartiles) <sup>a</sup>	26.5 (0.0; 53.5)		60.4 (22.8; 81.7)	
$\tau^2$ , median (first and third quartiles) <sup>a</sup>	0.031 (0.000; 0.141)		0.039 (0.005; 0.156)	
	Binary outcome		Continuous outcome	
Trial characteristics	CRTs	IRTs	CRTs	IRTs
	<i>(n</i> = 183)	<i>(n</i> = 734)	<i>(n</i> = 131)	<i>(n</i> = 410)
Year of publication, median (first and third quartiles)	2003 (1997; 2008)	2003 (1996; 2007)	2006 (2003; 2009)	2006 (2001; 2009)
Sample size, <sup>a</sup> median (first and third quartiles)	570 (213; 1764)	208 (83; 527)	139 (55; 291)	113 (56; 211)
Mean $\pm$ standard deviation (SD)	7 886 $\pm$ 43 120	1 589 $\pm$ 9 059	280 $\pm$ 424	197 $\pm$ 354
Cluster type, <i>n</i> (%)				
Clinical setting:	94 (51.4)		43 (32.8)	
Hospital	12 (6.6)		4 (3.0)	
Ward	11 (6.0)		3 (2.3)	
Health centre	13 (7.1)		1 (0.8)	
Residential care home	10 (5.5)		5 (3.8)	
Practice or health professional	43 (23.5)		24 (18.3)	
Other	5 (2.7)		6 (4.6)	
Non-clinical setting:	85 (46.4)		87 (66.4)	
School or classroom	20 (10.9)		63 (48.1)	
Family/household	12 (6.6)		4 (3.0)	
Village or geographical area	37 (20.2)		5 (3.8)	
Other	16 (8.7)		15 (11.5)	
Unclear	4 (2.2)		1 (0.8)	
Number of clusters, median (first and third quartiles), (range)	31 (12; 76), (2 to 68 146)		20 (10; 40), (4 to 531)	

<sup>a</sup>For CRTs, sample size was corrected for clustering.

CRT: cluster randomised trial, IRT: individually randomized trial.

and in less than one-third, randomized clusters corresponded to clinical settings (Table 1). For 45 of them we used the 0.03 generic value for the ICC to correct the sample size.

For MAs with a continuous outcome, 30 CRTs (23.4%) were at low risk of bias for blinding of

outcome assessment as compared with 161 IRTs (40.4%), although in most of these, the risk was assessed as unclear (Supplementary Table 2, available as Supplementary data at *IJE* online). For binary outcomes, no difference was observed between CRTs and IRTs in terms of risk of bias.

