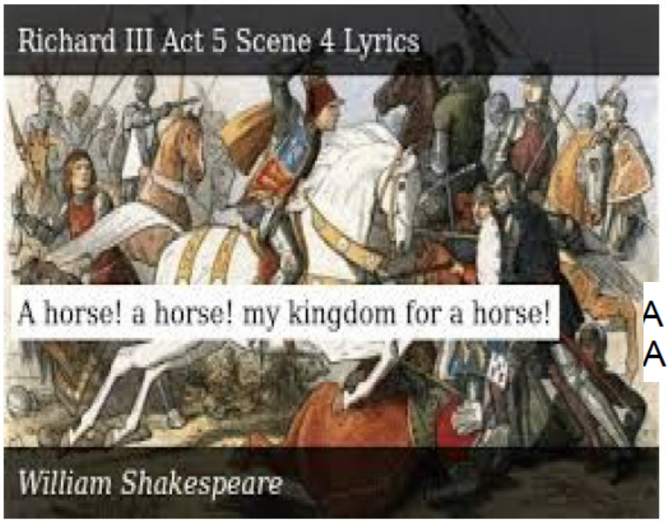
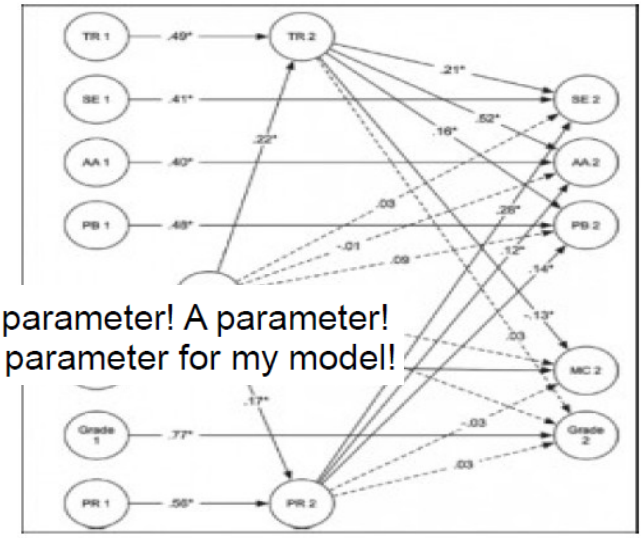
Lecture 23 Parameters for Models





1.All models depend on PARAMETERS

In most models results hinge critically on parameter values. If the Ro in contagion model is <1 the disease stops spreading (one person contaminates less than one person, who contaminates less than one person, so the population of infected levels off as geometric series ΣRo^n hits limit of 1/1-Ro as n--> infinity). If it is Ro =1 disease spreads slowly but goes on until everyone is contaminated/recovered; if Ro>1 can spread very fast.

If the elasticity of supply >elasticity of demand in cobweb model, market diverges from equilibrium cyclically; while it converges to equilibrium supply-demand crossing if demand elasticity is larger.

If elasticity of demand for minimum wage type labor < 1, increases in minimum raise low wage workers share of income, if elasticity > 1 it reduces income going to low wage workers.

If results matter whatever the parameter is in a model, you might have a “law of nature” or an identity, or in social science more likely have a model that is biased/missing some possible mechanism or response that could invalidate however you think social world works.

So, economists and scientists in every discipline seek to estimate key parameters in some way. Here are some ways to make estimate.

1- Ask experts – and hope the crowd of experts gets it right, per the people predicting the weight of Galton's ox and not the experts Meadows/Forester asked about future of world

2- Measure parameter in observational data if it is single simple statistic. You get the interest at bank and then can do all your present value discounting or compound interest stuff. (This can get complicated with many banks/currencies/exchange rates and terms of contract with bank)

3- Estimate with best causal statistical model – econometrics and more

4- Random controlled experiments – which should follow specified procedures

5- Survey people and ask “if price fell by 10%, how would your purchases change?” type questions. A bunch of complications – are people representative, if you posed question differently would you get different answer, etc

