Research Transparency & Reproducibility in Economics and Beyond

Edward Miguel // University of California, Berkeley

NBER Summer Institute Methods Lecture // July 2019

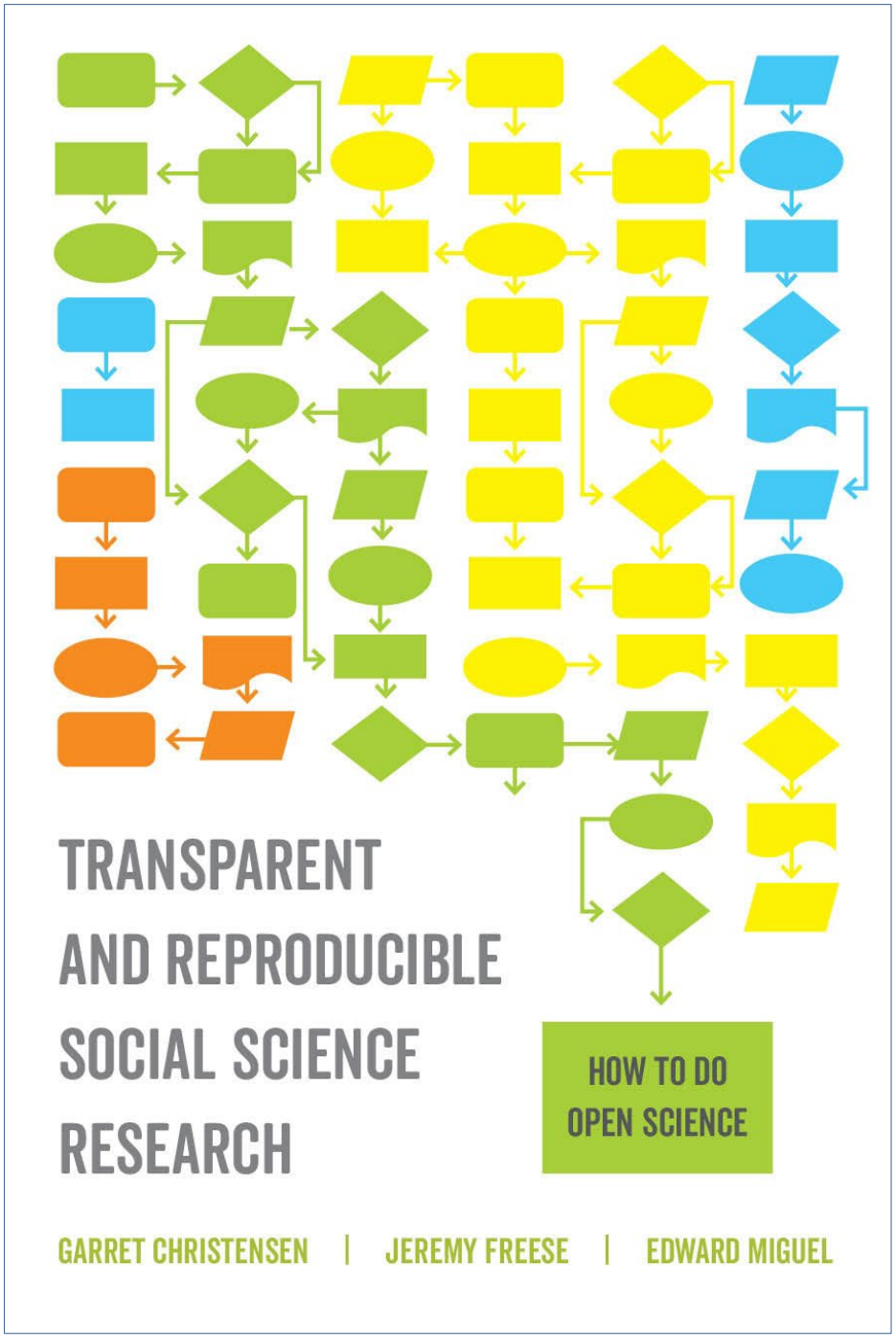
Overview

- Talk outline:

1. Introduction and overview
2. What are research transparency and open science?
3. Problems in economics and social science research
4. What does research transparency do?
5. Innovations in open science, with a focus on pre-registration
6. Looking forward

Overview

- Key references:
  - Casey, Glennerster and Miguel. (2012). “Reshaping Institutions: Evidence on Aid Impacts Using a Pre-analysis Plan”, *Quarterly Journal of Economics*, 127(4), 1755-1812.
  - Miguel et al. (2014). “Promoting Transparency in Social Science Research”, *Science*, 10.1126/science.1245317.
  - Christensen and Miguel. (2018). “Transparency, Reproducibility, and the Credibility of Economics Research”, *Journal of Economic Literature*, 56(3), 920-980.
  - Christensen et al. (2019). “Open Science Practices are on the Rise Across Four Social Science Disciplines”, working paper.
  - *Christensen, Freese and Miguel. (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science*, University of California Press.



>> Free advance book copies for the first 10 people who email me today at emiguel@berkeley.edu.

Overview

- Major topics covered in the book:
  - Ethical research
  - Publication bias
  - Specification searching
  - Study registration and pre-analysis plans
  - Meta-analysis and meta-regression
  - Multiple testing adjustments
  - Data sharing and differential privacy
  - Disclosure and other reporting standards
  - Replication
  - Reproducible coding, workflow

>> Online lectures also available via FutureLearn (UK), and the Berkeley Initiative for Transparency in the Social Sciences (bitss.org).

Overview

- Talk outline:

1. Introduction and overview
2. What are research transparency and open science?
3. Problems in economics and social science research
4. What does research transparency do?
5. Innovations in open science, with a focus on pre-registration
6. Looking forward

What is research transparency?

- Research transparency is advanced when scientific claims are independently verifiable, including through the promotion of free and open sharing of the process of conducting research, and the content and findings generated during research.
- Benefits for research quality and credibility: results that can be verified, and shown to be largely free of investigator bias, are more convincing.

>> Next step in the "credibility revolution" in Economics (Leamer 1983, Card and Krueger 1995, Angrist and Pischke 2010)

- A normative perspective: research transparency values resonate with the classical "scientific ethos" (Merton 1942).

Scientific norms (Merton 1942)

- Four core values:

1. Universalism

Scientific norms (Merton 1942)

- Four core values:

1. Universalism

- *"The acceptance or rejection of claims … is not to depend on the personal or social attributes of their protagonist."*

>> Research findings are fundamentally "impersonal".

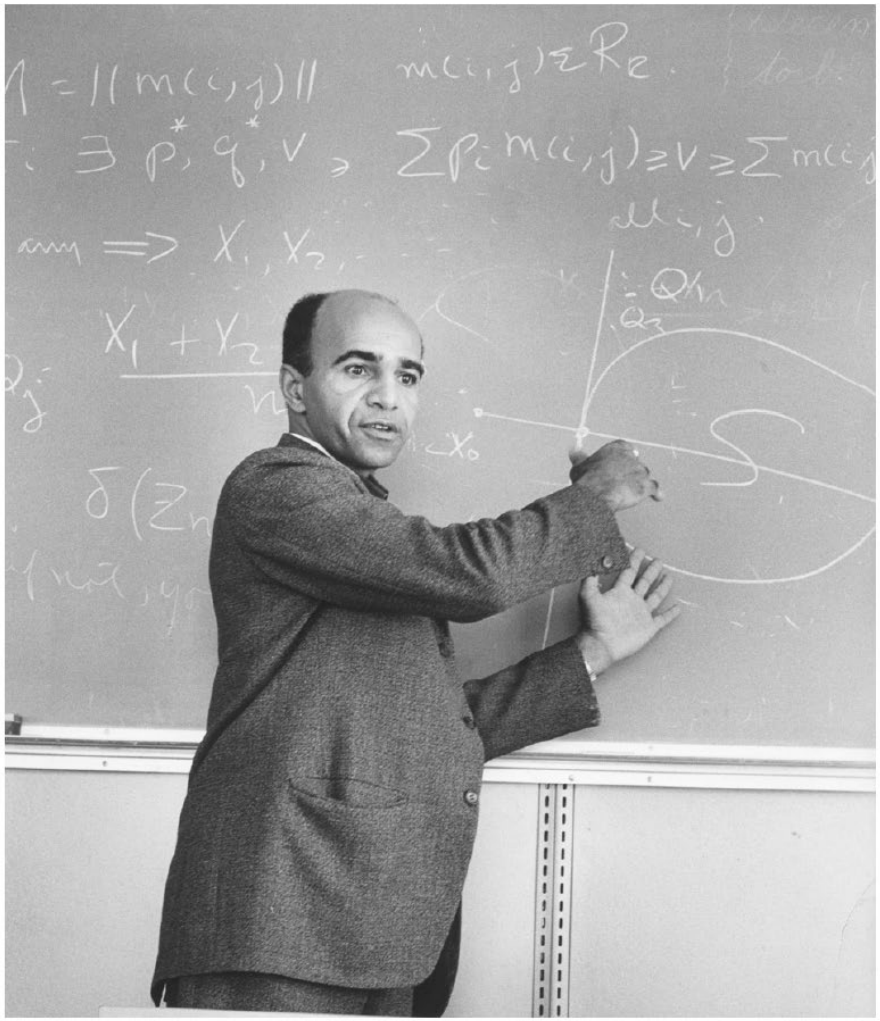
- *"Universalism finds further expression in the demand that careers be open to talents."*

>> Link between democracy, opportunity, and scientific progress?

Scientific norms (Merton 1942)

- Four core values:

1. Universalism



David Blackwell in the classroom. David Blackwell papers, BANC MSS 2001/79. Courtesy of the Bancroft Library, University of California, Berkeley.

Scientific norms (Merton 1942)

- Four core values:

1. Universalism
2. Communality

- *"The substantive findings of science are a product of social collaboration and are assigned to the community."*
- *"Secrecy is the antithesis of this norm; full and open communication its enactment."*

>> Open sharing of scientific knowledge

Scientific norms (Merton 1942)

- Four core values:

1. Universalism
2. Communality
3. Disinterestedness

- *"A passion for knowledge, idle curiosity, altruistic concern with the benefit to humanity, and a host of other special motives have been attributed to the scientist."*

>> Researchers should be motivated by identifying the truth rather than (selfish) professional or monetary motivations.

Scientific norms (Merton 1942)

- Four core values:

1. Universalism
2. Communality
3. Disinterestedness
4. Organized skepticism

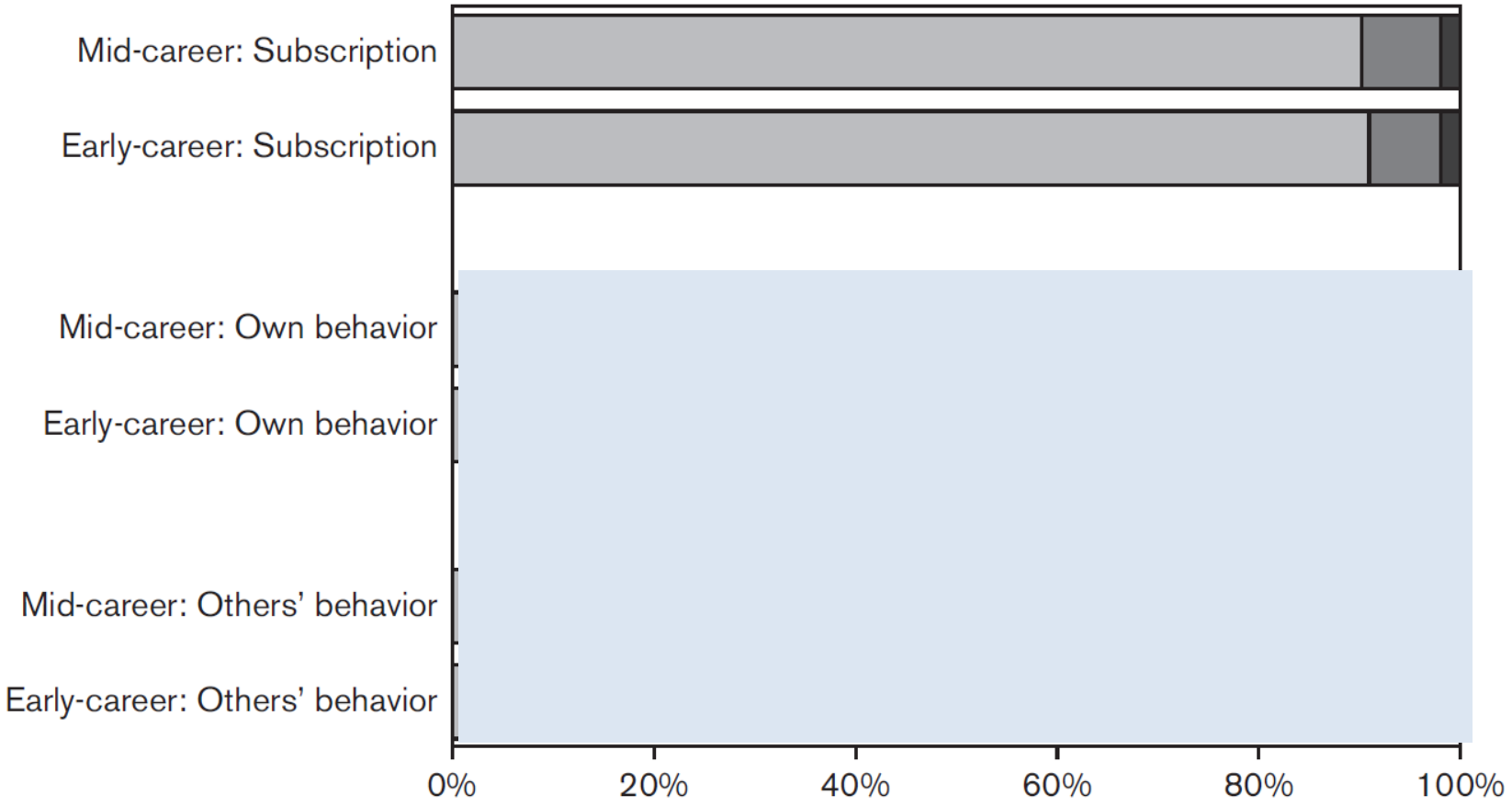
- *"Involving as it does the verifiability of results, scientific research is under the exacting scrutiny of fellow experts. … The activities of scientists are subject to rigorous policing, to a degree perhaps unparalleled in any other field of activity."*

>> The ability to verify data and scrutinize others' claims is critical for research credibility and progress.

Scientific norms (Merton 1942)

- How closely do scholars conform to these ideals?
- Anderson et al (2007) examine attitudes, beliefs and practices among N=3,247 early and mid-career U.S. researchers funded by NIH

>> In particular, attachment to Mertonian norms (e.g., Communality, Disinterestedness) vs. counter-norms (Secrecy, Self-interestedness), and beliefs about other scholars.



~90%

FIGURE 2.1. Attitudes, beliefs, and practices of early-career and mid-career U.S. researchers (*N* = 3,247) in regard to six pairs of scientific norms and counter-norms (see Table 2.1). Light gray indicates the proportion expressing more support for the norms, dark gray the proportion expressing roughly equal support for both the norms and the counter-norms, and black the proportion expressing more support for the counter-norms. Reprinted with permission from Anderson et al. (2007).

