INSTITUTE for REPLICATION

Email: instituteforreplication@gmail.com

X: https://twitter.com/I4Replication

Webpage: https://i4replication.org/

Reproducibility, Replication Packets, and Pre-analysis Plans

APEC Skills Workshop

Abel Brodeur
University of Ottawa
Institute for Replication

Taking Stock

- **Reproductions/replications in the social sciences:**
  - Very small number of (individual) reproductions/replications published
    - » About 20 publications per year in economics (ref. Replication Network)
    - » Focus on experimental studies (Open Science Framework and Camerer et al., 2016 and 2018)
- **Why such a small number of reproductions/replications?**
  - Lack of incentives; Harmful for career?
- **Bad equilibrium and lack of norms/guidelines**
  - Only “negative” reproductions/replications are disseminated

This Presentation

- Pre-Analysis plan
- Reproduction practices at journals
- Best practices for creating a packet
- Institute for Replication

Rise of (Pre-)Registration in the Social Sciences

- **RCTs have become increasingly prominent in the social sciences**
  - This talk is about economics, but similar pattern in poli sci and other (mostly non-experimental) disciplines
- **American Economic Association launched AEA RCT Registry in 2013**
  - As of 2020, +2,000 trials have been registered
- **Content vary tremendously**
  - In practice, the elements that are required by the platform are skeletal
  - Option to include a Pre-Analysis Plan (PAP)

Definitions and Lack of Understanding

- **Lots of ambiguity and lack of transparency about differences between registration and pre-registration**
  - Caused in part because some journals make it compulsory to register your study on AEA RCT registry
- **But it gets worse…**
- **In practice, pre-registration and pre-registration with PAP are distinct and separable things**
  - Not saying this is the way it should be… simply describing what is happening
  - Obviously, things are different in psychology and medicine where pre-registration implies a PAP. Not here!

Brodeur et al. (2024): Journal Political Economy: Micro

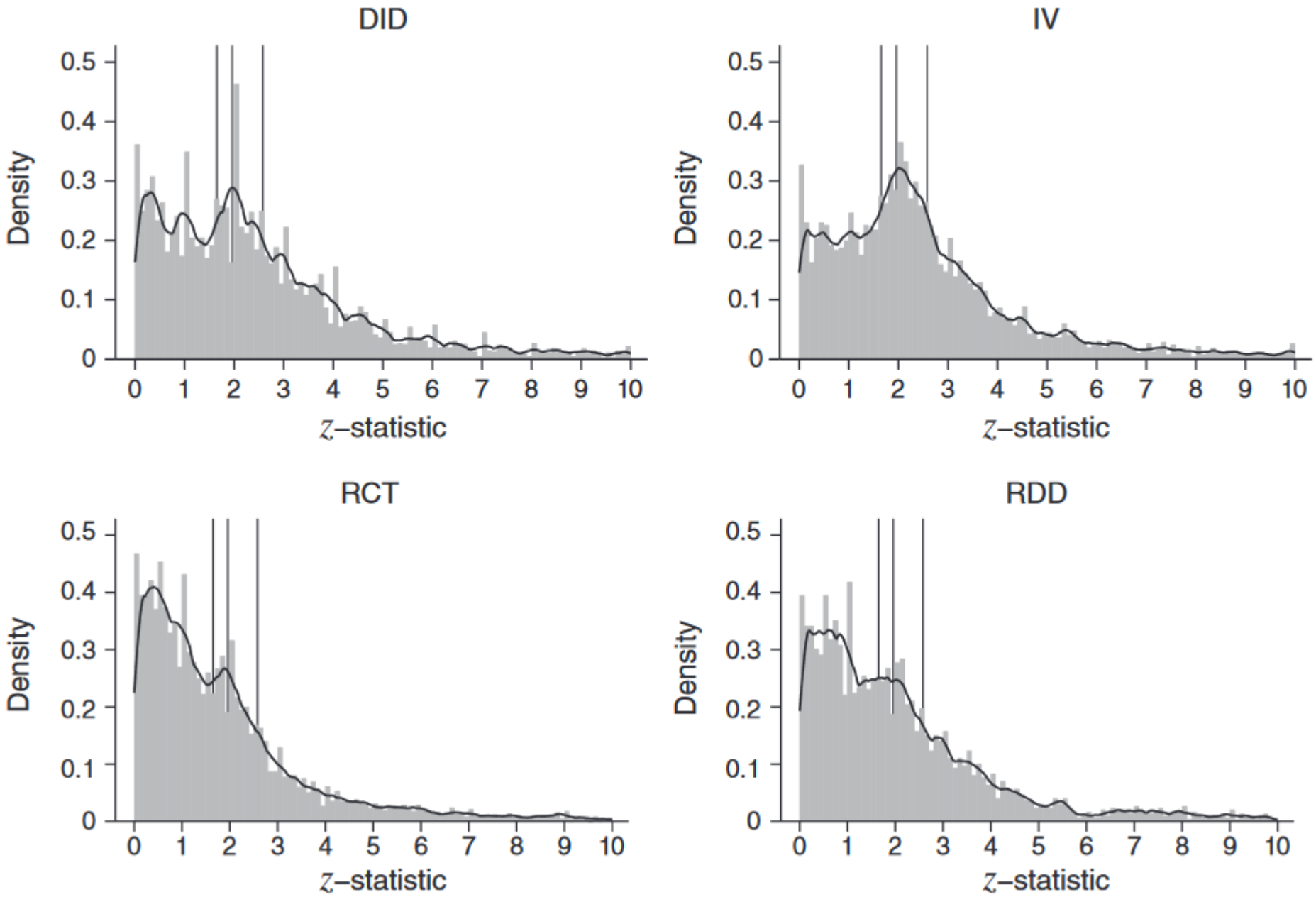
- **Universe of test statistics from RCTs published in 15 leading economics journals from 2018 through 2021 (314 articles)**

