



UNIVERSITY OF
LEICESTER

Department of
Health Sciences

Identifying publication bias in meta-analysis of continuous outcomes

Suzanne Freeman^{1,2} & Alex Sutton^{1,2}

Cochrane Learning Live Webinar

9th July 2020

suzanne.freeman@leicester.ac.uk

¹ Biostatistics Research Group, Department of Health Sciences, University of Leicester, UK

² NIHR Complex Reviews Support Unit

FUNDED BY

NIHR | National Institute
for Health Research

Outline

- Introduction
- Motivating example: Postoperative pain
- Proposed new approach for identifying funnel plot asymmetry
- Simulation study
- Conclusions
- References

“The conclusions of the review may be compromised when decisions about **how, when and where to report results** of eligible studies are **influenced by the nature and direction of the results**”

Cochrane Handbook for Systematic Reviews of Interventions, Second Edition

Introduction

- Cochrane Handbook recommends assessing meta-analyses for evidence of publication bias
 - The GRADE approach, grades the certainty in the estimate of the treatment effect in a meta-analysis (high, moderate, low or very low)
 - If publication bias present, downgrade the certainty of evidence
- To date, little work on assessing publication bias in meta-analyses of continuous outcomes

Publication Bias

- Studies with positive findings are more likely to be published and are published more quickly than studies with negative findings
- Often assessed using Egger's linear regression test and visualised using funnel plots
- Egger's test = regression of effect size on its standard error weighted by inverse variance

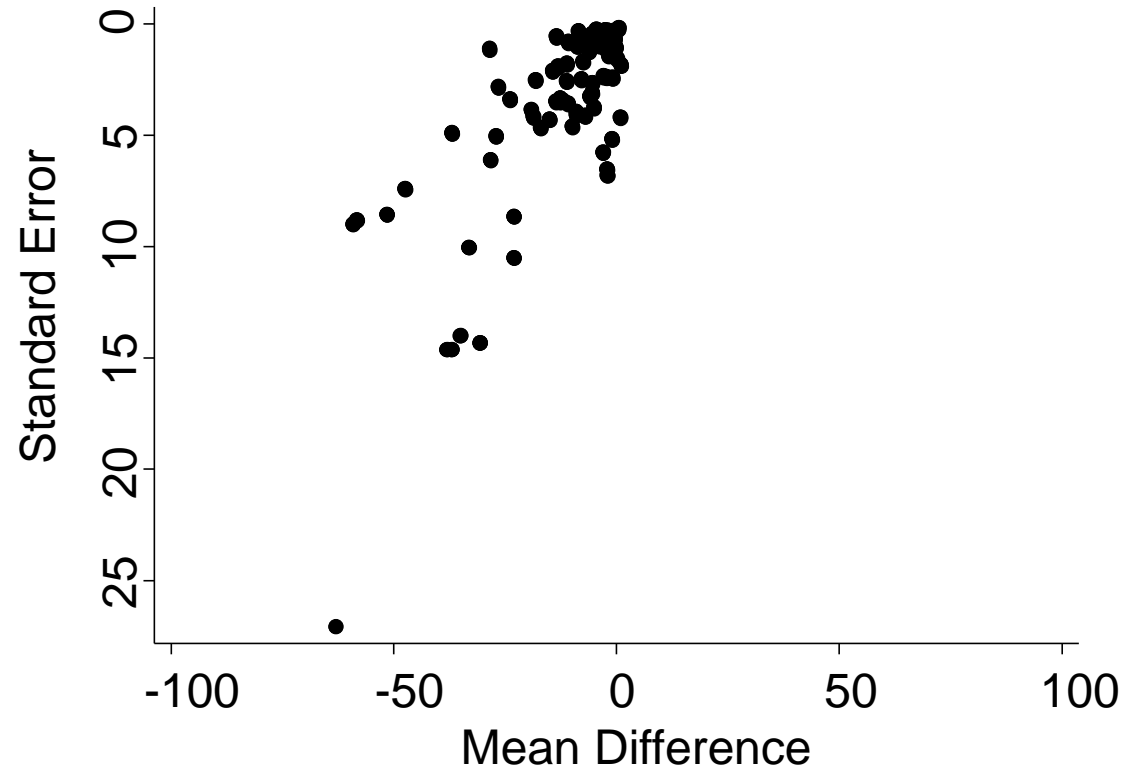
Egger's test

- Egger's test = regression of observed treatment effect size from trial i (y_i) on its standard error (se_i) weighted by inverse variance

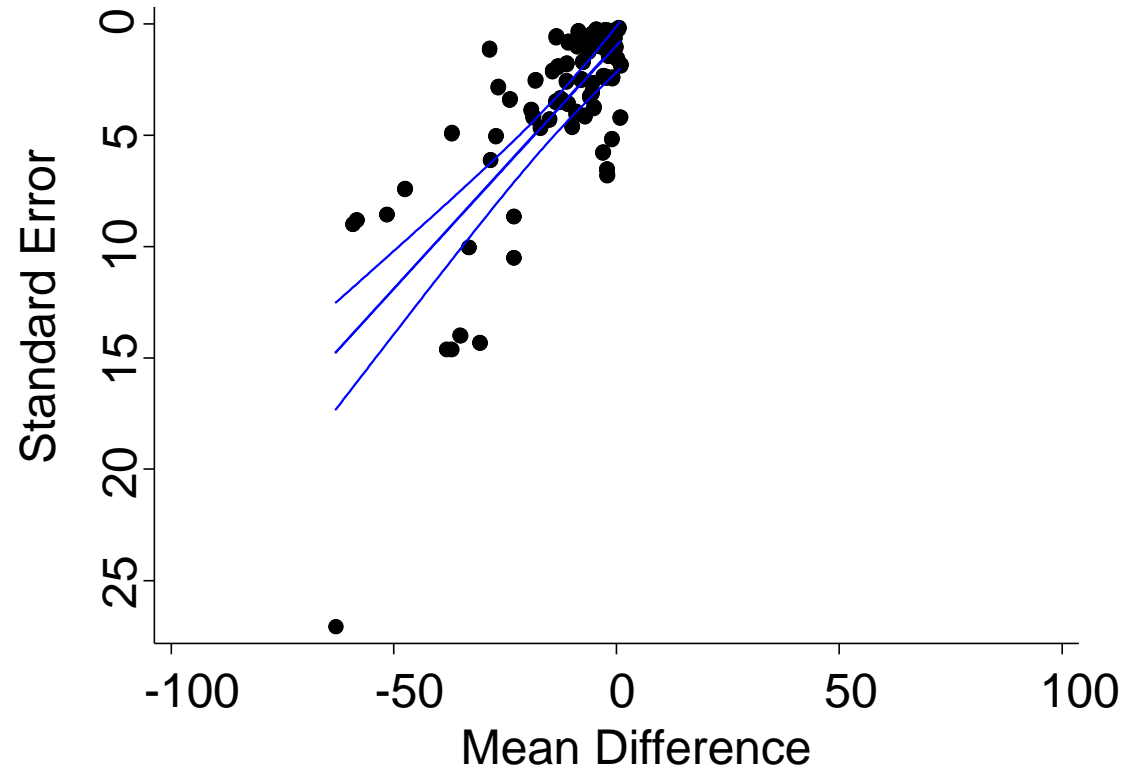
$$y_i = \alpha + \beta * se_i + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, se_i^2 * \varphi)$$

- φ is a multiplicative dispersion parameter estimated from the data which allows for heterogeneity inflation