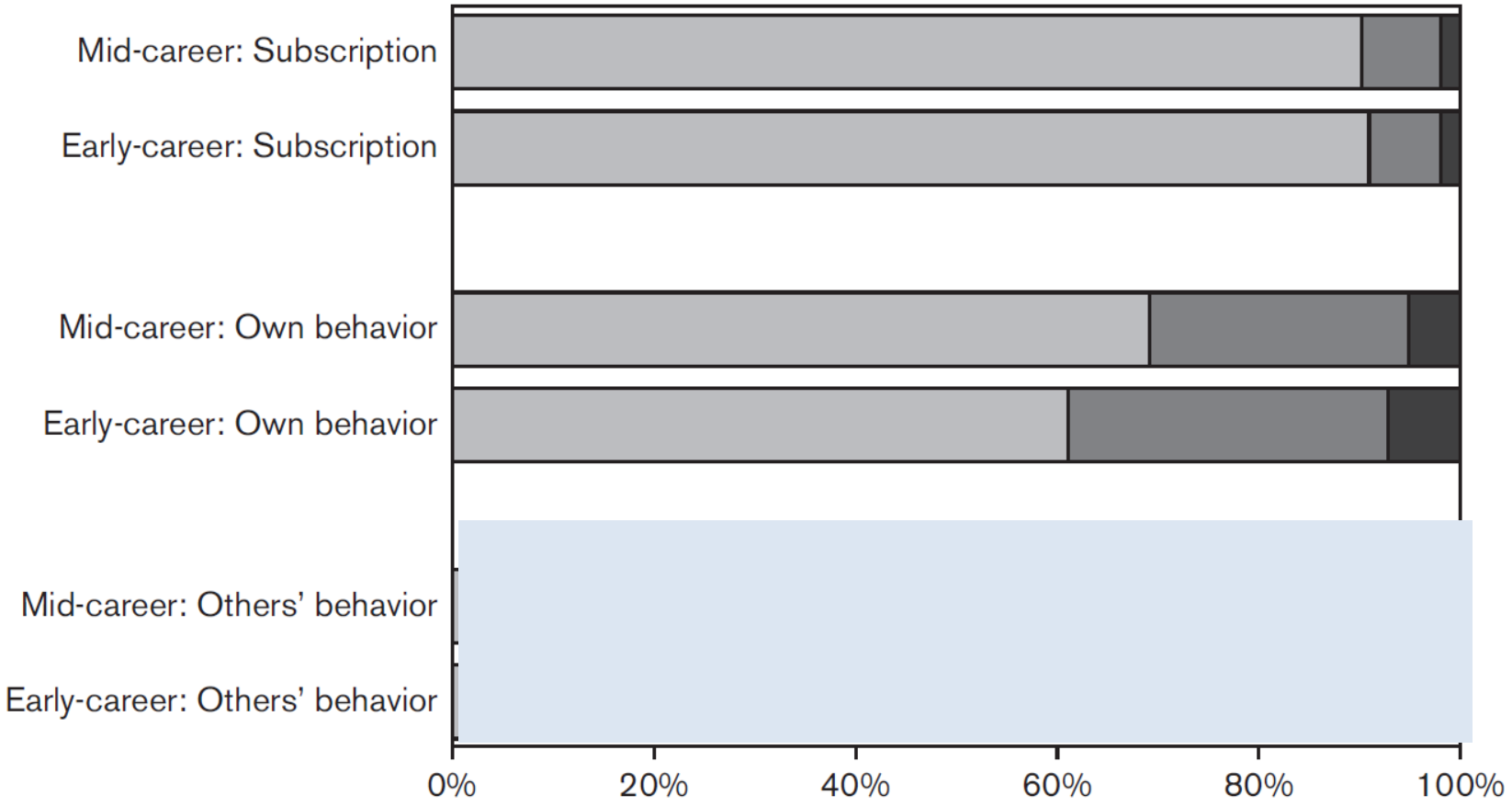


FIGURE 2.1. Attitudes, beliefs, and practices of early-career and mid-career U.S. researchers (*N* = 3,247) in regard to six pairs of scientific norms and counter-norms (see Table 2.1). Light gray indicates the proportion expressing more support for the norms, dark gray the proportion expressing roughly equal support for both the norms and the counter-norms, and black the proportion expressing more support for the counter-norms. Reprinted with permission from Anderson et al. (2007).

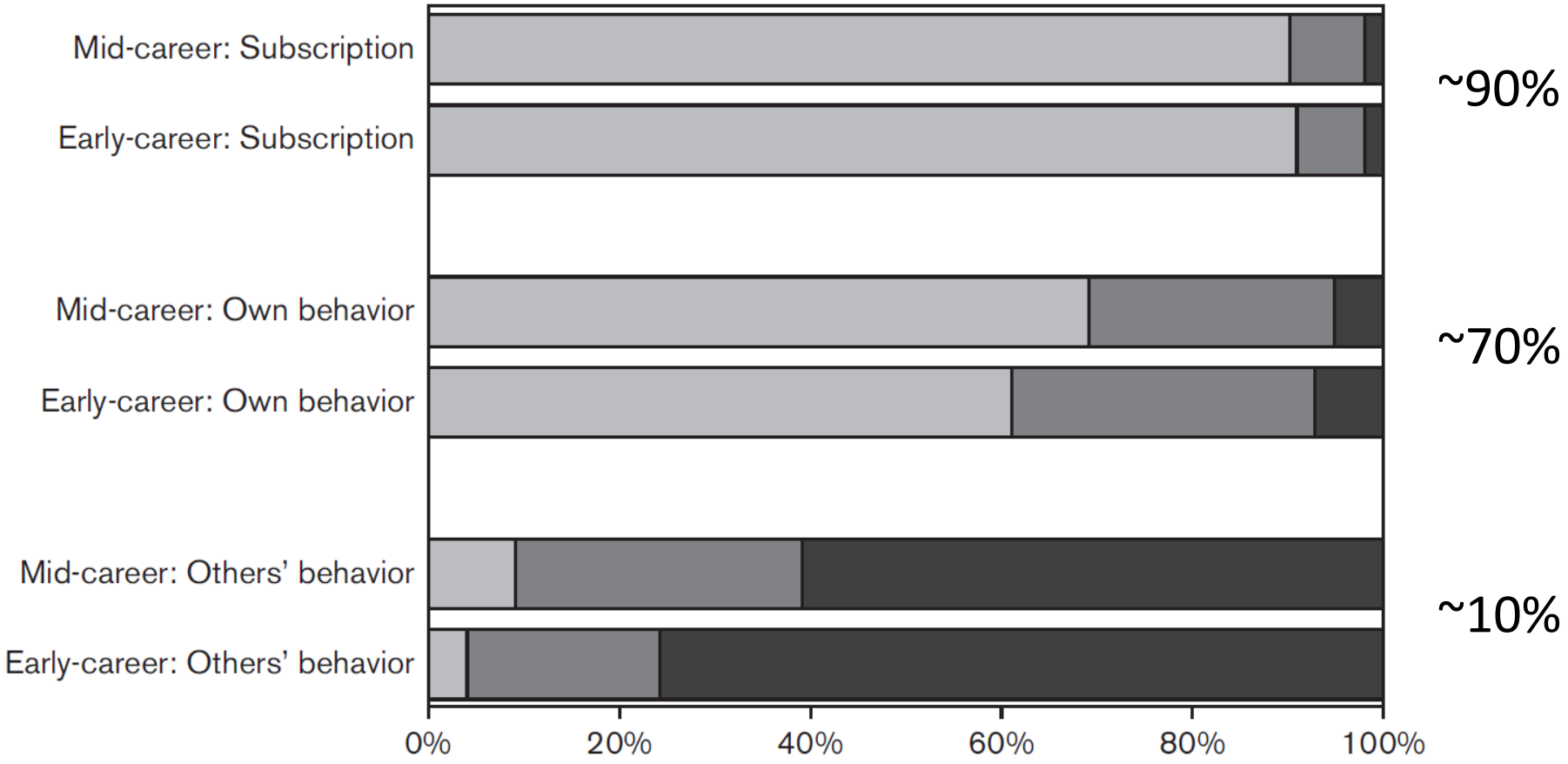


FIGURE 2.1. Attitudes, beliefs, and practices of early-career and mid-career U.S. researchers (*N* = 3,247) in regard to six pairs of scientific norms and counter-norms (see Table 2.1). Light gray indicates the proportion expressing more support for the norms, dark gray the proportion expressing roughly equal support for both the norms and the counter-norms, and black the proportion expressing more support for the counter-norms. Reprinted with permission from Anderson et al. (2007).

Scientific norms (Merton 1942)

- How closely do scholars conform to these ideals?
- Anderson et al (2007) examine attitudes, beliefs and practices among N=3,247 early and mid-career U.S. researchers funded by NIH

>> In particular, attachment to Mertonian norms (e.g., Communality, Disinterestedness) vs. counter-norms (Secrecy, Self-interestedness), and beliefs about other scholars.

- Finding: many researchers subscribe to Mertonian research norms but believe that most other scholars do not follow them, leading to what the authors call "normative dissonance"

>> How can research be brought back in line with the scientific ethos?

>> And how severe are real-world problems?

Overview

- Talk outline:

1. Introduction and overview
2. What are research transparency and open science?
3. Problems in economics and social science research
4. What does research transparency do?
5. Innovations in open science, with a focus on pre-registration
6. Looking forward

Threats to validity of research

- There is now a large body of evidence documenting problems with the process of scientific research across many disciplines.
- Three leading concerns:

(1) Fraud

(2) Publication bias

(3) Failure to replicate

Threats to validity of research

- There is now a large body of evidence documenting problems with the process of scientific research across many disciplines.

(1) Fraud: corrosive, even if rare: undermines public trust in science

- Widely publicized cases of leading scholars fabricating data in social psychology (Simonsohn 2013) and political science (Broockman, Kalla and Aranow 2015), e.g., case of psychologist Diederik Stapel.

>> Fraud in psychology: discovered when data showed "too little" variation (in means across treatment arms).

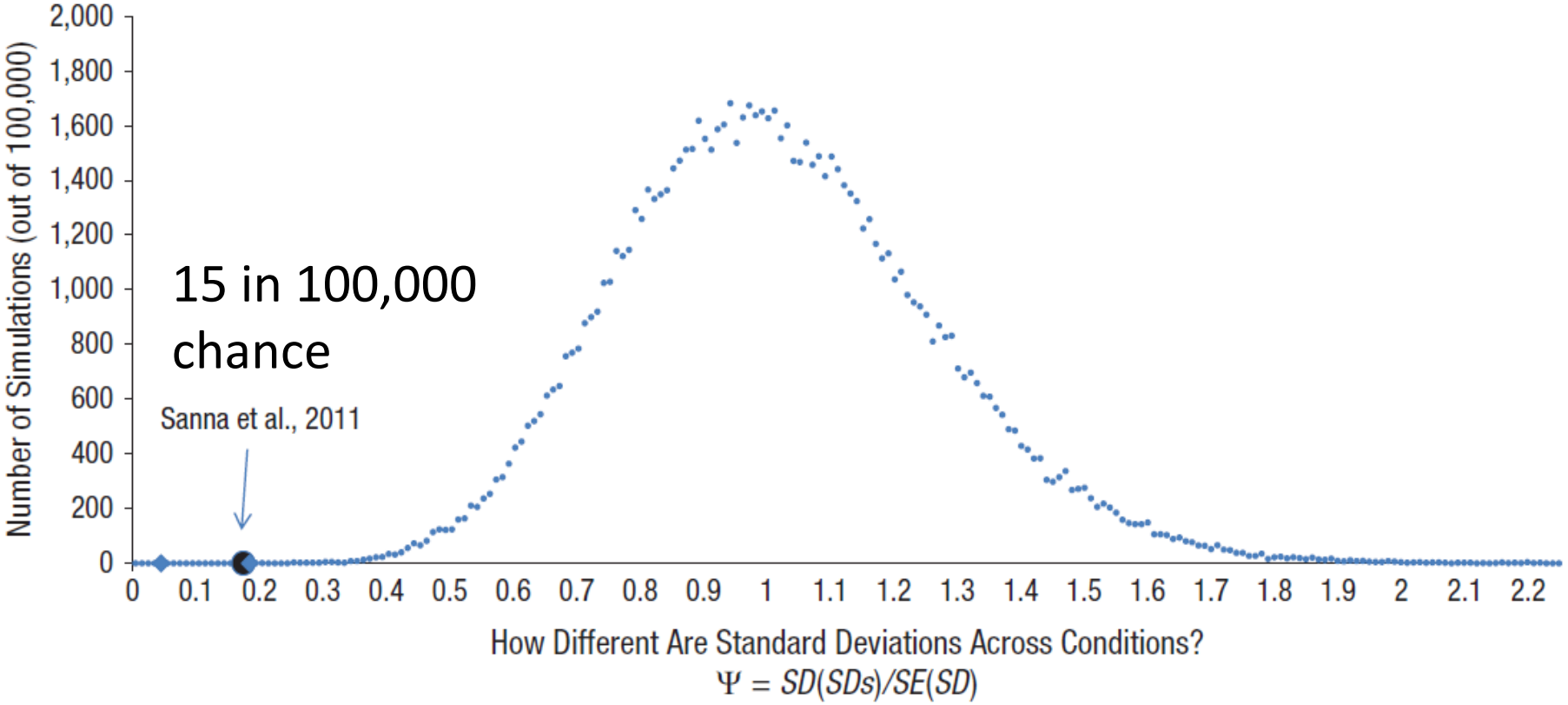


Fig. 2. Illustration of the extreme improbability of the similarity of the standard deviations reported for Sanna's three experiments on the embodiment of morality (see the text). Each condition in each experiment was simulated by drawing from a normal distribution with a mean equal to the sample mean for that condition and a standard deviation equal to the pooled standard deviation across all conditions. The standard deviation of the standard deviations for each simulated experiment was divided by the standard error of the pooled standard deviation. The graph shows the average for this value (Ψ) across the three experiments for both Sanna's reported data and 100,000 simulations of the experiments. Only 15 of the simulations yielded values that were as extreme as those for the published results.

Simonsohn (2013)

Threats to validity of research

- There is now a large body of evidence documenting problems with the process of scientific research across many disciplines.

(1) Fraud: corrosive, even if rare: undermines public trust in science

- Widely publicized cases of leading scholars fabricating data in social psychology (Simonsohn 2013) and political science (Broockman, Kalla and Aranow 2015), e.g., case of psychologist Diederik Stapel.

- Many instances of fraud in science: Gregor Mendel (probably) fabricated his famous data on the phenotypes of peas, his data showing far less variation than would prevail by chance (Fisher 1936)

>> Open data and code help uncover research fraud:

"Fictitious data can seldom survive careful scrutiny." Ronald Fisher (1936).

Threats to validity of research

(2) Publication bias: comes in many forms and is widely documented across nearly all scientific fields

- Missing studies / the “file-drawer problem” (Rosenthal 1979): studies with “null” (not statistically significant) findings, or certain other characteristics (e.g., controversial, run against conventional wisdom), are less likely to be published, leading to biased bodies of evidence

>> Null findings were rarely published in psychology then…

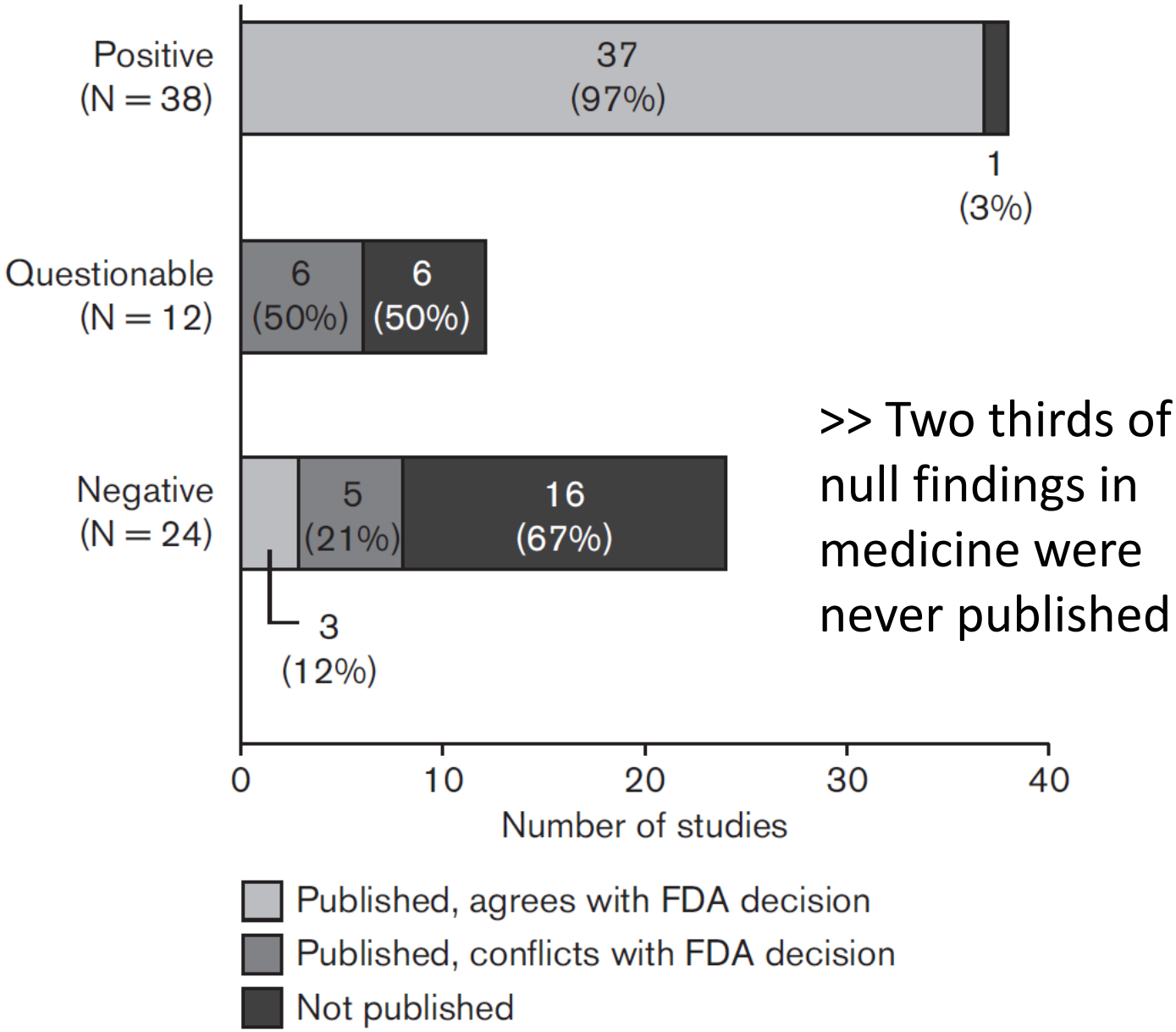
TABLE 3.2 TESTS OF SIGNIFICANCE IN FOUR PSYCHOLOGY JOURNALS

Journal and year	Articles reporting tests of significance	Articles that reject H_0 with $p < .05$	Articles that fail to reject H_0	Articles that are replications of previous studies
Experimental Psychology (1955)	106	105	1	0
Comparative and Physiological Psychology (1956)	94	91	3	0
Clinical Psychology (1955)	62	59	3	0
Social Psychology (1955)	32	31	1	0

NOTES: Data are from Sterling (1959). H_0 is the null hypothesis of no effect.