- **Articles and researchers' characteristics do not predict well who pre-register…**

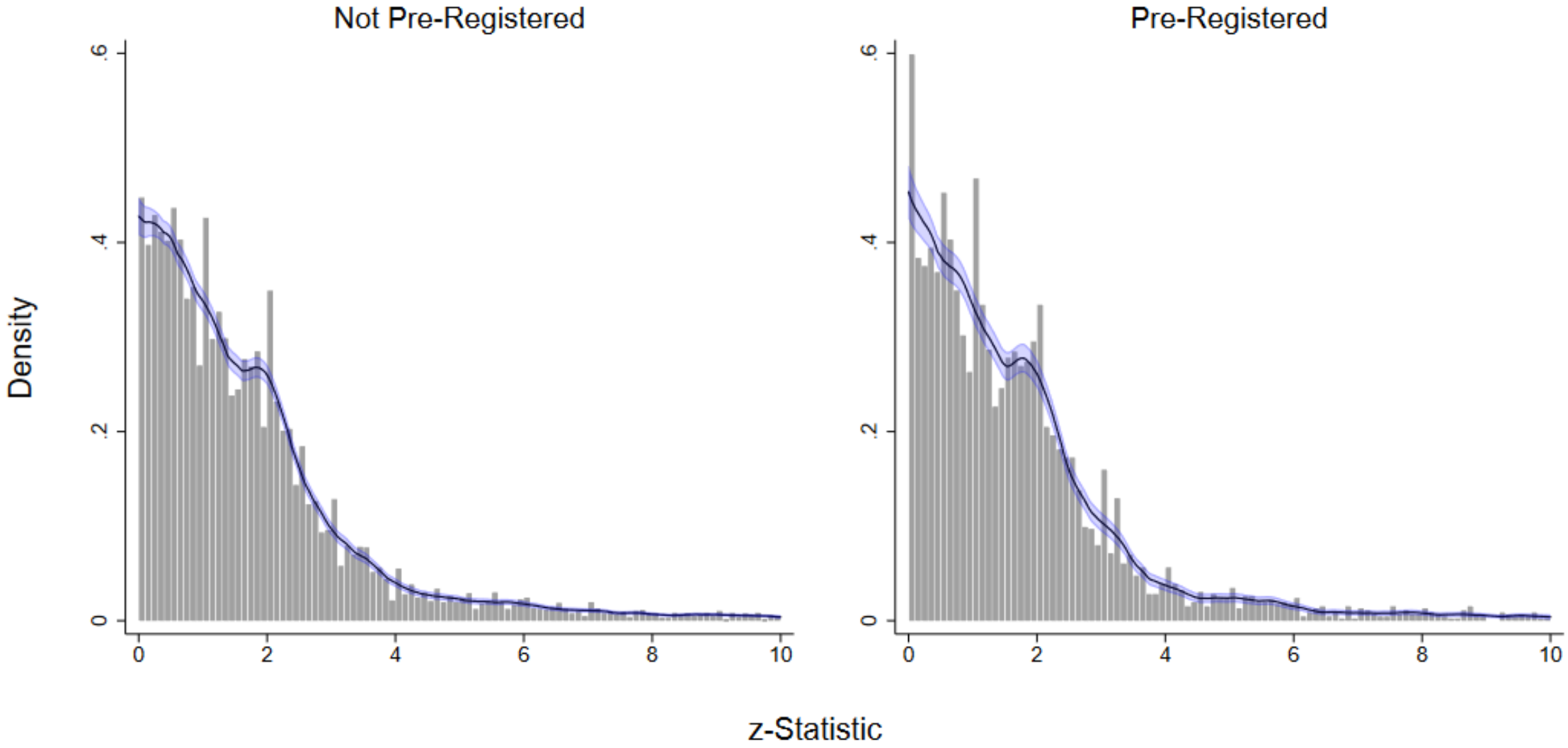
- **Test whether pre-registration reduces p-hacking/publication bias**
  - Note that RCTs are less p-hacked than non-experimental methods!

