

Adjusting for bias in unblinded randomized controlled trials.

A F Schmidt* [a], R H H Groenwold [b].

- a. Institute of Cardiovascular Science, Faculty of Population Health, University College London, London WC1E 6BT, United Kingdom.
- b. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, the Netherlands.

* Contact: 0044 (0)20 3549 5625

E-mail address: amand.schmidt@ucl.ac.uk (A.F.Schmidt).

Running title: Treatment effect estimators for unblinded RCTs.

Word count text: 6 074

Word count abstract: 188

Number of references: 29

Number of tables: 3

Number of figures: 6

(Web)appendix: 3

Abstract

It may not always be possible to blind participants of a randomized controlled trial for treatment allocation. As a result, estimators of the actual treatment effect may be biased. In this paper we will extend a novel method, originally introduced in genetic research, for instrumental variable meta-analysis, adjusting for bias due to unblinding of trial participants. Using simulation studies, this novel method, “Egger Correction for non-Adherence” (Egger-CA), is introduced and compared to the performance of the “intention to treat”, “as treated”, and conventional “instrumental variable” estimators. Scenarios considered (time-varying) non-adherence, confounding, and between study heterogeneity. The effect of treatment on a binary endpoint was quantified by means of a risk difference. In all scenarios with unblinded treatment allocation, the Egger-CA method was the least biased estimator. However, unless the variation in adherence was relatively large, precision was lacking, and power did not surpass 0.50. As a comparison, in a meta-analysis of blinded RCTs power of the conventional IV estimator was 1.00 versus at most 0.14 for the Egger-CA estimator. Due to this lack of precision and power we suggest to use this method mainly as a sensitivity analysis.

Keywords Statistics, Randomized controlled trials, Monte Carlo Method, Bias, Treatment Effectiveness, Instrumental Variable

Introduction

Randomized controlled trials (RCTs) are the gold standard to evaluate effects of (medical) interventions, with meta-analyses of RCTs often seen as the highest level of evidence(1). Such meta-analyses of RCTs frequently include a risk of bias assessment(2), evaluating the risk of bias of each included study based on e.g. blinding of participants and outcome assessors for the allocated treatment regime. However, blinding may not always be feasible. For example, in trials of epidural anesthesia vs. non-epidural anesthesia or no anesthesia among women in labor, participants were not blinded for their allocated treatment(3).

In RCTs any non-adherence to treatment allocation may lead to a mismatch between the effect of treatment allocation and the effect of actual treatment use. In part this may be related to “contamination” when subjects switch study interventions. However, non-adherence may also lead to subjects using different non-study interventions or no treatment at all. By blinding RCT participants to allocated treatments any non-adherence and possible contamination is expected to be random, and one can estimate the effect of allocating treatment on the outcome (i.e., the intention-to-treat [ITT] effect). Depending on the degree of non-adherence the ITT effect can be very similar to the effect of actually taking treatment (when adherence is close to 100%). In unblinded RCTs, one typically cannot assume that non-adherence and contamination are random, and treatment

allocation itself may have a “direct effect” on the outcome, independent of its effect on treatment (adherence). Hence conventional estimators of the treatment effect (such as the ITT estimator) may be biased.

Recently, a novel method was introduced in genetic “Mendelian randomization” [MR] studies(4,5) which can possibly be used in RCTs to correct for any bias due to unblinded treatment allocation as well. Briefly, in MR studies genetic variation may be seen as treatment allocation in the sense that genes are randomly distributed in a population (like randomization in a trial), and may be related to a biomarker (i.e., treatment use in RCTs), which is associated with a disease. Similar to a blinded RCT, if all the effects of a gene on the outcome are mediated by the biomarker, a direct effect of the gene on the outcome is absent. However, in many cases a gene has multiple effects on the outcome, possibly mediated by different biomarkers, and a direct effect cannot be ruled out. To correct for this so-called pleiotropy Bowden et al.,(5) introduced a novel version of the Egger-test(6) termed “MR-Egger”.

In this paper, we will extend the MR-Egger method to a meta-analysis of unblinded RCTs with the aim to estimate the effect that taking treatment has on the outcome. For this we will first review the typical estimators used in a single RCT. Second, we will extend the MR-Egger estimator to meta-analyses of RCTs with binary events. Using simulations we

will evaluate this novel method and compare its performance to the ITT, as-treated, and instrumental variable estimators. Finally, we will apply this method on data from a clinical example of epidural anesthesia for women in labor.

Methods

Notation and outcome model

In the following we assume that data are collected from an RCT allocating $i = 1, \dots, n$ subjects to a certain treatment, e.g., epidural anesthesia $z_i = 1$, or non-epidural anesthesia or no anesthesia $z_i = 0$, with the intention to investigate its effect on a binary endpoint $y_i = \{1,0\}$, e.g., caesarian section or vaginal birth. Possibly contrary to treatment allocation, subjects either receive epidural analgesia $x_i = 1$ or not $x_i = 0$. Note that treatment allocation occurs at time $t = 0$, \mathbf{X} occurs somewhere between the start of follow-up $t = 0$ and $t = T$, the end of follow-up. For now, we consider that x_i does not change over time; we will relax this assumption in a later described scenario. In addition, we assume that taking treatment, not only depends on \mathbf{Z} , but also depends on unmeasured (confounding) factors \mathbf{U} which also affects the (binary) outcome \mathbf{Y} . This scenario is presented in a causal directed acyclic graph (DAG in Figure 1A), representing a blinded RCT, where we note that the outcome risk is based on $Prob(\mathbf{Y} = 1|\mathbf{X}, \mathbf{U})$. Contrary to this blinded RCT, in Figure 1B we present a DAG for an unblinded trial were \mathbf{Z} has a *direct* effect on \mathbf{Y} , hence the outcome risk is based on $Prob(\mathbf{Y} = 1|\mathbf{Z}, \mathbf{X}, \mathbf{U})$. Throughout this paper we assume the

interest is in estimating the causal effect of \mathbf{X} (treatment) on \mathbf{Y} . Capital letters indicate a column matrix of the sort $\mathbf{X} = \{x_i, \dots, x_n\}^T$.

Intention To Treat estimator

Most, if not all, RCTs are analyzed using the Intention To Treat (ITT) estimator which evaluates if outcomes differ across levels of \mathbf{Z} . As such, the ITT effect measures the effect of assigning treatment, not the effect of actually taking treatment. In unblinded RCTs (Figure 1B) the ITT effect can potentially be decomposed in a *direct* effect ($\hat{\tau}$) of treatment allocation on the outcome and an *indirect* effect ($\hat{\alpha}$) of treatment allocation on the outcome which is mediated through the actual treatment received (i.e., the effect of taking treatment). We note that if in an unblinded trial all reasons for treatment compliance were observed (i.e., \mathbf{Z} and \mathbf{U}) one could estimate the *direct* effect and subtract this from the ITT effect to calculate the *indirect effect* of treatment allocation on the outcome, see equations 1A and 1B. However, because RCTs are typically initiated because one is unwilling to condition on \mathbf{U} the above procedure is unlikely sufficient in empirical settings. Instead, by blinding participants for their allocated treatment the direct effect is expected to be zero and the ITT effect reflects the *indirect* effect only. If, additionally, adherence is (close to) 100% this ITT effect will coincide with the effect of actually *taking* treatment.

The ITT effect: $[Prob(\mathbf{Y} = 1|\mathbf{Z} = 1) - Prob(\mathbf{Y} = 1|\mathbf{Z} = 0)] = \hat{\alpha} + \hat{\tau}$ [equation 1A],

the direct effect: $[Prob(Y = 1|Z = 1, X = x, U = u) -$

$Prob(Y = 1|Z = 0, X = x, U = u)] = \hat{\tau}$ [equation 1B]

the indirect effect: $[Prob(Y = 1|Z = 1) - Prob(Y = 1|Z = 0)] -$

$[Prob(Y = 1|Z = 1, X = x, U = u) - Prob(Y = 1|Z = 0, X = x, U = u)] = \hat{\alpha}$.

As Treated estimator

In addition to the ITT effect some RCTs report the “As Treated” (AT) effect which estimates the actual effect of taking treatment. Typically, the decision to take treatment X depends on covariates (U), therefore the unbiased AT effect equals:

the “conditional” (AT) effect: $Prob(Y = 1|X = 1, U = u) - Prob(Y = 1|X = 0, U = u)$.

Because, U is likely to be only partially unobserved the AT effect is typically implemented without conditioning on (all relevant) covariates and is therefore expected to be biased:

the “crude” (AT) effect: $Prob(Y = 1|X = 1) - Prob(Y = 1|X = 0)$.

Instrumental Variable estimator.

An unbiased AT effect can be estimated without the need to condition on (unobserved) covariate U using the instrumental variable (IV) estimator (7):

$$\frac{Prob(\mathbf{Y} = 1|\mathbf{Z} = 1) - Prob(\mathbf{Y} = 1|\mathbf{Z} = 0)}{Prob(\mathbf{X} = 1|\mathbf{Z} = 1) - Prob(\mathbf{X} = 1|\mathbf{Z} = 0)}. \text{ [equation 2]}$$

This is the ratio of the ITT effect and the degree of allocation compliance, or *adherence*; with the denominator being 1 if there is complete adherence to treatment allocation. We note that the denominator can be interpreted as the difference between the proportion of subjects who adhere to treatment allocation when $\mathbf{Z} = 1$, and the proportion of “contaminated” subjects who received $\mathbf{X} = 1$ despite being allocated to $\mathbf{Z} = 0$. Hereafter, we will simply refer to this quantity as adherence (or non-adherence).

This IV effect estimate is unbiased if, amongst other assumptions (see below), there is no direct effect of treatment allocation on the outcome. As described before, in an unblinded RCT a direct effect $\hat{\tau}$ may be present (i.e., the exclusion restriction assumption may be violated) and in such settings the unbiased IV effect can only be estimated by subtracting $\hat{\tau}$ from the numerator of equation 2. However, following equation 1B, $\hat{\tau}$ itself can only be estimated unbiasedly if we can condition on \mathbf{U} , resulting in the same limitation as the other discussed estimators that condition on \mathbf{U} .

Formally the IV effect is unbiased if the following 3 assumptions hold:

- (i). The instrumental variable **Z** is strongly related to the exposure of interest **X**.
- (ii). The instrumental variable is independent of potential confounders **U**.
- (iii). Conditional on **X** and **U**, the instrumental variable does not have an effect on **Y** (i.e., the exclusion restriction assumption).

For the IV estimate to reflect the "average causal effect" we need to additionally assume that the effect of **X** on **Y** is homogeneous(7,8) i.e., the effect of treatment on the outcome is the same for all individual subjects. Alternatively, to identify the "local average treatment effect" we must assume that participants will not categorically take the opposite treatment as indicated by **Z** (i.e., no 'defiers')(9).

We note that besides the ratio estimator, other IV estimators exist (10,11) which make similar assumptions. Furthermore, the standard error of these estimators should obviously account for variance in both the numerator and the denominator, for example using the delta method(12,13).

Estimating the unbiased instrumental variable effect in a meta-analysis of unblinded RCTs

If results from multiple $j = 1, \dots, J$ unblinded RCTs are available, or alternatively from J clusters within a single RCT, the effect of taking treatment can be estimated using a novel meta-regression(14) method developed for genetic MR studies(5).

Let us assume that for each RCT aggregate estimates are available on the ITT effect $\hat{\theta}_j = \hat{\alpha}_j + \hat{\tau}_j$, and the effect of treatment allocation on actual treatment use (i.e., adherence) $\hat{\omega}_j = Prob(\mathbf{X} = 1|\mathbf{Z} = 1) - Prob(\mathbf{X} = 1|\mathbf{Z} = 0)$, and we re-parameterize these such that the following restrictions hold:

- (a). $\hat{\theta}_j^* = sgn(\hat{\omega}_j) \hat{\theta}_j$ with sgn equal to the sign function.
- (b). $\hat{\omega}_j^* = |Prob(\mathbf{X} = 1|\mathbf{Z} = 1) - Prob(\mathbf{X} = 1|\mathbf{Z} = 0)|$.

Given that an RCT where $Prob(\mathbf{X} = 1|\mathbf{Z} = 1) < Prob(\mathbf{X} = 1|\mathbf{Z} = 0)$ (i.e., negative adherence) seems unlikely to occur in practice and also does not occur in this paper, we drop the starred notation.

Next, by regressing the study effects $\hat{\theta}_j$ on study specific adherence, $\hat{\omega}_j$, the unbiased effect of taking treatment can be recovered by fitting

$$\hat{\theta}_j = \beta_0 + \beta_1 \hat{\omega}_j + \varepsilon_j \text{ [equation 3],}$$

weighted by the inverse of the variance of $\hat{\theta}_j$. Here $\hat{\beta}_1$ reflects the unbiased IV estimate of taking treatment and $\hat{\beta}_0$, the direct effect of treatment allocation on the outcome Y independent of \mathbf{X} and \mathbf{U} , and ϵ_j the residual variance. The intuition behind this regression is that if \mathbf{Z} does not have an effect on \mathbf{Y} except through its effect on treatment, the effect of \mathbf{Z} on the outcome when adherence is zero, $\hat{\beta}_0$, should also be zero. In other words if $\beta_0 = 0$, this means $\tau = 0$ is zero as well, and there is no direct effect of treatment allocation on the outcome (with the usual constraint that proving the strict null is impossible). Using these estimates we can explore, 1) if IV assumption (iii) is violated by testing if $\beta_0 = 0$, and 2) if the IV effect adjusted for a possible direct effect, τ_j , differs from zero i.e., test if $\beta_1 = 0$.

The major assumption of this approach is that the *direct* effect τ_j does not depend on adherence ω_j ; following Bowden et al., we will call this the InSIDE assumption (Instrument

Strength Independent of Direct Effect). Under this InSIDE assumption $\hat{\beta}_1 = \frac{cov(\hat{\theta}_j, \hat{\omega}_j)}{var(\hat{\omega}_j)} =$

$\frac{cov(\hat{\alpha}_j, \hat{\omega}_j)}{var(\hat{\omega}_j)} + \frac{cov(\hat{\tau}_j, \hat{\omega}_j)}{var(\hat{\omega}_j)}$ and the expected value of the numerator in the last term will equal

zero. This InSIDE assumption is violated for example when treatment allocation \mathbf{Z} also affects confounders such as \mathbf{U} ; a slightly different version of IV assumption (ii).

Bowden et al. coined this method MR-Egger, because of the close resemblance to the **Egger test** for small study heterogeneity(6,15). Here we extend this method beyond

Mendelian randomization and refer to this method as the **Egger Correction for non-Adherence (Egger-CA)**. To gain more insight in the Egger-CA method we will briefly discuss the Egger test and Egger-CA methods in a graphical manner.