Significant

Null



>> Two thirds of null findings in medicine were never published

FIGURE 5.1. Publication bias in clinical trials of antidepressants (*N* = 74 studies). Reprinted with permission from Turner et al. (2008).

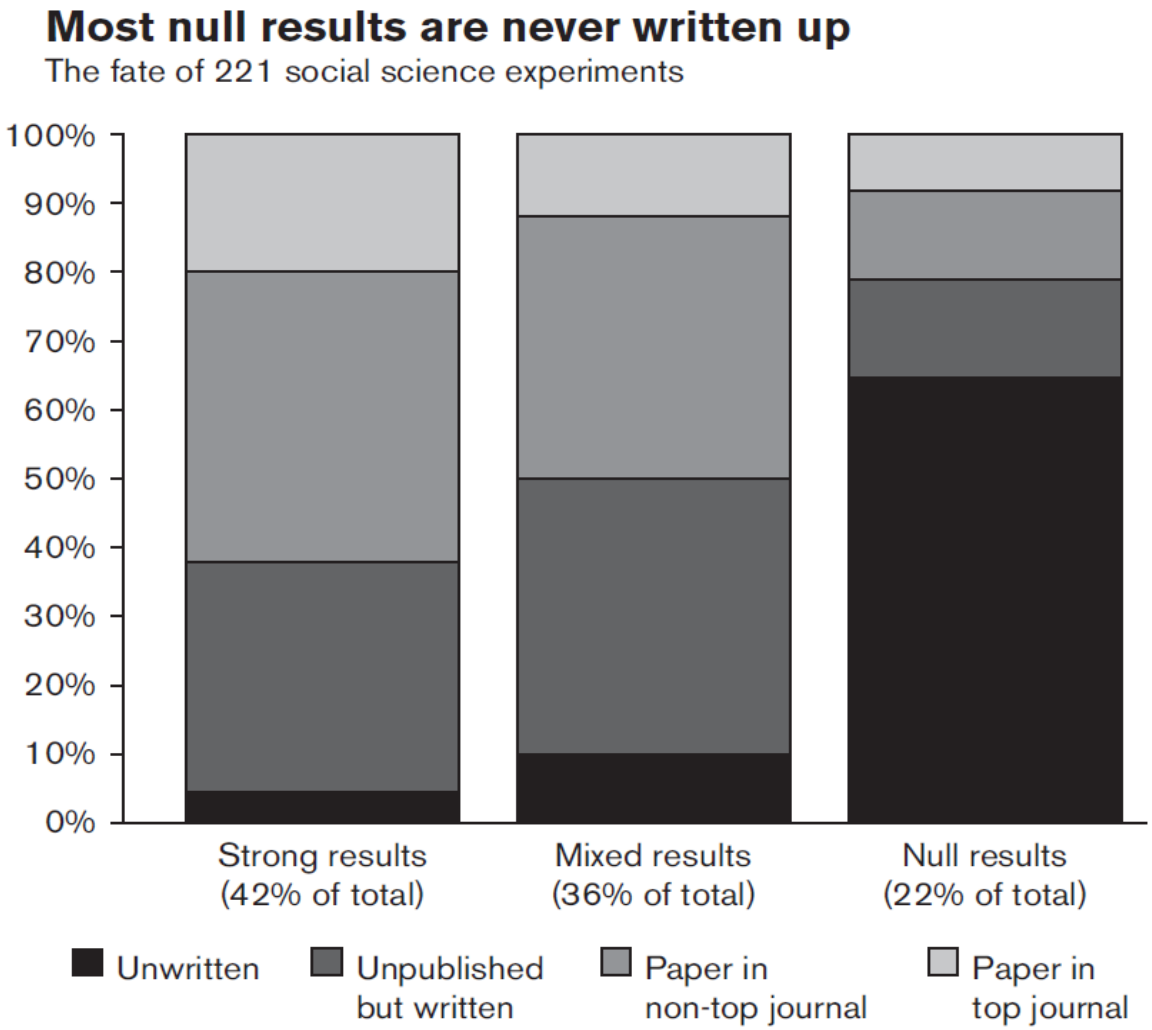


FIGURE 3.1. Publication rates and rates of writing-up of results from experiments with strong, mixed, and null results. These 221 experiments represent nearly the complete universe of studies conducted by the Time-sharing Experiments for the Social Sciences. The figure is from Mervis (2014), based on data from Franco, Malhotra, and Simonovits (2014). Reprinted with permission from AAAS.

>> Two thirds of null findings in the social sciences not written up.

>> Significant results are 3-4x more likely to be published.

Threats to validity of research

(2) Publication bias:

- Missing studies / the “file-drawer problem” (Rosenthal 1979): studies with “null” (not statistically significant) findings, or certain other characteristics (e.g., controversial, run against conventional wisdom), are less likely to be published, leading to biased bodies of evidence
- The lack of published null results is due to a combination of authors failing to submit these studies, and editor and referee decisions

>> May lead to wasted effort as other scholars re-do research that (unbeknownst to them) was already carried out, as well as inappropriate public policy decisions based only on the published evidence.

Threats to validity of research

(2) Publication bias:

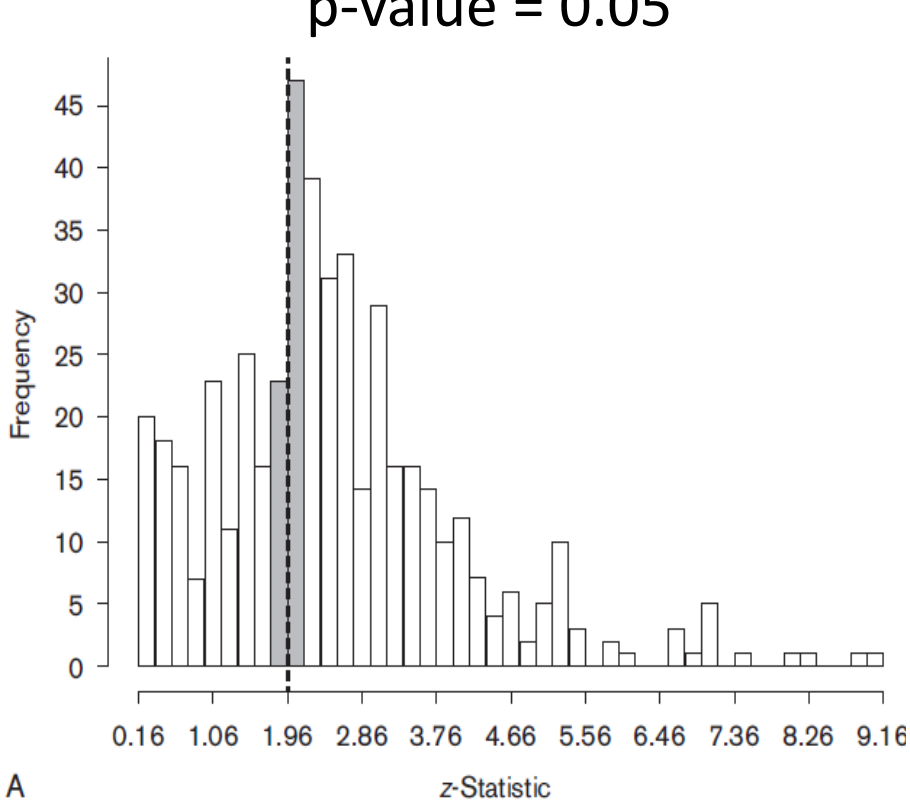
- Missing studies / the “file-drawer problem” (Rosenthal 1979)

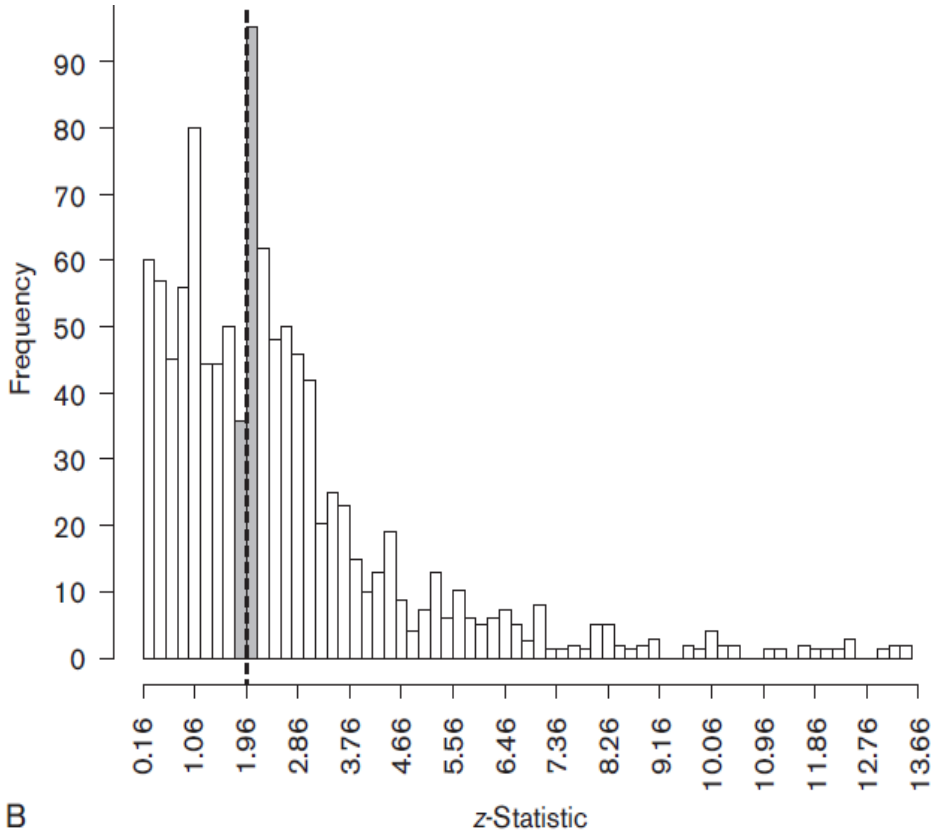
- Author manipulation of results/“p-hacking” (Gerber and Malhotra 2008): excess mass of studies with p-values just below “critical” thresholds, i.e., 0.05, and other statistical patterns that would not be generated in the absence of author bias and/or publication bias.

Sociology journals

Political Science journals

p-value = 0.05

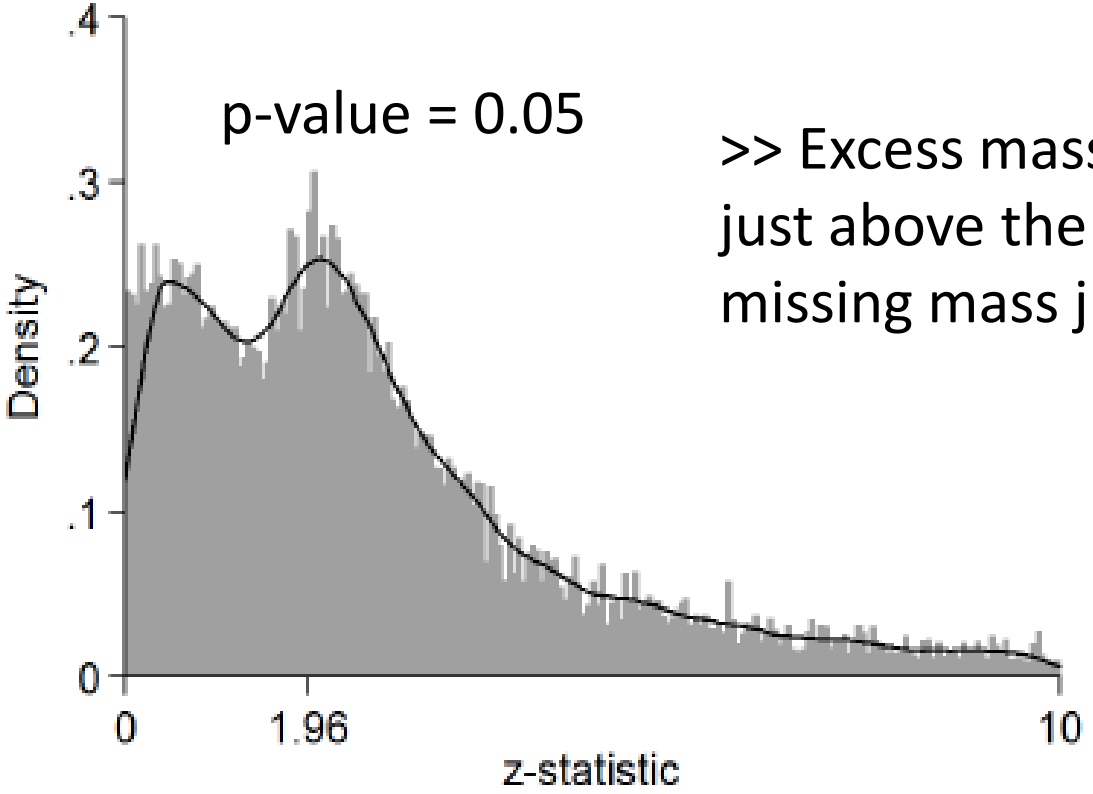




>> Excess mass of values just above the critical 1.96 level, missing mass just below.

FIGURE 3.2. Collection of *Z*-statistics from top sociology and political science journals: histograms from (**A**) *American Sociological Review*, *American Journal of Sociology*, and *Sociological Quarterly* (two-tailed) and (**B**) *American Political Science Review* and *American Journal of Political Science* (two-tailed). Width of bars (0.20) approximately represents 10 percent caliper. Dotted line represents critical *Z*-statistic (1.96) associated with $p = .05$ significance level for one-tailed tests. Reprinted with permission from Gerber and Malhotra (2008a, 2008b).

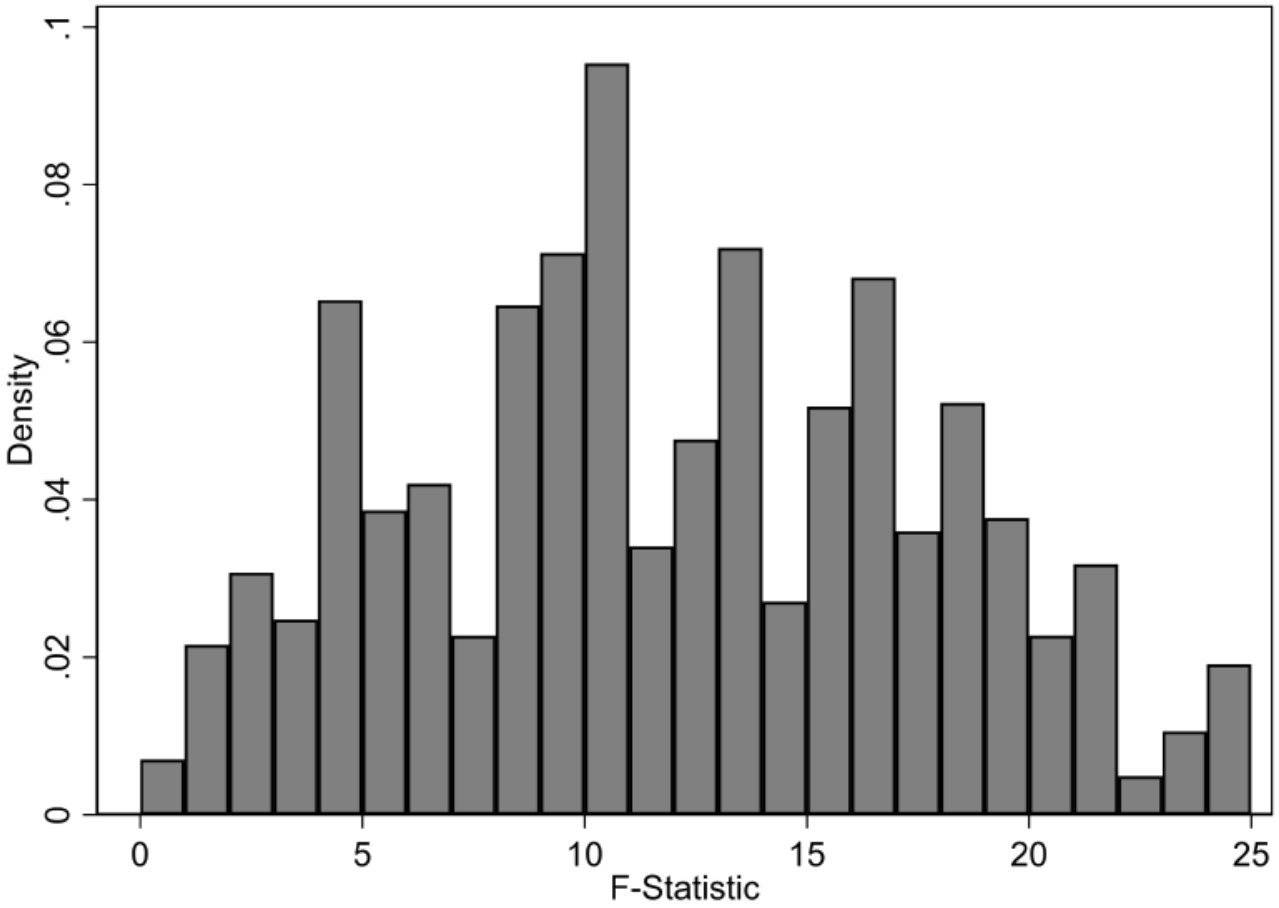
Economics journals: AER, QJE JPE (2005-2011)



>> Excess mass of values
just above the critical 1.96 level,
missing mass just below.