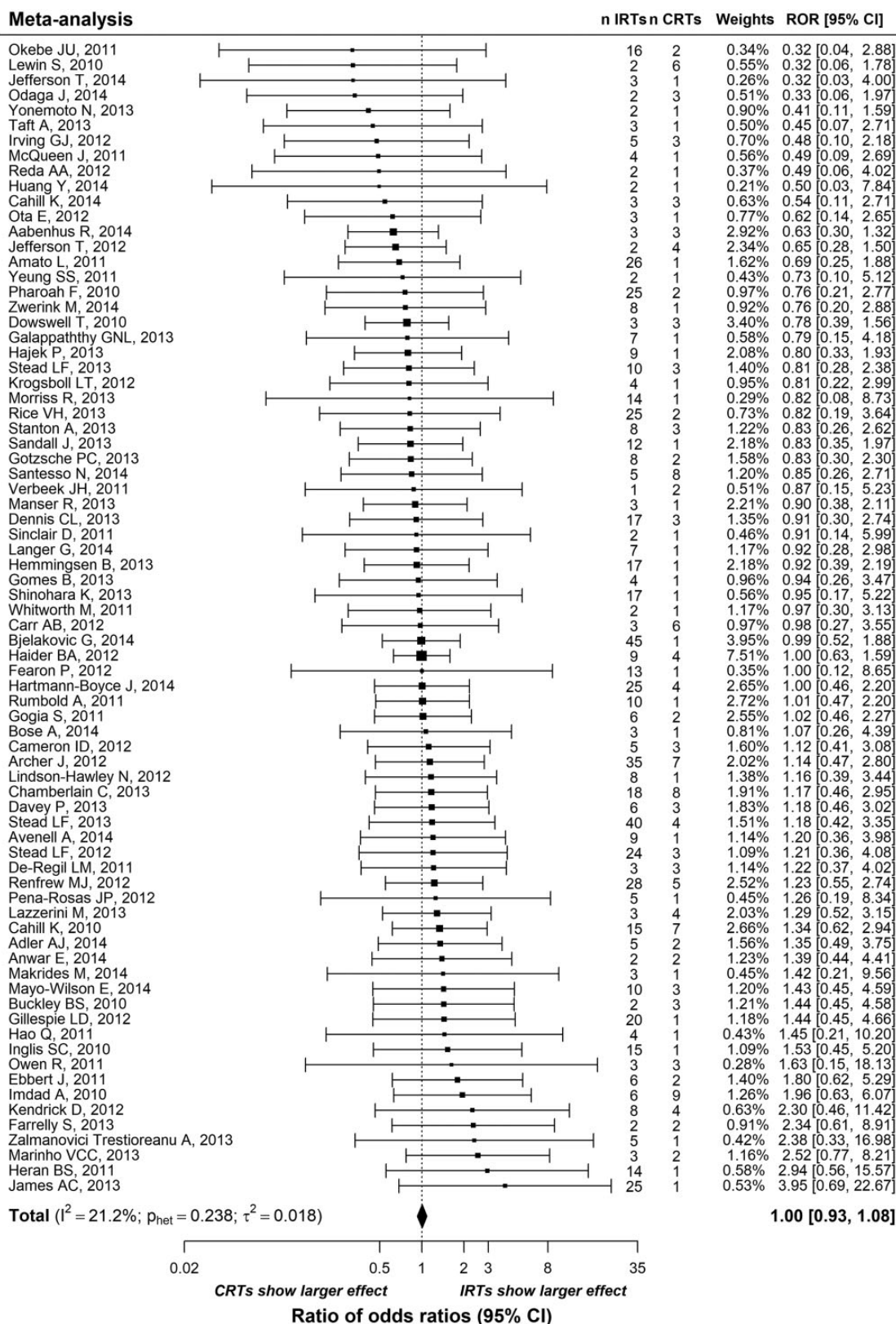


Figure 2. Differences in intervention effect estimates between cluster and individually randomized trials with a binary outcome.

**Table 2.** Difference in intervention effect estimates between cluster and individually randomized trials for binary and continuous outcomes

Binary outcome	<i>n</i> *	ROR		<i>P</i> -value	Heterogeneity		<i>P</i> -value
		Estimate	95% CI		<i>I</i> <sup>2</sup> (95% CI)	$\tau^2$ (95% CI)	
Global	76	1.00	(0.93 to 1.08)	0.238	21.2 (0.0 to 41.4)	0.018 (0.000 to 0.047)	
Subgroup analyses							
Pharmacological	25	1.02	(0.94 to 1.10)	0.405	0.0 (0.0 to 64.9)	0.000 (0.000 to 0.103)	0.360
Non-pharmacological	51	0.98	(0.89 to 1.08)	0.218	20.9 (0.0 to 43.9)	0.023 (0.000 to 0.067)	
Subjective	51	0.99	(0.89 to 1.10)	0.090	26.9 (0.0 to 48.9)	0.035 (0.000 to 0.902)	0.496
Objective	25	1.00	(0.93 to 1.08)	0.738	0.0 (0.0 to 49.3)	0.000 (0.000 to 0.047)	
Active	24	1.02	(0.89 to 1.15)	0.657	6.0 (0.0 to 43.0)	0.006 (0.000 to 0.072)	0.929
Inactive	52	1.01	(0.91 to 1.11)	0.114	29.0 (0.0 to 57.6)	0.025 (0.000 to 0.083)	
Adjusted on risk of bias of:							
Generation of random sequence	60	1.03	(0.94 to 1.12)	0.005	37.8 (5.6 to 63.6)	0.034 (0.003 to 0.097)	
Allocation concealment	60	1.01	(0.92 to 1.11)	0.012	37.1 (2.6 to 59.2)	0.034 (0.002 to 0.084)	
Blinding for participants	31	1.01	(0.87 to 1.18)	0.001	53.0 (14.2 to 80.1)	0.062 (0.009 to 0.220)	
Blinding for the outcome assessor	44	0.99	(0.89 to 1.11)	0.014	40.8 (2.6 to 63.0)	0.040 (0.002 to 0.099)	
Adjusted on trial sample size	69	0.98	(0.89 to 1.07)	0.056	27.1 (0.0 to 58.3)	0.028 (0.000 to 0.105)	