First we note, that the meta-analyzed ITT effect of J RCTs can be estimated by fitting a linear regression model of the sort $\hat{\theta}_j = \delta_0 + \varepsilon_j$, weighted by the inverse of the variance in $\hat{\theta}_j$, where $\hat{\delta}_0$ represents the meta-analyzed ITT estimate. Heterogeneity in ITT effects due to small sample size studies can be explored by fitting the linear regression model:

$\hat{\theta}_j/SE(\hat{\theta}_j) = \delta_0 + \delta_1\hat{P}_j + \varepsilon_j$ where $\hat{P}_j = 1/SE(\hat{\theta}_j)$, i.e., the precision of the study specific estimate of $\hat{\theta}_j$. Hence $\hat{\delta}_0$ reflects the standardized ITT effect in a study without any precision and can be used to explore if there is a sample size dependent trend in treatment effects, and $\hat{\delta}_1$ reflects the standardized ITT effect in an infinitely large RCT; Figure 2, left panel dashed line.

Similarly, by fitting an intercept only model (i.e., weighted average) using the study specific IV estimate (equation 2) as the dependent variable, the meta-analyzed IV estimate can be obtained (Figure 2, right panel, horizontal line). Finally, the Egger-CA estimate is obtained by regressing the ITT effect estimates on the adherence estimates (equation 3), where the slope of the regression line estimates the ITT effect if adherence was 100%. This ITT effect with perfect adherence is also the only setting when the ITT effect estimate equals

the effect of *taking* treatment (Figure 2 right panel dashed dotted line). Also note that by regressing the ITT effect estimates on the adherence estimates and suppressing the intercept (i.e., forcing the intercept to be 0) the IV effect of equation 2 is estimated.

Simulation study

Using simulations, we evaluated the “crude” **AT** estimator (unconditional on **U**), the **ITT** estimator (equation 1A), the **IV** estimator (equation 2), and the **Egger-CA** estimator [**ECA**] (equation 3). For the ECA method, we evaluated both the treatment effect (i.e., indirect effect) and the direct effect estimates. The ECA method was implemented with three different variance estimators: 1) dividing the usual OLS standard error estimates by the error of the residual variance when this is larger than 1 [**ECA-OLS**], 2) a standard error estimator based on a parametric bootstrap procedure, sampling the ITT and adherence estimates from a normal distribution [**ECA-PB**]. Finally, given that all the observed variables considered are categorical, the individual patient data (IPD) can easily be reconstructed from the available aggregate data, and 3) the variance of the ECA can also be estimated by performing a study specific non-parametric bootstrap [**ECA-NPB**] in order to estimate the percentile confidence intervals(16). Both bootstrap estimators were implemented with 1 000 replications.

We note that the ECA-OLS estimator as implemented here entails a multiplicative random effects model(17). Bowden et.al., suggest that this multiplicative random effects model

may be preferred because it does not up-weight outlying estimates, which may be outlying because the InSIDE assumption is more severely violated. Contrary to the more commonly used additive random effects model, a multiplicative random effects model has the desired property of increasing the standard error without influencing the point estimates. All other estimators (including the ECA-PB and ECA-NPB) are fitted assuming a fixed effect model.

Data-generating process

Data were simulated to mimic an aggregate data meta-analysis of J studies, in which the number of subjects per study ranged between 400 and 5 000 ($n_j \sim \text{unif}(400, 5\,000)$), randomization followed a 1:1 allocation scheme $z_j \sim \text{bernoulli}(0.50)$, a confounder \mathbf{U} was generated as $u_j \sim \text{bernoulli}(0.40 + \mu_z z_j)$, treatment status as $x_j \sim \text{bernoulli}(p_j)$, where $p_j = q_{0,j}(1 - z_j) + q_{1,j}z_j$. Non-adherence was generated based on \mathbf{U} : $q_{0,j} = 0 + \text{unif}(0, \mu_0)u_j$ and $q_{1,j} = 1 + \text{unif}(\mu_1, 0)u_j$, with $\mu_1 \leq 0$ and $\mu_0 \geq 0$. In other words $\omega_j = q_{1,j} - q_{0,j} = [1 + \text{unif}(\mu_1, 0)u_j] - [0 + \text{unif}(0, \mu_0)u_j]$. Hence the minimum $\omega_j = 1 + (u_1 - u_0)E(\mu_j)$ and the maximum $\omega_j = 1$, see Appendix Table 1 for the scenario specific values. Finally, a binary outcome was generated based on \mathbf{X} , \mathbf{Z} , and \mathbf{U} : $y_j \sim \text{bernoulli}(\mu_{y,j} + \mu_{yx}x_j + \tau z_j + \mu_{yu}u_j)$. An overview of simulation parameters is given in Table 1.

Simulation scenarios

In the following, simulation **scenario I** is described in full, to increase readability changes from preceding scenarios are discussed for the remaining scenarios (with the complete scenarios details presented in Table 1 and Figure 3).

In **scenario I** we simulated $J = 20$ perfectly blinded RCTs with $\mu_{y,j} \sim N(0.15, 0.025^2)$ $\mu_z = 0.00$, $\mu_{yx} = 0.10$, $\tau = 0.00$, $\mu_{yu} = 0.10$, $\mu_0 = 0.050$, and $\mu_1 = \{-0.050, -0.200, -0.350\}$. In **scenarios II and III** unblinded RCTs were simulated by setting τ to 0.15 or -0.15 respectively, and additionally for scenario **III** $\mu_{y,j} \sim N(0.35, 0.025^2)$ instead of the previously defined distribution. In **scenario IV** $\mu_0 = \{0.050, 0.200, 0.350\}$, and $\mu_1 = -0.200$. In **scenario V** performance under the null hypothesis was evaluated by setting $\mu_{yx} = 0.00$.

To evaluate violation of the InSIDE assumption, the following parameter values were used in **scenario VI**: $\mu_z \sim unif(0, \varphi)$, with $\varphi = \{0.10, 0.25, 0.50\}$, $\tau_j \sim unif(-0.15, 0)$, $\mu_{yx} = -0.10$, $\mu_0 = 0.200$, and $\mu_1 = -0.200$. Between study heterogeneity was simulated by $\mu_{yx,j} \sim N(-0.10, \sigma_{yx}^2)$ with $\sigma_{yx} = \{0.025, 0.03, 0.05\}$, $\mu_z = 0$, and $\mu_{y,j} \sim N(0.46, 0.025^2)$

(**scenario VII**). To simulate time varying treatments, let $t = 1, \dots, 5$, $\mu_{yx,t} = -0.10/5$, $y_j \sim bernoulli(\mu_{y,j} + \sum_{t=1}^5 \mu_{yx,t} x_{j,t} + \tau z_j + \mu_{yu} u_j)$, where $x_{j,t}$ is generated by applying $x_j \sim bernoulli(p)$ five times and $\mu_{y,j} \sim N(0.35, 0.025^2)$ in **scenario VIII**. To evaluate power, scenario II was repeated using $J = 40$ **scenario IX**, $J = 80$ **scenario X**, and $J = 20$ with $n_j \sim unif(400, 10\ 000)$ in **scenario XI**. Up until here the between study difference in

adherence was kept relatively modest (appendix Table 1), in **scenario XII** we allow ω to vary between 1 and 0.30 by setting $\mu_1/\mu_0 = -0.50/0.50$ and increasing the prevalence of **U** from 0.40 to 0.70.

We note that in scenario VII we introduced a between study difference on an additive scale. To properly model this we re-analyse results from this scenario using an additive random effects model as described by Thompson and Sharp(17). Furthermore, the multiple treatment effects in scenario VIII were meta-analysed together resulting in an overall effect of treatment on the outcome and an overall effect of treatment allocation on adherence.

All simulations were repeated 5 000 times and were performed using the statistical package R version 3.1.2 for Unix(18). The number of replications was chosen to ensure sufficient precision to detect small deviations from the nominal coverage rate of 0.95 (the 95% lower and upper bounds are 0.944 and 0.956, given a true coverage of 0.95)(19). Please contact the lead author for a copy of the simulation and analysis scripts.

Performance metrics

Performance of the treatment effect estimators (AT, ITT, IV, EIV) was evaluated based on the following metrics. Bias was defined as $\overline{RD} - \text{True RD}$, with \overline{RD} equal to the mean

estimated risk difference (\widehat{RD}) of the effect of treatment on the outcome; empirical SE (ESE), estimated by taking the standard deviation of the distribution of \widehat{RD} ; the mean SE; the root mean square error $RMSE = \sqrt{(RD - \text{True RD})^2 + ESE^2}$; the coverage rate, defined as the proportion of times the 95% confidence interval included the true RD; the proportion of rejected null-hypotheses (i.e., depending on whether μ_{yx} equals 0 the type 1 error or power; using an alpha of 0.05); and the number of models that did not converge. Given that estimation of a direct effect of treatment allocation (τ) is usually not of interest we evaluate the three ECA intercept tests only on ESE, mean SE, rejection rate, number of failed models, and the mean point estimate.

Results of the simulation study

Figure 4 depicts the bias and coverage in the first three simulation scenarios of a meta-analysis of 20 RCTs. The top row of Figure 4 depicts the performance of the treatment effect estimators in a meta-analysis of perfectly blinded RCTs under different levels of adherence. In this scenario, unless non-adherence was minimal, the AT estimator was a biased estimator of the effect of *taking* treatment. Even at low levels of non-adherence (e.g., range $\omega = 1, 0.96$) the ITT was a biased estimator of *taking* treatment. Instead both IV estimators provided an unbiased estimate of the effect of *taking* treatment. Coverage of both the ITT and the AT estimator was lower than 0.95 unless adherence was close to perfect, whereas the coverage of the IV and ECA estimators was 0.95 or higher. Precision

of the ECA method was, however, much lower than for the IV estimator (a minimal ESE of 0.004 for the IV estimator versus 0.093 for the ECA estimator). In this scenario the type 1 error rate of the Egger-CA intercept tests was 0.05 (Appendix Table 2).

The middle and bottom panels of Figure 4, depict performance when allocated treatment could not be blinded. In this setting all estimators were biased except for the ECA estimator, which had a maximum bias of 0.009 compared to a minimum bias of 0.132 for the other estimators. Note that in scenario II bias in the ITT was positive (minimum of 0.142), indicating the ITT effect overestimated the treatment effect. Coverage was zero for all estimators except for the ECA which was close to 0.95 in all scenarios, however precision of the ECA estimator was again low (minimal ESE was 0.098 for the ECA compared to for example, 0.004 for the ITT estimator in scenario II). Power for the ECA (Appendix Tables 3 and 4 “rejection rate”) was at most 0.169, which decreased with increasing adherence. The RMSE for the ECA estimate was typically higher than for the other estimator, except when adherence was at its lowest (i.e., range $\omega = 1, 0.84$), in which case RMSE 1.098 for the ECA versus a minimal RMSE of 0.132 for the AT estimator.

Power to detect a direct effect of treatment allocation (i.e., intercept test) was highest, but still low in an absolute sense (scenario II 0.396), at an adherence range of $\omega = 1, 0.84$ and

when using the ECA-NPB method. For the same scenario and adherence, power to detect a direct effect was 0.368 for the ECA-PB, and 0.351 for the ECA-OLS. Similar results as for scenarios II and III were found in Scenario IV when the non-adherence was the largest in the control arms instead of the intervention arm (Appendix Table 5).

Figure 5 depicts the results of scenario V where treatment did not have an effect on the outcome. Due to the fact that treatment allocation was not blinded the ITT, AT, and IV estimators were all severely biased (minimum bias -0.130), while the ECA estimator remained unbiased (maximum bias -0.003). Precision and coverage were similar as in previous scenarios. The type 1 error rates (Appendix Table 6) were 1.00 for the ITT, AT, and IV estimators, and 0.05 or smaller, for the ECA estimator. Power of to detect a direct effect was at most 0.488 for the ECA-NPB (Appendix Table 6).

Violating the InSIDE assumption (Figure 6, Scenario VI Appendix Table 7), did not impact bias of the ECA (which was at most -0.001), coverage, however, was reduced to as low as 0.208 for the ECA-PB or 0.536 for the ECA-NPB. Coverage of the ECA-OLS estimator was, however, never lower than 0.90. Bias of the other estimators was of the same magnitude as before.

Violating the homogeneity assumption (Scenario VII, Appendix Table 8A) markedly decreased coverage of all the ECA methods, with the ECA-OLS showing the best coverage which did not drop below 0.900. Bias was close to -0.10 for all ECA methods, however, bias for the other estimators was always larger (i.e., more negative) than this. Using an (additive) random effects model (Appendix Table 8B) improved coverage of both the ECA-OLS and ECA-PB estimators to values above 0.90. Coverage of the ECA-NPB however remained markedly lower than 0.95 and decreased as adherence decreased.

Similar results as in scenario II-IV were observed when adherence was measured five times, i.e., treatment compliance was allowed to change five times, and meta-analysed in an average treatment effect (Appendix Table 9).

Power was more comprehensively explored in scenarios IX, X, and XI, (see Appendix tables 10-12) showing that even in meta-analyses of 80 studies power to detect a treatment effect using the ECA estimators was around 0.55; and only when adherence ranged between 1 and 0.84 (Table 2). Increasing the average sample size from 2 700 to 5 200 subjects per study of a 20 study meta-analysis had similar impact as increasing the number of RCTs to 40 studies (with an average sample size of 2 700); both resulted in a maximum power of about 0.50. Power to detect a direct effect was more responsive to increments in the number of studies or sample size, increasing to 0.60 or higher for all

three scenarios when mean adherence ranged between 1 and 0.84, and as high as 0.900 when 80 studies were included in the meta-analysis. Finally, in scenario XII we let the adherence range from 1 to 0.30 and we observed a power of 0.88 or higher for both the ECA treatment effect estimator and the direct effect ECA test, Appendix Table 13.

Example - the effect of epidural anaesthesia in labouring women

To illustrate the Egger-CA method we considered a meta-analysis by Bannister-Tyrrel et al.(3) which included RCTs comparing epidural anesthesia to non-epidural anesthesia or no anesthesia among women in labor; the primary endpoint was caesarean section incidence. Due to the inclusion of a no anesthesia group some of these RCTs may have had difficulty blinding participants completely which in turn may bias treatment effect estimates. The meta-analysis included 9 RCTs, for which details on non-adherence and contamination were reported (Appendix Table 14).

Adherence ranged between 0.42 and 1.00 in this subset of RCTs, with the proportion of caesarean sections ranging between 0.04 and 0.16. Of the conventional effect estimates, only the as treated (AT) estimate (RD 0.104; 95%CI 0.087;0.121) showed an increased risk of caesarean due to epidural anaesthesia, while excluding a potential protective effect (Table 3). However, as described above, the AT estimate may be biased. The ECA direct effect estimate of 0.0004 seemed to indicate that treatment allocation did not affect the

outcome except through its effect on treatment adherence. Indeed, there was minimal difference between the ECA and IV estimates: Egger-CA estimate (RD 0.017, 95%CI -0.086; 0.121) versus the conventional IV estimate (RD 0.018, 95%CI -0.007; 0.044). However, the confidence intervals around the direct effect were wide (e.g., -0.064; 0.065). Finally, we note that these estimates were dominated by two large studies (see Appendix Figure). This empirical example together with an R analysis script is included in Appendices 2 and 3.