Brodeur et al (2016)

Figure 4: Instrumental Variable: First Stage F-Statistics



Notes: This figure displays a histogram of first stage F-Statistics of instrumental variables for $F \in [0, 25]$.

>> Excess mass above "critical" IV 1st stage F-test value of 10 in economics.

Brodeur et al (2019)

Threats to validity of research

(2) Publication bias:

- Missing studies / the “file-drawer problem” (Rosenthal 1979)

- Author manipulation of results/“p-hacking” (Gerber and Malhotra 2008): excess mass of studies with p-values just below “critical” thresholds, i.e., 0.05, and other statistical patterns that would not be generated in the absence of author bias and/or publication bias.

>> Role of author manipulation: 40% of significant coefficient estimates in *AJPS* are sensitive to covariate adjustment (Lenz and Sahn 2017);

and >45% of failed “sniff tests” (e.g., balance tables, IV over-ID tests) in economics appear to go unreported (Snyder and Zhuo 2018).

Threats to validity of research

(2) Publication bias:

- How can we correct for publication bias in bodies of evidence?
- Long tradition of attempting to "undo" effects of publication bias (e.g., Rosenthal 1979 "Fail-safe N"; Stanley 2008 "Precision-Effect test"; Simonsohn, Nelson & Simmons 2014 "P-curve"; McCrary et al 2015)
- Under assumptions on data generating process, can back out true effects (Hedges 1992), e.g., re-weight existing estimates by inverse of the publication probability as a function of the test statistic, p(Z).

Threats to validity of research

(2) Publication bias:

- How can we correct for publication bias in bodies of evidence?
- Long tradition of attempting to "undo" effects of publication bias (e.g., Rosenthal 1979 "Fail-safe N"; Stanley 2008 "Precision-Effect test"; Simonsohn, Nelson & Simmons 2014 "P-curve"; McCrary et al 2015)
- Under assumptions on data generating process, can back out true effects (Hedges 1992), e.g., re-weight existing estimates by inverse of the publication probability as a function of the test statistic, p(Z).

>> Simplest test: examine relationship between estimated effect sizes ($\widehat{\beta_i}$) and associated standard errors ($\widehat{\sigma_j}$) in a funnel plot (or related plot).

$$\widehat{\beta_i} = b_0 + b_1\left(1/\widehat{\sigma_j}\right) + e_j$$

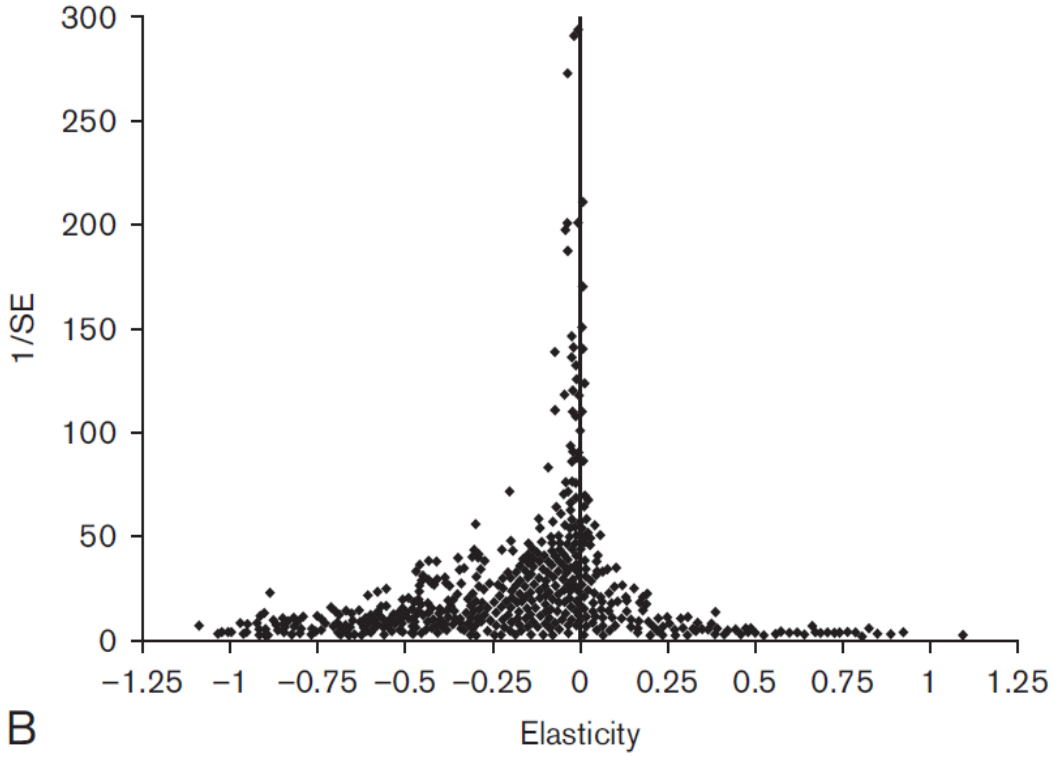


FIGURE 5.4. Examples of funnel graphs from the union and minimum wage literature in labor economics. The more symmetric graph (**A**) is from the union productivity literature, while the more skewed graph (**B**) shows estimates from the minimum wage literature. On each graph, the estimate is on the horizontal axis and precision on the vertical axis, with each dot representing an estimate from a study or paper. Panel A was constructed by the authors using data from Doucouliagos and Laroche (2003). Panel B is from Doucouliagos and Stanley (2009) and is used by permission (© Blackwell Publishing Ltd/ London School of Economics 2009).

>> Funnel plot of estimates, minimum wage literature.
(a) Precise estimates are all clustered close to zero, indicating no effect.
(b) Imprecise estimates are asymmetric, with most underpowered studies showing negative employment effects, in a manner highly suggestive of publication bias.

Threats to validity of research

(2) Publication bias:

- Implications for the credibility of published research? (Ioannidis 2005).
- Positive predictive value of research (PPV) in research field i, likelihood that a claimed significant relationship is actually true:

$$PPV_i = \frac{(1-\beta)R_i + u\beta R_i}{\{(1-\beta)R_i + u\beta R_i\} + \alpha + u(1-\alpha)}$$

where statistical power is $(1-\beta)$, significance level α, author bias u, and R_i is ratio of true to null relationships in field i (e.g., development)

Threats to validity of research

(2) Publication bias:

- Implications for the credibility of published research? (Ioannidis 2005).
- Positive predictive value of research (PPV) in research field i, likelihood that a claimed significant relationship is actually true:

$$PPV_i = \frac{(1-\beta)R_i + u\beta R_i}{\{(1-\beta)R_i + u\beta R_i\} + \alpha + u(1-\alpha)}$$

where statistical power is $(1-\beta)$, significance level α, author bias u, and R_i is ratio of true to null relationships in field i (e.g., development)

>> In an arguably realistic case of 50% power, 5% significance, one third of tested relationships exist (R_i=0.5), and 30% author bias: PPV=0.49, so more than half of all results are “false positives”.

Threats to validity of research

(3) Lack of replicability: computational reproducibility ("verification") is often challenging in social science research (Dewald et al 1986)

- Journal data posting requirements – starting with the AER in 2005 – have helped, but recent attempts indicate that still only about a third (Galiani et al 2018) to a half (Chang and Li 2015) of empirical economics papers can be readily reproduced.

>> New AEA data and code posting requirements (Nosek et al 2015)

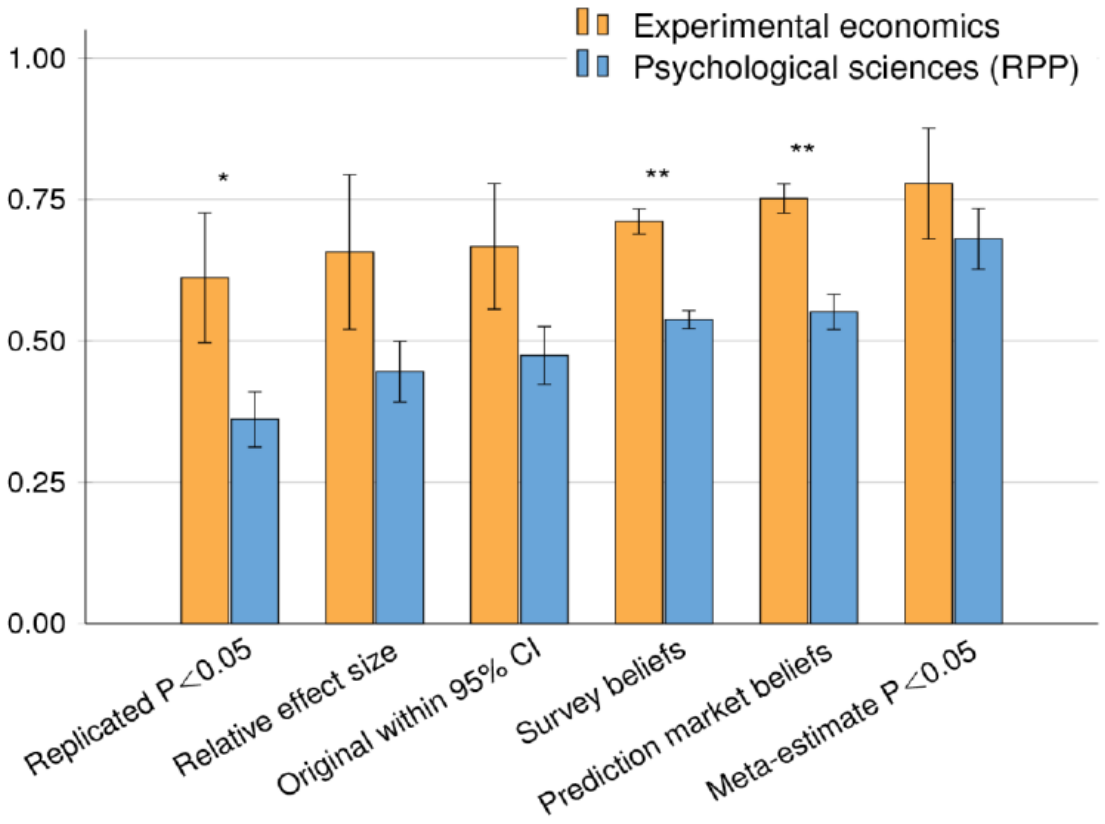
Threats to validity of research

(3) Lack of replicability: computational reproducibility ("verification") is often challenging in social science research (Dewald et al 1986)

- Journal data posting requirements – starting with the AER in 2005 – have helped, but recent attempts indicate that still only about a third (Galiani et al 2018) to a half (Chang and Li 2015) of empirical economics papers can be readily reproduced.

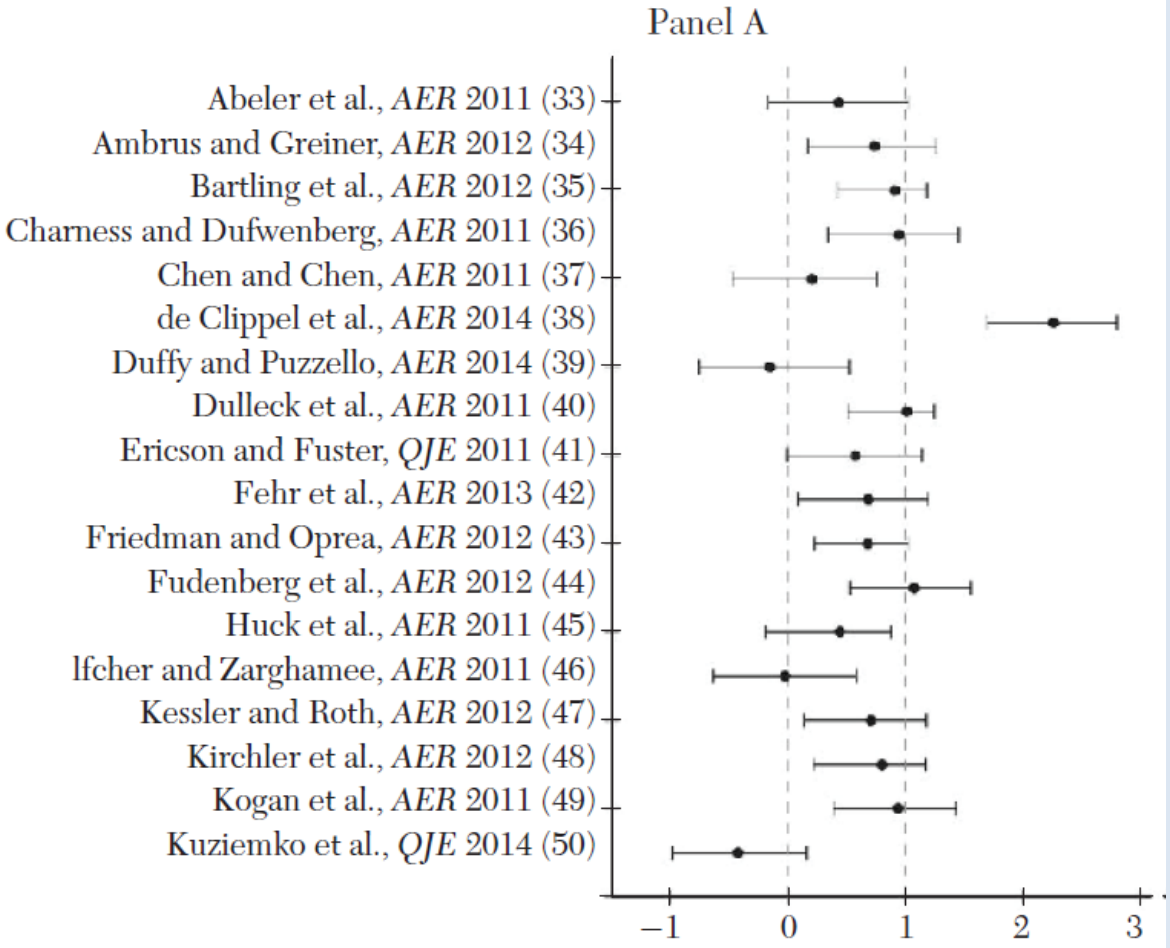
>> New AEA data and code posting requirements (Nosek et al 2015)

- Beyond verification, many prominent findings fail to replicate in lab experimental settings (Open Science Collaboration 2015, Camerer et al 2016): pre-registered and well-powered replication studies reject original study point estimates over half the time in psychology, and in a third of experimental economics papers published in the AER, QJE.



Partial replicability of lab experimental results in economics and psychology, Camerer et al (2016)

Fig. 4. A comparison of different reproducibility indicators between experimental economics and psychological sciences (the Reproducibility Project Psychology). Error bars denotes ±se. The reproducibility is higher for experimental economics for all six reproducibility indicators; this difference is significant for three of the reproducibility indicators. The average difference in reproducibility across the six indicators is 19 percentage points. See the Supplementary Materials for details about the statistical tests. *P<0.05 for the difference between experimental economics and psychological sciences, **P<0.01 for the difference between experimental economics and psychological sciences.



>> Inflated estimates: replications typically have far smaller estimated effects, often by half, Camerer et al (2016)

Figure 3. Replicability in Experimental Economics

Notes: Figure from Camerer et al. (2016). Reprinted with permission from AAAS. Panel A: Plotted are 95% CIs of replication effect sizes (standardized to correlation coefficients). The standardized effect sizes are normalized so that 1 equals the original effect size (fig. S1 in Camerer et al. 2016 shows a non-normalized version). Eleven replications have a significant effect in the same direction as in the original study [61.1%; 95% CI = (36.2%, 86.1%)]. The 95% CI of the replication effect size includes the original effect size for twelve replications [66.7%; 95% CI = (42.5%, 90.8%)]; if one also includes the study in which the entire 95% CI exceeds the original effect size, this increases to thirteen replications [72.2%; 95% CI = (49.3%, 95.1%)].

Overview

- Talk outline:

1. Introduction and overview
2. What are research transparency and open science?
3. Problems in economics and social science research
4. What does research transparency do?
5. Innovations in open science, with a focus on pre-registration
6. Looking forward

Impacts of transparency

- The frontier in meta-science has shifted from solely focusing on documenting problems, to studying the impact of open science practices and other interventions on the research process.
- Four major themes:

(1) Open data and replications help diagnose, correct for publication bias

(2) Experimental methods are associated with less publication bias

(3) Study registration may increase reporting of null results

(4) Journal editorial practices are influential

Impacts of transparency

(1) Open data and replications help diagnose, correct for publication bias

- New estimates of how many studies "disappear", e.g., Franco et al (2014), produce more realistic assumptions about what the publication probability is for studies with different results / statistical significance.
- Andrews and Kasy (2019) can validate their model of study selection into publication with the Camerer et al (2016) and Open Science Collaboration (2015) replication results, to get a handle on the degree of estimate "inflation" and publication bias

Impacts of transparency

(1) Open data and replications help diagnose, correct for publication bias

- New estimates of how many studies "disappear", e.g., Franco et al (2014), produce more realistic assumptions about what the publication probability is for studies with different results / statistical significance.
- Andrews and Kasy (2019) can validate their model of study selection into publication with the Camerer et al (2016) and Open Science Collaboration (2015) replication results, to get a handle on the degree of estimate "inflation" and publication bias

>> Their key contribution: examine the joint distribution of estimates from original studies, Z, and replication studies, Z^r to estimate publication probability as a function of significance. Assume publication prob. is 1 for studies with $|Z| \geq 1.96$, and estimate β_p for $|Z| < 1.96$.