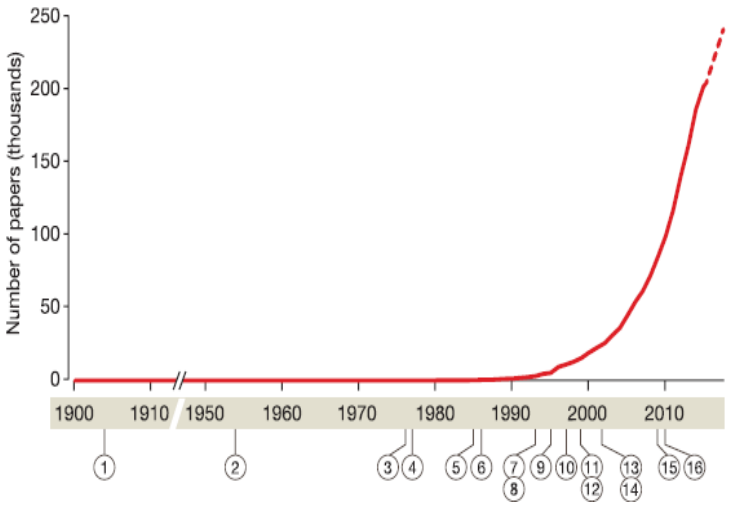
When researchers use these different modes to try to determine parameter, they will almost surely make some different decisions – ask different experts; choose different measures (mean vs median?) or observational data (CPS vs ACS); use different statistical tests; make different experimental decisions; phrase survey questions differently. The result is that on many issues, we have many studies, of differing quality, with different estimated parameters for impacts of some variable on the outcomes of concern. Which do you choose?

In ancient times, people would read them, give narrative and say what they felt was a reasonable summary

But this could be biased and cannot readily deal with dozens and dozens of studies.

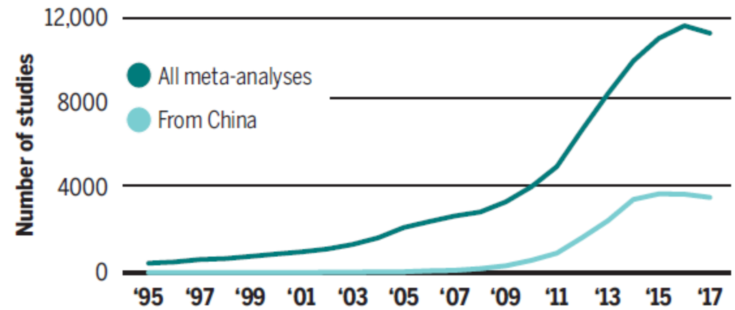
“Seven studies say they got a big number for the elasticity but 12 got a small number and they use samples with different sizes and report different standard errors around their estimates. WHAT DO I DO?

To the rescue **META-STATISTICS** **Can meta-analysis rescue knowledge from an ocean of studies?**



Eager for answers

Virtually unknown until the 1990s, meta-analyses have recently become increasingly popular. More than 11,000 were published last year, one-third of them by authors from China.



1. **1904** First (medical) meta-analysis published (effect of innoculation against typhoid) (ref. 83)
2. **1954** First meta-analytic methods formalized (fixed- and random-effects models) (ref. 86)
3. **1976** Term 'meta-analysis' coined (ref. 95)
4. **1977** First social science meta-analysis published (efficacy of psychotherapy) (ref. 87)
5. **1985** Statistics textbook dedicated to meta-analytic methods released (ref. 16)
6. **1986** Method for calculating between-study variance developed (ref. 96)
7. **1993** Review of 302 social science meta-analyses on treatment efficacy published (ref. 97)
8. **1993** Cochrane Collaboration established
9. **1995** Term 'systematic review' introduced (ref. 98)
10. **1997** Methods for assessing publication bias introduced (funnel plot and Egger's test) (ref. 19)
11. **1999** QUOROM (Quality of Reporting of Meta-analyses) standards developed (ref. 99)
12. **1999** Campbell Collaboration established
13. **2002** Heterogeneity index I^2 proposed (ref. 100)
14. **2002** Term 'network meta-analysis' coined (ref. 74)
15. **2009** PRISMA guidelines established (ref. 12)
16. **2010** *metafor* (free and comprehensive R package for meta-analysis) released (ref. 17)

Box 1 Figure | Milestones in the history of meta-analysis. The red line shows the number of papers from a Scopus search; the dashed component indicates the expected future trajectory. The milestone publications[12,16,17,19,74,83,86,87,95–100] are chosen on the basis of two main criteria—precedence and influence (for these criteria, we relied heavily on refs 93 and 94).