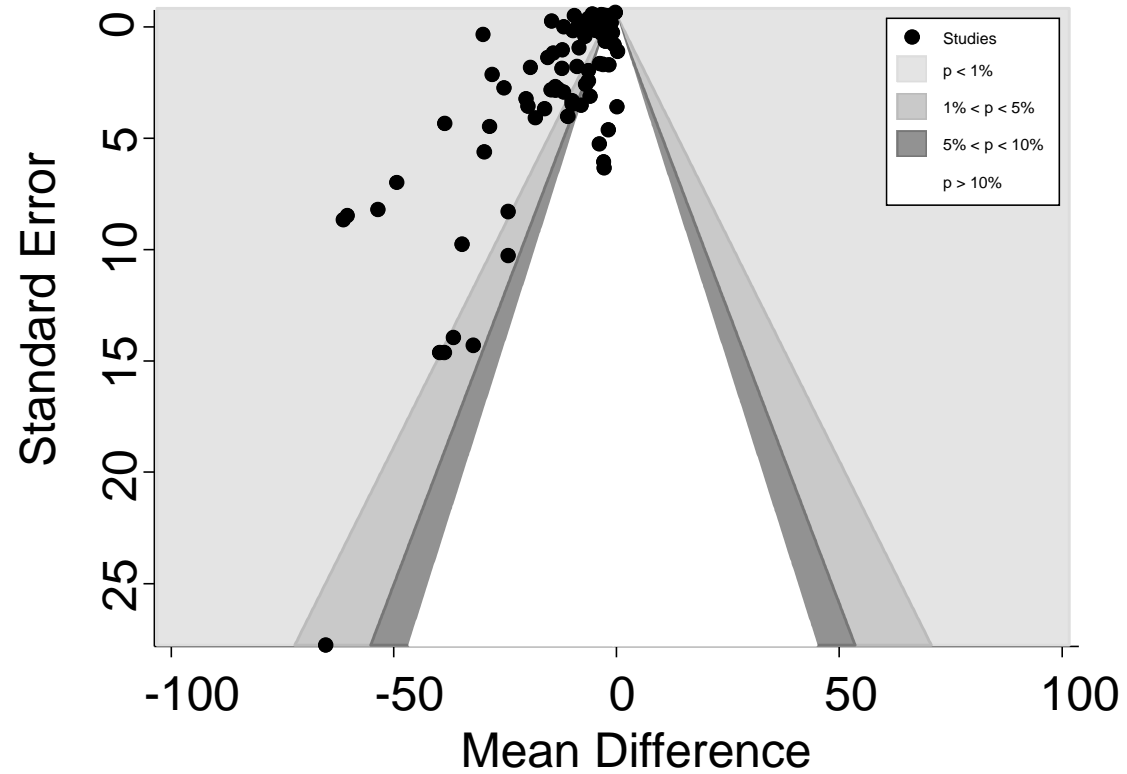
Funnel Plots



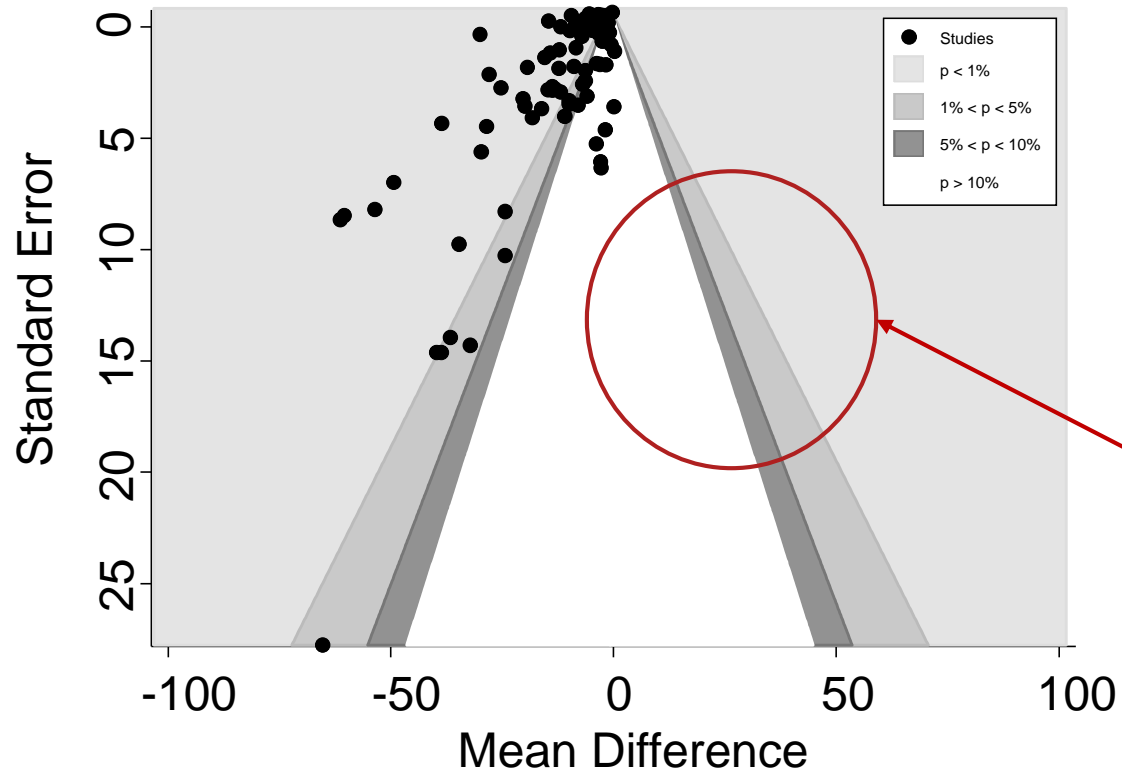
Funnel Plots



Funnel Plots



Funnel Plots

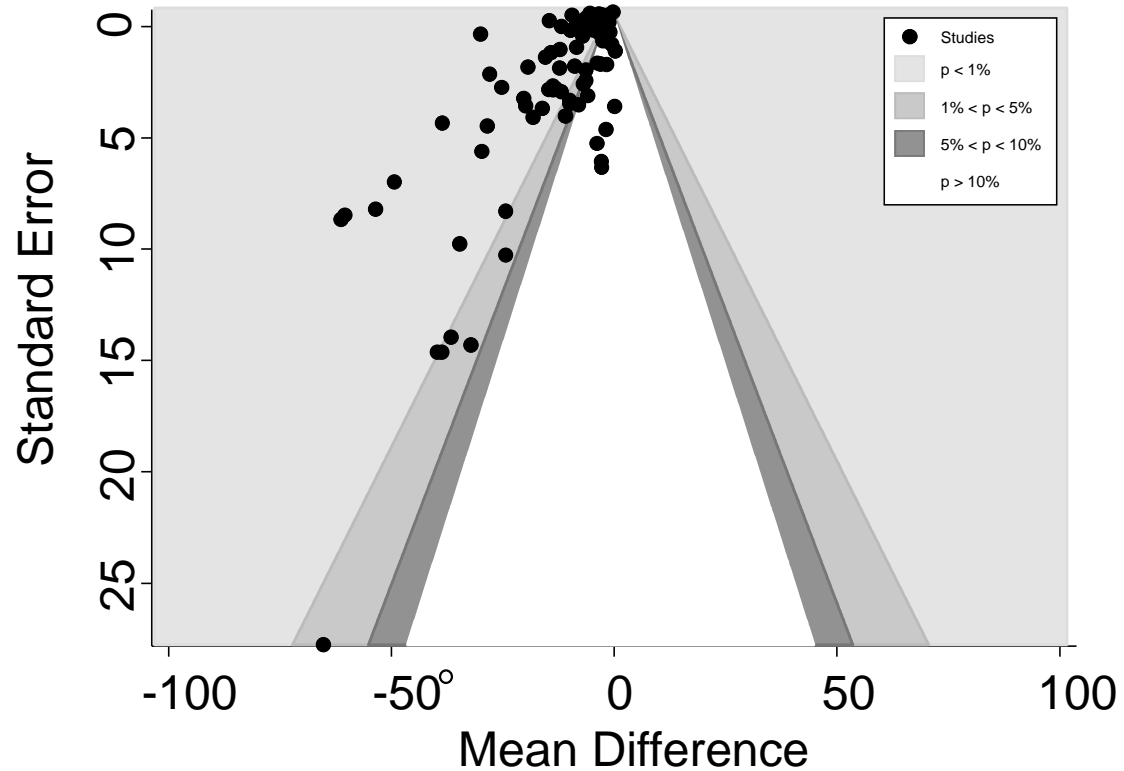


Studies with non-significant results appear to be missing

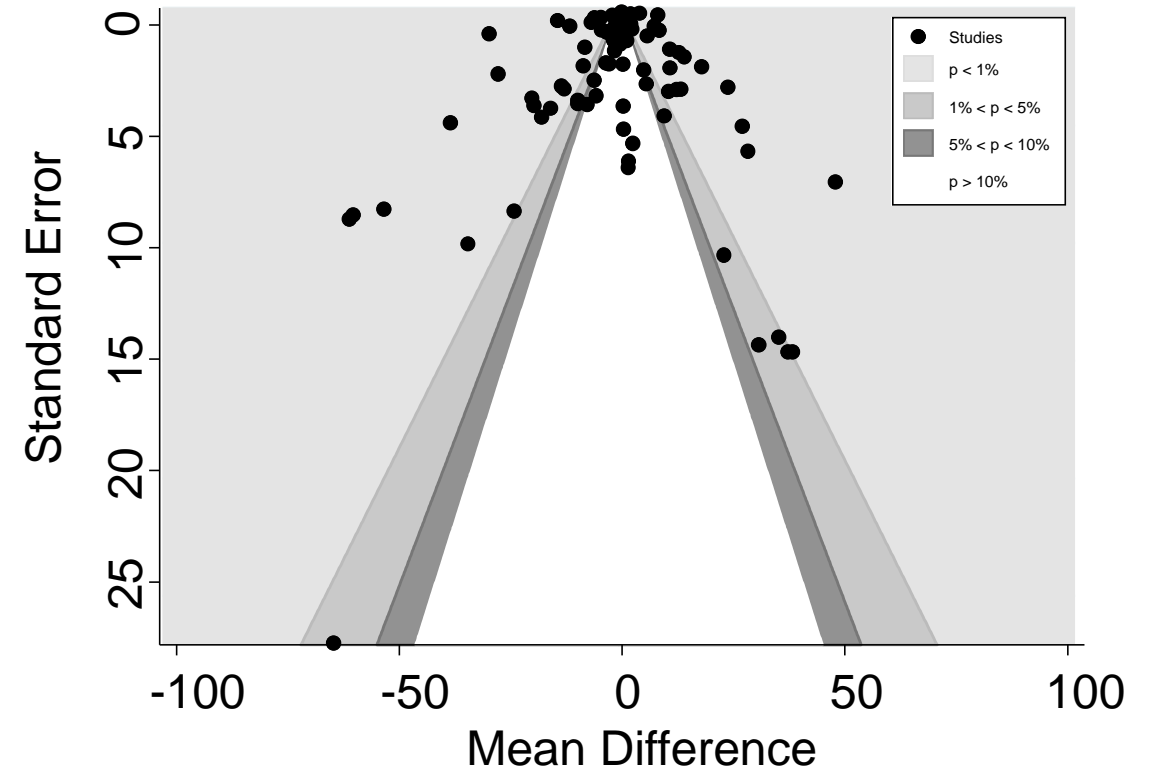
Funnel Plots



Asymmetry



No asymmetry

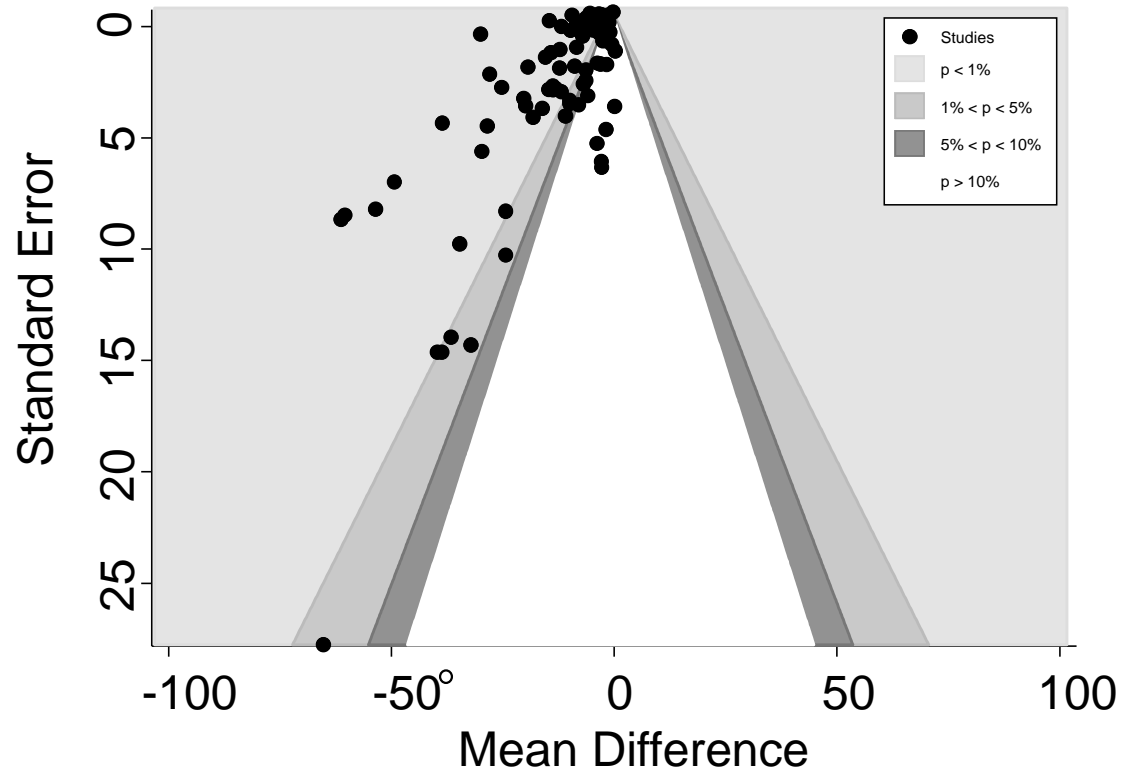


Note: Funnel plot asymmetry may not be caused by publication bias

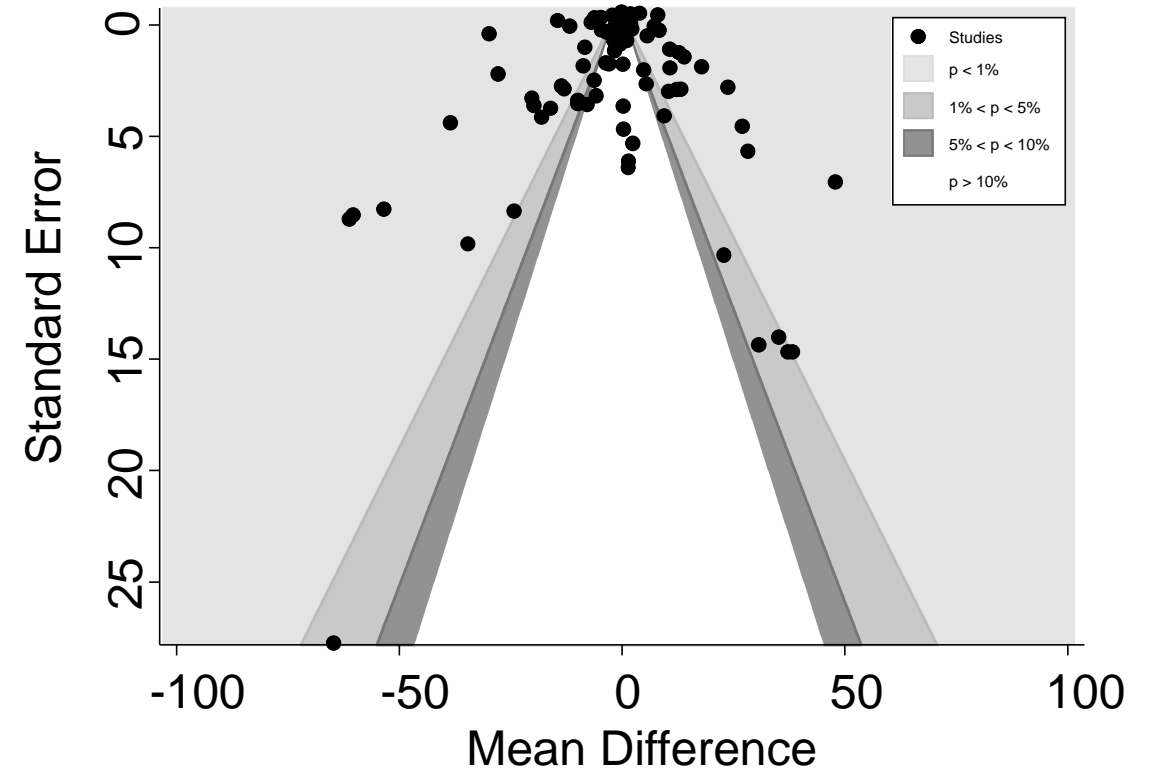
Funnel Plots



Asymmetry



No asymmetry