Discussion

When interest lies in the effects of actually *taking* treatment, estimates from unblinded RCTs may be biased because treatment allocation may affect the outcome independent of the effect of *taking* treatment. In this study we have extended a method used in genetics (5) to a meta-analysis setting of unblinded RCTs. Additionally, we have presented simulation results, showing that under various scenarios the Egger-CA method was unbiased and had superior performance compared to other more established estimators such as the as-treated (AT), intention to treat (ITT), and instrumental variable estimator (IV) without Egger adjustment. While bias and coverage of the Egger-CA estimator was appropriate, power was usually lower than 0.50, and precision was lacking throughout. Finally, we showed that if the Egger-CA assumptions are violated, bias was lower than for other estimators, and implementing the Egger-CA using the OLS variance estimator

markedly improved performance as compared to (non-) parametric bootstrap variance estimators.

While we focussed on unblinded RCTs, the discussed method can be more generally used as a sensitivity analysis for instrumental variable analyses. The Egger-CA method not only adjusts for the presence of a direct effect of treatment allocation on the outcome, but also includes an intercept test to detect such a direct effect. In RCTs it may be known beforehand if the instrument (i.e., treatment allocation) has a direct effect on the outcome or not (e.g., placebo controlled vs open label), in which case the intercept test is of less value. In other settings, such as post marketing surveillance(20), however this distinction may be less clear, and the intercept test can be a valuable test to explore the validity of an IV analysis.

Bowden introduced the Egger-CA method for a single study with multiple genetic instruments. Here we extend this to a meta-analysis of multiple studies (RCTs), each with a single instrument (treatment allocation). Given that most contemporary RCTs are implemented in multiple centres, the Egger-CA method could also be applied in a single study with multiple centres. Returning to the meta-analysis setting for the moment, if the included sample of RCTs contains both blinded and unblinded studies, we feel the best solution would be to perform a sensitivity analysis excluding unblinded RCTs. By focusing

on blinded RCTs one can apply the ITT and IV estimators which have more power than the Egger-CA estimator. Instead the Egger-CA estimator is most valuable when an intervention is impossible to allocate in a blinded fashion which potentially invalidates the other treatment effect estimators. We wish to highlight that both Egger-CA and conventional IV methods estimate the effect of taking treatment 100% of the time versus never taking the treatment. It should be noted however, that these methods can intuitively be used to estimate the effect of any percentage of adherence, which may be more relevant for clinical practice where 100% adherence can be unrealistic.

Throughout the Egger-CA method using the OLS variance estimator performed best, especially when the InSIDE assumption was violated. This is not surprising because, based on Bowden et al., the OLS estimator was implemented under a multiplicative random effects model, while (based on the same publication) the (non-) parametric bootstrap variance estimators were implemented under a fixed effect model. All three ECA variance estimators can be implemented using a fixed effect model, a multiplicative or additive (as done in simulation scenario VII) random effects model. The better performance of a multiplicative random effects model can be explained by noting that violations of the InSIDE assumptions are likely study specific and hence induce between study variance. Ignoring the between study variance and incorrectly using a fixed effect model will result in a lower than nominal (95%) coverage rate. Modelling this between

study variance (due to violations of the InSIDE assumption) using an additive random effects model will (correctly) inflate the coverage (and variance) but has the undesired side effect of also affecting (biasing) the point estimate. Instead, a multiplicative random effects model only influences the variance, not the point estimate, and is therefore more suitable in this setting. In applied settings it is often unknown whether between study variance is due to violations of the InSIDE assumption or due to other (biological) sources, with the latter preferably modelled using an additive random effects model(18). Therefore, we recommended to model any between study variance of the ECA estimate using both an additive and a multiplicative model. Regarding an additive random effects model we note the unexpected low coverage of the non-parametric bootstrap percentile confidence interval. Until this issue is resolved we advise applied researcher to preferentially use the OLS variance estimator.

In addition to the InSIDE assumption, Bowden et al, (21) recently showed that the ECA method is only unbiased when there is negligible sampling error in $\hat{\omega}_j$, or more specifically there is no measurement error. This assumption is similar to assuming the predictor variable in a regression is measured without error, hence, Bowden coined this the “NO Measurement Error” (NOME) assumption. In a two-sample genetic MR setting (when the correlation between θ_j and ω_j is zero) Bowden showed that violating the NOME assumption increased the type 1 error of the direct effect, while the treatment effect test

was unaffected. How this translates to the current one-sample settings remains to be explored. Particularly, this future work should focus on evaluating and correcting for the NOME assumption violation while accounting for the ω_j and θ_j correlation. Such research should also compare performance to the ITT, AT, and IV estimators included here. While the IV estimator makes the same NOME assumption as the ECA (albeit with more robustness), the AT and ITT similarly assume the exposure variable is measured without error. For the ITT, the exposure variable is treatment allocation and one would hope this no measurement error assumption is true in most cases. However, for the AT the exposure variable is treatment received, and depending on the treatment, measurement error *of the actual treatment received* (e.g., based on self-reported data) may be meaningfully large. Finally, in this simulation study we have not adjusted for the confounder (**U**). However, as noted before, likely at least part of the confounders will be observed, hence it may be interesting to further explore the behaviour of the ECA when adjusting for the observed subset of confounders. An important consideration for this future research would be the behaviour of the ECA when the number of observed confounders differs between studies. Further relevant parameters include the (multivariable) correlation between the observed and unobserved confounders.

Besides the Egger-CA method, several other methods have been suggested to adjust for the presence of a direct effect of an instrument on the outcome. Recently, Bowden et

al.(22) proposed a weighted median method, which performs better than the Egger-CA method if a fraction of the instruments are correct. A similar principle was applied by Kang et al.(23), implemented using a penalization method and assuming at least 50% of the instrument were correct. Under the same InSIDE assumption Kolesár introduced the “modified-biased-corrected-two-stage-least squares” estimator which unbiasedly estimates the effect of treatment in a single study with multiple instruments(24).

The simulations presented here are naturally limited and the following points merit discussion. First, we evaluated the ITT estimator on its ability to estimate the effect of taking treatment, however, it is well known that the ITT effect only equals the effect of *taking* treatment when adherence is 100%. Nevertheless, we evaluated the ITT effect as such to underline how different the ITT effect can be from the effect of actually *taking* treatment. We also exemplified a pervasive false believe that the ITT effect always underestimates the effect of *taking* treatment, which has been described before(25). Furthermore, as shown analytically and emphasized by our simulation, the ITT effect in an unblinded RCT is no longer ensured to be independent of covariables and consequently the second term (the direct effect) in equation 1A does not necessarily equal zero. A second limitation is that we focussed on the risk difference scale only, and ignored relative effect estimators such as the risk ratio or the odd ratio. The rational for this is given by Didelez who showed that these relative effect estimators additionally assume that

treatment is normally distributed conditional on the instrumental variable and confounders(11). Given that potential confounding variables are typically unobserved this assumption is difficult to evaluate in empirical data, hence we decided against using these effect estimators. Furthermore, relative effect measures are often seen as less informative for determining an individual's harm or benefit from an intervention(26). Additionally, the marginal and conditional effects do not generally agree when using non-collapsible effect estimators such as the odds ratio and hazard ratio(27). Our interest here is to estimate the unconditional (marginal) treatment effect, hence simulating a conditional effect that does not equal the marginal effect will complicate matters. Third, in scenario VIII we allowed treatment adherence to differ at 5 time points and meta-analysed the effect of treatment allocation on treatment adherence; implicitly assuming the same effect across time on treatment adherence. This equal effect across time is likely unrealistic in empirical settings, where it seems plausible that the effect of treatment allocation on treatment adherence changes over time (e.g., decreases with time). Time-varying effects can be estimated, within study, using for example, mixed effect models(28). Fourth, in most simulations the minimum adherence (Appendix Table 1) was typically not smaller than 0.70 and sometimes even limited to the range 1, 0.96, which may not reflect potential greater variability found in empirical data (see e.g., the motivating example). Because perfect adherence does not result in unbiased treatment effect estimates in unblinded RCTs, we felt it important to also evaluate performance when variation in adherence was minimal.

However, if in empirical settings variation in adherence is larger, higher power and precision may be expected, for example similar to scenario XII. Even though perfect adherence does not guarantee an absence of bias, we named this procedure “Egger Correction for non-Adherence” reflecting that some degree of non-adherence is necessary for this method to provide estimates. Fifth, the risk difference parameters used in the simulations were admittedly large, and perhaps larger than what would be found in most empirical studies. These parameters were nevertheless chosen to emphasize difference between estimators. Finally, while applying a random effects models in the presence of between study variance has desirable statistical properties, we wish to echo Greenland’s comment(29) in the sense that this heterogeneity in treatment effects should be explored not ignored. If, however, this heterogeneity cannot be explored, e.g., because of a lack of individual participant data, using a random effects method is preferred over using a fixed effect method. In the presence of unexplained heterogeneity the effect estimate can no longer be interpreted as the “average causal effect”, instead, in the absence of “defiers” (i.e., subjects that always defy allocated treatment), the “local average treatment effect”(11) can be identified.

In conclusion we introduced a novel effect estimator, Egger-CA, which unbiasedly estimates the effect of taking treatment in meta-analyses of unblinded RCTs. In this setting the as treated, intention to treat, or the conventional instrumental variable effect estimators

are all potentially biased. Unless variation in adherence is large, power and precision is expected to be low, hence we recommend using the Egger-CA as a sensitivity analyses, not as a replacement for the more traditional effect estimators.

Conflict of interest statement

None of the authors of this paper has a financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.

Author contributions

AFS and RHHG contributed to the idea and design of the study. AFS performed the analyses and drafted the manuscript. RHHG provided guidance during initial planning of the paper and during critical revision. AFS had full access to all of the data and takes responsibility for the integrity of the data presented.

Funding

AFS is funded by UCLH NIHR Biomedical Research Centre (BRC10200) and is a UCL Springboard Population Health Sciences Fellow.

Prior postings and presentations

This study and its results have not been published previously. An abstract based on this work was presented at the 2016 ISCB conference and has been accepted for the 2016 Cochrane colloquium.

References

1. Grobbee DE, Hoes AW. *Intervention Research: Unintended Effects*. Vol. 2nd. Burlington: Jones and Bartlett Learning; 2015. 181-214 p.
2. Higgins JPT, Altman DG, Sterne JAC. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. 2011. p.
3. Bannister-Tyrrell M, Miladinovic B, Roberts CL, Ford JB. Adjustment for compliance behavior in trials of epidural analgesia in labor using instrumental variable meta-analysis. *J Clin Epidemiol*. 2015;68(5):525–33.
4. Hingorani A, Humphries S. Nature's randomised trials. *Lancet*. 2005;366(9501):1906–8.
5. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512–25.
6. Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Vol. 315, *BMJ: British Medical Journal*. 1997. p. 629–34.
7. Schmidt AF, Groenwold RHH, Knol MJ, Hoes AW, Nielen M, Roes KCB, et al. Exploring interaction effects in small samples increases rates of false-positive and false-negative findings: Results from a systematic review and simulation study. *Journal of Clinical Epidemiology* 2014 p. 821–9.
8. Schmidt AF, Klungel OH, Nielen M, de Boer A, Groenwold RHH, Hoes AW. Tailoring treatments using treatment effect modification. *Pharmacoepidemiology and Drug Safety* 2016 p. 355–62.
9. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006;17(4):360–72.
10. Palmer TM, Sterne JA, Harbord RM, Lawlor DA, Sheehan NA, Meng S, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *AmJEpidemiol*. 2011;173(12):1392–403.
11. Didelez V, Meng S, Sheehan NA. Assumptions of IV Methods for Observational Epidemiology. *Stat Sci*. 2010;25(1).
12. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29(4):722–9.
13. Schmidt AF, Hingorani AD, Jefferis BJ, White J, Groenwold RHH, Dudbridge F. Comparison of variance estimators for meta-analysis of instrumental variable estimates. *Int J Epidemiol*. 2016;
14. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559–73.
15. Rücker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics*.

- 2011;12(1):122–42.
16. Carpenter J, Bithell J. Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Stat Med.* 2000;19(9):1141–64.
 17. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: A comparison of methods. Vol. 18, *Statistics in Medicine*. 1999. p. 2693–708.
 18. RCore Team. R. A language and environment for statistical computing. [Internet]. R Foundation for Statistical Computing, Vienna, Austria. Vienna, Austria: R Foundation for Statistical Computing; 2013. Available from: <http://www.r-project.org/>
 19. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med.* 2006;25(24):4279–92.
 20. Schmidt AF, Klungel OH, Groenwold RH. Adjusting for Confounding in Early Postlaunch Settings: Going Beyond Logistic Regression Models. Vol. 27, *Epidemiology*. p. 133–42.
 21. Bowden J, Del Greco M. F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *Int J Epidemiol* . 2016 Sep;
 22. Bowden J, Davey-Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol.* 2016;40(4).
 23. Kang H, Zhang A, Cai TT, Small DS. Instrumental Variables Estimation With Some Invalid Instruments and its Application to Mendelian Randomization [Internet]. *Journal of the American Statistical Association*. Taylor & Francis; p. 0. Available from: <http://dx.doi.org/10.1080/01621459.2014.994705>
 24. Kolesár M, Chetty R, Friedman J, Glaeser E, Imbens GW. Identification and Inference With Many Invalid Instruments. *J Bus Econ Stat.* 2015;33(4):474–84.
 25. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials.* 2012;9(1):48–55.
 26. Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat Med.* 1998;17(8):857–72.
 27. Burgess S. Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. Vol. 32, *Stat.Med.* p. 4726–47.
 28. Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using Eigen and S4 [R Package] [Internet]. 2015. p. Available from: <https://github.com/lme4/lme4>
 29. Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol.* 1994;140(3):290–6.

Figure captions

Figure 1 Directed acyclic graphs of (A) a blinded and (B) an unblinded randomized controlled trial.

[Figure 1 here]

Nb. Node Z represents treatment allocation, node X treatment, node U unmeasured confounders of the X to Y relation, and Y the outcome.

Figure 2 Graphical representation of the Egger test, and the IV and the Egger-CA estimates.

[Figure 2 here]

Nb. Dots represent study specific estimates. Left graph, the dashed line indicates the standardized ITT effect “corrected” for small study heterogeneity (i.e., results from an Egger analysis). Right graph, the dashed dotted line represents the IV effect corrected for a possible direct effect of allocation on the outcome (i.e., results from an Egger-CA analysis), the long dashed line represents the IV estimate uncorrected for this possible direct effect.

Figure 3 Directed acyclic graph of an unblinded randomized controlled trial with treatment allocation Z affecting confounder U.