Continuous outcome	<i>n</i> *	DSMD		<i>P</i> -value	Heterogeneity		<i>P</i> -value
		Estimate	95% CI		<i>I</i> <sup>2</sup> (95% CI)	$\tau^2$ (95% CI)	
Global	45	0.13	(0.06 to 0.19)	0.221	21.7 (0.0 to 47.4)	0.009 (0.000 to 0.029)	
Subgroup analyses							
Pharmacological	6	-0.03	(-0.12 to 0.07)	0.435	0.0 (0.0 to 90.6)	0.000 (0.000 to 0.436)	0.016
Non pharmacological	39	0.15	(0.08 to 0.21)	0.515	7.5 (0.0 to 43.2)	0.003 (0.000 to 0.027)	
Subjective	34	0.15	(0.08 to 0.22)	0.398	11.1 (0.0 to 52.5)	0.005 (0.000 to 0.040)	0.118
Objective	11	0.05	(-0.08 to 0.17)	0.420	20.5 (0.0 to 74.2)	0.008 (0.000 to 0.091)	
Active	11	0.25	(0.15 to 0.36)	0.877	0.0 (0.0 to 57.2)	0.000 (0.000 to 0.049)	0.006
Inactive	34	0.08	(0.01 to 0.15)	0.352	15.4 (0.0 to 54.5)	0.006 (0.000 to 0.037)	
Adjusted on risk of bias of:							
Generation of random sequence	32	0.12	(0.05 to 0.19)	0.583	8.8 (0.0 to 48.0)	0.003 (0.000 to 0.031)	
Allocation concealment	36	0.11	(0.03 to 0.19)	0.116	29.3 (0.0 to 60.9)	0.013 (0.000 to 0.050)	
Blinding for participants	16	0.11	(0.00 to 0.22)	0.065	38.3 (0.0 to 78.9)	0.016 (0.000 to 0.094)	
Blinding for the outcome assessor	23	0.22	(0.03 to 0.41)	<0.0001	84.6 (66.7 to 93.0)	0.134 (0.049 to 0.328)	
Adjusted on trial sample size	38	0.06	(-0.02 to 0.13)	0.060	24.1 (0.0 to 75.1)	0.011 (0.000 to 0.102)	

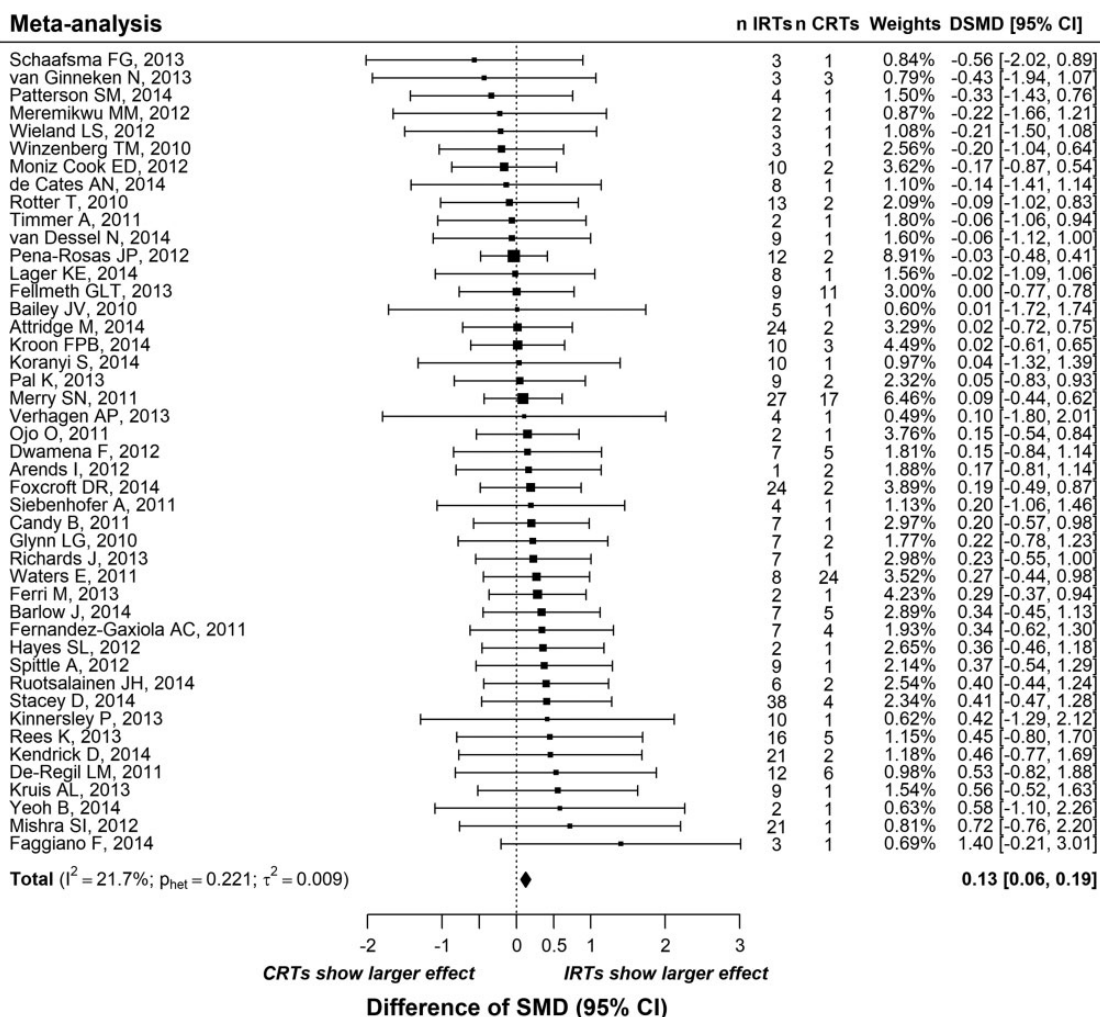
*n*\*, number of MAs included in the analysis.