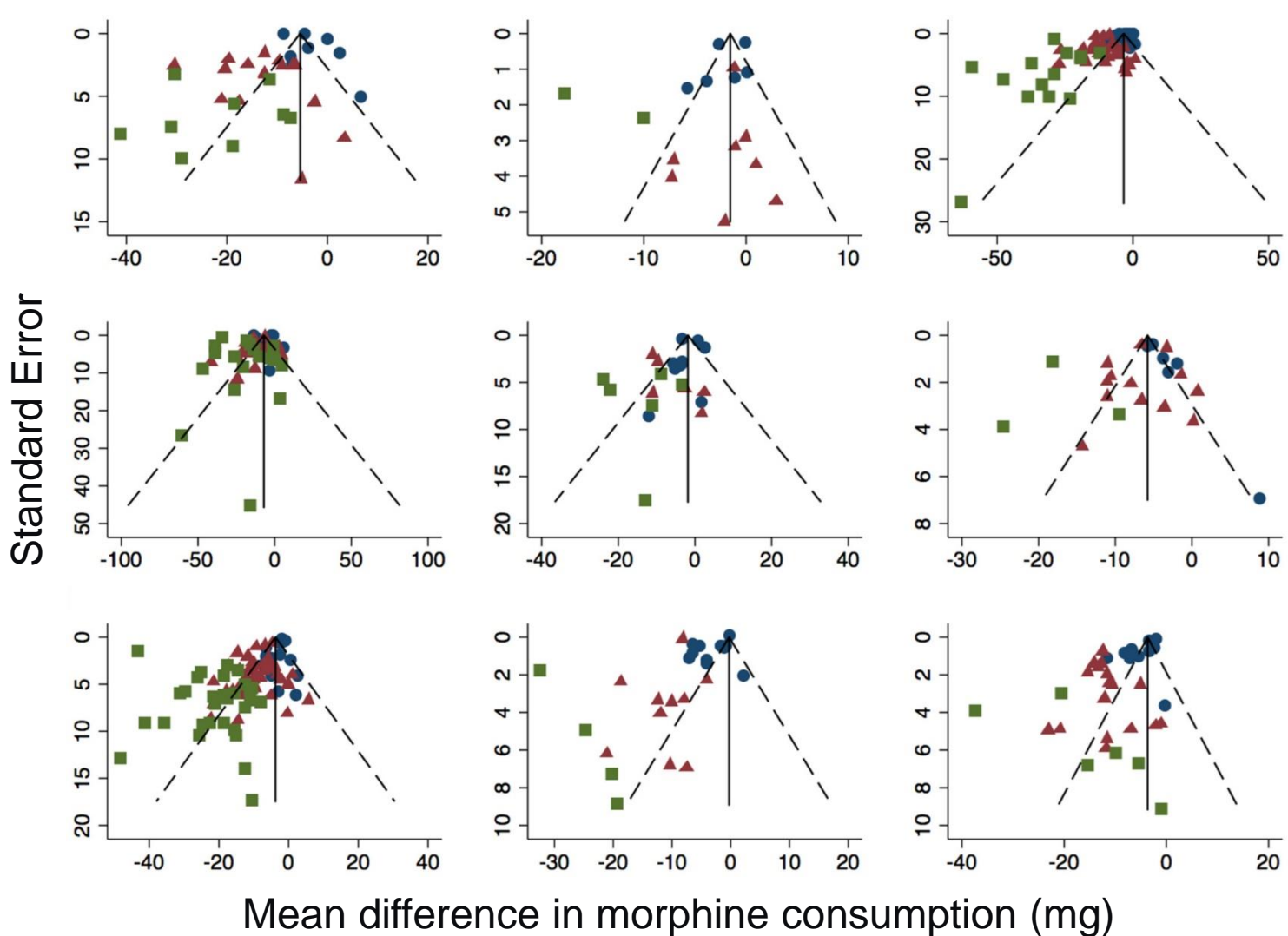


Assumption: Asymmetry in funnel plot = evidence of publication bias

Postoperative Pain Example

- Large systematic review (339 RCTs; 25,348 patients) compared nine different postoperative analgesics to placebo in a series of pairwise meta-analyses
- Outcome: 24 hour morphine consumption (in milligrams)

Postoperative Pain Funnel Plots

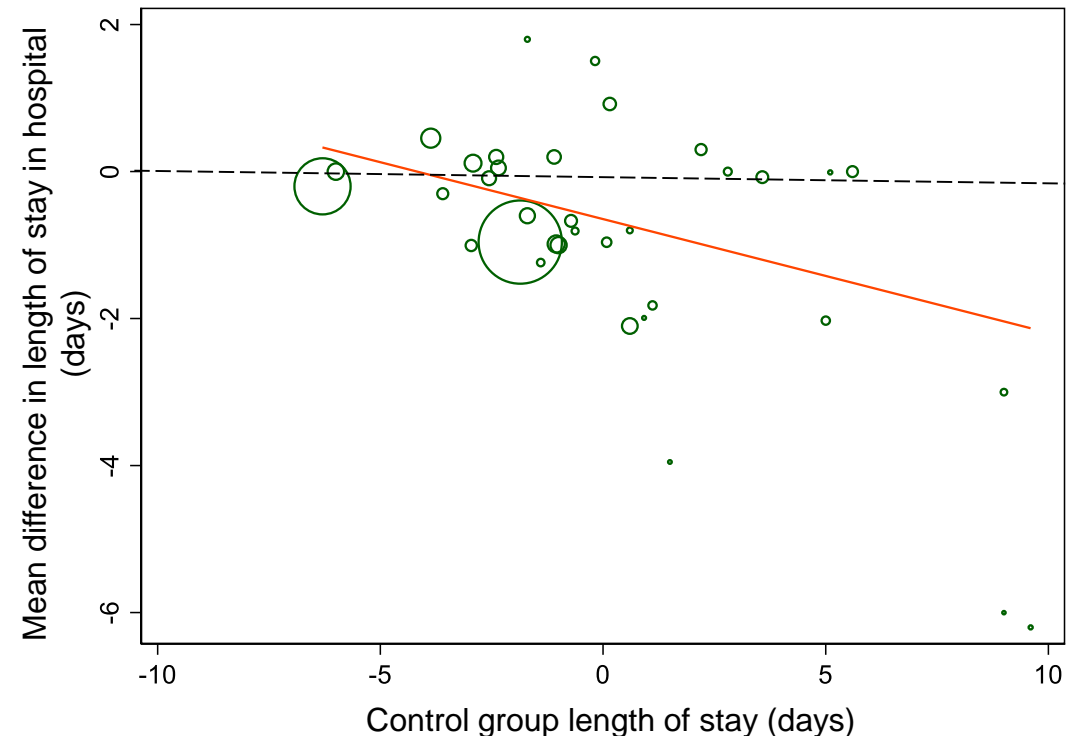


- Control group morphine consumption <20 mg
- ▲ Control group morphine consumption 20-50 mg