[Figure 3 here]

Nb. Node Z represents treatment allocation, node X treatment, node U unmeasured confounders of the X to Y relation, and Y the outcome. μ_z is the effect of treatment allocation Z on U, ω_j represents the effect of Z on treatment (X), τ is the direct effect of Z on Y (the outcome), μ_{yx} represents the treatment effect on the outcome and μ_{yu} the U effect on Y.

Figure 4 Bias and coverage of different treatment effect estimators in a meta-analysis of randomized controlled trials (simulation scenarios I, II, and III).

[Figure 4 here]

Nb. AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. In the bar plot the dark grey bar depicts coverage under an adherence range of 1,0.96, the lighter grey for adherence range of 1,0.90, and the lightest grey bar an adherence range of 1,0.84. Simulations

were repeated 5,000 times. In scenario I treatment allocation was blinded, in scenarios II and III treatment allocation was unblinded and allocation had a direct effect on the outcome.

Figure 5 Performance of different treatment effect estimators where treatment has no effect on the outcome in a meta-analysis of randomized controlled trials (simulation scenario V).

[Figure 5 here]

Nb. AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. In the bar plot the dark grey bar depicts coverage under an adherence range of 1,0.90, the lighter grey for adherence range of 1,0.84, and the lightest grey bar an adherence range of 1,0.78. Simulations were repeated 5,000 times.

Figure 6 Performance of different treatment effect estimators when the InSIDE assumption is violated in a meta-analysis of randomized controlled trials (simulation scenario VII).

[Figure 6 here]

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. In the bar plot the dark grey bar depicts coverage under a violation of the InSIDE assumption of 0.05, the lighter grey 0.125, and the lightest grey 0.25. Simulations were repeated 5,000 times.

Appendix Table 1 adherence range (i.e., ω).

	Sub-scenario 1	Sub-scenario 2	Sub-scenario 3
Scenario I	1, 0.96	1, 0.90	1, 0.84
Scenario II	1, 0.96	1, 0.90	1, 0.84
Scenario III	1, 0.96	1, 0.90	1, 0.84
Scenario IV	1, 0.90	1, 0.84	1, 0.78
Scenario V	1, 0.90	1, 0.84	1, 0.78
Scenario VI	1, 0.86	1, 0.83	1, 0.78
Scenario VII	1, 0.84	1, 0.84	1, 0.84
Scenario VIII	1, 0.84	NA	NA
Scenario IX	1, 0.96	1, 0.90	1, 0.84
Scenario X	1, 0.96	1, 0.90	1, 0.84
Scenario XI	1, 0.96	1, 0.90	1, 0.84
Scenario XII	1, 0.30	NA	NA

Appendix Table 2 simulation results for scenario I assessing performance of different treatment effect estimators in a meta-analysis of blinded RCTs.

	Treatment effect				Direct effect		
Adherence range (ω)	1, 0.96	1, 0.90	1, 0.84		1, 0.96	1, 0.90	1, 0.84
Risk difference							
AT	0.100	0.096	0.093				
ITT	0.098	0.095	0.092				
IV	0.100	0.100	0.100				
ECA-OLS	0.087	0.094	0.098		0.013	0.006	0.002
ECA-PB	0.087	0.094	0.098		0.013	0.006	0.002
ECA-NPB	0.087	0.094	0.098		0.013	0.006	0.002
Bias							
AT	0.000	-0.004	-0.007				
ITT	-0.002	-0.005	-0.008				
IV	0.000	0.000	0.000				
ECA-OLS	-0.013	-0.006	-0.002				
ECA-PB	-0.013	-0.006	-0.002				
ECA-NPB	-0.013	-0.006	-0.002				
Coverage							
AT	0.946	0.823	0.489				
ITT	0.910	0.723	0.397				
IV	0.946	0.951	0.954				
ECA-OLS	0.960	0.956	0.962				
ECA-PB	0.952	0.945	0.950				
ECA-NPB	0.979	0.957	0.957				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.046	0.078	0.169		0.039	0.044	0.037
ECA-PB	0.053	0.093	0.174		0.049	0.054	0.052
ECA-NPB	0.028	0.083	0.186		0.022	0.044	0.043
Mean SE							
AT	0.004	0.004	0.004				
ITT	0.004	0.004	0.004				
IV	0.004	0.004	0.004				
ECA-OLS	0.460	0.167	0.099		0.450	0.159	0.091
ECA-PB	0.452	0.165	0.097		0.443	0.157	0.089
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.004	0.004	0.004				
ITT	0.004	0.004	0.004				
IV	0.004	0.004	0.004				
ECA-OLS	0.442	0.161	0.093		0.433	0.153	0.085
ECA-PB	0.442	0.161	0.093		0.433	0.153	0.085
ECA-NPB	0.442	0.161	0.093		0.433	0.153	0.085
RMSE							
AT	0.004	0.005	0.008				
ITT	0.004	0.006	0.009				
IV	0.004	0.004	0.004				
ECA-OLS	0.442	0.161	0.093				
ECA-PB	0.442	0.161	0.093				
ECA-NPB	0.442	0.161	0.093				
Number of failed models							
AT	0	0	0				
ITT	0	0	0				
IV	0	0	0				
ECA-OLS	0	0	0		0	0	0
ECA-PB	0	0	0		0	0	0
ECA-NPB	0	0	0		0	0	0

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 3 simulation results for scenario II assessing performance of different treatment effect estimators in a meta-analysis of unblinded RCTs, where treatment allocation has a direct effect on the outcome ($\tau = 0.15$).

	Treatment effect				Direct effect		
Adherence range (ω)	1, 0.96	1, 0.90	1, 0.84		1, 0.96	1, 0.90	1, 0.84
Risk difference							
AT	0.247	0.239	0.232				
ITT	0.248	0.245	0.242				
IV	0.253	0.258	0.263				
ECA-OLS	0.109	0.098	0.101		0.142	0.152	0.149
ECA-PB	0.109	0.098	0.101		0.142	0.152	0.149
ECA-NPB	0.109	0.098	0.101		0.142	0.152	0.149
Bias							
AT	0.147	0.139	0.132				
ITT	0.148	0.145	0.142				
IV	0.153	0.158	0.163				
ECA-OLS	0.009	-0.002	-0.001				
ECA-PB	0.009	-0.002	-0.001				
ECA-NPB	0.009	-0.002	-0.001				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.959	0.960	0.962				
ECA-PB	0.950	0.948	0.955				
ECA-NPB	0.977	0.961	0.959				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.044	0.074	0.152		0.055	0.145	0.351
ECA-PB	0.056	0.097	0.165		0.062	0.162	0.368
ECA-NPB	0.028	0.079	0.169		0.034	0.160	0.396
Mean SE							
AT	0.004	0.004	0.004				
ITT	0.004	0.004	0.004				
IV	0.004	0.004	0.004				
ECA-OLS	0.485	0.177	0.104		0.475	0.168	0.096
ECA-PB	0.477	0.174	0.103		0.467	0.165	0.095
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.004	0.004	0.005				
ITT	0.004	0.004	0.004				
IV	0.004	0.004	0.005				
ECA-OLS	0.467	0.169	0.098		0.457	0.161	0.090
ECA-PB	0.467	0.169	0.098		0.457	0.161	0.090
ECA-NPB	0.467	0.169	0.098		0.457	0.161	0.090
RMSE							
AT	0.147	0.139	0.132				
ITT	0.148	0.145	0.142				
IV	0.153	0.158	0.163				
ECA-OLS	0.467	0.169	0.098				
ECA-PB	0.467	0.169	0.098				
ECA-NPB	0.467	0.169	0.098				
Number of failed models							
AT	0	0	0				
ITT	0	0	0				
IV	0	0	0				
ECA-OLS	0	0	0		0	0	0
ECA-PB	0	0	0		0	0	0
ECA-NPB	0	0	0		0	0	0

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 4 simulation results for scenario III assessing performance of different treatment effect estimators in a meta-analysis of unblinded RCTs, where treatment allocation has a direct effect on the outcome ($\tau = -0.15$).

	Treatment effect				Direct effect		
Adherence range (ω)	1, 0.96	1, 0.90	1, 0.84		1, 0.96	1, 0.90	1, 0.84
Risk difference							
AT	-0.047	-0.046	-0.046				
ITT	-0.052	-0.055	-0.058				
IV	-0.053	-0.058	-0.063				
ECA-OLS	0.091	0.095	0.099		-0.141	-0.145	-0.149
ECA-PB	0.091	0.095	0.099		-0.141	-0.145	-0.149
ECA-NPB	0.091	0.095	0.099		-0.141	-0.145	-0.149
Bias							
AT	-0.147	-0.146	-0.146				
ITT	-0.152	-0.155	-0.158				
IV	-0.153	-0.158	-0.163				
ECA-OLS	-0.009	-0.005	-0.001				
ECA-PB	-0.009	-0.005	-0.001				
ECA-NPB	-0.009	-0.005	-0.001				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.955	0.952	0.958				
ECA-PB	0.949	0.945	0.947				
ECA-NPB	0.977	0.955	0.954				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.048	0.076	0.138		0.051	0.123	0.307
ECA-PB	0.056	0.083	0.154		0.061	0.133	0.332
ECA-NPB	0.027	0.075	0.153		0.032	0.125	0.342
Mean SE							
AT	0.004	0.004	0.004				
ITT	0.004	0.004	0.004				
IV	0.004	0.004	0.005				
ECA-OLS	0.525	0.191	0.112		0.514	0.181	0.103
ECA-PB	0.512	0.188	0.110		0.502	0.179	0.102
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.004	0.004	0.004				
ITT	0.004	0.004	0.004				
IV	0.004	0.005	0.005				
ECA-OLS	0.507	0.185	0.107		0.497	0.176	0.099
ECA-PB	0.507	0.185	0.107		0.497	0.176	0.099
ECA-NPB	0.507	0.185	0.107		0.497	0.176	0.099
RMSE							
AT	0.147	0.146	0.146				
ITT	0.152	0.155	0.158				
IV	0.153	0.158	0.163				
ECA-OLS	0.507	0.185	0.107				
ECA-PB	0.507	0.185	0.107				
ECA-NPB	0.507	0.185	0.107				
Number of failed models							
AT	0	0	0				
ITT	0	0	0				
IV	0	0	0				
ECA-OLS	0	0	0		0	0	0
ECA-PB	0	0	0		0	0	0
ECA-NPB	0	0	0		0	0	0

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 5 simulation results for scenario IV assessing performance of different treatment effect estimators in a meta-analysis of unblinded RCTs, where treatment allocation has a direct effect on the outcome ($\tau = -0.15$).

	Treatment effect				Direct effect		
Adherence range (ω)	1, 0.90	1, 0.84	1, 0.78		1, 0.90	1, 0.84	1, 0.78
Risk difference							
AT	-0.046	-0.038	-0.030				
ITT	-0.055	-0.058	-0.061				
IV	-0.058	-0.063	-0.068				
ECA-OLS	0.096	0.102	0.101		-0.147	-0.152	-0.151
ECA-PB	0.096	0.102	0.101		-0.147	-0.152	-0.151
ECA-NPB	0.096	0.102	0.101		-0.147	-0.152	-0.151
Bias							
AT	-0.146	-0.138	-0.130				
ITT	-0.155	-0.158	-0.161				
IV	-0.158	-0.163	-0.168				
ECA-OLS	-0.004	0.002	0.001				
ECA-PB	-0.004	0.002	0.001				
ECA-NPB	-0.004	0.002	0.001				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.962	0.960	0.958				
ECA-PB	0.950	0.947	0.949				
ECA-NPB	0.962	0.960	0.956				
Rejection rate							
AT	1.000	1.000	0.986				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.067	0.116	0.169		0.121	0.229	0.413
ECA-PB	0.084	0.123	0.185		0.139	0.234	0.427
ECA-NPB	0.069	0.127	0.196		0.128	0.248	0.455
Mean SE							
AT	0.004	0.004	0.004				
ITT	0.004	0.004	0.004				
IV	0.004	0.005	0.005				
ECA-OLS	0.191	0.141	0.100		0.181	0.129	0.089
ECA-PB	0.189	0.139	0.099		0.179	0.128	0.088
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.004	0.004	0.005				
ITT	0.004	0.004	0.004				
IV	0.005	0.005	0.005				
ECA-OLS	0.181	0.138	0.095		0.172	0.127	0.084
ECA-PB	0.181	0.138	0.095		0.172	0.127	0.084
ECA-NPB	0.181	0.138	0.095		0.172	0.127	0.084
RMSE							
AT	0.146	0.138	0.130				
ITT	0.155	0.158	0.161				
IV	0.158	0.163	0.168				
ECA-OLS	0.181	0.138	0.095				
ECA-PB	0.181	0.138	0.095				
ECA-NPB	0.181	0.138	0.095				
Number of failed models							
AT	0	0	0				
ITT	0	0	0				
IV	0	0	0				
ECA-OLS	0	0	0		0	0	0
ECA-PB	0	0	0		0	0	0
ECA-NPB	0	0	0		0	0	0

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 6 simulation results for scenario V assessing performance of different treatment effect estimators in a meta-analysis of unblinded RCTs, where treatment has no effect on the outcome ($\mu_{yx} = 0.00$).

	Treatment effect				Direct effect		
Adherence range (ω)	<u>1, 0.90</u>	<u>1, 0.84</u>	<u>1, 0.78</u>		<u>1, 0.90</u>	<u>1, 0.84</u>	<u>1, 0.78</u>
Risk difference							
AT	-0.146	-0.138	-0.130				
ITT	-0.150	-0.150	-0.150				
IV	-0.158	-0.163	-0.168				
ECA-OLS	-0.003	-0.001	-0.001		-0.147	-0.149	-0.149
ECA-PB	-0.003	-0.001	-0.001		-0.147	-0.149	-0.149
ECA-NPB	-0.003	-0.001	-0.001		-0.147	-0.149	-0.149
Bias							
AT	-0.146	-0.138	-0.123				
ITT	-0.150	-0.150	-0.150				
IV	-0.158	-0.163	-0.168				
ECA-OLS	-0.003	-0.001	0.001				
ECA-PB	-0.003	-0.001	0.001				
ECA-NPB	-0.003	-0.001	0.001				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.955	0.955	0.962				
ECA-PB	0.951	0.946	0.949				
ECA-NPB	0.955	0.956	0.960				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.045	0.045	0.038		0.135	0.238	0.449
ECA-PB	0.049	0.054	0.051		0.151	0.259	0.460
ECA-NPB	0.045	0.044	0.040		0.143	0.262	0.488
Mean SE							
AT	0.004	0.004	0.004				
ITT	0.004	0.004	0.004				
IV	0.004	0.004	0.004				
ECA-OLS	0.181	0.134	0.094		0.172	0.123	0.084
ECA-PB	0.179	0.132	0.094		0.170	0.122	0.083
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.004	0.004	0.005				
ITT	0.004	0.004	0.004				
IV	0.004	0.004	0.005				
ECA-OLS	0.175	0.131	0.090		0.166	0.121	0.080
ECA-PB	0.175	0.131	0.090		0.166	0.121	0.080
ECA-NPB	0.175	0.131	0.090		0.166	0.121	0.080
RMSE							
AT	0.147	0.138	0.131				
ITT	0.150	0.150	0.150				
IV	0.158	0.163	0.168				
ECA-OLS	0.175	0.131	0.090				
ECA-PB	0.175	0.131	0.090				
ECA-NPB	0.175	0.131	0.090				
Number of failed models							
AT	0	0	0				
ITT	0	0	0				
IV	0	0	0				
ECA-OLS	0	0	0		0	0	0
ECA-PB	0	0	0		0	0	0
ECA-NPB	0	0	0		0	0	0

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 7 simulation results for scenario VI assessing performance of different treatment effect estimators in a meta-analysis of unblinded RCTs, where treatment allocation has an effect on the outcome and on a confounder.