What it is: Method of combining results from many studies following "standard" statistical procedures into a single estimate that summarizes results in way that older "narrative summaries" could not do, while allowing for analysts comments. Widely used in different ways across fields, from economics to statistics, and to medical research. If you have a model and need parameters that will give the model best shot of explaining world/simulating how different policies might affect things, best practice is to find a meta-analysis of many studies and use appropriate meta-summary for you parameter. As you can see above, very popular.

You find studies; you turn key stats into single parameter, elasticity or B-coefficient, and you get SE on the estimate; You pool the estimates and come up with an average. Sounds simple and that "preponderance of evidence" calculated by statistics would give you right answer. Founders of meta-analyses thought this was indeed the saving grace in world where many people producing many papers, with some ideally random differences, would come to nice consensus of effects.

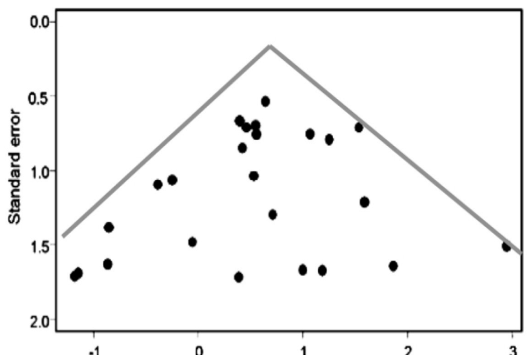
Economics Example to see where problems arise: New Congress wants to raise the minimum wage but not to create lots of lost jobs. We know that if we add a 0 after current minimum, lots of people will lose jobs, so we are thinking of modest increase – maybe to $12.00 per hour or $15.00 per hour. If elasticity of demand is small, and job loss is small, we are happy to raise minimum by bigger amount. You are hired to summarize the estimates. For

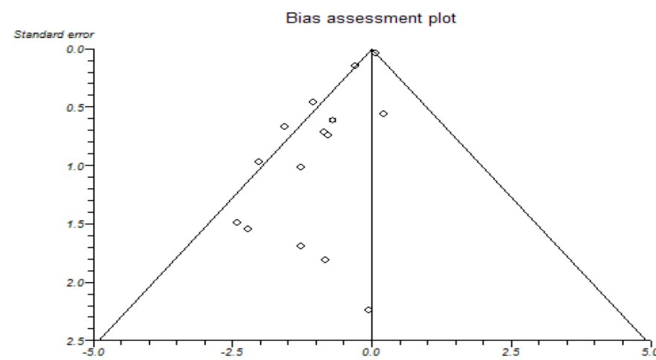
simplicity, say there are 2 studies of the effect of an increase in minimum wages of 10% on employment. Study 1 says the elasticity of demand for labor is about -0.10 – and thus predicts 1% loss of jobs. – small by most standards. Study 2 estimates elasticity of -0.50, says 5% loss of jobs – five times as many job losers – not so small.
You go through the 2 studies and note some key facts

	# observations	Estimate	Std error	"Precision"= $1/SE^2$
Study 1	80	-0.1	0.02	2500
Study 2	20	-0.5	0.2	25

You could simply average the 2 studies – -0.3
You could average by # of observations -0.18
You could take wtd average by precision, -0,12. The smaller std error → much greater precision so weight more.

But just 2 studies? Say there are 20-30 estimates. If they truly estimate the same parameter and have different estimates main reason might be # observations that would affect STD Errors/precision or maybe random differences in specification, maybe with more conservative folk specifying in ways that produce bigger elasticities and more progressive specifying in ways that produce smaller elasticities. BUT what if you see this?





Here STD error//Precision gets smaller and dispersion of estimates declines, with no obvious pattern. Could also be # observations

STD Error gets smaller and estimates get bigger
Looks like missing positive high std error results
PUBLICATION BIAS???

If have a theory/bias/incentive to produce a certain result, maybe not report those that go against your desired outcome … or maybe journals won't publish the results … or maybe top journals won't. Relation of size of effect on std error in first would have 0 correlation while relation of effect on std error would be positive (Egger's regression)

Actual meta-analysis of time series studies of minimum wage by Card and Krueger (1995, AEA Proceedings) found clear evidence of publication bias favoring studies that found a statistically significant negative employment effect. If studies were "right" as a group, the more recent ones with more data should have similar coefficients as the earlier ones, lower std errors and higher t-statistics. Instead studies had a t-statistic of about two, just above the level of statistical significance at the .05 level.