- Control group morphine consumption >50 mg

Baseline Risk

- “Average risk of a patient to experience the outcome of interest if they have not been treated” (Achana et al. 2013)
- Often measured as the response in the control group
- Trials with a greater outcome in the control group offer greater potential for larger absolute reductions in the outcome
 - Treatment effect size is dependent on the outcome in the control group

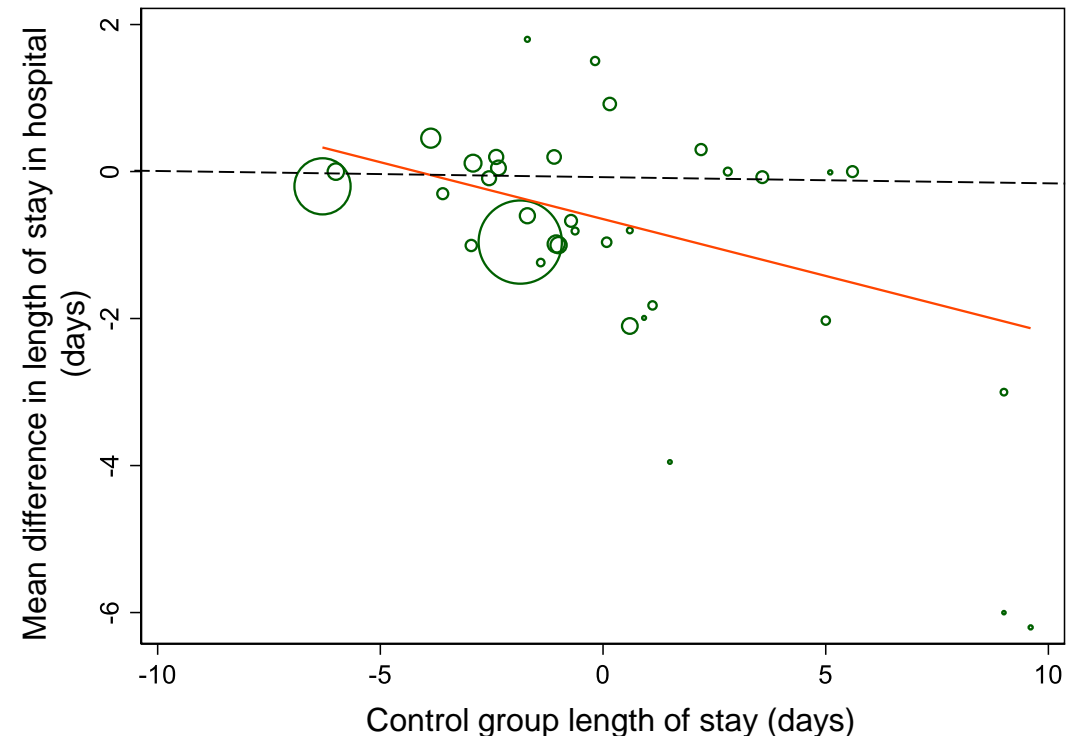
Example: Postoperative length of stay in hospital