But First…

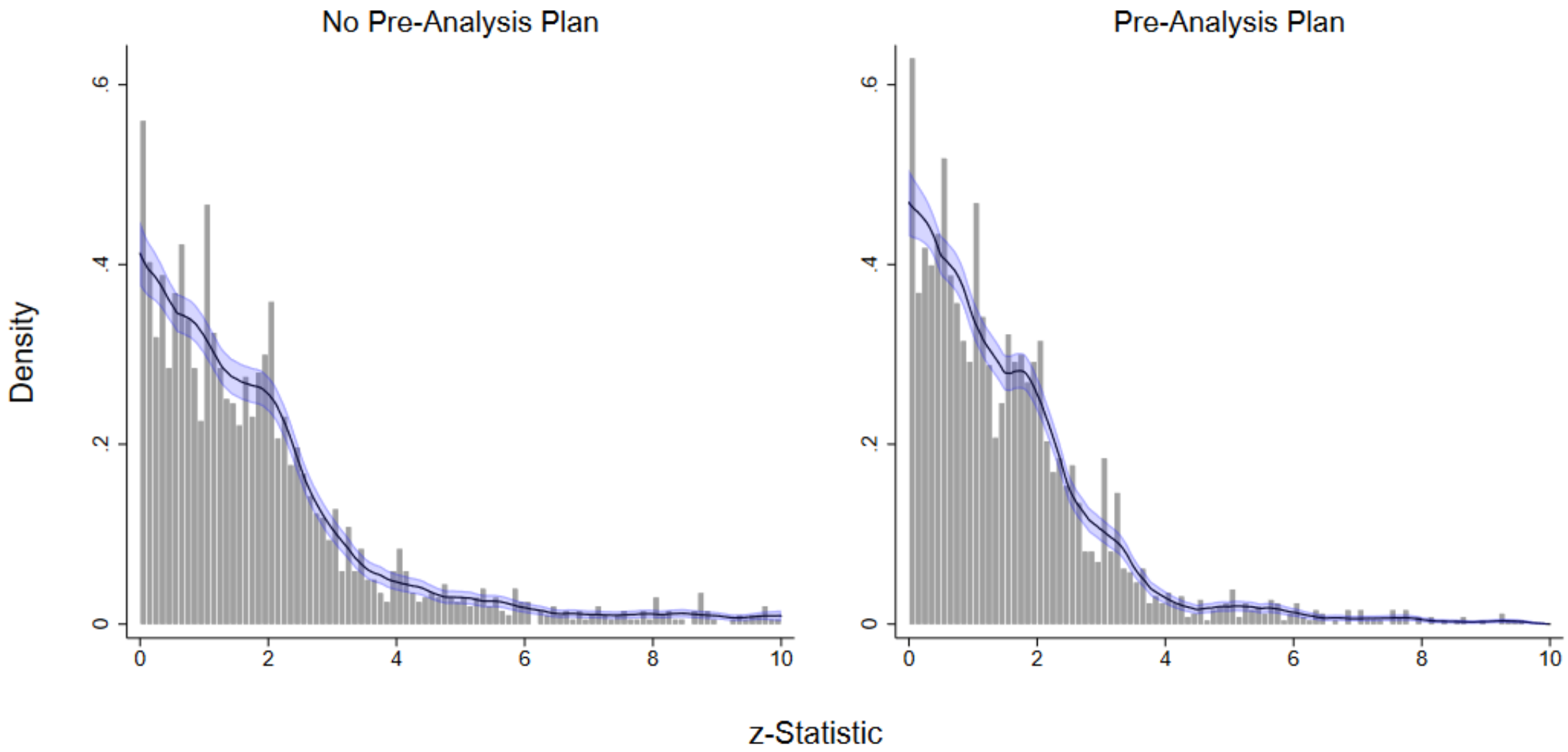
- Are RCTs less p-hacked than other methods…