Impacts of transparency

(1) Open data and replications help diagnose, correct for publication bias

- In Andrews and Kasy (2019), the marginal density of (Z, Z^r), $f_{Z,Z^r}(z, z^r)$, should be symmetric in the absence of publication bias, so the degree of asymmetry in estimates reveals the relative probability of publication as a function of the results:

$$\frac{f_{Z,Z^r}(b,a)}{f_{Z,Z^r}(a,b)} = \frac{p(b)}{p(a)}$$

>> Let "a" denote a significant $|Z| \geq 1.96$, and "b" not significant. An estimate for β_p is (roughly) the share of cases where the original estimate is not significant but the replication is significant, divided by the share where the original estimate is significant but the replication is not.

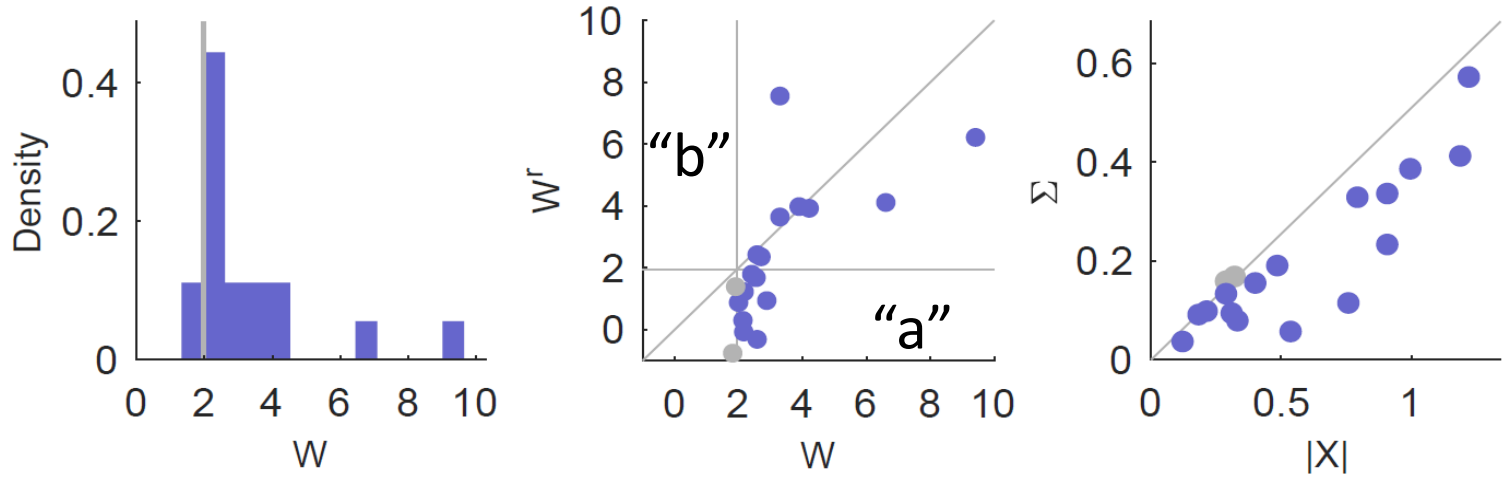
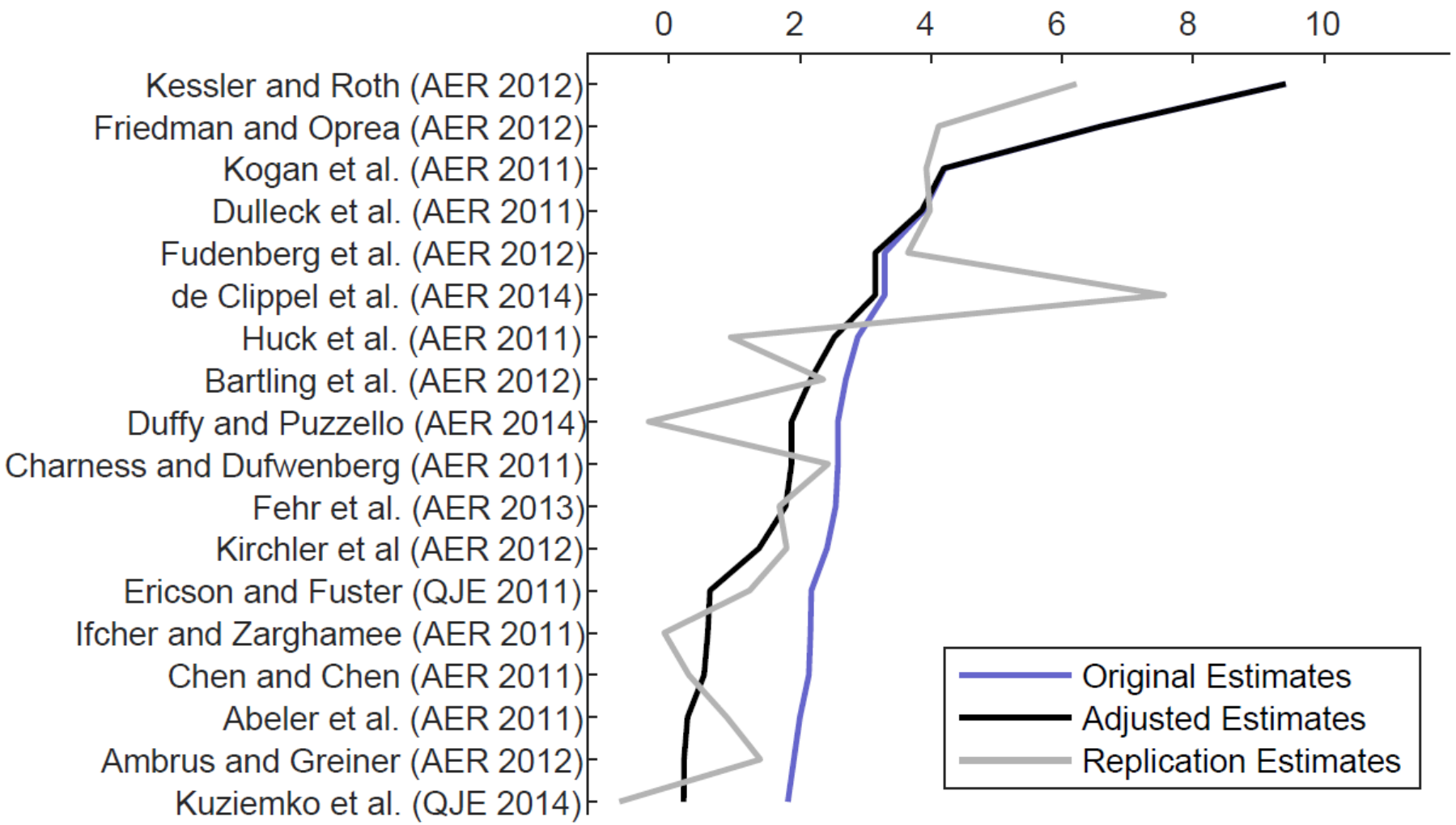


Figure 5: The left panel shows a binned density plot for the normalized z-statistics $W = |X|/\Sigma$ using data from Camerer et al. (2016). The grey line marks $W = 1.96$. The middle panel plots the z-statistics W from the initial study against the estimate W^r from the replication study. The grey lines mark W and $W^r = 1.96$, as well as $W = W^r$. The right panel plots the initial estimate $|X| = W \cdot \Sigma$ against its standard error Σ. The grey line marks $|X|/\Sigma = 1.96$.

>> Andrews and Kasy (2019), based on experimental economics studies in Camerer et al (2016). The middle panel indicates that β_p=0.03, implying that significant results are 30 times more likely than null results to be published in top Economics journals.



>> Andrews and Kasy (2019) publication selection correction tracks observed replication estimates for experimental economics studies in Camerer et al (2016)

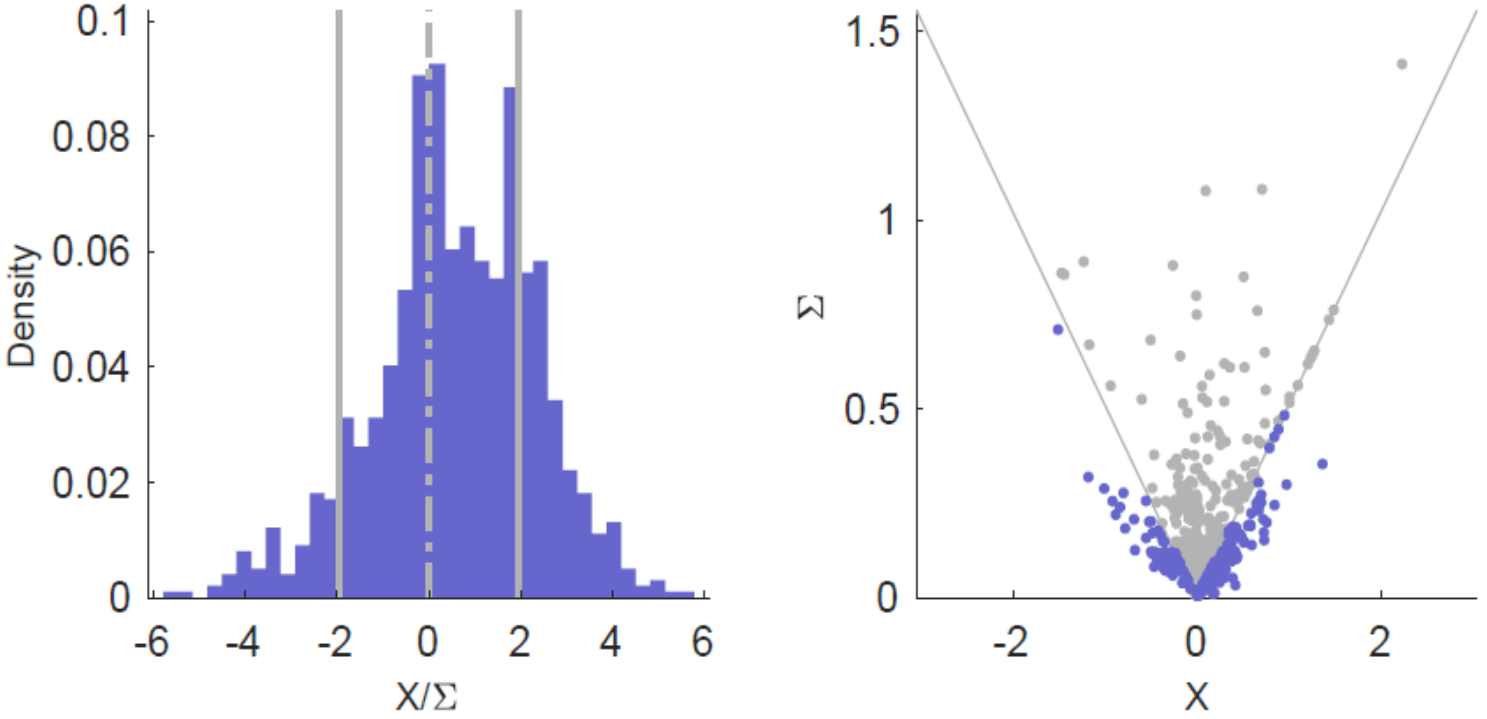


Figure 9: The left panel shows a binned density plot for the z-statistics X/Σ in the Wolfson and Belman (2015) data. The solid grey lines mark $|X|/\Sigma = 1.96$, while the dash-dotted grey line marks $X/\Sigma = 0$. The right panel plots the estimate X against its standard error Σ. The grey lines mark $|X|/\Sigma = 1.96$.

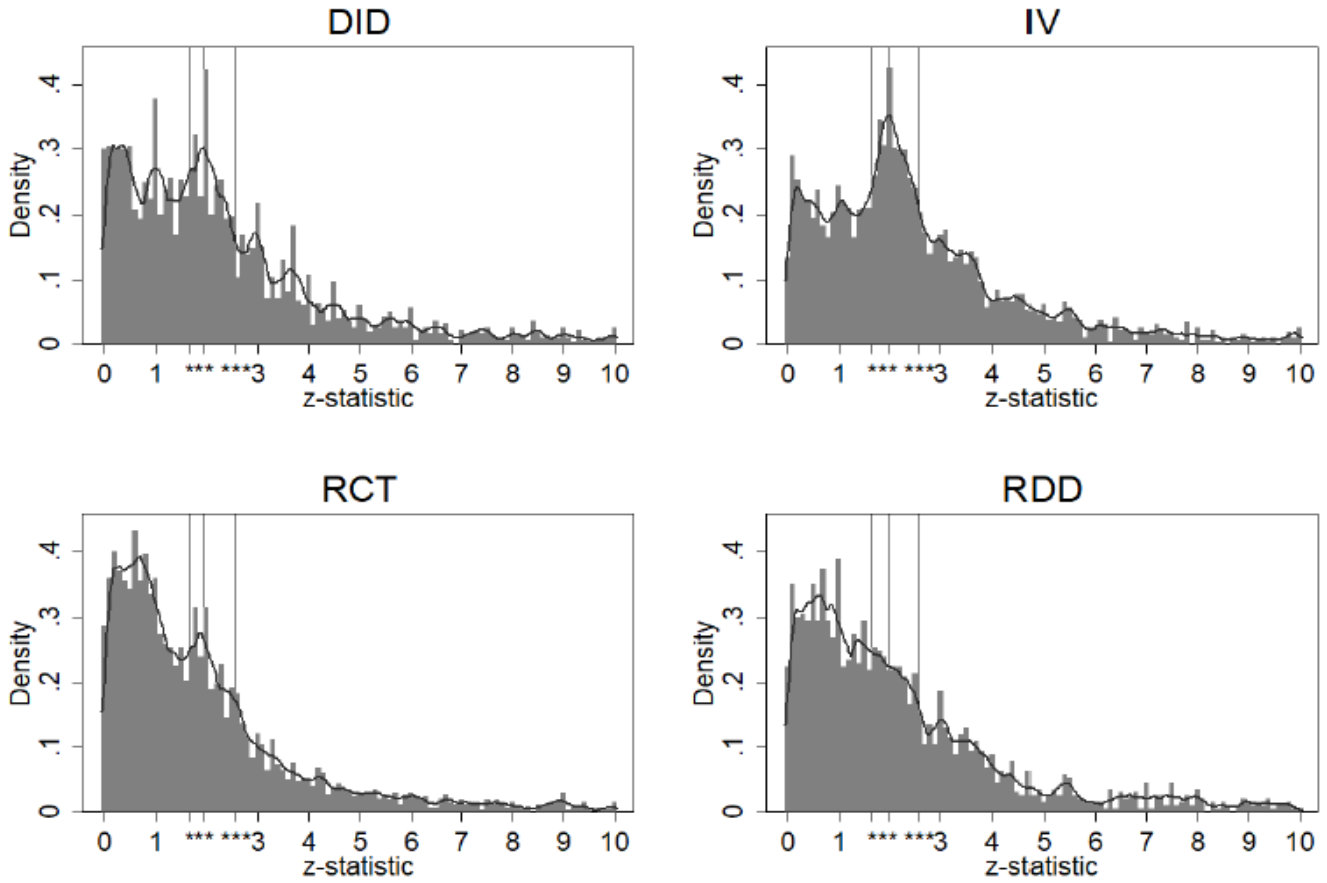
>> Minimum wage literature: a spike at Z=2, and many "missing" underpowered null studies. Estimated β_p from 0.01 to 0.3, implying significant results are 3 to 100 times more likely to be published.

Impacts of transparency

(2) Experimental methods are associated with less publication bias

- The familiar p-value "spike" near 0.05 is significantly less pronounced for studies that use RCT and RDD methods, compared to DD, IV (Brodeur et al 2019)
- Field experiment studies are also far more likely to report null findings.

Figure 2: z-Statistics by Method



>> Brodeur et al (2019), “Methods Matter”.

Notes: This figure displays histograms of test statistics for $z \in [0, 10]$. Test statistics are partitioned by identification method: difference-in-differences (DID), instrumental variables (IV), randomized control trial (RCT) and regression discontinuity design (RDD). Bins are 0.1 wide. Reference lines are displayed at the conventional two-tailed significance levels. We have also superimposed an Epanechnikov kernel. Test statistics have been de-rounded.

Impacts of transparency

(3) Study registration may increase reporting of null results

- The AEA RCT registry dates from 2013, and similar tools now in other social sciences (EGAP in political science, AsPredicted and OSF in Psychology), although not in sociology.
- Early studies with pre-analysis plans include Finkelstein et al (2012) in health economics, and Casey et al (2012) in development.