### Differences in intervention effect estimates between CRTs and IRTs

For MAs with a binary outcome, intervention effect estimates did not differ between CRTs and IRTs. The combined ROR was estimated at 1.00 (95% CI 0.93 to 1.08) (Figure 2 and Table 2). Heterogeneity was low across MAs ( $I^2 = 21.2\%$ ;  $P = 0.238$ ; between-meta-analyses variance  $\tau^2 = 0.018$ ). Subgroup and adjusted analyses led to consistent results with a combined ROR very close to 1.00, whatever the analysis (Table 2 and Supplementary Figures 2–4, available as Supplementary data at *IJE* online). The results were also robust across all the performed sensitivity analyses (see Supplementary Tables 3–5, available as Supplementary data at *IJE* online).

For MAs with a continuous outcome, intervention effect estimates were more favourable for IRTs, with a combined DSMD of 0.13 (95% CI 0.06 to 0.19) (Figure 3). Although statistically significant, this difference is small according to Cohen's classification of effect sizes<sup>23</sup>. Heterogeneity was low across MAs ( $I^2 = 21.7\%$ ;  $P = 0.221$ ; between-meta-analyses variance  $\tau^2 = 0.009$ ). Subgroup analyses led to inconsistent results among subgroups. The combined DSMD was significant for non-pharmacological interventions but was lower and non-significant for pharmacological interventions: 0.15 (0.08 to 0.21) vs -0.03 (-0.12 to 0.07) (interaction  $P$ -value = 0.016). Similarly, the effect of cluster randomization on intervention effect estimates was larger for subjective than for objective outcomes, although the interaction





IRT: individually randomised trial; CRT: cluster randomised trial; DSMD: difference of standardised mean differences; CI: confidence interval

Figure 3. Differences in intervention effect estimates between cluster and individually randomized trials with a continuous outcome.

was not significant: 0.15 (0.08 to 0.22) vs 0.05 (-0.08 to 0.17) (interaction  $P$ -value = 0.118). Finally, the DSMD was significantly lower for inactive compared with active control interventions: 0.08 (0.01 to 0.15) vs 0.25 (0.15 to 0.36) (interaction  $P$ -value = 0.006). Adjusting for the effective trial sample size led to a smaller difference of 0.06 (-0.02 to 0.13), which was not significant. Adjusting for risk of bias items did not affect the results, except for blinding of outcome assessors, with a higher DSMD, estimated to be 0.22 (0.03 to 0.41). The choice of the ICC value to adjust for clustering when the trials values were not known does not impact the results (results are presented in [Supplementary Table 3](#), available as [Supplementary data](#) at *IJE* online).

### Discussion

In this meta-epidemiological study, we selected 121 MAs: 76 (917 trials) with a binary and 45 (541 trials) with continuous

outcomes. For binary outcomes, the ratio of odds ratios was 1.00 (95% CI 0.93 to 1.08), indicating that intervention effect estimates did not systematically differ between CRTs and IRTs. Consistent results were observed in all subgroup and adjusted analyses. For continuous outcomes, intervention effect estimates were more favourable with individual randomization, although the difference was moderate (difference in standardized mean differences of 0.13, 95% CI 0.06 to 0.19). This difference was much smaller and not significant for the trial subgroup of pharmacological interventions or when adjusting on sample size.