DID
IV
RCT
RDD
Density
z–statistic
0
0.1
0.2
0.3
0.4
0.5
0 1 2 3 4 5 6 7 8 9 10

Extent of Bias by Pre-Registration



Extent of Bias: Pre-Registration with/without PAP



This Presentation

- **Pre-Analysis plan**
- **Reproduction practices at journals**
- **Best practices for creating a packet**
- **Institute for Replication**

Brodeur et al. (2024): EJ

Table 1: Journal Data-Sharing Policies

Journal	Policy	Announcement (Year)	# Articles	# Test Statistics	Data Collection (Year)
American Economic Review	Yes	2004	132	5,238	2002-2020
A. Econ. J.: Applied Econ.	Yes	2009	50	2,470	2015, 2018
A. Econ. J.: Econ. Policy	Yes	2009	42	1,251	2015, 2018
A. Econ. J.: Macroeconomics	Yes	2009	5	54	2015, 2018
Econometrica	Yes	2004	22	578	2002-2020
Economic Journal	Yes	2012	78	2,629	2002-2020
Economic Policy	Yes	2017	6	2,629	2015, 2018
Experimental Economics	Encourage		6	79	2015, 2018
J. of Applied Econometrics	Yes	1994	5	86	2015, 2018
J. of Development Economics	Yes	2014	64	2,818	2015, 2018
J. of Economic Growth	Encourage		8	100	2015, 2018
Journal of Finance	Only Code	2018	51	2,084	2002-2020
J. of Financial Economics	No		39	569	2015, 2018
J. of Finan. Intermediation	Encourage		16	185	2015, 2018
J. of Human Resources	Yes	2019	57	1,697	2002-2020
J. of International Econ.	No		19	488	2015, 2018
J. of Labor Economics	Yes	2010	39	1,114	2002-2020
J. of Political Economy	Yes	2005	51	1,854	2002-2020
J. of Public Economics	Encourage		74	2,605	2015, 2018
J. of Urban Economics	Encourage		26	660	2015, 2018
J. of the Euro. Econ. Ass.	Yes	2011	56	1,648	2002-2020
Quarterly Journal of Econ.	Yes	2016	71	3,951	2002-2020
Review of Economic Studies	Yes	2006	26	1,634	2002-2020
Review of Econ. & Stat.	Yes	2010	96	3,286	2002-2020
Review of Financial Studies	No		67	1,618	2002-2020

Effectiveness of Data Availability Policy

- **Data availability policy has no impact on p-hacking and pub bias**
  - A recent piece in JEEA finds the opposite result
- **Also, not much difference across data types**
  - But big differences across methods
- **Journal Development Economics**
  - Having a policy and (not) enforcing it…
  - Out of 75 studies, 47 did not provide a replication package. The remaining 28 studies can be categorized as follows: 13 report relying on confidential data; 14 provided a link to a replication package; and one provided only Stata codes and information on how to obtain the data. I contacted all of authors; 7 ended up providing a package.