TABLE 6.1 ERRONEOUS INTERPRETATIONS UNDER "CHERRY-PICKING"

Outcome variable	Treatment effect	Standard error	Mean in control group
A. Main institutional and social change or "software" effects			
Mean effect for family B (hypotheses 4–12; 155 unique outcomes)	0.028	0.020	0.00
B. GoBifo "weakened institutions"			
Attended meeting to decide what to do with the tarpaulin (tarp)	–0.04+	0.02	0.81
Everybody had equal say in deciding how to use the tarp	–0.11+	0.06	0.51
Community used the tarp (verified by physical assessment)	–0.08+	0.04	0.90
Community can show research team the tarp	–0.12*	0.05	0.84
Respondent would like to be a member of the Village Development Committee	–0.04*	0.02	0.36
Respondent voted in the local government election (2008)	–0.04*	0.02	0.85
C. GoBifo "strengthened institutions"			
Community teachers have been trained	0.12+	0.07	0.47
Respondent is a member of a women's group	0.06**	0.02	0.24
Someone took minutes at the most recent community meeting	0.14*	0.06	0.30
Building materials stored in a public place when not in use	0.25*	0.10	0.13
Chiefdom official did not have the most influence over tarp use	0.06*	0.03	0.54
Respondent agrees with "Responsible young people can be good leaders"	0.04*	0.02	0.76
Correctly able to name the year of the next general elections	0.04*	0.02	0.19

SOURCE: Adapted from Casey, Glennerster, and Miguel (2012: tables II and VI).
NOTES: Significance levels (per comparison p-value): + $p < .10$, * $p < .05$, ** $p < .01$, with robust standard errors.

Casey et al (2012): pre-specified results indicate no significant or meaningful institutional impacts of a community-driven development program in Sierra Leone (Panel A). But the existence of many outcome measures allows for "cherry-picking" of negative (Panel B) or positive (Panel C) subsets of findings.

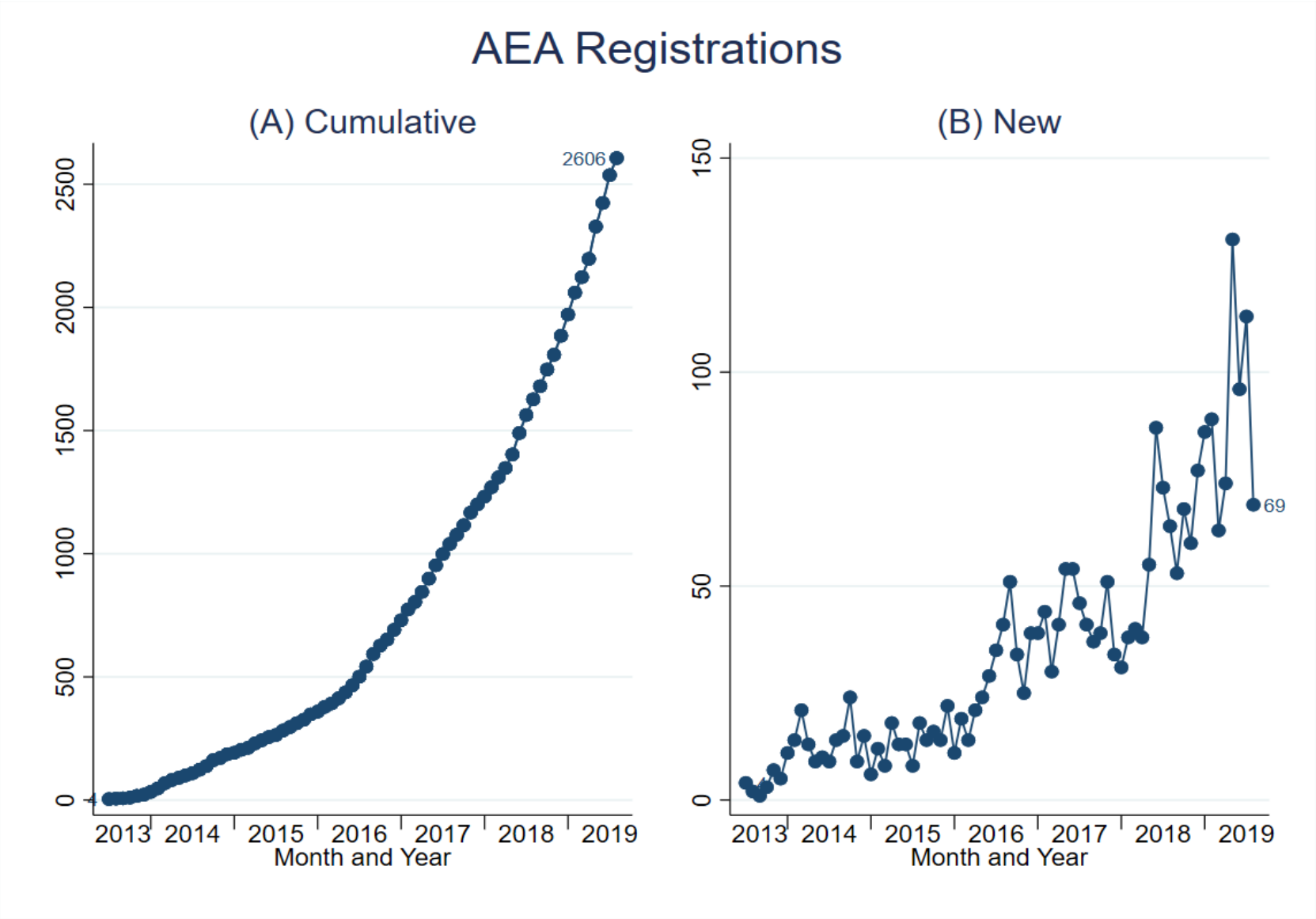
>> How many published results are just some version of the data-mined Panel B or Panel C here?

Impacts of transparency

(3) Study registration may increase reporting of null results

- The AEA RCT registry dates from 2013, and similar tools now in other social sciences (EGAP in political science, AsPredicted and OSF in Psychology), although not in sociology.
- Early studies with pre-analysis plans include Finkelstein et al (2012) in health economics, and Casey et al (2012) in development.

>> Still early to assess aggregate effects of registration and pre-analysis plans on economics (subject of ongoing work by Ofosu and Posner 2019)



N = 2606 registered studies from May 2013 to May 2019

In all, 34% have posted a pre-analysis plan (PAP), rising over time.

(http://dx.doi.org/10.7910/DVN/FU07FC)

Impacts of transparency

(3) Study registration may increase reporting of null results

- Major medical trial registries, e.g., Clinicaltrials.gov, since 2000

>> Lessons from medicine: registries make it possible to document how much hypotheses “shift” in the published paper (Mathieu et al 2009), and appear to increase reporting of null results (Kaplan and Irvin 2015)

Kaplan and Irvin (2015)

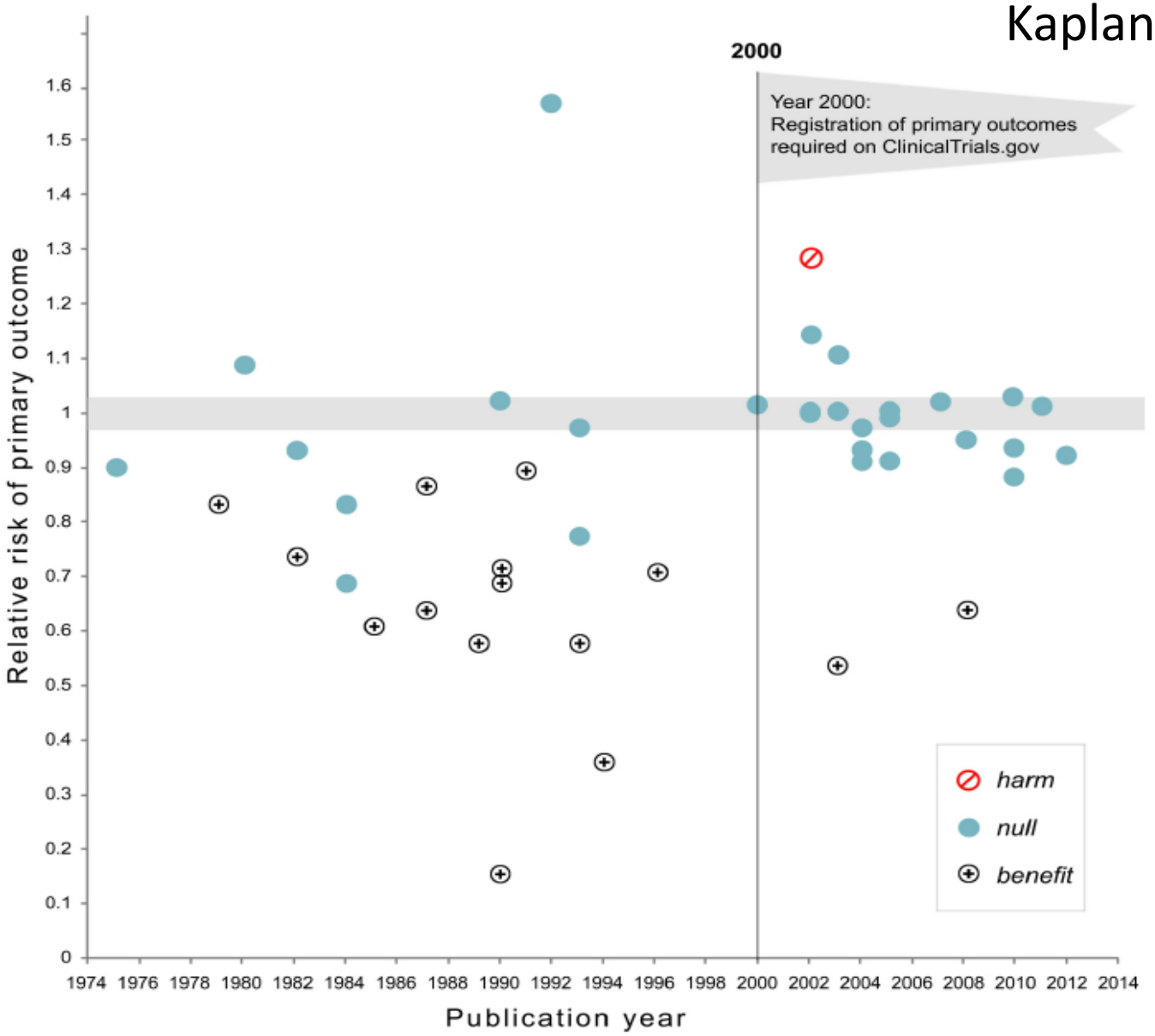


Fig 1. Relative risk of showing benefit or harm of treatment by year of publication for large NHLBI trials on pharmaceutical and dietary supplement interventions. Positive trials are indicated by the plus signs while trials showing harm are indicated by a diagonal line within a circle. Prior to 2000 when trials were not registered in clinical trials.gov, there was substantial variability in outcome. Following the imposition of the requirement that trials preregister in clinical trials.gov the relative risk on primary outcomes showed considerably less variability around 1.0.

doi:10.1371/journal.pone.0132382.g001

Impacts of transparency

(4) Journal editorial practices are influential

- A change in research "culture" and norms will likely be needed to move the scholarly community towards more open science practices.

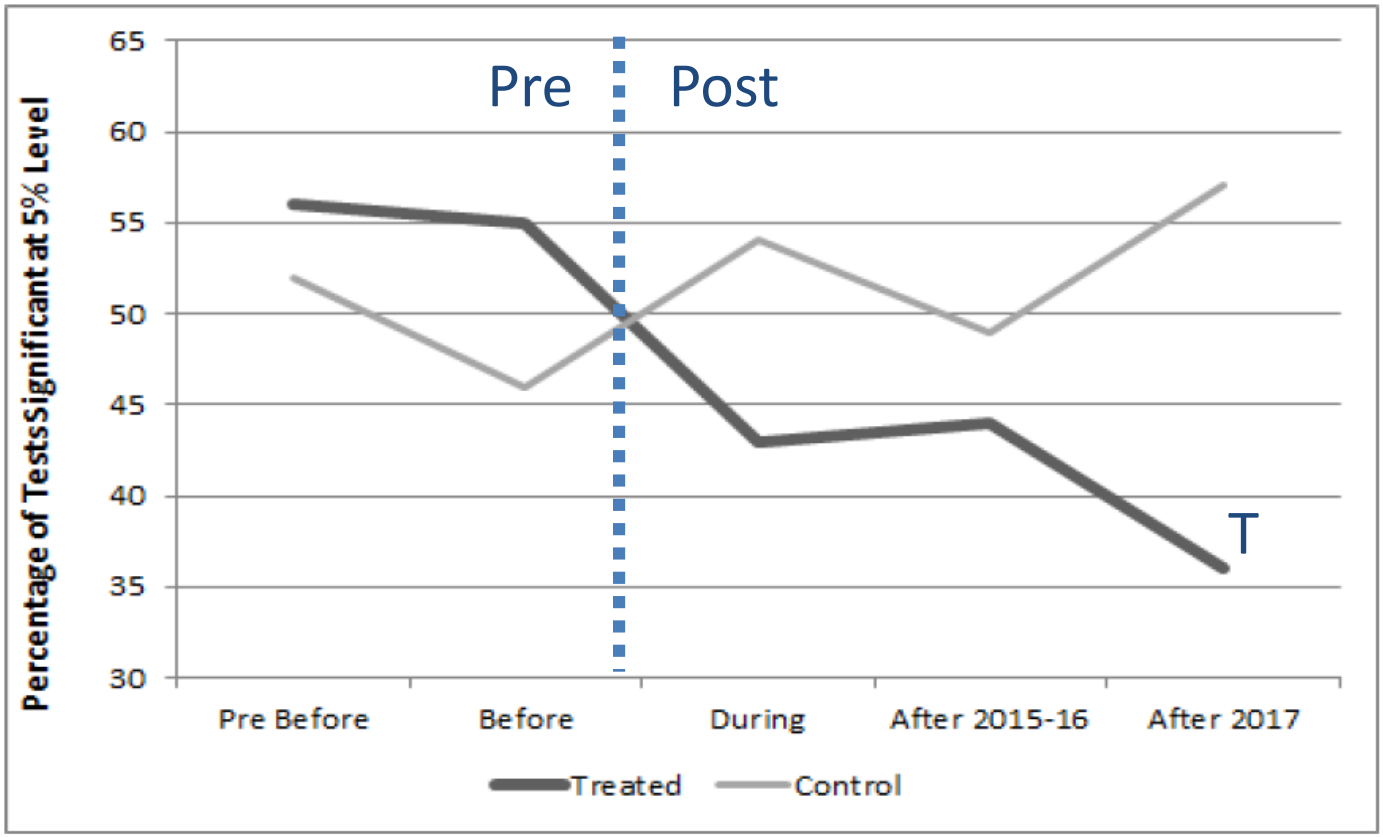
Impacts of transparency

(4) Journal editorial practices are influential

- A change in research "culture" and norms will likely be needed to move the scholarly community towards more open science practices.
- Journal policies and practices can play a role in this shift. E.g., eight (8) health economics journals' issued an editorial statement in 2015 aimed at reducing specification searching, and reminded referees to accept well-designed studies *"regardless of whether such studies' empirical findings do or do not reject null hypotheses."*

>> In difference-in-differences (vs. two non-health applied micro economics journals), the share of null results in these journals increased by 18 percentage points (Blanco-Perez and Brodeur 2019).

Figure 3: Percentage of tests significant at the 5% level.



Sources: treated journals include the *European Journal of Health Economics*, *Health Economics*, *Health Economics Review*, the *International Journal of Health Economics and Management* and the *Journal of Health Economics*. Control journals include *Labour Economics* and the *Journal of Public Economics*. Percentage of tests significant at the 5% level by categories. Pre Before the editorial category includes papers that were published one year before the category Before. Before the editorial category includes papers that were submitted and published before the statement on negative findings. During the editorial category includes papers that were submitted before the statement on negative findings, but published after. After the editorial categories include papers submitted and published (respectively in 2015–16 and 2017) after the statement on negative findings.

Blanco-Perez and Brodeur (2019)

Overview

- Talk outline:

1. Introduction and overview
2. What are research transparency and open science?
3. Problems in economics and social science research
4. What does research transparency do?
5. Innovations in open science, with a focus on pre-registration
6. Looking forward

Advancing pre-registration

- As discussed, pre-registration of research designs and analysis plans could have a range of benefits for bodies of research, including:

(i) creating a paper trail of unpublished studies, for use in meta-analysis;

(ii) constraining the extent of data mining and tendentious reporting;

(iii) generating correctly sized statistical tests;

(iv) forcing scholars to more carefully think through design beforehand.

Advancing pre-registration

- As discussed, pre-registration of research designs and analysis plans could have a range of benefits for bodies of research, including:

(i) creating a paper trail of unpublished studies, for use in meta-analysis;

(ii) constraining the extent of data mining and tendentious reporting;

(iii) generating correctly sized statistical tests;

(iv) forcing scholars to more carefully think through design beforehand.

- Other issues raised regarding PAPs (Olken 2015):

(i) time cost: PAP's shift the work of formulating analysis earlier in time;

(ii) length: norms are still evolving regarding the degree of detail in PAP's;

(iii) flexibility: all papers present some analysis that goes beyond the PAP – which is a good thing.

Advancing pre-registration

- Recent work on pre-registration has been fertile. Four innovations:

(1) Forecasting study results

(2) Pre-specifying the research process

(3) Pre-registration of prospective observational studies

(4) *Adopting pre-results review / registered reports

Advancing pre-registration

- Recent work on pre-registration has been fertile. Four innovations:

(1) Forecasting study results

- A common ex post justification for not publishing null findings is that they are boring, or "we already knew that".
- But did we? DellaVigna and Pope (2018) collect data on expert forecasts of lab experimental tasks, and find that those with higher academic rank and more scholarly citations do not perform better.