FIGURE 1. RELATION OF ESTIMATED t RATIOS TO SAMPLE SIZE

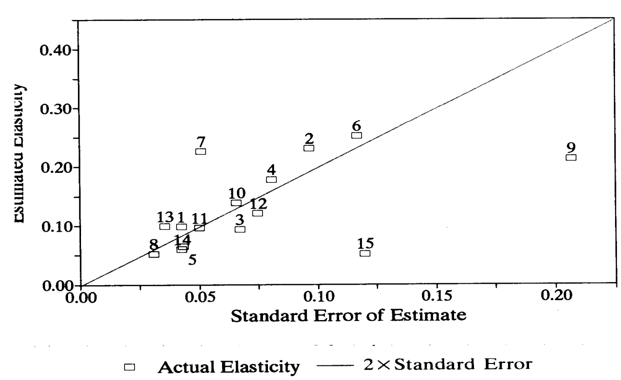


FIGURE 2. RELATION OF ESTIMATED EMPLOYMENT ELASTICITY TO STANDARD ERROR

What do you do to correct this bias? 1) Can do some stats by filling in missing data. If you truly believe that lots of positive unreported estimates and that right model is the full funnel plot, stick in the missing; 2) search for the missing studies in, say, UG papers or Master's theses or … regression output not reported.

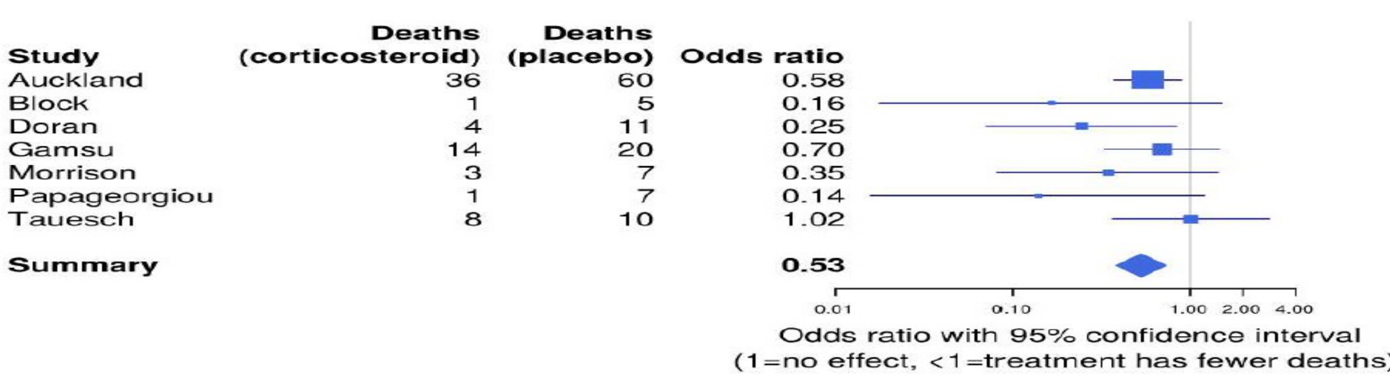
A 2007 meta-analyses by Neumark and Wascher of 96 studies found "A sizable majority of the studies surveyed in this monograph give a relatively consistent (although not always statistically significant) indication of negative employment effects of minimum wages. In addition, among the papers we view as providing the most credible evidence, almost all point to negative employment effects, both for the United States as well as for many other countries … we see very few - if any - studies that provide convincing evidence of positive employment effects of minimum wages, especially from those studies that focus on the broader groups (rather than a narrow industry) for which the competitive model predicts disemployment effects. Second, the studies that focus on the least-skilled groups provide relatively overwhelming evidence of stronger disemployment effects for these groups."

BUT THE KEYS IS SIZE OF PARAMETER and SIZE of INCREASE IN MINIMUM.

In 2008, Doucouliagos and T.D. Stanley did a similar meta-analysis of 64 U.S. studies on disemployment effects and concluded that Card and Krueger's initial claim of publication bias still holds, but with little or no evidence of a negative association between minimum wages and employment remains. Despite continued debate, the evidence of publication bias moved discussion forward …

Meta in Medicine ---Area where meta-analysis most extensively use is Medicine where RCT are "gold standard" but in fact they often vary in quality and depend on small samples. Say you have 100 doctors who experiment with 12 patients on Dr T's new cure for "mental health misery " due to stay-at-home policies to fight spread of virus *(the days just run into each other.....day on day....tweet on tweet....this way lies madness – Peter Doherty)* . If each doctor had RCT with 6 mental misery people control and 3 "naturally recover" while 4 of 6 in the treated group is cured by Dr. T's miracle patent medicine, each doctor will conclude T medicine is not statistically significant. With 50% natural recovery rate, chance of 4 of 6 recovering is 23% so could be easily due to chance. But pooling results for all 1200 patients gives overwhelming support to hypothesis that Dr T's cure works – virtually impossible that 400 out of 600 would get better with ½ probability for each.