	Treatment effect				Direct effect		
Average effect of Z on confounder U (μ_z)	0.05	0.125	0.25		0.05	0.125	0.25
Risk difference							
AT	-0.171	-0.170	-0.168				
ITT	-0.169	-0.167	-0.165				
IV	-0.185	-0.184	-0.184				
ECA-OLS	-0.100	-0.100	-0.101		-0.077	-0.076	-0.074
ECA-PB	-0.100	-0.100	-0.101		-0.077	-0.076	-0.074
ECA-NPB	-0.100	-0.100	-0.101		-0.077	-0.076	-0.074
Bias							
AT	-0.071	-0.070	-0.068				
ITT	-0.069	-0.067	-0.065				
IV	-0.085	-0.084	-0.084				
ECA-OLS	0.000	-0.000	-0.001				
ECA-PB	0.000	-0.000	-0.001				
ECA-NPB	0.000	-0.000	-0.001				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.903	0.914	0.908				
ECA-PB	0.208	0.226	0.223				
ECA-NPB	0.557	0.551	0.536				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.113	0.104	0.113		0.106	0.101	0.112
ECA-PB	0.790	0.790	0.799		0.792	0.783	0.774
ECA-NPB	0.106	0.465	0.489		0.452	0.467	0.467
Mean SE							
AT	0.004	0.004	0.004				
ITT	0.004	0.004	0.004				
IV	0.004	0.004	0.004				
ECA-OLS	0.309	0.281	0.239		0.283	0.255	0.215
ECA-PB	0.052	0.047	0.039		0.048	0.042	0.035
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.011	0.011	0.011				
ITT	0.012	0.012	0.012				
IV	0.013	0.013	0.013				
ECA-OLS	0.347	0.312	0.266		0.318	0.283	0.239
ECA-PB	0.347	0.312	0.266		0.318	0.284	0.239
ECA-NPB	0.347	0.312	0.266		0.318	0.284	0.239
RMSE							
AT	0.072	0.071	0.069				
ITT	0.070	0.068	0.066				
IV	0.086	0.085	0.085				
ECA-OLS	0.347	0.312	0.266				
ECA-PB	0.347	0.312	0.266				
ECA-NPB	0.347	0.312	0.266				
Number of failed models							
AT	0	0	0				
ITT	0	0	0				
IV	0	0	0				
ECA-OLS	0	0	0		0	0	0
ECA-PB	0	0	0		0	0	0
ECA-NPB	0	0	0		0	0	0

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 8A simulation results for scenario VII assessing performance of different treatment effect estimators in a fixed effect meta-analysis of unblinded RCTs.

	Treatment effect				Direct effect		
Between study standard deviation in μ_{yx}	<u>0.025</u>	<u>0.030</u>	<u>0.050</u>		<u>0.025</u>	<u>0.030</u>	<u>0.050</u>
Risk difference							
AT	-0.239	-0.239	-0.241				
ITT	-0.243	-0.243	-0.244				
IV	-0.264	-0.264	-0.266				
ECA-OLS	-0.104	-0.102	-0.111		-0.147	-0.149	-0.142
ECA-PB	-0.104	-0.102	-0.111		-0.147	-0.149	-0.142
ECA-NPB	-0.104	-0.102	-0.111		-0.147	-0.149	-0.142
Bias							
AT	-0.139	-0.139	-0.141				
ITT	-0.143	-0.143	-0.144				
IV	-0.164	-0.164	-0.166				
ECA-OLS	-0.004	-0.002	-0.011				
ECA-PB	-0.004	-0.002	-0.011				
ECA-NPB	-0.004	-0.002	-0.011				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.919	0.914	0.904				
ECA-PB	0.551	0.455	0.219				
ECA-NPB	0.783	0.721	0.572				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.116	0.115	0.107		0.160	0.150	0.115
ECA-PB	0.487	0.590	0.780		0.554	0.626	0.786
ECA-NPB	0.266	0.315	0.460		0.344	0.377	0.468
Mean SE							
AT	0.004	0.004	0.004				
ITT	0.004	0.004	0.004				
IV	0.005	0.004	0.004				
ECA-OLS	0.206	0.232	0.346		0.190	0.213	0.318
ECA-PB	0.088	0.079	0.058		0.081	0.073	0.053
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.008	0.009	0.014				
ITT	0.007	0.008	0.012				
IV	0.008	0.009	0.013				
ECA-OLS	0.225	0.257	0.387		0.207	0.236	0.356
ECA-PB	0.225	0.257	0.387		0.207	0.236	0.356
ECA-NPB	0.225	0.257	0.387		0.207	0.236	0.356
RMSE							
AT	0.139	0.139	0.141				
ITT	0.143	0.143	0.144				
IV	0.164	0.164	0.166				
ECA-OLS	0.225	0.257	0.387				
ECA-PB	0.225	0.257	0.387				
ECA-NPB	0.225	0.257	0.387				
Number of failed models							
AT	0	0	9				
ITT	0	0	9				
IV	0	0	9				
ECA-OLS	0	0	9		0	0	9
ECA-PB	0	0	9		0	0	9
ECA-NPB	0	0	9		0	0	9

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 8B simulation results for scenario VII assessing performance of different treatment effect estimators in a random effects meta-analysis of unblinded RCTs.

	Treatment effect				Direct effect		
Between study standard deviation in μ_{yx}	<u>0.025</u>	<u>0.030</u>	<u>0.050</u>		<u>0.025</u>	<u>0.030</u>	<u>0.050</u>
Risk difference							
AT	-0.239	-0.238	-0.239				
ITT	-0.242	-0.242	-0.242				
IV	-0.264	-0.264	-0.264				
ECA-OLS	-0.099	-0.102	-0.099		-0.151	-0.148	-0.151
ECA-PB	-0.099	-0.102	-0.099		-0.151	-0.148	-0.151
ECA-NPB	-0.099	-0.102	-0.099		-0.151	-0.148	-0.151
Bias							
AT	-0.139	-0.138	-0.139				
ITT	-0.142	-0.142	-0.142				
IV	-0.164	-0.164	-0.164				
ECA-OLS	-0.001	-0.002	0.001				
ECA-PB	-0.001	-0.002	0.001				
ECA-NPB	-0.001	-0.002	0.001				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.933	0.923	0.933				
ECA-PB	0.936	0.925	0.934				
ECA-NPB	0.834	0.791	0.649				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.099	0.098	0.079		0.144	0.125	0.097
ECA-PB	0.095	0.092	0.076		0.143	0.125	0.097
ECA-NPB	0.221	0.253	0.361		0.309	0.323	0.397
Mean SE							
AT	0.007	0.008	0.012				
ITT	0.007	0.008	0.011				
IV	0.007	0.008	0.012				
ECA-OLS	0.211	0.237	0.349		0.194	0.218	0.322
ECA-PB	0.213	0.239	0.351		0.196	0.220	0.323
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.007	0.008	0.012				
ITT	0.007	0.008	0.011				
IV	0.008	0.009	0.013				
ECA-OLS	0.220	0.248	0.363		0.202	0.229	0.334
ECA-PB	0.220	0.248	0.363		0.202	0.229	0.334
ECA-NPB	0.220	0.248	0.363		0.202	0.229	0.334
RMSE							
AT	0.139	0.139	0.139				
ITT	0.143	0.143	0.143				
IV	0.164	0.164	0.164				
ECA-OLS	0.220	0.249	0.363				
ECA-PB	0.220	0.249	0.363				
ECA-NPB	0.220	0.249	0.363				
Number of failed models							
AT	0	0	12				
ITT	0	0	12				
IV	0	0	12				
ECA-OLS	0	0	12		0	0	12
ECA-PB	0	0	12		0	0	12
ECA-NPB	0	0	12		0	0	12

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by the residual standard error; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 9 simulation results for scenario VIII assessing performance of different treatment effect estimators in a fixed effect meta-analysis of unblinded RCTs, with 5 measurements of treatment adherence.

	Risk difference	Bias	Coverage	Rejection rate	Mean SE	ESE	RMSE	Number of failed models
Treatment effect								
AT	-0.227	-0.127	0.000	1.000	0.002	0.004	0.127	5
ITT	-0.242	-0.142	0.000	1.000	0.004	0.004	0.142	5
IV	-0.263	-0.163	0.000	1.000	0.002	0.004	0.163	5
ECA-OLS	-0.101	-0.001	0.955	0.114	0.127	0.122	0.122	5
ECA-PB	-0.101	-0.001	0.952	0.138	0.125	0.122	0.122	5
ECA-NPB	-0.101	-0.001	0.948	0.138	NA	0.122	0.122	5
Direct effect								
ECA-OLS	-0.149			0.256	0.117	0.113		5
ECA-PB	-0.149			0.274	0.115	0.113		5
ECA-NPB	-0.149			0.293	NA	0.113		5

Adherence range

(ω): 1, 0.84

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 10 simulation results for scenario IX assessing performance of different treatment effect estimators in a meta-analysis of unblinded RCTs, including 40 studies instead of 20.

	Treatment effect				Direct effect		
Adherence range (ω)	1, 0.96	1, 0.90	1, 0.84		1, 0.96	1, 0.90	1, 0.84
Risk difference							
AT	0.247	0.239	0.232				
ITT	0.248	0.245	0.242				
IV	0.253	0.258	0.263				
ECA-OLS	0.106	0.098	0.098		0.144	0.152	0.152
ECA-PB	0.106	0.098	0.098		0.144	0.152	0.152
ECA-NPB	0.106	0.098	0.098		0.144	0.152	0.152
Bias							
AT	0.147	0.139	0.132				
ITT	0.148	0.145	0.142				
IV	0.153	0.158	0.163				
ECA-OLS	0.006	-0.002	-0.002				
ECA-PB	0.006	-0.002	-0.002				
ECA-NPB	0.006	-0.002	-0.002				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.958	0.953	0.958				
ECA-PB	0.950	0.946	0.951				
ECA-NPB	0.977	0.956	0.957				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.054	0.121	0.283		0.072	0.264	0.662
ECA-PB	0.067	0.138	0.300		0.083	0.285	0.665
ECA-NPB	0.033	0.127	0.296		0.050	0.293	0.699
Mean SE							
AT	0.003	0.003	0.003				
ITT	0.003	0.003	0.003				
IV	0.003	0.003	0.003				
ECA-OLS	0.325	0.119	0.070		0.319	0.113	0.065
ECA-PB	0.315	0.116	0.069		0.309	0.110	0.064
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.003	0.003	0.003				
ITT	0.003	0.003	0.003				
IV	0.003	0.003	0.003				
ECA-OLS	0.315	0.115	0.067		0.309	0.110	0.062
ECA-PB	0.315	0.115	0.067		0.309	0.110	0.062
ECA-NPB	0.315	0.115	0.067		0.309	0.110	0.062
RMSE							
AT	0.147	0.139	0.132				
ITT	0.148	0.145	0.142				
IV	0.153	0.158	0.163				
ECA-OLS	0.315	0.115	0.068				
ECA-PB	0.315	0.115	0.068				
ECA-NPB	0.315	0.115	0.068				
Number of failed models							
AT	0	0	0				
ITT	0	0	0				
IV	0	0	0				
ECA-OLS	0	0	0		0	0	0
ECA-PB	0	0	0		0	0	0
ECA-NPB	0	0	0		0	0	0

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 11 simulation results for scenario X assessing performance of different treatment effect estimators in a meta-analysis of unblinded RCTs, including 80 studies instead of 20.

	Treatment effect				Direct effect		
Adherence range (ω)	1, 0.96	1, 0.90	1, 0.84		1, 0.96	1, 0.90	1, 0.84
Risk difference							
AT	0.247	0.239	0.217				
ITT	0.248	0.245	0.242				
IV	0.253	0.258	0.263				
ECA-OLS	0.102	0.100	0.098		0.149	0.150	0.152
ECA-PB	0.102	0.100	0.098		0.149	0.150	0.152
ECA-NPB	0.102	0.100	0.098		0.149	0.150	0.152
Bias							
AT	0.147	0.139	0.132				
ITT	0.148	0.145	0.142				
IV	0.153	0.158	0.163				
ECA-OLS	0.002	0.000	-0.002				
ECA-PB	0.002	0.000	-0.002				
ECA-NPB	0.002	0.000	-0.002				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.957	0.958	0.957				
ECA-PB	0.946	0.949	0.949				
ECA-NPB	0.975	0.961	0.956				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.071	0.228	0.532		0.101	0.484	0.929
ECA-PB	0.084	0.239	0.548		0.115	0.505	0.934
ECA-NPB	0.046	0.228	0.541		0.080	0.520	0.946
Mean SE							
AT	0.002	0.002	0.002				
ITT	0.002	0.002	0.002				
IV	0.002	0.002	0.002				
ECA-OLS	0.223	0.082	0.048		0.219	0.078	0.045
ECA-PB	0.216	0.080	0.047		0.212	0.076	0.044
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.002	0.002	0.002				
ITT	0.002	0.002	0.002				
IV	0.002	0.002	0.002				
ECA-OLS	0.220	0.079	0.047		0.215	0.075	0.043
ECA-PB	0.220	0.079	0.047		0.215	0.075	0.043
ECA-NPB	0.220	0.079	0.047		0.215	0.075	0.043
RMSE							
AT	0.147	0.139	0.132				
ITT	0.148	0.145	0.142				
IV	0.153	0.158	0.163				
ECA-OLS	0.220	0.079	0.047				
ECA-PB	0.220	0.079	0.047				
ECA-NPB	0.220	0.079	0.047				
Number of failed models							
AT	0	0	0				
ITT	0	0	0				
IV	0	0	0				
ECA-OLS	0	0	0		0	0	0
ECA-PB	0	0	0		0	0	0
ECA-NPB	0	0	0		0	0	0

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 12 simulation results for scenario XI assessing performance of different treatment effect estimators in a meta-analysis of unblinded RCTs, including 20 studies with an expected sample size of 5,200 subjects instead of 2,700.