>> Systematically collecting the research community's "priors" on the likely impact of a treatment or intervention could be useful in quantifying how much "news" there is in a set of empirical results – making clear that certain null results are in fact unexpected (Vivalt and Coville 2019).

Advancing pre-registration

- Recent work on pre-registration has been fertile. Four innovations:

(2) Pre-specifying the research process

- PAP's can be useful in limiting unprincipled data mining, but could come at a cost if researchers choose the "wrong" specification, tests.

>> Two ways to improve on a "pure" PAP have been proposed:

(i) A hybrid approach: major hypotheses are included in the PAP, while others (e.g., heterogeneity) are examined by splitting the data into an "exploratory sample" and then verified on a (larger) "confirmatory sample" (Anderson and Magruder 2017).

(ii) Incorporate machine learning into PAP's, to reduce the risk of regression mis-specification and improve the construction of outcome measures (Ludwig, Mullainathan and Spiess 2019).

Advancing pre-registration

- Recent work on pre-registration has been fertile. Four innovations:

(3) Pre-registration of prospective observational studies

- The share of observational papers remains high (80%, Burlig 2018). How to incorporate pre-registration into the majority of empirical work?
- Prospective non-experimental studies can also utilize PAPs, and in fact several have done so, e.g., studies designed before an election, or the release of a new round of data.

>> Pre-registration of non-prospective observational work is more problematic since it is difficult to establish whether authors had prior data access, although some health scholars support it (Dal Re et al 2014)

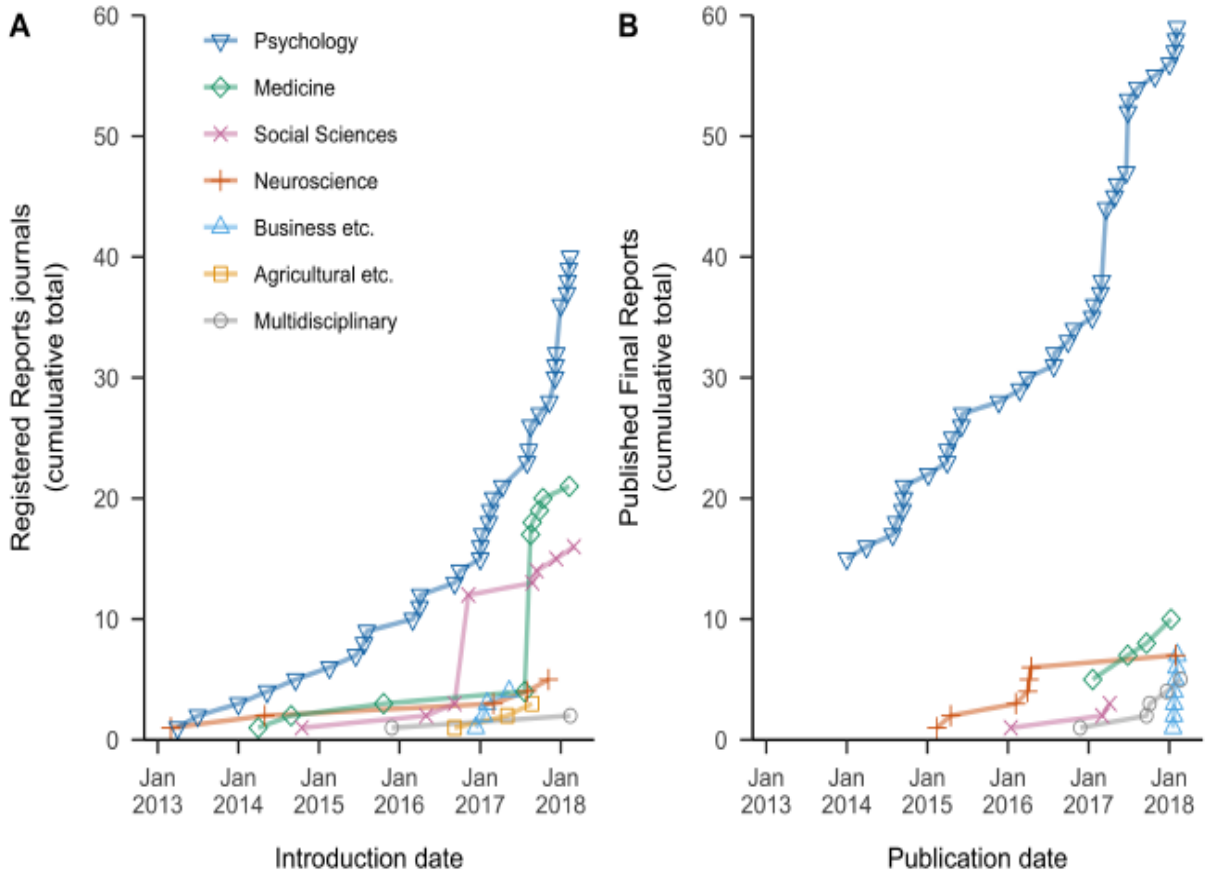
Advancing pre-registration

- Recent work on pre-registration has been fertile. Four innovations:

(4) *Adopting pre-results review / registered reports

- A growing trend in other social sciences – and especially psychology, cognitive science – is pre-results review, also called registered reports.
- Research would ideally be judged on the quality of the question, data and analysis, and not if results are significant or conform to theory.
- Granting agencies (e.g., NSF panels, Gates Foundation, graduate student travel funds, etc.) already make these "calls" all the time.

>> An idealized publication pipeline for pre-results review, adapted from the Center for Open Science, https://cos.io/rr/.



>> Hardwicke and Ioannidis (2018). Journals with pre-results review has since risen to 203, including most top psychology journals, https://cos.io/rr/

Advancing pre-registration

- Recent work on pre-registration has been fertile. Four innovations:

(4) *Adopting pre-results review / registered reports

- The earliest example (to my knowledge) of both a pre-analysis plan and pre-results review in economics is Neumark (2001).

>> According to David I. Levine (Berkeley), Alan Krueger had the idea in 1996 for various participants in the minimum-wage literature to pre-specify their analysis before the next Federal wage increase, and as editor of *Industrial Relations*, Levine would commit to publish results (Levine 2001).

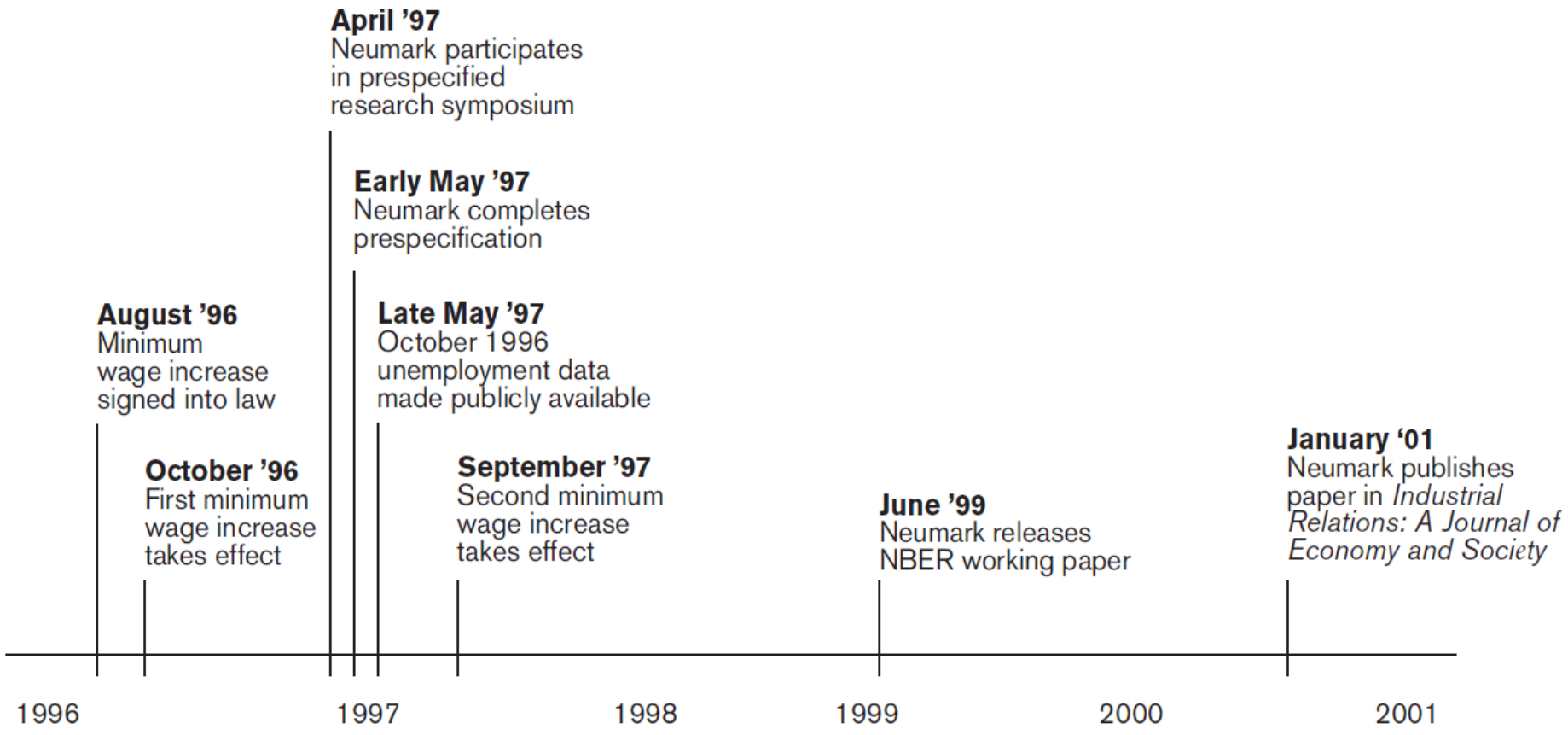


FIGURE 6.2. Timeline of events in Neumark's (2001) minimum wage study.

Advancing pre-registration

- Recent work on pre-registration has been fertile. Four innovations:

(4) *Adopting pre-results review / registered reports

- *Journal of Development Economics* launched a pilot of pre-results review in March 2018, led by editors Andrew Foster and Dean Karlan, and with support from BITSS (which created guidelines for authors and referees, answers to FAQs, etc.)
- Positive response to the pilot: 46 "proposals" have been submitted through the track, with 5 already receiving in-principle acceptance.
- Interviews with authors and referees have not brought up red flags.

>> *JDE* recently made pre-results review a standard permanent submission track, and *Experimental Economics* is launching a pilot.

Overview

- Talk outline:

1. Introduction and overview
2. What are research transparency and open science?
3. Problems in economics and social science research
4. What does research transparency do?
5. Innovations in open science, with a focus on pre-registration
6. Looking forward

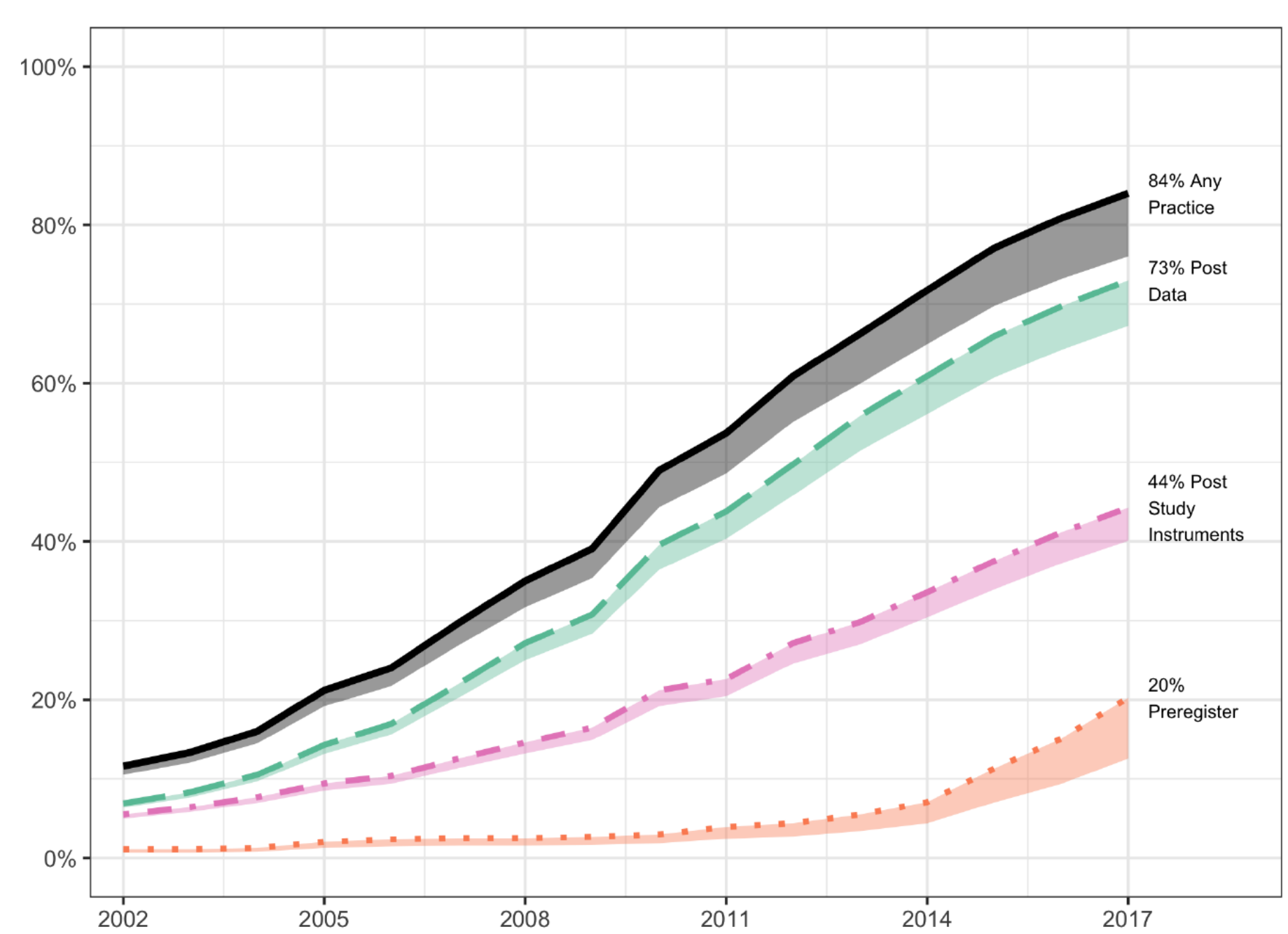
Looking forward

- New evidence that the last 10-15 years really have been a period of rapid methodological change in economics, other social sciences, with the rise of a suite of open science practices (Christensen et al 2019).

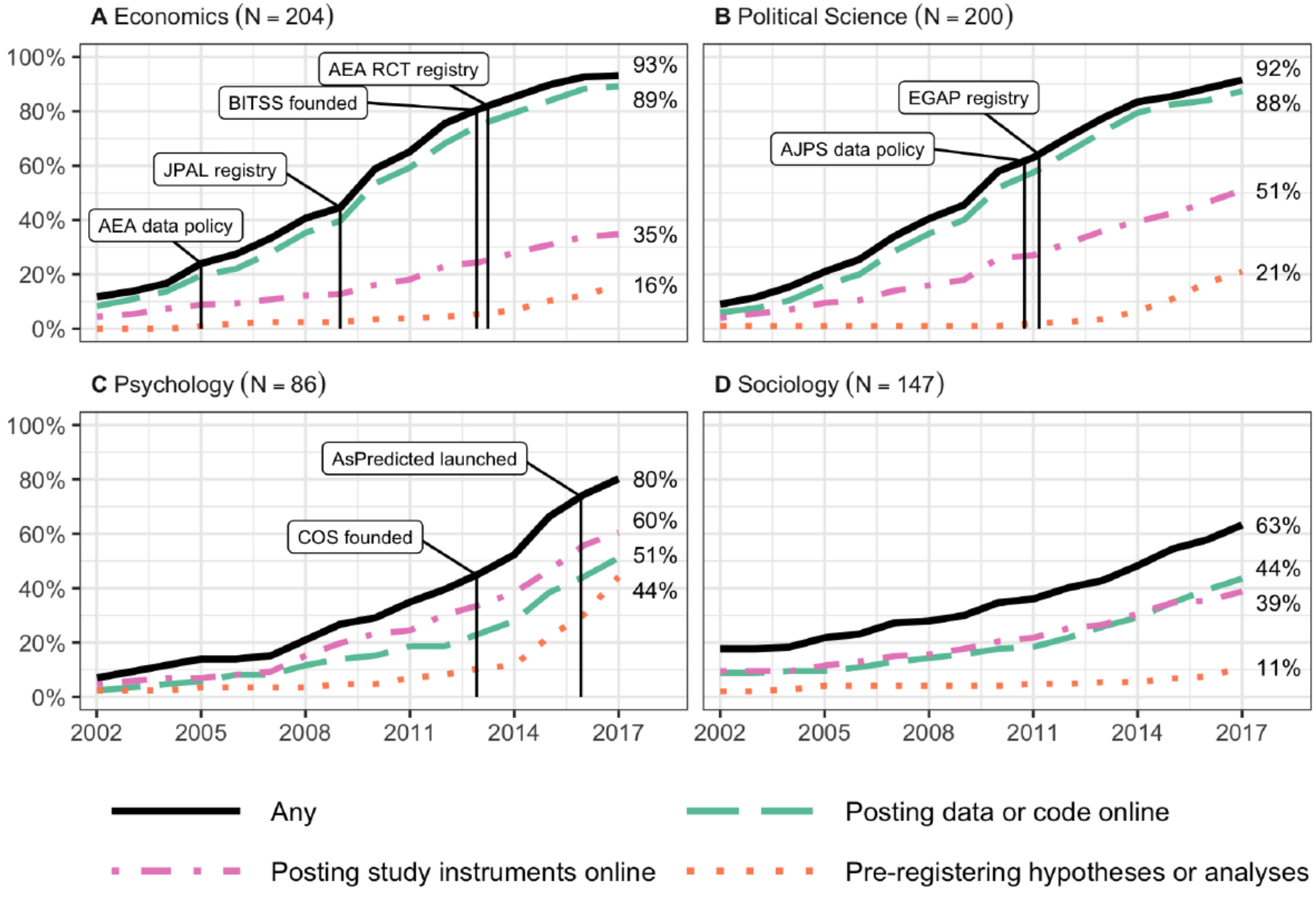
Looking forward

- New evidence that the last 10-15 years really have been a period of rapid methodological change in economics, other social sciences, with the rise of a suite of open science practices (Christensen et al 2019).
- Representative sample of active researchers (publishing in top-10 journals during 2014-16) and graduate students (in top-20 North American departments) across economics, political science, psychology, and sociology; 46% survey response rate, N=2,799.