Big player in meta in medicine is Cochrane a multinational organization headquartered in London that conducts systematic reviews of health care interventions and diagnostic tests, which are published in the Cochrane Library. The Cochrane logo is based on this meta-analyses that led to doctors using corticosteroid widely Their website for this month highlighted: Organised inpatient (stroke unit) care for stroke: network meta-analysis; Antidepressants for smoking cessation: Interventions for the management of malignant pleural effusions: a network meta-analysis



Meta studies now guide "thousands of treatment guidelines and social policies.
Harvard PhD Thesis of Maryaline Catillon shows how you get sample for study of the meta-analyses and gives key aspect of the Cochrane Reviews – reviewers code the methods of RCTs by adequacy and inadquates.

Figure 1: Data flow

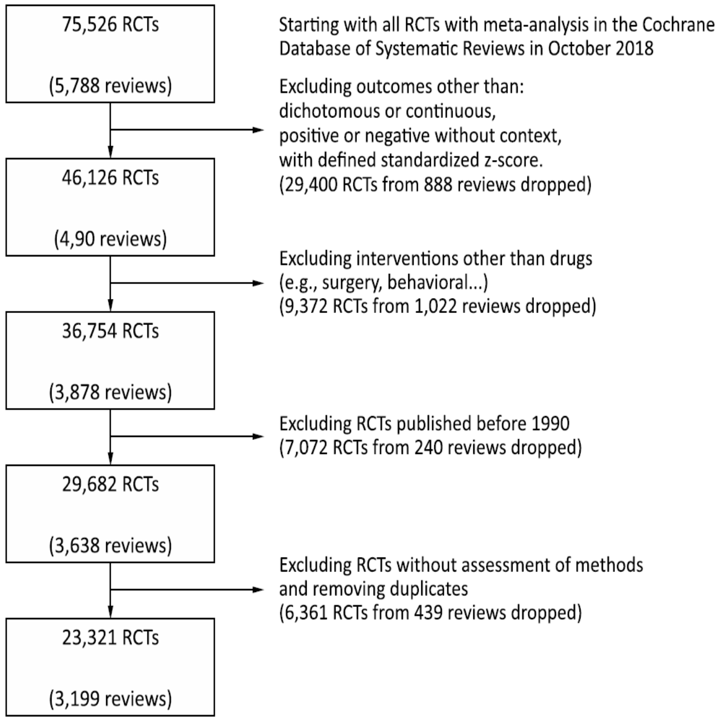


Table 2: Descriptive statist

	RCT	
	Mean	St
Methods		
Adequate*	0.17	
Inadequate* (ordered by frequency)	0.45	
Blinding participants and personnel	0.23	
Incomplete outcome data	0.16	
Selective reporting	0.11	
Blinding outcome assessment	0.10	
Allocation concealment	0.05	
Random sequence generation	0.03	
Unclear*	0.37	
Number of inadequacies	0.68	
Number of unclear methods	1.74	
Number inadequate or unclear	2.42	

Inadequate studies more likely to produce positive results, but get fewer citations and end up in less prestigious journals. But positive results may be more likely to be published.

Deworming Wars: Economists vs Public Health types.
Does Mass Deworming Affect Child Nutrition?Meta-Analysis, Cost-Effectiveness, and Statistical Power
Kevin Croke Joan Hamory Hicks Eric Hsu Michael Kremer Edward Miguel World Bank

Table 2: Tests of the Hypothesis of Common Zero Effect, Adding Updates Individually

Study	Effect Size	P-value
1. TMSDG	**.061**	**.089***
2. TMSDG (using prior Cochrane classifications)	**.092**	**.009*****
3. Sur 2005	.092	.006***
4. Willett 1979	.077	.021**
5. Joseph 2015	.054	.066*
6. Awasthi 2001	.086	.006***
7. Ostwald 1984	.066	.069*
8. Gateff 1972	.082	.019**
9. Liu 2015	.06	.089*
10. Stephenson 1993	.092	.009***
11. Wiria 2013	.062	.084*
12. Ndibazza 2012	.054	.105
13. Gupta 1982a	.06	.09*
14. Gupta 1982b	.065	.063*
15. Hall 2006 (ANCOVA)	.073	.032**
16. Miguel 2004	.056	.121
17. Full Sample (rows 3-16)	**.111**	**<0.001*****