Baseline Risk

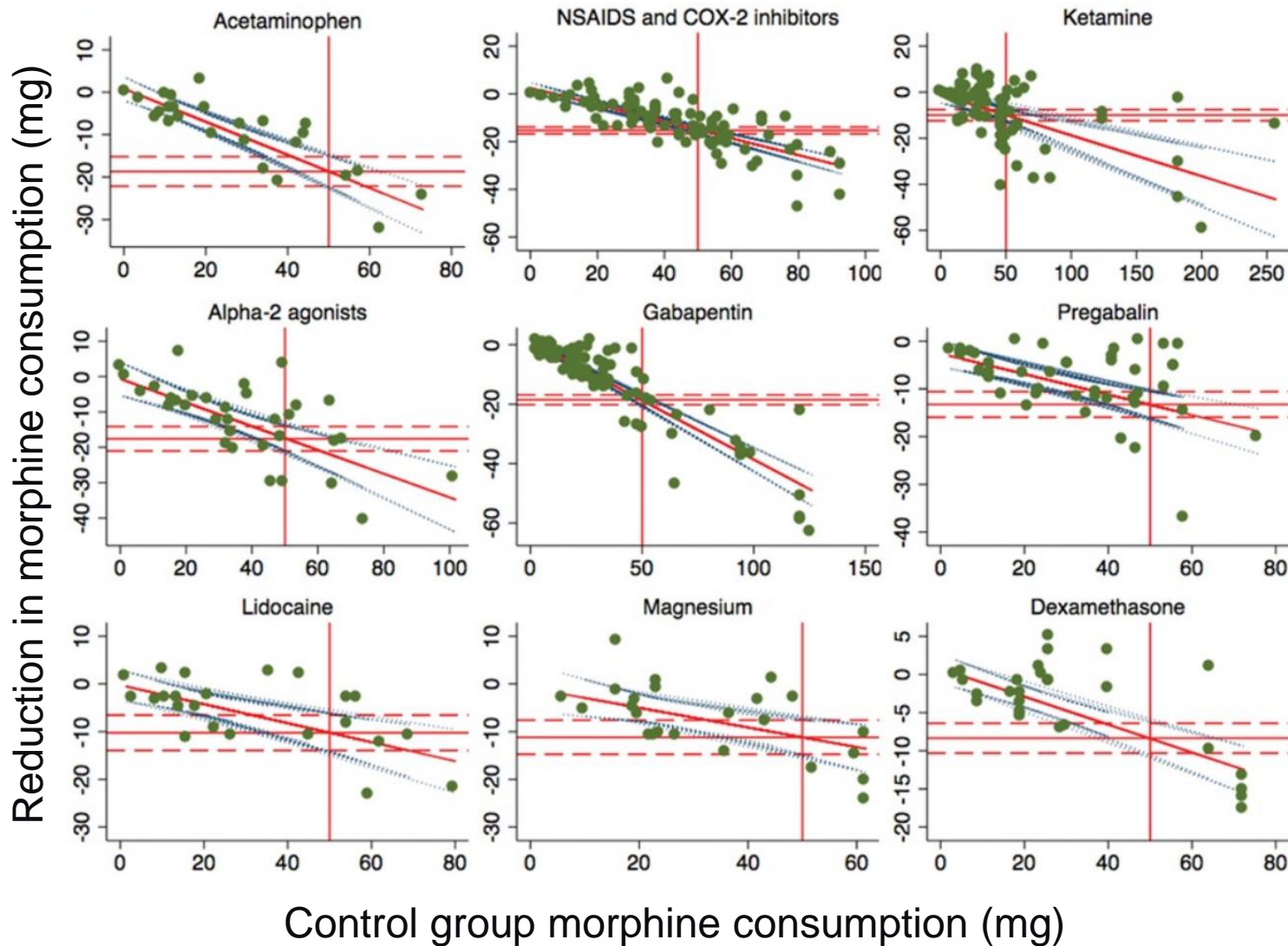
- “Average risk of a patient to experience the outcome of interest if they have not been treated” (Achana et al. 2013)
- Often measured as the response in the control group
- Trials with a greater outcome in the control group offer greater potential for larger absolute reductions in the outcome
 - Treatment effect size is dependent on the outcome in the control group

Example: Postoperative length of stay in hospital



Baseline risk = Treatment interaction with baseline response

Postoperative Pain Example

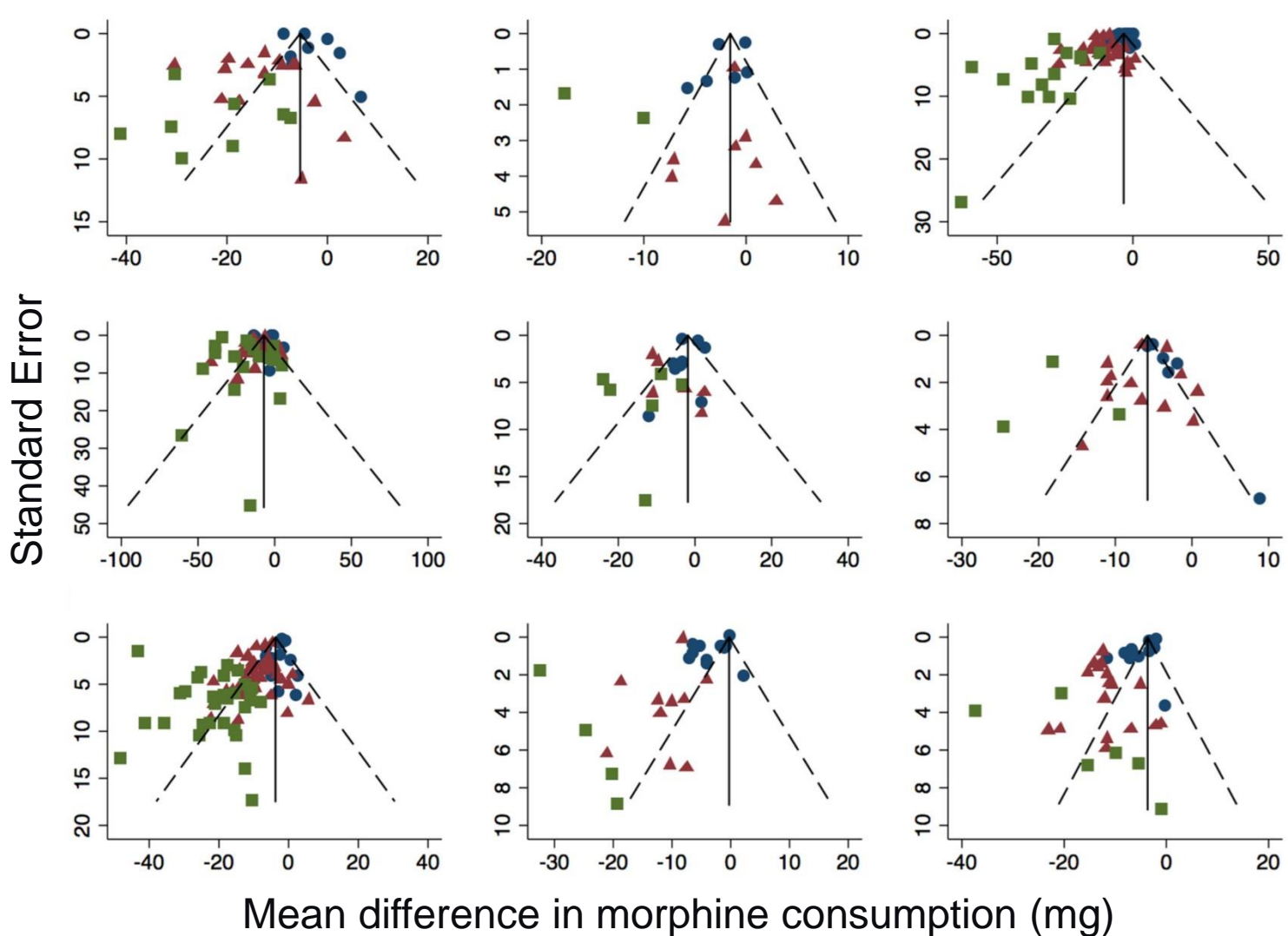


Conclusion: Control group morphine consumption was a significant cause of heterogeneity for all nine postoperative analgesics

The problem with Egger's test in the presence of baseline risk?