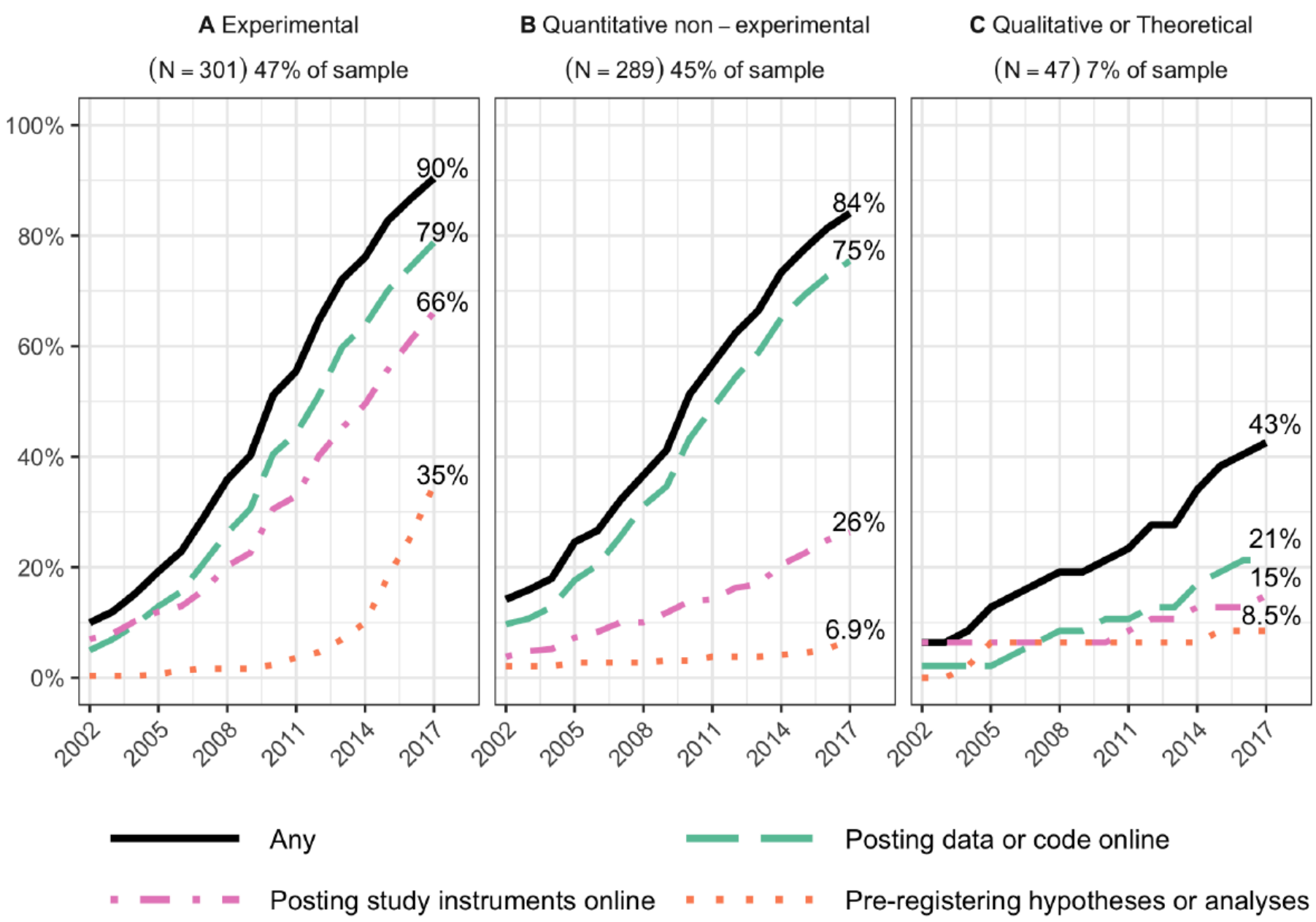
>> Rapid rise in adoption of sharing of data and materials, and more recently in pre-registration.



Christensen et al (2019)



Christensen et al (2019)



Christensen et al (2019)

Looking forward

- New evidence that the last 10-15 years really have been a period of rapid methodological change in economics, other social sciences, with the rise of a suite of open science practices (Christensen et al 2019).
- Representative sample of active researchers (publishing in top-10 journals during 2014-16) and graduate students (in top-20 North American departments) across economics, political science, psychology, and sociology; 46% survey response rate, N=2,799.

>> Rapid rise in adoption of sharing of data and materials, and more recently in pre-registration.

- Changing norms: >80% of development economists surveyed support pre-specifying analyses. Beliefs about others' practices, attitudes substantially lag actual adoption, for data sharing and pre-registration.

Looking forward

- The move towards research transparency and reproducibility is the logical next step in the credibility revolution in economics – and it is already well underway.
- A growing body of evidence indicates that publication bias remains a first-order problem in economics and other social sciences, and threatens the reliability of published bodies of literature.

Looking forward

- The move towards research transparency and reproducibility is the logical next step in the credibility revolution in economics – and it is already well underway.
- A growing body of evidence indicates that publication bias remains a first-order problem in economics and other social sciences, and threatens the reliability of published bodies of literature.
- Meaningful change is now possible via improved methods (e.g., RCT's), practices (open data, pre-analysis plans, replications), and donor + journal policies (e.g., pre-results review, editor statements).

>> Cultural change is needed, too: we should focus more on the quality of the research question, design, and data than on the findings.

Looking forward

- The move towards research transparency and reproducibility is the logical next step in the credibility revolution in economics – and it is already well underway.
- A growing body of evidence indicates that publication bias remains a first-order problem in economics and other social sciences, and threatens the reliability of published bodies of literature.
- Meaningful change is now possible via improved methods (e.g., RCT's), practices (open data, pre-analysis plans, replications), and donor + journal policies (e.g., pre-results review, editor statements).

>> Cultural change is needed, too: we should focus more on the quality of the research question, design, and data than on the findings.

>> Questions and discussion

References

- Anderson, Michael L., and Jeremy Magruder. (2017). “Split-Sample Strategies for Avoiding False Discoveries”, unpublished working paper, University of California, Berkeley.
- Anderson, M. S., Martinson, B. C., and De Vries, R. (2007). “Normative Dissonance in Science: Results from a National Survey of U.S. Scientists,” *Journal of Empirical Research on Human Research Ethics*, 2(4), 3-14.
- Andrews, Isaiah, and Maximilian Kasy. (2019). “Identification of and correction for publication bias”, forthcoming *American Economic Review*.
- Angrist, Joshua D., and Jörn-Steffen Pischke. (2010). “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics,” *Journal of Economic Perspectives*, 24(2), 3-30.
- Blanco-Perez, Cristina, and Abel Brodeur. (2019). “Publication Bias and Editorial Statement on Negative Findings”, forthcoming *Economic Journal*.
- Brodeur, A., Cook, N. and Heyes, A. (2019. “Methods Matter: P-Hacking, and Causal Inference in Economics,” unpublished working paper, University of Ottawa.
- Brodeur, A., Le, M., Sangnier, M. and Zylberberg, Y. (2016). “Star Wars: The Empirics Strike Back,” *American Economic Journal: Applied Economics,* 8(1), 1-32.
- Broockman, D., Kalla, J., and Aranow, P. (2015). “Irregularities in LaCour (2014),” unpublished manuscript, Stanford University. http://stanford.edu/~dbroock/broockman_kalla_aronow_lg_irregularities.pdf.
- Burlig, Fiona. (2018). “Improving transparency in observational social science research: A pre-analysis plan approach”, *Economic Letters*,168, 56-60.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., and Wu, H. (2016). “Evaluating replicability of laboratory experiments in economics,” *Science*, 351(6280):1433-1436.

References

(continued)

- Card, D. and Krueger, A. B. (1995). "Time-series minimum-wage studies: A meta-analysis," *American Economic Review*, 85(2), 238-243.
- Casey, Katherine, Rachel Glennerster and Edward Miguel. (2012). "Reshaping Institutions: Evidence on Aid Impacts Using a Pre-analysis Plan", *Quarterly Journal of Economics*, 127(4), 1755-1812.
- Chang, Andrew C., and Phillip Li. (2015). "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say 'Usually Not'." Federal Reserve Board Finance and Economics Discussion Paper 2015-083.
- Christensen, Garret, Jeremy Freese and Edward Miguel. (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science*, University of California Press.
- Christensen, Garret and Edward Miguel. (2018). "Transparency, Reproducibility, and the Credibility of Economics Research", *Journal of Economic Literature*, 56(3), 920-980.
- Christensen, Garret, Zenan Wang, Elizabeth Levy Paluck, Nicholas Swanson, David Birke, Edward Miguel, and Rebecca Littman. (2019). "Open Science Practices are on the Rise Across Four Social Science Disciplines", unpublished working paper, University of California, Berkeley.
- Dal-Ré, R., Ioannidis, J. P., Bracken, M. B., Buffler, P. A., et al. (2014). "Making prospective registration of observational research a reality." *Science Translational Medicine*, 6, 224cm1, https://doi.org/10.1126/scitranslmed.3007513.
- DellaVigna, Stefano, and Devin Pope. (2018). "Predicting Experimental Results: Who Knows What?", *Journal of Political Economy*, 126, 2410-2456.
- Dewald, William G., Jerry G. Thursby, and Richard G. Anderson. (1986). "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project," *American Economic Review*, 76(4), 587–603.
- Doucouliagos, Chris, and Patrice Laroche. (2003). "What Do Unions Do to Productivity? A Meta-analysis," *Industrial Relations*, 42(4), 650–91.

References

(continued)

- Doucouliagos, Hristos, and T. D. Stanley. (2009). “Publication Selection Bias in Minimum-Wage Research? A Meta-regression Analysis.,” *British Journal of Industrial Relations*, 47(2), 406–28.
- Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., et al. (2012). “The Oregon health insurance experiment: evidence from the first year,” *Quarterly Journal of Economics*, 127(3), 1057–1106, https://doi.org/10.1093/qje/qjs020.
- Fisher, Ronald A. (1936). "Has Mendel's work been rediscovered?“ *Annals of Science* 1(2), 115–126, doi:10.1080/00033793600200111.
- Franco, A., Malhotra, N., and Simonovits, G. (2014). “Publication bias in the social sciences: unlocking the file drawer,” *Science*, 345, 1502–1505. https://doi.org/10.1126/science.1255484.
- Galiani, Sebastian, Paul Gertler, and Mauricio Romero. (2018). “How to Make Replication the Norm”, *Nature*, 554, 417-19.
- Gerber, A., and Malhotra, N. (2008a). “Do statistical reporting standards affect what is published? Publication bias in two leading political science journals,” *Quarterly Journal of Political Science*, 3(3), 313–326, https://doi.org/10.1561/100.00008024.
- Gerber, A., and Malhotra, N. (2008b). “Publication bias in empirical sociological research: do arbitrary significance levels distort published results?” *Sociological Methods & Research*, 37(1), 3–30, https://doi.org/10.1177/0049124108318973.
- Hardwicke, Tom E., and John P.A. Ioannidis. (2018). “Mapping the universe of registered reports,” *Nature Human Behaviour*, 2, 793-796.
- Hedges, L. V. (1992). “Modeling publication selection effects in meta-analysis,” *Statistical Science*, 246-255.
- Ioannidis, J. P. A. (2005). “Why most published research findings are false.” *PLoS Medicine*, 2(8), e124, https://doi.org/10.1371/journal.pmed.0020124.

References

(continued)

- Kaplan RM, Irvin VL. (2015). "Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time," *PLoS ONE*, 10(8), e0132382, doi:10.1371/journal.pone.0132382.
- Leamer, E. E. (1983). "Let's take the con out of econometrics", *American Economic Review*, 73(1), 31–43.
- Lenz, Gabriel, and Alexander Sahn. (2017). "Achieving Statistical Significance with Covariates and without Transparency," unpublished working paper, University of California, Berkeley.
- Levine, D. I. (2001). Editor's introduction to "The Unemployment Effects of Minimum Wages: Evidence from a Prespecified Research Design," *Industrial Relations*, 40(2), 161–162, https://doi.org/10.1111/0019–8676.00204.
- Ludwig, Jens, Sendhil Mullainathan, and Jann Spiess. (2019). "Augmenting Pre-Analysis Plans with Machine Learning", unpublished working paper.
- Mathieu, S., Boutron, I., Moher, D., Altman, D. G., and Ravaud, P. (2009). "Comparison of registered and published primary outcomes in randomized controlled trials." *Journal of the American Medical Association*, 302(9), 977–984, https://doi.org/10.1001/jama.2009.1242.
- McCrary, J., Christensen, G., and Fanelli, D. (2015). "Conservative tests under satisficing models of publication bias," *PLoS One*, 11(2), e0149590.
- Merton, R. K. (1942). "A note on science and democracy," *Journal of Legal and Political Sociology*, 1, 115–126.
- Miguel, E., Camerer, C., Casey, K., Cohen, J., et al. (2014). "Promoting transparency in social science research," *Science*, 343, 30–31, https://doi.org/10.1126/science.1245317.
- Neumark, D. (2001). "The employment eff ects of minimum wages: evidence from a prespecified research design," *Industrial Relations*, 40(1), 121–144, https://doi.org/10.1111/0019–8676.00199.

References

(continued)

- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., et al. (2015). “Promoting an open research culture,” *Science*, 348, 1422–1425, https://doi.org/10.1126/science.aab2374.
- Ofosu, George, and Daniel Posner. (2019). “Evidence on the Use of Pre-analysis plans in Economics and Political Science”, unpublished working paper, UCLA.
- Olken, B. A. (2015). “Promises and perils of pre-analysis plans,” *Journal of Economic Perspectives*, 29(3), 61–80, https://doi.org/10.1257/jep.29.3.61.
- Open Science Collaboration (2015). “Estimating the reproducibility of psychological science,” *Science*, 349, aac4716, https://doi.org/10.1126/science.aac4716.
- Rosenthal, Robert. (1979). “The file drawer problem and tolerance for null results.,” *Psychological Bulletin*, 86(3), 638–641, https://doi.org/10.1037/0033–2909.86.3.638.
- Sanna, L. J., Chang, E. C., Miceli, P. M., & Lundberg, K. B. (2011). “Rising up to higher virtues: Experiencing elevated physical height uplifts prosocial actions [Retracted article],” *Journal of Experimental Social Psychology*, 47, 472–476.
- Simonsohn, U. (2013). “Just post it: the lesson from two cases of fabricated data detected by statistics alone,” *Psychological Science*, 24(10), 1875–1888, https://doi.org/10.1177/0956797613480366.
- Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014a). “P-curve: a key to the file-drawer,” *Journal of Experimental Psychology: General*, 143(2), 534-547, https://doi.org/10.1037/a0033242.
- Snyder, Christopher, and Ran Zhuo. (2018). “Sniff Tests in Economics: Aggregate Distribution of Their Probability Values and Implications for Publication Bias,” NBER Working Paper #25058.

References

(continued)

- Stanley, T. D. (2008). “Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection,” *Oxford Bulletin of Economics and Statistics*, 70(1), 103-127, https://doi.org/10.1111/j.1468–0084.2007.00487.x.
- Sterling, T. D. (1959). “Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa,” *Journal of the American Statistical Association*, 54(285), 30–34, https://doi.org/10.1080/01621459.1959.10501497.
- Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., and Rosenthal, R. (2008). “Selective publication of antidepressant trials and its influence on apparent efficacy,” *New England Journal of Medicine*, 358(3), 252–260, https://doi.org/10.1056/NEJMsa065779.
- Vivalt, E., and A. Coville. (2019). “The Implications of Variance Neglect for the Formation and Estimation of Subjective Expectations”, unpublished working paper, Australian National University.
- Wolfson, P. J. and Belman, D. (2015). “15 years of research on us employment and the minimum wage,” available at SSRN 2705499.