### Strengths and limitations of the study

We selected a large sample of MAs covering a wide range of medical and educational areas, which provides good generalizability of our results. We nevertheless restricted our study to Cochrane MAs: to identify potentially eligible MAs, we had to access the full text of the systematic reviews

because the abstracts of reports rarely specify the inclusion of both CRTs and IRTs. Restricting our study to Cochrane reviews may limit the generalizability of our results. This study was conducted using trial-level summaries of the intervention effect. Therefore, no information was available regarding patients' non-adherence or loss to follow-up, which might have had an impact on the trials' results, if these issues were to affect CRTs differently from IRTs. However, our aim was to assess whether there exists systematic differences between CRTs and IRTs. Further studies using individual patient data would be needed to investigate the specific effect of each component that differs between CRTs and IRTs. Such studies would probably need to restrict the focus to a specific medical area, which differs from the philosophy of meta-epidemiological studies.

We discarded studies not randomized, such as quasi-randomized trials or before–after studies, so as to obtain well-defined groups for comparison. We handled clustering, thus making sure that our results were not distorted by over-weighted CRTs. Finally we explored both binary and continuous outcomes in the same study (although independently) which, except for the Alexander *et al.*<sup>24</sup> or Smail-Faugeron *et al.*<sup>25</sup> studies, is uncommon.

### Relation to previous work

To our knowledge and in view of the Dechartres *et al.*<sup>26</sup> recently published systematic review of meta-epidemiological studies, our study is the first meta-epidemiological study to compare intervention effect estimates between CRTs and IRTs. However, our results are consistent with Selvaraj and Prasad's,<sup>11</sup> who showed that the proportions of statistically significant findings were similar in CRTs and IRTs.

### Possible mechanisms

CRTs and IRTs differ in several ways. CRTs may face recruitment bias, but they may benefit from a 'herd effect'; IRTs may suffer from group contamination. All these elements may lead to larger intervention effect estimates in CRTs than in IRTs. Besides, most of the interventions assessed in CRTs do not allow for any form of blinding, which invites both performance and detection bias.<sup>6</sup> This feature has been shown to be associated with an over-estimation of intervention effects.<sup>7</sup> Conversely, CRTs are considered more pragmatic,<sup>27</sup> and allow the estimation of the effectiveness, rather than the efficacy, as in many IRTs. Effectiveness is usually smaller than efficacy, mainly because of non-compliance. Pragmatic trials also nearly always involve several centres and they are usually larger. These characteristics are important because the intervention effect estimates have been shown to be lower in multi-centre than single-centre trials,<sup>28,29</sup> and in larger trial sample

sizes.<sup>30</sup> Therefore, antagonist mechanisms may occur and might counterbalance each other. In the end, although CRTs and IRTs may look as if they are similar but just conducted as CRTs or IRTs, very different mechanisms—sometimes antagonist—may apply and contribute to systematic differences in intervention effect estimates between CRTs and IRTs.

### Discrepancy between binary and continuous outcomes

The finding that there is no difference between CRTs and IRTs for binary outcomes suggests that the different mechanisms are not very strong, or non-existent or that they compensate for each other, and this result held in all considered subgroups. For continuous outcomes, the observed 0.13 difference in standardized mean differences invites the two following comments. First, although significant, the observed difference can be considered moderate in view of previously reported differences in standardized mean differences.<sup>26</sup> Second, one could have expected a difference in the opposite sense in view of the underlying mechanisms (i.e. larger intervention effect estimates in CRTs than in IRTs). A potential explanation is that there are probably many single-centre IRTs, with low median size (113 participants), whereas CRTs are intrinsically multicentre studies, most randomizing practices, schools or classrooms.

The discrepancy we observed between MAs with binary and continuous outcomes is not new, and others have urged caution when extrapolating results of meta-epidemiological studies of binary outcomes to situations of continuous outcomes.<sup>24,26</sup> We found several differences between trials and MAs according to whether the outcome was continuous or binary: (i) the sample size was smaller in trials with continuous outcomes; (ii) heterogeneity was higher (median  $I^2$  of 60.4 compared with 26.5); (iii) blinding was less frequent; (iv) outcomes were more frequently subjective (64.4% of MAs with a continuous outcome vs 27.6% with a binary outcome when focusing on only 'subjective outcome'; and (v) the settings differed, with cluster trials with a continuous outcome being more likely to have non-clinical settings. All these differences may explain the discrepancy we observed.