Notes: This table presents meta-analysis treatment effect estimates for the impact of multiple-dose mass deworming on weight using a fixed effects model, and the p-values associated with a test of the null hypothesis of a common zero effect across all studies included in the sample. Row (1) includes the TMSDG sample of studies (described in the notes of Figure 1). Row (2) adds Stephenson 1993

SYSTEMA T I C REVIEW https://onlinelibrary.wiley.com/doi/pdf/10.1002/cl2.1058

Mass deworming for improving health and cognition of children in endemic helminth areas: A systematic review and individual participant data network meta-analysis

Data Collection and Analysis: We conducted NMA with individual participant data (IPD), using a frequentist approach for random-effects NMA. The covariates were: age, sex, weight, height, haemoglobin and infection intensity. The effect estimate chosen was the mean difference for the continuous outcome of interest.

Results: We received data from 19 randomized controlled trials with 31,945 participants. Overall risk of bias was low. There were no statistically significant subgroup effects across any of the potential effect modifiers. However, analyses showed that there may be greater effects on weight for moderate to heavily infected children (very low certainty evidence).

Authors' Conclusions: This analysis reinforces the case against mass deworming at a population-level, finding little effect on nutritional status or cognition. However, children with heavier intensity infections may benefit more. We urge the global community to adopt calls to make data available in open repositories to facilitate IPD analyses such as this, which aim to assess effects for the most vulnerable individuals.

MORE META WARS

21 SEPTEMBER 2018 • SCIENCE sciencemag.org VOL 361 ISSUE 6408

Video games → violence problem of missing – Ferguson and Kilburn have used several statistical methods to measure publication bias and correct for it. Without publication bias, the results would have been distributed symmetrically, and the plot would look like an inverted funnel,centered on the mean. But it didn't; the plot was lopsided. To correct for this bias, they essentially added a "missing"—and supposedly unpublished—study for each study that lacked a counterpart on the other side of the mean. With that and other corrections, the evidence that games and movies made people more aggressive evaporated, they concluded in their 2009 meta-analysis.

Bushman et al tried to find all unpublished studies, mainly by asking the authors of published studies whether they had failed to publish others and checking Ph.D. Theses. for chapters not published in scientific journals. They then included what they had and claimed video games did cause violence. Applying a statistical method to show that the results of the studies were now distributed evenly around the mean reassured them that they had overcome publication bias. The apparent link between video games and aggression persisted.The debate became heated. Ferguson accused his opponents of only collecting unpublished studies with desirable results and "overestimating and overadvertising"the effect—which Bushman and Andersonsaid was "a red herring.

Than another researcher entered and reexamined Bushman's 2010 meta-analysis...Based on those results, they concluded in a 2017 paper in the Psychological Bulletin that Bushman and Anderson hadn't managed to collect all unpublished studies, and that publication bias still played a role. After correcting for that bias, the relationship between violent games and aggression turned out to be "very small," they said. Bushman and Anderson reject Hilgard's analysis andstand by the results of their meta-analysis.

Alternative way to deal with "slippery science" TESTS : Comparing meta-analyses and preregistered multiple-laboratory replication projects Amanda Kvarven Eirik Strømland and Magnus Johannesson NATURE HUMAN BEHAVIOR Dec 23, 2019

Many researchers rely on meta-analysis to summarize research evidence. However, there is a concern that publication bias and selective reporting may lead to biased meta-analytic effect sizes.

We compare the results of meta-analyses to large-scale preregistered replications in psychology carried out at multiple laboratories. The multiple-laboratory replications provide precisely estimated effect sizes that do not suffer from publication bias or selective reporting.

We searched the literature and identified 15 meta-analyses on the same topics as multiple-laboratory replications. We find that meta-analytic effect sizes are significantly different from replication effect sizes for 12 out of the 15 meta-replication pairs. These differences are systematic and, on average, **meta-analytic effect sizes are almost three times as large as replication effect sizes.**

We also implement three methods of correcting meta-analysis for bias, but these methods do not substantively improve the meta-analytic results.