- Reductions in morphine consumption are dependent on the amount of morphine consumed in the control group (i.e. baseline risk)
- Studies with higher baseline risks will have larger standard deviations
- If treatment effect estimates are also dependent on baseline risk then this may cause correlation between mean differences (x-axis) and standard errors (y-axis) which could result in funnel plot asymmetry even in the absence of publication bias

Postoperative Pain Funnel Plots



- Control group morphine consumption <20 mg
- ▲ Control group morphine consumption 20-50 mg
- Control group morphine consumption >50 mg

Proposed New Approach

- Two steps:
 1. Fit meta-regression including baseline risk as a study-level covariate
 2. Regression test of the residuals
- Plot residuals on x-axis against standard error, sample size or inverse sample size on y-axis

Step 1: Meta-regression including BR

Residuals are generated from:

$$\hat{y}_i = \alpha + \beta \hat{\mu}_i + u_i + \hat{\sigma}_i \varepsilon_i$$

Where:

\hat{y}_i is the observed mean difference from study i

α and β are the intercept and slope of the regression, respectively

$\hat{\mu}_i$ is the observed mean response of the control arm in study i

u_i is a random effect term, $u_i \sim N(0, \tau^2)$

$\hat{\sigma}_i \varepsilon_i$ is the random error term where $\hat{\sigma}_i^2$ is the sample variance of \hat{y}_i

Step 2: Regression Test

- Regression test on residuals following adjustment for BR with **standard error** as the predictor
- Regression test on residuals following adjustment for BR with **sample size** as the predictor
- Regression test on residuals following adjustment for BR with **inverse sample size** as the predictor

Step 2: Regression Test

- Regression test on residuals following adjustment for BR with **standard error** as the predictor

$$residual_i = \alpha + \beta(se(residual_i)) + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, se(residual_i)^2 \varphi)$$

- Regression test on residuals following adjustment for BR with **sample size** as the predictor

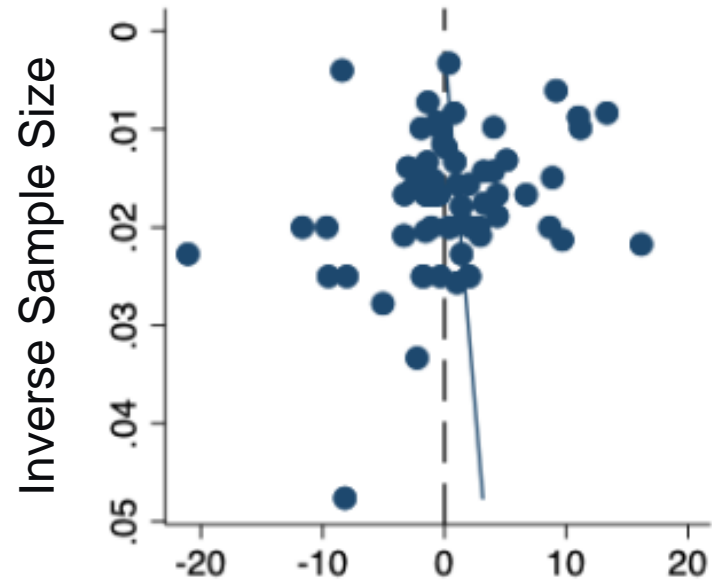
$$residual_i = \alpha + \beta(sample.size_i) + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, s.e.(residual_i)^2 \varphi)$$

- Regression test on residuals following adjustment for BR with **inverse sample size** as the predictor

$$residual_i = \alpha + \beta(1/sample.size_i) + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, s.e.(residual_i)^2 \varphi)$$

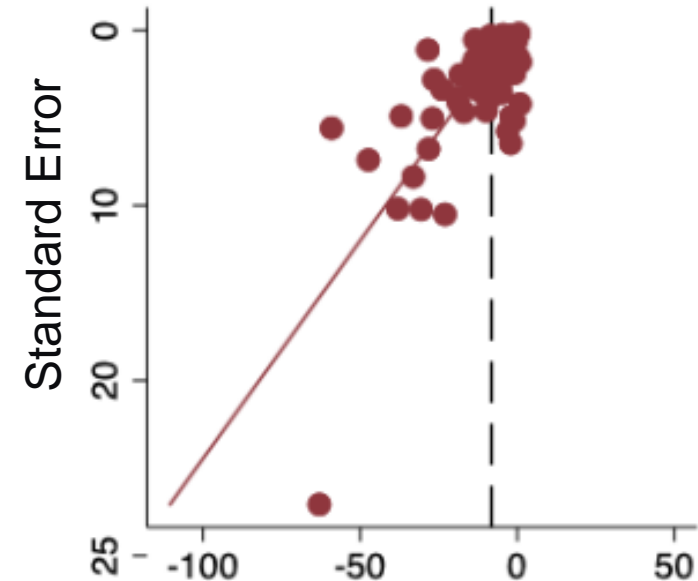
Postoperative Pain Example - Gabapentin

Proposed New Approach
 $p=0.55$



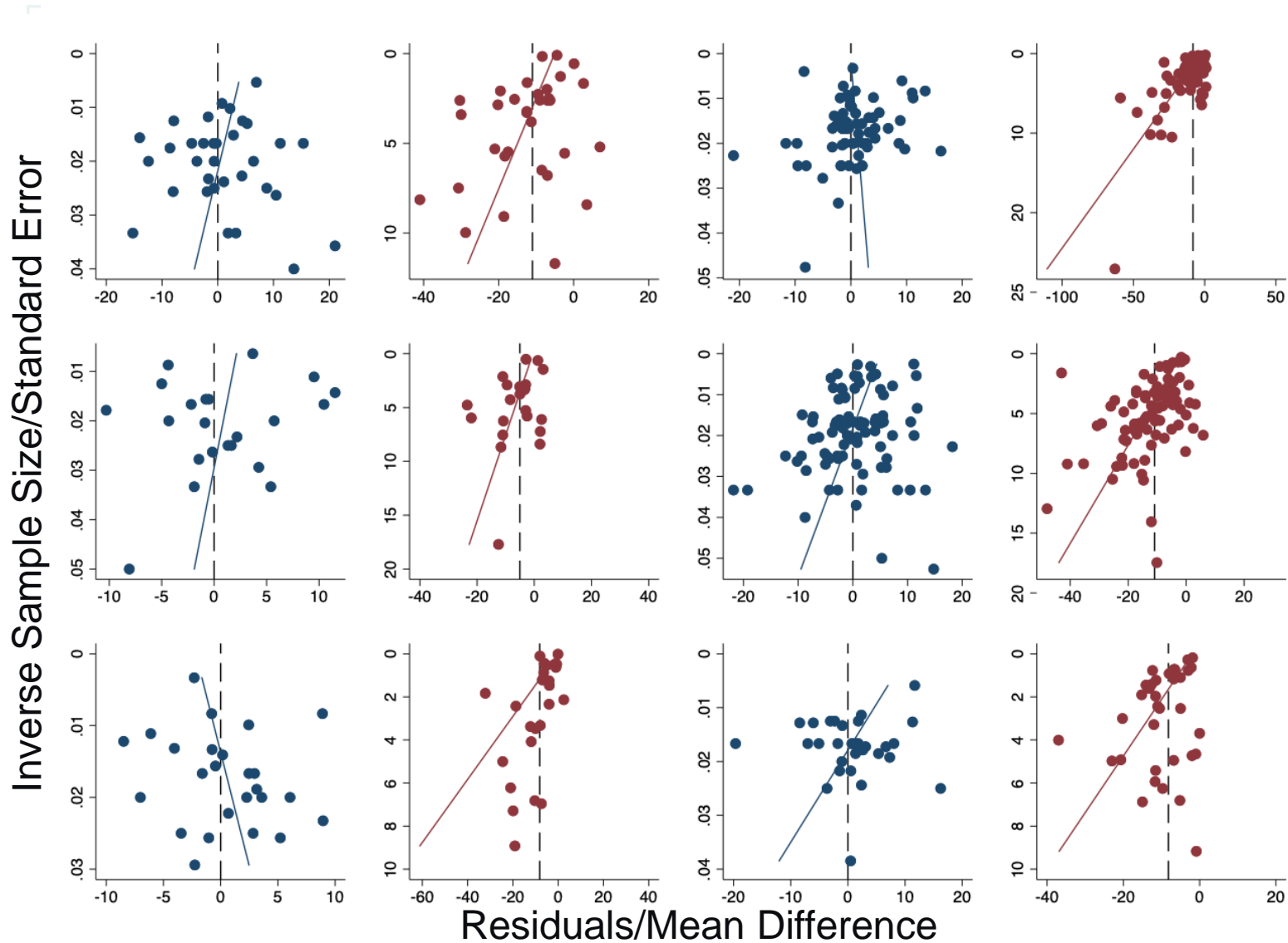
Residuals from meta-regression accounting for baseline risk