	Treatment effect				Direct effect		
Adherence range (ω)	<u>1, 0.96</u>	<u>1, 0.90</u>	<u>1, 0.84</u>		<u>1, 0.96</u>	<u>1, 0.90</u>	<u>1, 0.84</u>
Risk difference							
AT	0.247	0.239	0.232				
ITT	0.248	0.245	0.242				
IV	0.253	0.258	0.263				
ECA-OLS	0.088	0.097	0.099		0.162	0.153	0.150
ECA-PB	0.088	0.097	0.099		0.162	0.153	0.150
ECA-NPB	0.088	0.097	0.099		0.162	0.153	0.150
Bias							
AT	0.147	0.139	0.132				
ITT	0.148	0.145	0.142				
IV	0.153	0.158	0.163				
ECA-OLS	-0.012	-0.003	-0.001				
ECA-PB	-0.012	-0.003	-0.001				
ECA-NPB	-0.012	-0.003	-0.001				
Coverage							
AT	0.000	0.000	0.000				
ITT	0.000	0.000	0.000				
IV	0.000	0.000	0.000				
ECA-OLS	0.960	0.960	0.956				
ECA-PB	0.951	0.952	0.952				
ECA-NPB	0.971	0.958	0.950				
Rejection rate							
AT	1.000	1.000	1.000				
ITT	1.000	1.000	1.000				
IV	1.000	1.000	1.000				
ECA-OLS	0.049	0.107	0.266		0.062	0.250	0.600
ECA-PB	0.054	0.119	0.280		0.074	0.263	0.600
ECA-NPB	0.039	0.123	0.298		0.055	0.278	0.642
Mean SE							
AT	0.003	0.003	0.003				
ITT	0.003	0.003	0.003				
IV	0.003	0.003	0.003				
ECA-OLS	0.365	0.130	0.076		0.358	0.123	0.070
ECA-PB	0.364	0.128	0.076		0.357	0.122	0.069
ECA-NPB	NA	NA	NA		NA	NA	NA
ESE							
AT	0.003	0.003	0.004				
ITT	0.003	0.003	0.003				
IV	0.003	0.003	0.004				
ECA-OLS	0.350	0.124	0.072		0.343	0.118	0.067
ECA-PB	0.350	0.124	0.072		0.343	0.118	0.067
ECA-NPB	0.350	0.124	0.072		0.343	0.118	0.067
RMSE							
AT	0.147	0.139	0.132				
ITT	0.148	0.145	0.142				
IV	0.153	0.158	0.163				
ECA-OLS	0.351	0.124	0.072				
ECA-PB	0.351	0.124	0.072				
ECA-NPB	0.351	0.124	0.072				
Number of failed models							
AT	0	0	0				
ITT	0	0	0				
IV	0	0	0				
ECA-OLS	0	0	0		0	0	0
ECA-PB	0	0	0		0	0	0
ECA-NPB	0	0	0		0	0	0

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 13 simulation results for scenario XII assessing performance of different treatment effect estimators in a fixed effect meta-analysis of unblinded RCTs, including 20 studies with an expected sample size of 2,700 subjects, with a higher variation in adherence.

	Risk difference	Bias	Coverage	Rejection rate	Mean SE	ESE	RMSE	Number of failed models
Treatment effect								
AT	0.200	0.100	0.000	1.000	0.004	0.007	0.100	0
ITT	0.215	0.115	0.000	1.000	0.004	0.005	0.115	0
IV	0.320	0.220	0.000	1.000	0.006	0.013	0.221	0
ECA-OLS	0.100	0.000	0.959	0.881	0.032	0.030	0.030	0
ECA-PB	0.100	0.000	0.948	0.879	0.031	0.030	0.030	0
ECA-NPB	0.100	0.000	0.948	0.907	NA	0.030	0.030	0
Direct effect								
ECA-OLS	0.150			0.999	0.021	0.020		0
ECA-PB	0.150			1.000	0.021	0.020		0
ECA-NPB	0.150			1.000	NA	0.020		0
Adherence range								
<i>E(ω): 1, 0.30</i>								

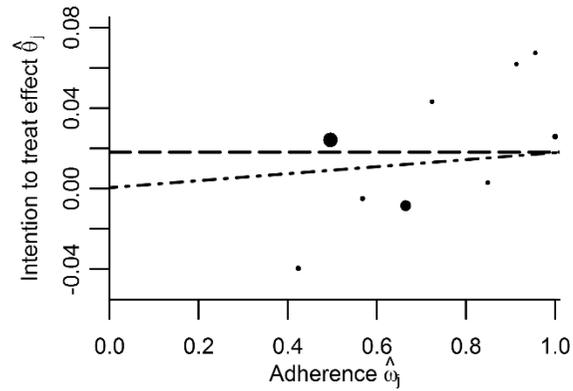
AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

Appendix Table 14 Adherence and caesarian section incidence in randomized controlled trials of epidural anesthesia compared to non-epidural anesthesia or no anesthesia (control group) during labor.

Author	Allocated epidural			Allocated control		
	Received epidural	Did not receive epidural	Caesarian sections	Received epidural	Did not receive epidural	Caesarian sections
Bofill 1997 (1)	47	2	5	12	39	3
Clark 1998 (2)	147	9	15	84	78	22
Halpern 2004 (3)	124	0	12	51	67	12
Head 2002 (4)	53	3	10	2	58	7
Jain 2003 (5)	43	2	9	0	83	11
Nafisi 2006 (6)	197	0	24	0	198	19
Ramin 1995 (7)	432	232	41	103	563	25
Sharma 1997 (8)	243	115	13	5	352	16
Volmanen 2008 (9)	24	1	1	3	24	1

Based on data presented by Bannister-Tyrrell et.al. (10)

Figure meta-analysis of epidural anesthesia compared to non-epidural anesthesia or no anesthesia (control group) during labor on caesarian section incidence.



Nb. Dots (scaled by precision) represent study specific estimates. The dashed dotted line represents the IV effect corrected for a possible direct effect of allocation on the outcome (i.e., results from an Egger-CA analysis), the long dashed line represents the IV estimate uncorrected for this possible direct effect.

Reference List

- (1) Bofill JA, Vincent RD, Ross EL, Martin RW, Norman PF, Werhan CF, et al. Nulliparous active labor, epidural analgesia, and cesarean delivery for dystocia. *Am J Obstet Gynecol* 1997 Dec;177(6):1465-70.
- (2) Clark A, Carr D, Loyd G, Cook V, Spinnato J. The influence of epidural analgesia on cesarean delivery rates: a randomized, prospective clinical trial. *Am J Obstet Gynecol* 1998 Dec;179(6 Pt 1):1527-33.
- (3) Halpern SH, Muir H, Breen TW, Campbell DC, Barrett J, Liston R, et al. A multicenter randomized controlled trial comparing patient-controlled epidural with intravenous analgesia for pain relief in labor. *Anesth Analg* 2004 Nov;99(5):1532-8.
- (4) Head BB, Owen J, Vincent RD, Jr., Shih G, Chestnut DH, Hauth JC. A randomized trial of intrapartum analgesia in women with severe preeclampsia. *Obstet Gynecol* 2002 Mar;99(3):452-7.
- (5) Jain S, Arya VK, Gopalan S, Jain V. Analgesic efficacy of intramuscular opioids versus epidural analgesia in labor. *Int J Gynaecol Obstet* 2003 Oct;83(1):19-27.
- (6) Nafisi S. Effects of epidural lidocaine analgesia on labor and delivery: a randomized, prospective, controlled trial. *BMC Anesthesiol* 2006;6:15.
- (7) Ramin SM, Gambling DR, Lucas MJ, Sharma SK, Sidawi JE, Leveno KJ. Randomized trial of epidural versus intravenous analgesia during labor. *Obstet Gynecol* 1995 Nov;86(5):783-9.
- (8) Sharma SK, Sidawi JE, Ramin SM, Lucas MJ, Leveno KJ, Cunningham FG. Cesarean delivery: a randomized trial of epidural versus patient-controlled meperidine analgesia during labor. *Anesthesiology* 1997 Sep;87(3):487-94.
- (9) Volmanen P, Sarvela J, Akural EI, Raudaskoski T, Korttila K, Alahuhta S. Intravenous remifentanyl vs. epidural levobupivacaine with fentanyl for pain relief in early labour: a randomised, controlled, double-blinded study. *Acta Anaesthesiol Scand* 2008 Feb;52(2):249-55.
- (10) Bannister-Tyrrell M, Miladinovic B, Roberts CL, Ford JB. Adjustment for compliance behavior in trials of epidural analgesia in labor using instrumental variable meta-analysis. *J Clin Epidemiol* 2015 May;68(5):525-33.

Table 1 Simulation scenarios on a meta-analysis of randomized controlled trials with different levels of non-adherence, and a potential direct effect of treatment allocation on a binary outcome*.

Parameters	Scenario I	Scenario II	Scenario III	Scenario IV	Scenario V	Scenario VI	Scenario VII	Scenario VIII
Number of studies J	20	20	20	20	20	20	20	20
$E(n_j)$ sample size	2 700	2 700	2 700	2 700	2 700	2 700	2 700	2 700
Effect of Z on outcome Y (τ)	0.00	0.15	-0.15	-0.15	-0.15	$\sim unif(-0.15, 0)$	-0.15	-0.15
Effect of Z on confounder U (μ_z)	0.00	0.00	0.00	0.00	0.00	$\sim unif(0, \varphi)$ $\varphi = \{0.10, 0.25, 0.50\}$	0.00	0.00
Mean effect of X on outcome Y (μ_{yx})	0.10	0.10	0.10	0.10	0.00	-0.10	-0.10	5 * -0.02
Between study standard deviation in μ_{yx}	NA	NA	NA	NA	NA	NA	{0.025, 0.030, 0.050}	NA
$Prob(X = 1 Z = 1)$ (min/max)	{0.95/1.00, 0.80/1.00, 0.65/1.00}	{0.95/1.00, 0.80/1.00, 0.65/1.00}	{0.95/1.00, 0.80/1.00, 0.65/1.00}	0.80/1.00	0.80/1.00	0.80/1.00	0.80/1.00	0.80/1.00
$Prob(X = 0 Z = 1)$ (min/max)	0.00/0.05	0.00/0.05	0.00/0.05	{0.00/0.05, 0.00/0.20, 0.00/0.35}	{0.00/0.05, 0.00/0.20, 0.00/0.35}	0.00/0.20	0.00/0.20	0.00/0.20
Non-adherence coefficient per treatment arm (μ_1/μ_0)	{-0.050, -0.200, -0.350}/0.050	{-0.050, -0.200, -0.350}/0.050	{-0.050, -0.200, -0.350}/0.050	-0.200/{0.050, 0.200, 0.350}	-0.200/{0.050, 0.200, 0.350}	-0.200/0.200	-0.200/0.200	-0.200/0.200
Effect of U on outcome Y (μ_{yu})	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Mean baseline risk of Y ($\mu_{y,j}$)	0.15	0.15	0.35	0.35	0.35	0.35	0.46	0.35
Between study standard deviation in baseline risk.	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025

* Changes from the previous scenario (on the left) are presented in bold. NA = not applicable.

Table 2 Empirical power of different effect estimators of the effect of treatment on an outcome, and effect estimators of the direct effect of randomization on an outcome.

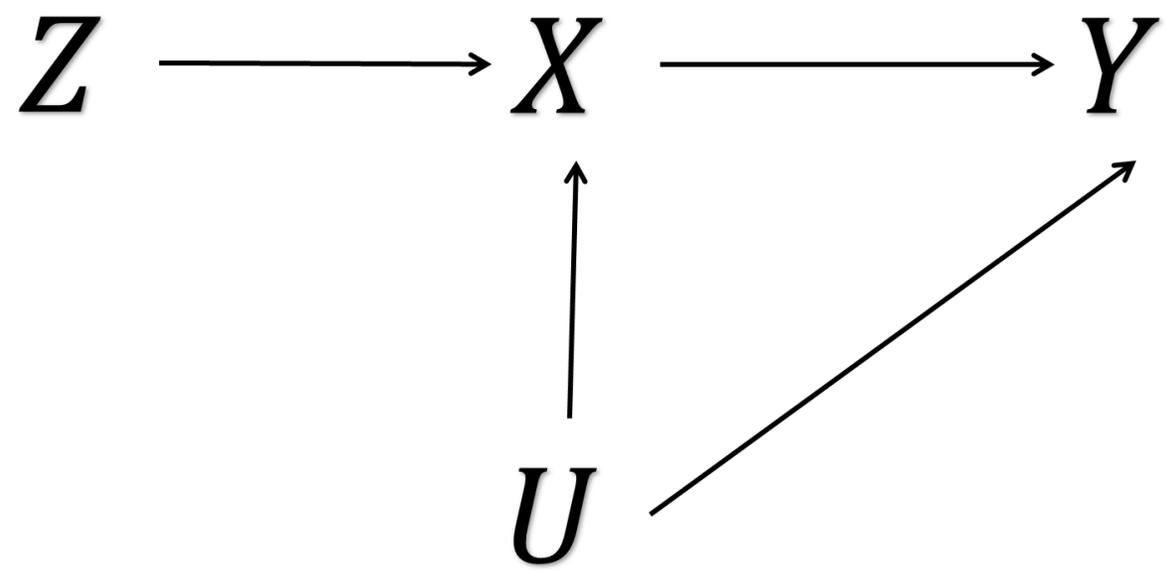
	Scenario IX $k = 40; E(n_j) = 2\,700$	Scenario X $k = 80; E(n_j) = 2\,700$	Scenario XI $k = 20; E(n_j) = 5\,200$
<u>Treatment effect</u>			
Adherence range (ω): 1, 0.96			
AT	1.000	1.000	1.000
ITT	1.000	1.000	1.000
IV	1.000	1.000	1.000
ECA-OLS	0.054	0.071	0.049
ECA-PB	0.067	0.084	0.054
ECA-NPB	0.033	0.046	0.039
Adherence range (ω): 1, 0.90			
AT	1.000	1.000	1.000
ITT	1.000	1.000	1.000
IV	1.000	1.000	1.000
ECA-OLS	0.121	0.228	0.107
ECA-PB	0.138	0.239	0.119
ECA-NPB	0.127	0.228	0.123
Adherence range (ω): 1, 0.84			
AT	1.000	1.000	1.000
ITT	1.000	1.000	1.000
IV	1.000	1.000	1.000
ECA-OLS	0.283	0.532	0.266
ECA-PB	0.300	0.548	0.280
ECA-NPB	0.296	0.541	0.298
<u>Direct effect</u>			
Adherence range (ω): 1, 0.96			
ECA-OLS	0.072	0.101	0.062
ECA-PB	0.083	0.115	0.074
ECA-NPB	0.050	0.080	0.055
Adherence range (ω): 1, 0.90			
ECA-OLS	0.264	0.484	0.250
ECA-PB	0.285	0.505	0.263
ECA-NPB	0.293	0.520	0.278
Adherence range (ω): 1, 0.84			
ECA-OLS	0.662	0.929	0.600
ECA-PB	0.665	0.934	0.600
ECA-NPB	0.699	0.946	0.642