Finally, from a statistical point of view, we cannot exclude some form of meta-confounding. We indeed adjusted analyses, but doing so led to discarding some MAs (notably those with only three trials), and we adjusted on only one covariate at a time.

### Conclusions and implications

For binary outcomes, CRTs and IRTs produced the same intervention effect estimates, but intervention effect estimates were marginally more favourable (i.e. either more beneficial or less detrimental) for IRTs with continuous outcomes. However, this result was not observed for trials

assessing a pharmacological intervention or with an objective outcome. More work is needed, in particular to understand how the type of intervention, outcome, setting or trial sample size affects the results.

## Supplementary Data

Supplementary data are available at *IJE* online.

## Funding

This work was supported by a grant from the French Ministry of Health (PREPS 13-0015).

**Conflict of interest:** None declared.

## References

- Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London: Arnold, 2000.
- Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomized trials: review of recent trials published in three general medical journals. *BMJ* 2003;327:785–89.
- Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomized trials: what can be done? *BMJ* 2009;339:b4006.
- Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomized trials: systematic review of recent trials. *BMJ* 2008;336:876–80.
- Giraudeau B, Ravaud P. Preventing bias in cluster randomized trials. *PLoS Med* 2009;6:e1000065.
- Caille A, Kerry S, Tavernier E, Leyrat C, Eldridge S, Giraudeau B. Timeline cluster: a graphical tool to identify risk of bias in cluster randomized trials. *BMJ* 2016;354:i4291.
- Savović J, Jones HE, Altman DG *et al*. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429–38.
- Santesso N, Carrasco-Labra A, Brignardello-Petersen R. Hip protectors for preventing hip fractures in older people. Cochrane Database of Systematic Reviews. 2014. <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001255.pub5/abstract> (7 Sep 2016, date last accessed).
- Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomized trials. *BMC Med Res Methodol* 2005;5:10.
- Gilbody S, Bower P, Torgerson D, Richards D. Cluster randomized trials produced similar results to individually randomized trials in a meta-analysis of enhanced care for depression. *J Clin Epidemiol* 2008;61:160–68.
- Selvaraj S, Prasad V. Characteristics of cluster randomized trials: are they living up to the randomized trial? *JAMA Intern Med* 2013;173:313–15.
- Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. London: Cochrane Collaboration, 2011.
- Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. *Evid Based Med* 2017;22:139–42.
- Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. *Clinical Trials* 2005;2:99–107.
- Viechtbauer W. Approximate confidence intervals for standardized effect sizes in the two-independent and two-dependent samples design. *J Educ Behav Stat* 2007;32:39–60.
- Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in “meta-epidemiological” research. *Stat Med* 2002;21:1513–24.
- Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.
- Veroniki AA, Jackson D, Viechtbauer W *et al*. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods* 2016;7:55–79.
- Higgins JPT, Altman DG, Gotzsche PC *et al*. The Cochrane Collaboration’s tool for assessing risk of bias in randomized trials. *BMJ* 2011;343:d5928.
- Giraudeau B, Higgins JPT, Tavernier E, Trinquart L. Sample size calculation for meta-epidemiological studies. *Stat Med* 2016;35:239–50.
- Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4:e78.
- Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *Int J Epidemiol* 2012;41:818–27.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Abingdon, UK: Routledge, 2013.
- Alexander PE, Bonner AJ, Agarwal A *et al*. Sensitivity subgroup analysis based on single-center vs. multi-center trial status when interpreting meta-analyses pooled estimates: the logical way forward. *J Clin Epidemiol* 2016;74:80–92.
- Smail-Faugeron V, Fron-Chabouis H, Courson F, Durieux P. Comparison of intervention effects in split-mouth and parallel-arm randomized controlled trials: a meta-epidemiological study. *BMC Med Res Methodol* 2014;14:64.
- Dechartres A, Trinquart L, Faber T, Ravaud P. Empirical evaluation of which trial characteristics are associated with treatment effect estimates. *J Clin Epidemiol* 2016;77:24–37.
- Freemantle N, Strack T. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. *J Clin Epidemiol* 2010;63:1053–58.
- Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. *Ann Intern Med* 2011;155:39–51.
- Bafeta A, Dechartres A, Trinquart L, Yavchitz A, Boutron I, Ravaud P. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. *BMJ* 2012;344:e813.
- Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ* 2013;346:f2304.