Traditional Funnel Plot
 $p<0.001$



Mean Difference

Postoperative Pain Example



Drug	Residuals p-value	Egger's Test p-value
Alpha-2 agonists	0.17	0.02
Gabapentin	0.55	<0.001
Lidocaine	0.60	0.02
NSAIDs	0.002	<0.001
Paracetamol	0.20	0.02
Pregabalin	0.01	<0.001

Simulation study: Aims

1. Evaluate the proposed new two-step approach to establish its performance
2. Compare performance between different measures of study 'size' (standard error, sample size and inverse sample size)

Data Generating Mechanisms

- Four scenarios, two settings:
 - 30 trials (Scenarios 1, 2, 3 & 4)
 - 15 trials (Scenarios 5, 6, 7 & 8)

	No BR	BR
No PB	Scenario 1 Scenario 5	Scenario 2 Scenario 6
PB	Scenario 3 Scenario 7	Scenario 4 Scenario 8

Data Generation

No baseline risk

- Individual patients responses in control arm assumed to be normally distributed

$$c_i \sim N(\mu_c, \sigma^2)$$

- Individual patient responses in the treatment arm assumed to be normally distributed but with a treatment effect added

$$t_i \sim N(\mu_c + trt.diff, \sigma^2)$$

Data Generation

Baseline risk

- Individual patients responses in control arm assumed to be normally distributed with the standard deviation assumed to depend on the mean response of the control arm

$$c_i \sim N(\mu_c, (\sigma + (0.5 * \mu_c))^2)$$

- In the treatment arm both the treatment effect and standard deviation are assumed to depend on the mean response of the control arm

$$t_i \sim N(\mu_c + trt.diff - (b.interaction * \mu_c), (\sigma + (0.5 * \mu_c))^2)$$

Data Generation

Publication bias

- Following generation of the control and treatment arms the estimated mean difference and associated variance were calculated for each trial
- All trials with a treatment effect for which $p > 0.05$ excluded from the meta-analysis with further trials generated until the pre-specified number of trials was reached

Methods

1. Egger's regression test on **observed** data (**standard error** as the predictor)
2. Regression test on **residuals** following adjustment for BR with **standard error** as the predictor
3. Regression test on **observed** data with **sample size** as the predictor
4. Regression test on **residuals** following adjustment for BR with **sample size** as the predictor
5. Regression test on **observed** data with **inverse sample size** as the predictor
6. Regression test on **residuals** following adjustment for BR with **inverse sample size** as the predictor

Other Details

- Estimand: Proportion of times the p-value for the funnel plot asymmetry test is ≤ 0.05
- Number of simulations = 10,000 meta-analyses per scenario
- μ_c took one of the values 20, 25, 30, 35, 40, 45 or 50
- Trial arm size took one of the values 15, 25 or 50 patients
- Data analysed using `metafor` package in R v3.6.0

Results: 30 trials

Six methods

Proportion of times $p \leq 0.05$ for asymmetry test

Four scenarios

Scenario:	Publication Bias Test Using Standard Error		Publication Bias Test Using Sample Size		Publication Bias Test Using Inverse Sample Size	
	Naïve / Conventional on observed data (Egger's test)	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)

- 1. no BR, no PB
- 2. BR, no PB
- 3. no BR, PB
- 4. BR & PB

1. no BR, no PB	0.0444	0.0438	0.0532	0.0568	0.0558	0.0585
2. BR, no PB	0.6026	0.0780	0.0089	0.0563	0.0116	0.0602
3. no BR, PB	0.6324	0.6249	0.5892	0.5806	0.6267	0.6252
4. BR & PB	0.9860	0.0067	0.3547	0.3595	0.3319	0.4018

Results: 30 trials – No BR or PB

Proportion of times $p \leq 0.05$ for asymmetry test

Scenario:	Publication Bias Test Using Standard Error		Publication Bias Test Using Sample Size		Publication Bias Test Using Inverse Sample Size	
	Naïve / Conventional on observed data (Egger's test)	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)
	1. no BR, no PB	0.0444	0.0438	0.0532	0.0568	0.0558
2. BR, no PB	0.6026	0.0780	0.0089	0.0563	0.0116	0.0602
3. no BR, PB	0.6324	0.6249	0.5892	0.5806	0.6267	0.6252
4. BR & PB	0.9860	0.0067	0.3547	0.3595	0.3319	0.4018

- All tests for funnel plot asymmetry produce a significant result approximately 5% of the time

Results: 30 trials – BR only

Proportion of times $p \leq 0.05$ for asymmetry test

Scenario:	Publication Bias Test Using Standard Error		Publication Bias Test Using Sample Size		Publication Bias Test Using Inverse Sample Size	
	Naïve / Conventional on observed data (Egger's test)	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)
	1. no BR, no PB	0.0444	0.0438	0.0532	0.0568	0.0558
2. BR, no PB	0.6026	0.0780	0.0089	0.0563	0.0116	0.0602
3. no BR, PB	0.6324	0.6249	0.5892	0.5806	0.6267	0.6252
4. BR & PB	0.9860	0.0067	0.3547	0.3595	0.3319	0.4018

- Egger's test incorrectly identifies funnel plot asymmetry 60% of the time

Results: 30 trials – BR only

Proportion of times $p \leq 0.05$ for asymmetry test

Scenario:	Publication Bias Test Using Standard Error		Publication Bias Test Using Sample Size		Publication Bias Test Using Inverse Sample Size	
	Naïve / Conventional on observed data (Egger's test)	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)
	1. no BR, no PB	0.0444	0.0438	0.0532	0.0568	0.0558
2. BR, no PB	0.6026	0.0780	0.0089	0.0563	0.0116	0.0602
3. no BR, PB	0.6324	0.6249	0.5892	0.5806	0.6267	0.6252
4. BR & PB	0.9860	0.0067	0.3547	0.3595	0.3319	0.4018

- Egger's test incorrectly identifies funnel plot asymmetry 60% of the time
- After accounting for BR, sample size/inverse sample size correctly identifies funnel plot asymmetry 5-6% of the time

Results: 30 trials – PB only

Proportion of times $p \leq 0.05$ for asymmetry test

Scenario:	Publication Bias Test Using Standard Error		Publication Bias Test Using Sample Size		Publication Bias Test Using Inverse Sample Size	
	Naïve / Conventional on observed data (Egger's test)	Two-stage (Test meta- regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta- regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)
	1. no BR, no PB	0.0444	0.0438	0.0532	0.0568	0.0558
2. BR, no PB	0.6026	0.0780	0.0089	0.0563	0.0116	0.0602
3. no BR, PB	0.6324	0.6249	0.5892	0.5806	0.6267	0.6252
4. BR & PB	0.9860	0.0067	0.3547	0.3595	0.3319	0.4018

- All tests correctly identify evidence of PB approximately the same number of times

Results: 30 trials – BR & PB

Proportion of times $p \leq 0.05$ for asymmetry test

Scenario:	Publication Bias Test Using Standard Error		Publication Bias Test Using Sample Size		Publication Bias Test Using Inverse Sample Size	
	Naïve / Conventional on observed data (Egger's test)	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)
	1. no BR, no PB	0.0444	0.0438	0.0532	0.0568	0.0558
2. BR, no PB	0.6026	0.0780	0.0089	0.0563	0.0116	0.0602
3. no BR, PB	0.6324	0.6249	0.5892	0.5806	0.6267	0.6252
4. BR & PB	0.9860	0.0067	0.3547	0.3595	0.3319	0.4018

- Egger's test identified asymmetry 99% of the time (but with only BR it identifies asymmetry 60% of the time anyway)

Results: 30 trials – BR & PB

Proportion of times $p \leq 0.05$ for asymmetry test

Scenario:	Publication Bias Test Using Standard Error		Publication Bias Test Using Sample Size		Publication Bias Test Using Inverse Sample Size	
	Naïve / Conventional on observed data (Egger's test)	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)
	1. no BR, no PB	0.0444	0.0438	0.0532	0.0568	0.0558
2. BR, no PB	0.6026	0.0780	0.0089	0.0563	0.0116	0.0602
3. no BR, PB	0.6324	0.6249	0.5892	0.5806	0.6267	0.6252
4. BR & PB	0.9860	0.0067	0.3547	0.3595	0.3319	0.4018

- After adjustment for baseline risk the asymmetry test on the residuals with sample size and inverse sample size is significant 36-40% of the time

Results: 15 trials

Proportion of times $p \leq 0.05$ for asymmetry test. [se = standard error]