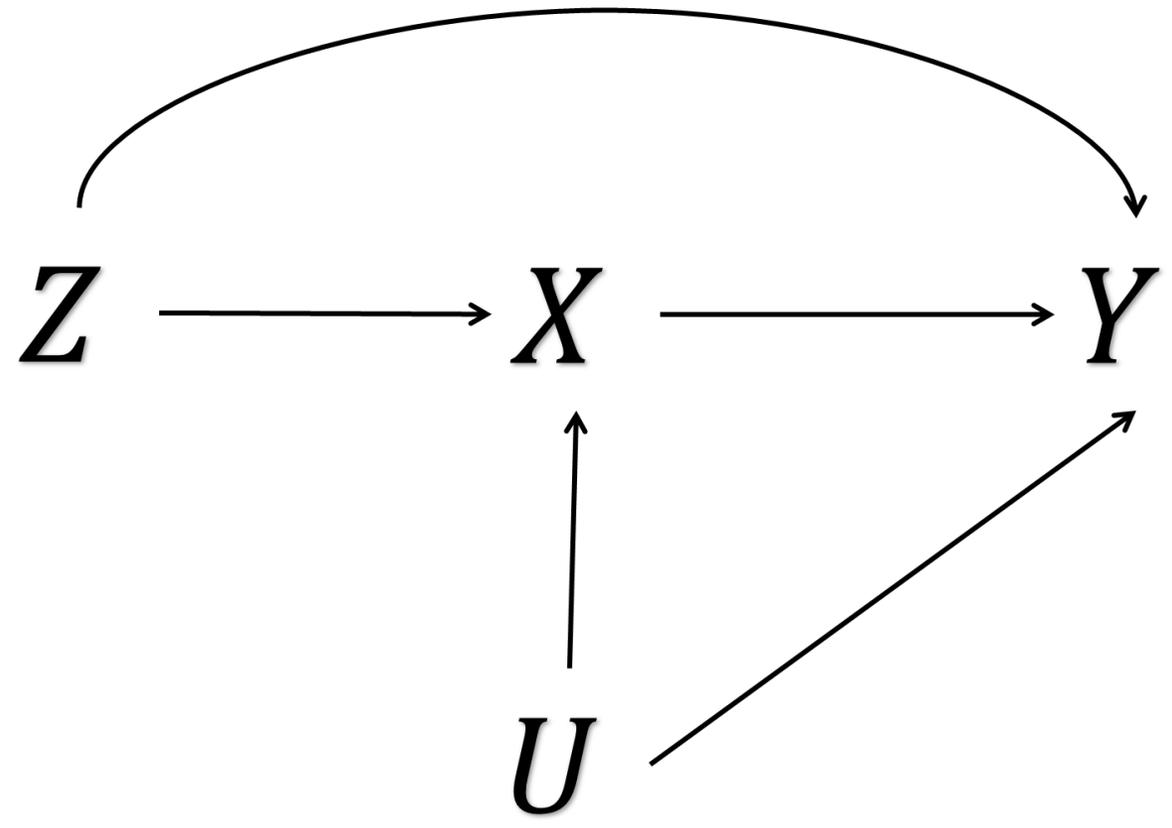
AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error. Simulations were repeated 5 000 times.

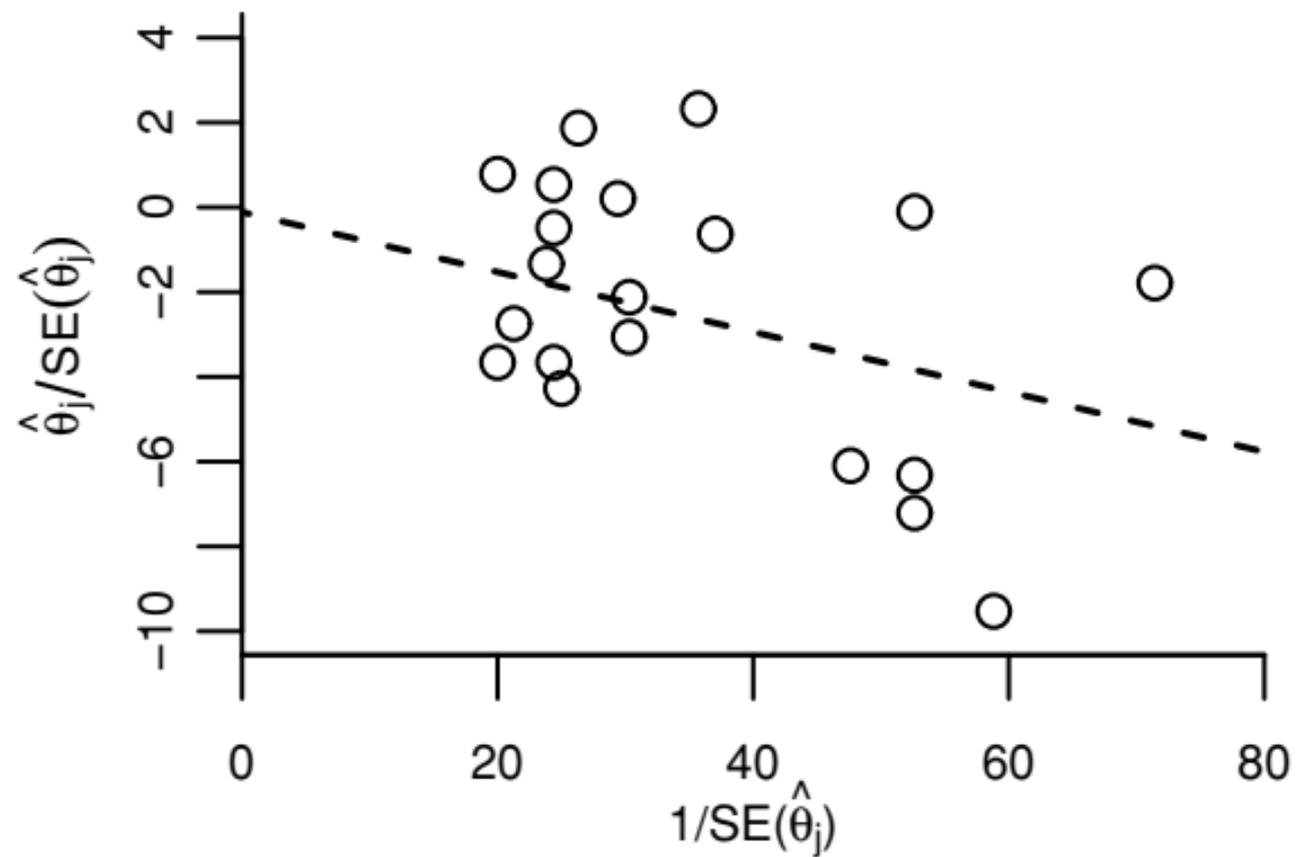
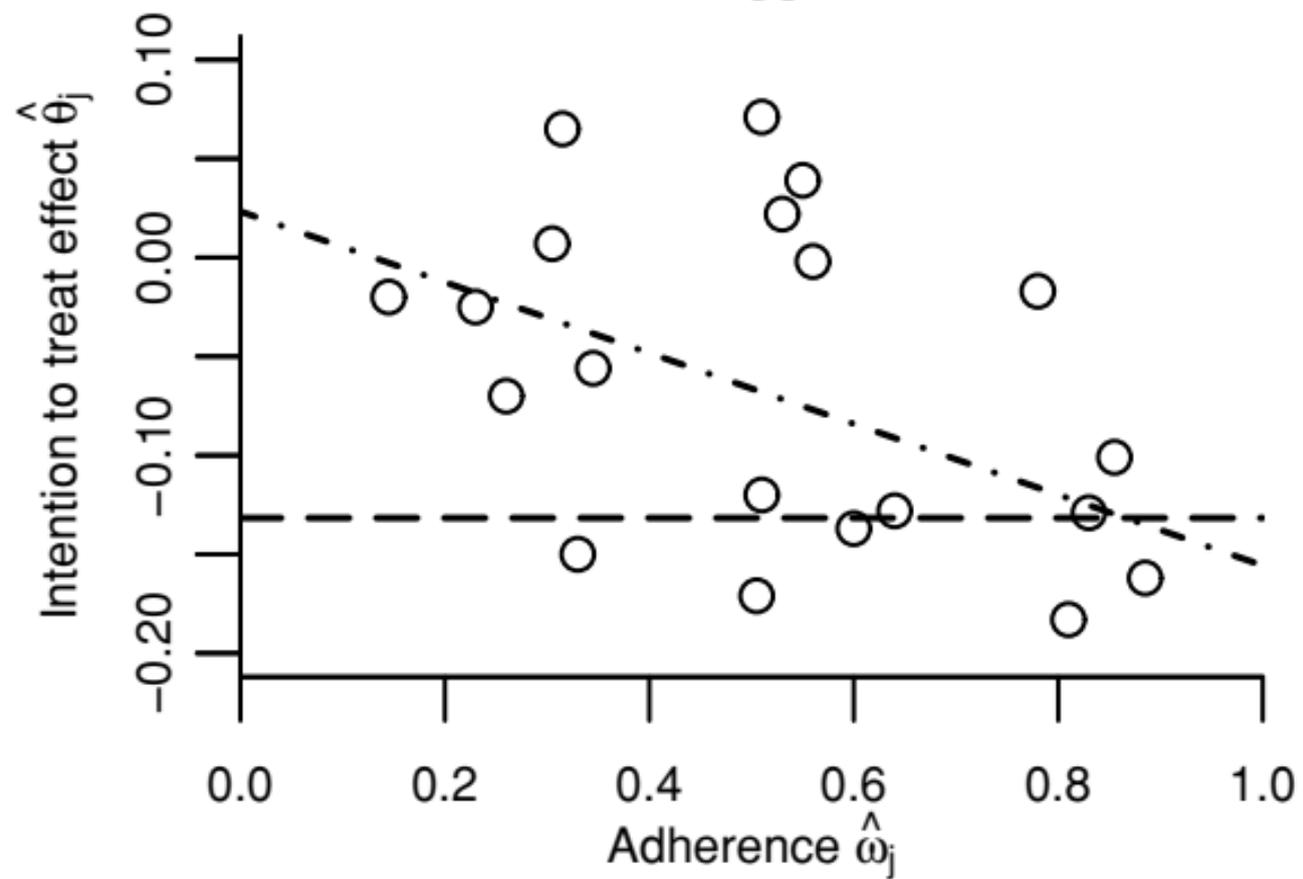
Table 3 Treatment and direct effect risk difference estimates (95%CI) of epidural anesthesia compared to non-epidural anesthesia or no anesthesia (control group) during labor on caesarian section incidence in a meta-analysis of 9 randomized controlled trials.

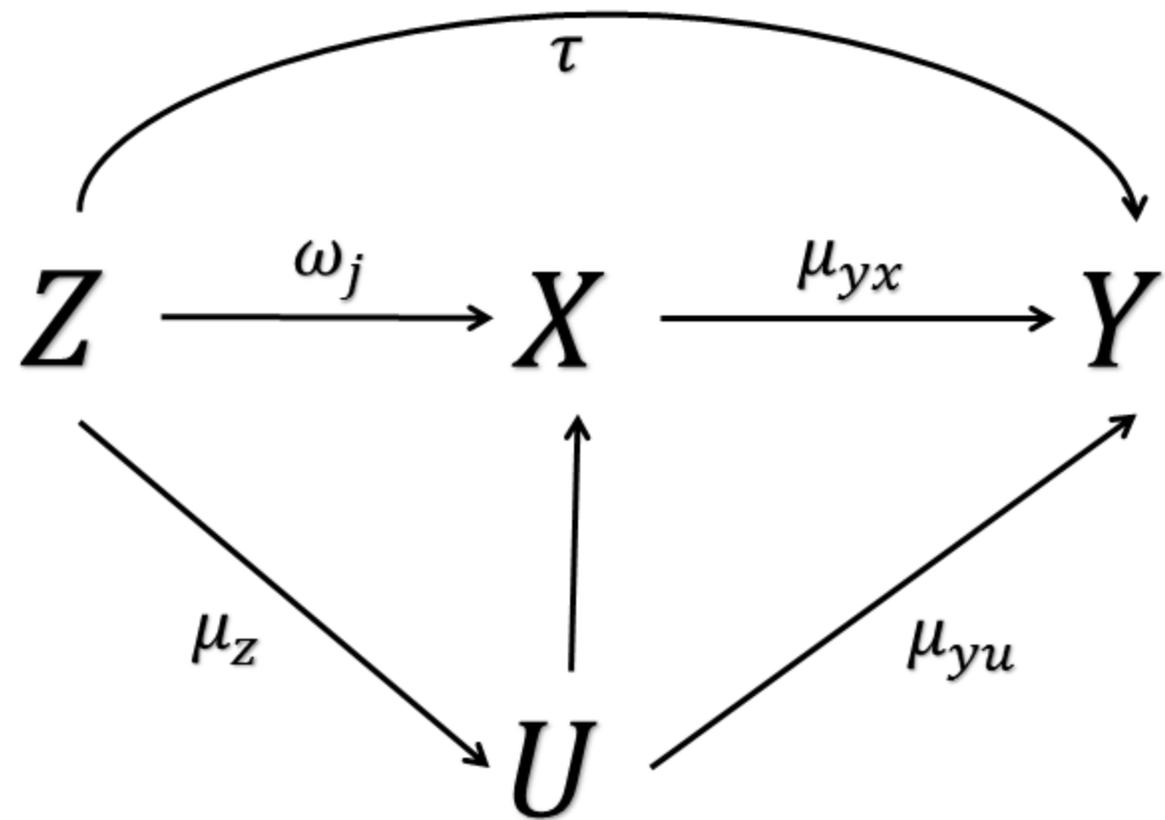
	<u>Treatment effect</u>	<u>Direct effect</u>
adherence range(ω): 1.0, 0.42		
AT	0.104(0.087;0.121)	
ITT	0.011(-0.005;0.027)	
IV	0.018(-0.007;0.044)	
ECA-OLS	0.017(-0.086;0.121)	0.0004(-0.064;0.065)
ECA-PB	0.017(-0.088;0.123)	0.0004(-0.066;0.067)
ECA-NPB	0.017(-0.080;0.123)	0.0004(-0.066;0.061)

AT, the As Treated effect estimator; ITT, the Intention to Treat effect estimator; IV, the Instrumental Variable effect estimator; ECA, Egger Correction for non-Adherence estimator. The ECA estimator was implemented using 3 standard error estimators: the Ordinary Least Square (OLS) variance estimator, estimated by dividing the standard errors of the slope coefficient by $\min(1, \text{"residual standard error"})$; the parametric bootstrap estimator (PB); the nonparametric bootstrap (NPB) estimator which was used to estimate the percentile confidence interval instead of the standard error.