Scenario:	Publication Bias Test Using Standard Error		Publication Bias Test Using Sample Size		Publication Bias Test Using Inverse Sample Size	
	Naïve / Conventional on observed data (Egger's test)	Two-stage (Test meta- regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta- regression residuals adjusting for baseline risk)	Naïve / Conventional on observed data	Two-stage (Test meta-regression residuals adjusting for baseline risk)
5. no BR, no PB	0.0466	0.0541	0.0551	0.0624	0.0551	0.0624
6. BR, no PB	0.3865	0.0587	0.0178	0.0624	0.0178	0.0624
7. No BR, PB	0.3109	0.3137	0.3012	0.2974	0.3149	0.3212
8. BR & PB	0.6977	0.0063	0.1126	0.1541	0.0955	0.1608

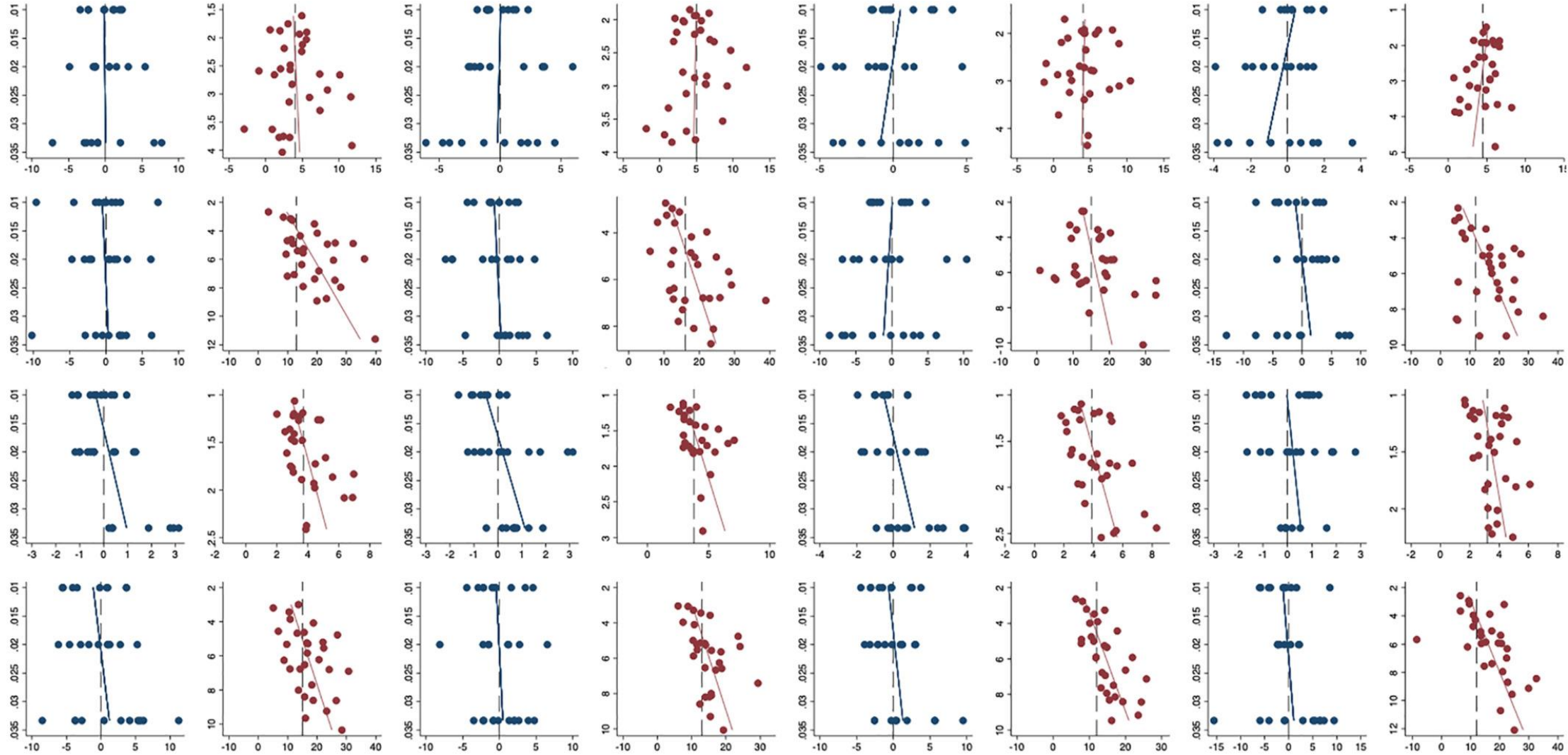
Reduced from 60%

Reduced from 36-40%

- Otherwise, similar results to 30 trials

Funnel Plots from Simulation Study

Inverse Sample Size/Standard Error



Scenario 1

Scenario 2

Scenario 3

Scenario 4

Residuals/Mean Difference

Summary

- Meta-analyses often inform clinical decision making
 - Important to assess for evidence of publication bias
 - Publication bias can lead to the downgrading of evidence using GRADE
- However, little work to date on publication bias using continuous outcomes
- In the presence of baseline risk meta-analyses of continuous outcomes can exhibit considerable heterogeneity
- Tendency in postoperative pain for trials with higher control group morphine consumption to have larger standard deviations
 - Dependency between the mean difference (larger with higher baseline risk) and the standard errors (larger with higher baseline risk)
 - Can result in asymmetric funnel plots even in the absence of publication bias (scenarios 2 & 6)

Summary

- We proposed a new approach
 1. Meta-regression accounting for baseline risk
 2. Regression test on the residuals with sample size or inverse sample size as the predictor
- Simulation study showed the new proposed approach:
 - Performs similarly to conventional methods when baseline risk is not present
 - Reduces type 1 errors in the presence of baseline risk
 - Low power to detect publication bias in the presence of baseline risk if there are 30 or less studies
 - Performance of inverse sample size marginally better than sample size

Limitations

- Postoperative pain data from previously published reviews – only a small number searched for unpublished studies
 - Our sample likely susceptible to publication bias
- We did not consider extra unexplainable heterogeneity on top of that induced by the dependency of outcome on baseline risk in simulation study
 - We suspect such extra variability would reduce the power of the regression testing
- We used a two-stage approach (i.e. regression model followed by regression test)
 - Would be possible to achieve similar results with a one-stage approach
 - But one-stage would remove ability to plot residuals
 - Unclear whether the regression should have additive or multiplicative heterogeneity parameters

Conclusions

- Traditional funnel plots are not reliable for detecting asymmetric study effects for morphine consumption
- For mean difference outcomes, we recommend, first accounting for baseline risk in a meta-regression and then assess funnel plot asymmetry by regressing the residuals on inverse sample size
 - Present the results using a funnel plot with inverse sample size (y-axis) and residuals (x-axis)
- More research needed for other continuous outcomes e.g. standardised mean differences, ratio of means

References

Achana F, Cooper NJ, Dias S, Lu G, Rice SJC, Kendrick D, Sutton AJ. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. *Statistics in Medicine* 2013; 32: 752-771

Doleman B, Sutton AJ, Sherwin M, Lund JN, Williams JP. Baseline morphine consumption may explain between-study heterogeneity in meta-analyses of adjuvant analgesics and improve precision and accuracy of effect estimates. *Anesth Analg* 2018; 126: 648-60.

Doleman B, Freeman SC, Lund J, Williams J, Sutton A. Funnel plots may show asymmetry in the absence of publication bias with continuous outcomes dependent on baseline risk: presentation of a new publication bias test for continuous outcomes. *Research Synthesis Methods* 2020; 11: 522-534

Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997, 315: 629-634.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd Edition. Chichester (UK): John Wiley & Sons, 2019.

Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR *et al*. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:bmj.d4002

Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Soft* 2010; 36(3): 1-48. URL: <http://www.jstatsoft.org/v36/i03/>.

Open Access paper available:

<http://dx.doi.org/10.1002/jrsm.1414>



Received: 30 January 2018 | Revised: 18 April 2020 | Accepted: 23 April 2020

DOI: 10.1002/jrsm.1414

RESEARCH ARTICLE

Research
Synthesis Methods **WILEY**

Funnel plots may show asymmetry in the absence of publication bias with continuous outcomes dependent on baseline risk: presentation of a new publication bias test

Brett Doleman¹  | **Suzanne C Freeman²**  | **Jonathan N Lund¹** |
John P Williams¹ | **Alex J Sutton²**

Funding: The Complex Reviews Support Unit (CRSU) is funded by the National Institute for Health Research (project number 14/178/29). The views and opinions expressed are those of the authors and do not necessarily reflect those of NIHR, NHS or the Department of Health.