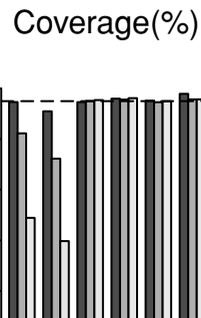
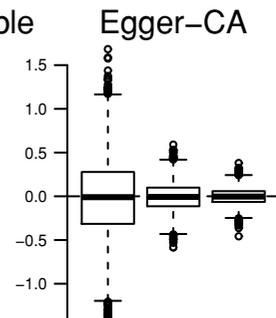
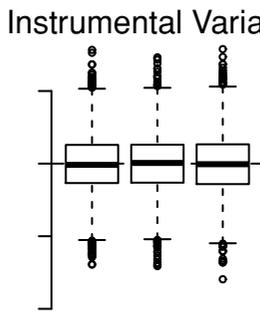
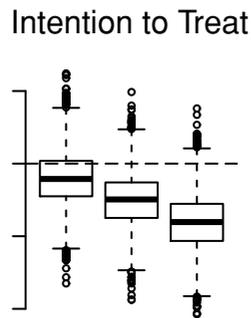
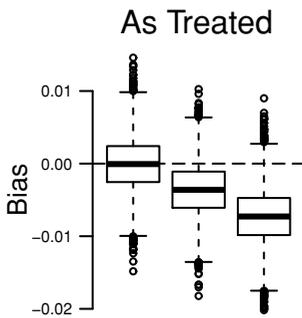




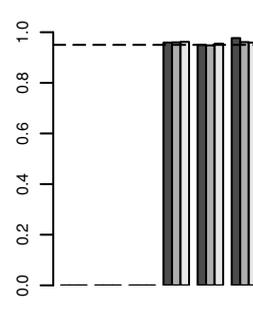
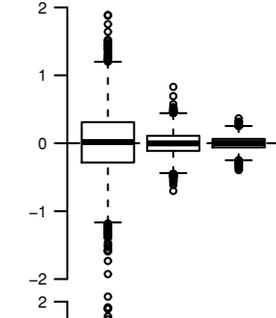
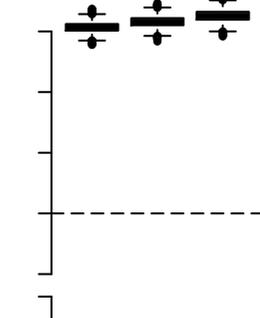
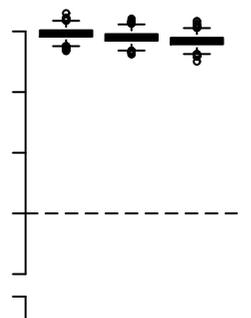
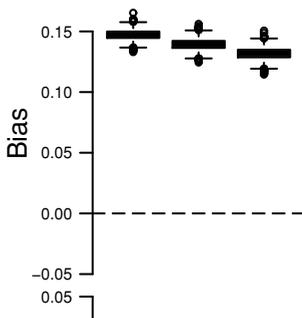
Egger test**IV and Egger-CA**



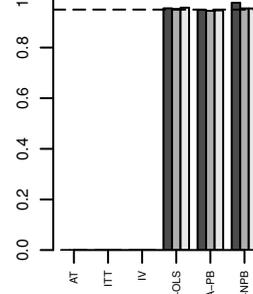
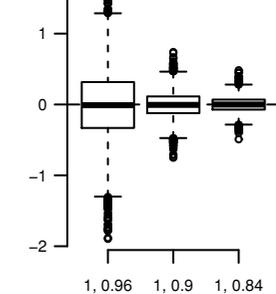
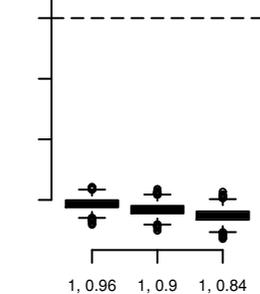
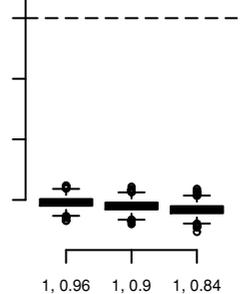
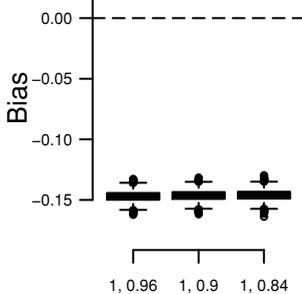
Scenario 1



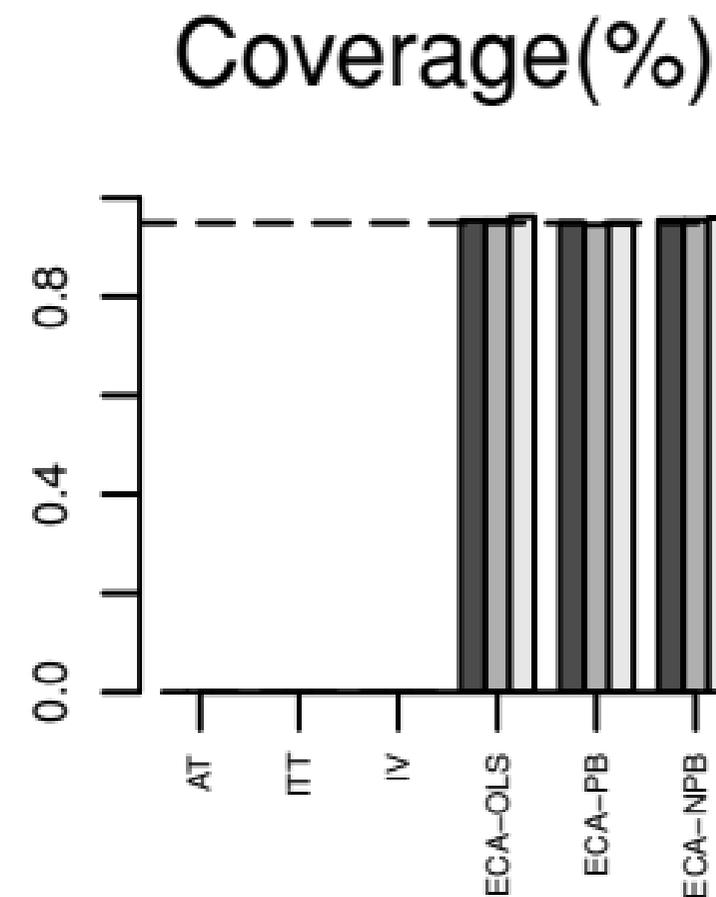
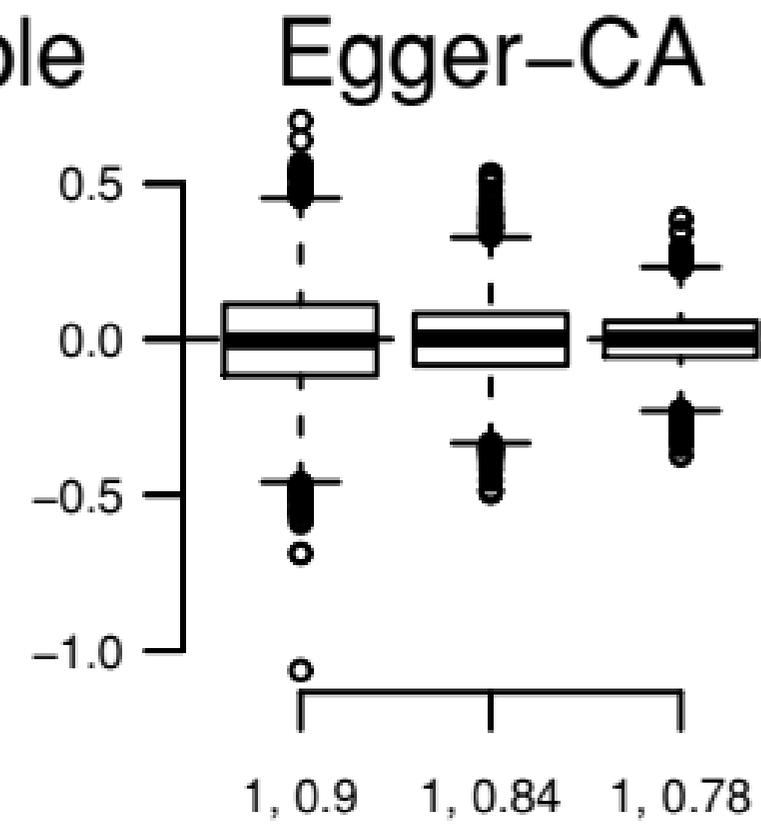
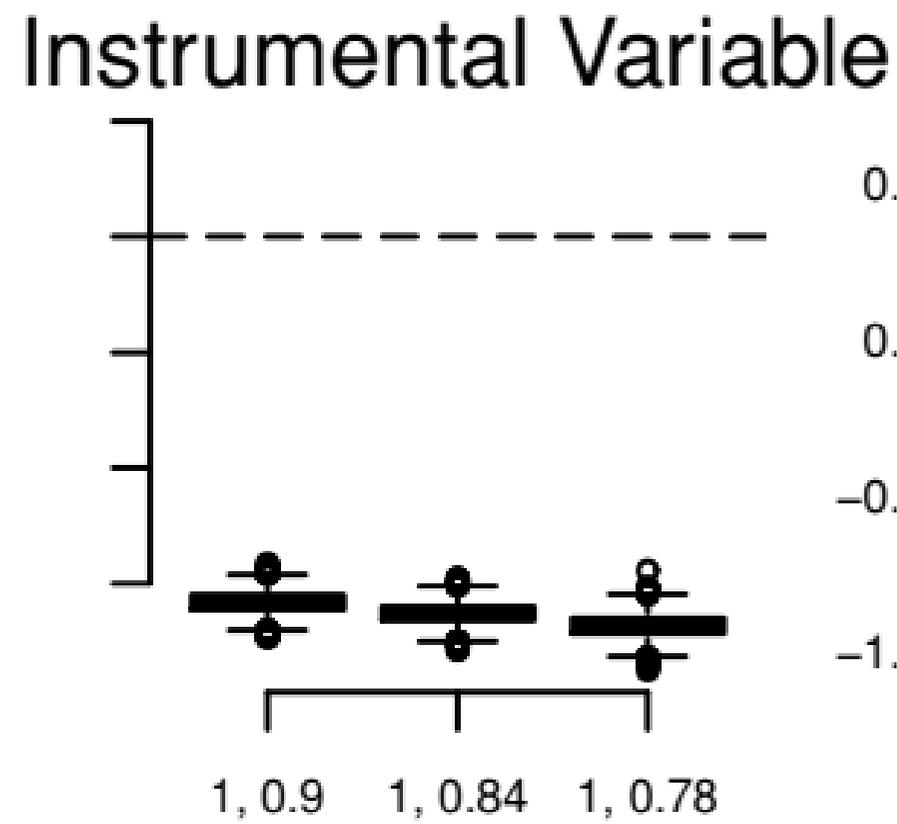
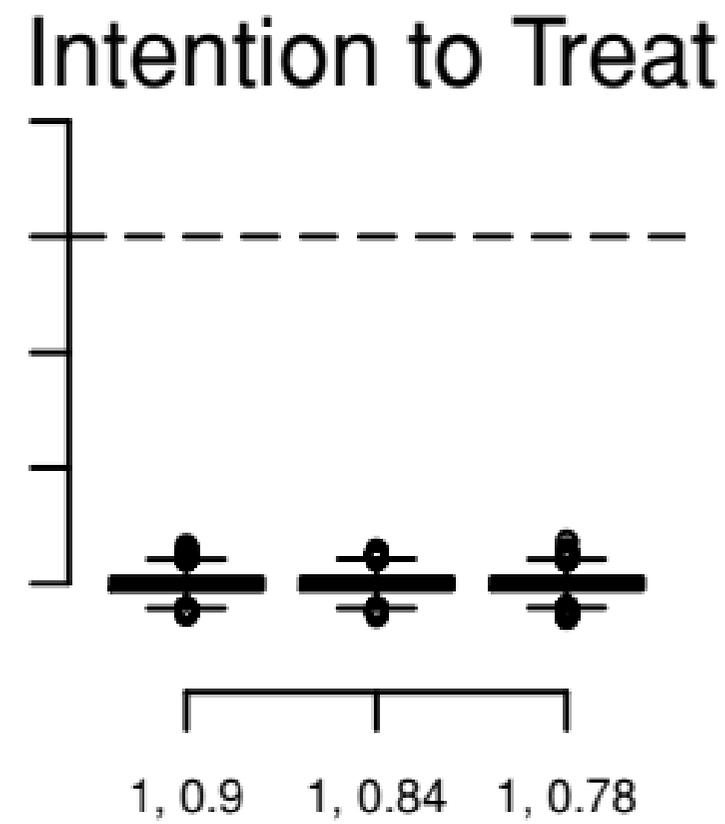
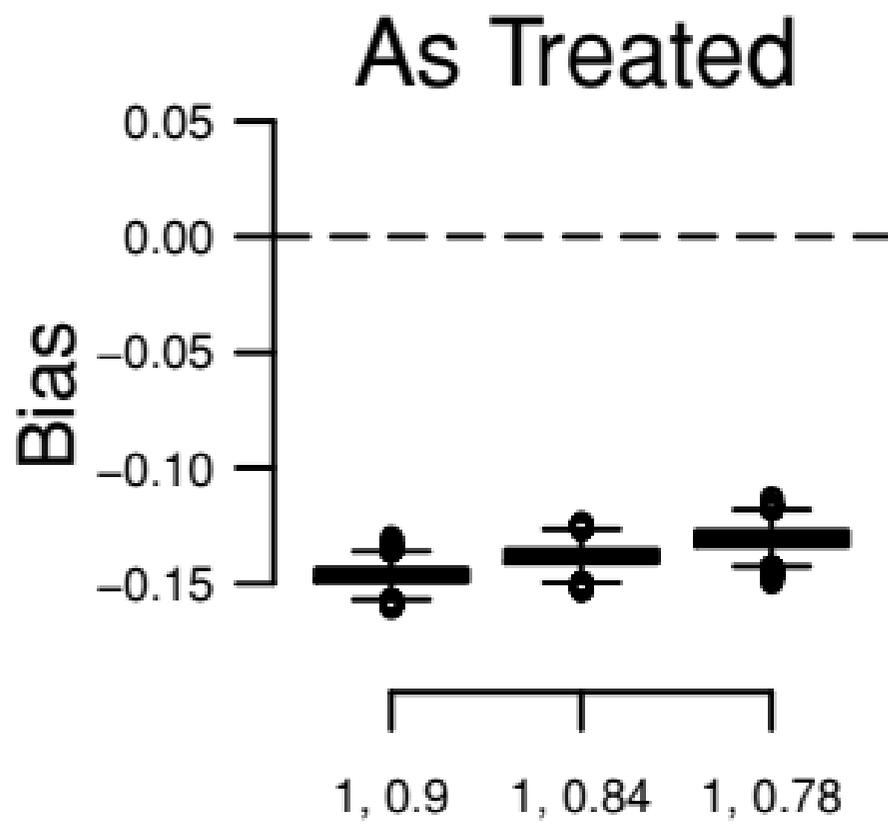
Scenario 2



Scenario 3



Adherence range



Adherence range