Computational Reproducibility at the Journal Stage

- **Data editors**
  - AEA journals, Econometrics society, Economic Journals, JEEA, Econ Inquiry, Canadian Journal of Economics, etc.
  - They do not check for coding errors
  - A researcher or RAs computationally reproduce the results (i.e., make sure codes run and produce results in the article)
  - At the conditionally accepted stage

This Presentation

- **Pre-Analysis plan**
- **Reproduction practices at journals**
- **Best practices for creating a packet**
- **Institute for Replication**

Make it Reproducible Day ONE

- **Template Readme**
  - https://social-science-data-editors.github.io/template_README/
- **Keep track of what you do**
- **Get someone else to check your codes**
  - Code review someone else in exchange

This Presentation

- **Pre-Analysis plan**
- **Reproduction practices at journals**
- **Best practices for creating a packet**
- **Institute for Replication**

Institute for Replication (I4R)

- **Launched in 2022**

- **Initial focus on economics and political science:**
  - New collaborations with Nature Human Behaviour and Psychological Science

- **Objectives:**
  - Mass reproduction and replication
  - Change norms through collaborations with editors, original authors and replicators

Which Studies Are Reproduced/Replicated?

- **Start with journals that have a data availability/code policy:**
  - Selected top economics and political science journals
  - List here: https://i4replication.org/reports.html
- **Only going forward (studies published in 2022-)**
- **Expand selection of journals**
    - » Psychological Science (2024-)
    - » Nature Human Behaviour (2023-)

I4R's Strategies for Generating Reproductions/Replications

- Identify studies to be reproduced/replicated
  - » Empirical studies published in selected leading journals
  - » Check if data and codes available
  - » Check if data can be accessed and by whom
  - » Then reproduce the results (or done by data editor)

- (1) **Editorial board selects replicators**
  - » Invitation to replicators sent by email
    - Similar to requesting referee reports
  - » Choice of replicators is based on knowledge of the literature and data, but also data access in some cases

I4R's Strategies for Generating Replications

- (2) **Replication Games**
  - » Team of 3-5 researchers with similar interests
    - Mix of PhD students, faculty and researchers
    - Assign study to reproduce/replicate 3 weeks before Games
    - Replication during/after Games: robustness or recoding
    - Start games with "We Will ~~Rock~~ Replicate You" song
  - » 25+ scheduled events for 2024:
    - London, Toronto, UCLA, UC Berkeley, Brown, Northwestern, Seattle, Cambridge, Sydney, Melbourne, Rotterdam, Munich…
    - About 700 participants for 2023

I4R's Strategies for Generating Replications

- (3) **Admin data, non-public data and lab experiments**
  - » Payments to replicators (USD 5,000)
    - Start this stream this Summer
    - Especially key in economics with large admin data sets that can only be accessed in data centers
    - Also lab replications with new data for experiments published in top economics journals

Replicators

- **Anonymous if wanted**
- **No incentives to show that the results do not reproduce/replicate**
  - Positive and negative replications are disseminated
- **Conflict of interest**
  - Cannot be colleague, recent collaborator, friend, etc.
- **They choose “how” to reproduce/replicate**
  - Different design / research question requires different specification check
    - » Identification of coding errors could lead to different checks
  - But general guidelines (with examples of specification checks) are provided to the replicators
  - Pre-analysis plan required

Once a Reproduction/Replication Is Completed

- **(1) Replicators provide report to the Institute**
  - Similar to a referee report (use a template)
  - May remain anonymous
- **(2) Reviewed by Chair and sent to original authors**
- **(3) Authors respond (if they want)**
- **(4) Publicly release as I4R discussion papers (or on OSF) simultaneously report and response**

Communication with Original Authors

- **Authors almost always respond:**
  - 95% of original authors that A.B. reached out to responded to his email, of which one author whose email bounced back
  - Of those that responded, 22% provided a short note (e.g., thanking replicators) or mentioned they could not respond (e.g., due to personal reasons or ongoing conflict in their country)
  - 54% provided feedback without a formal response
  - And 24% provided a formal response

- **Remaining disagreements for only 18% of articles in our sample**

Communication with Original Authors

- **Clarifications or help needed?**
  - We asked replicators whether their team or I4R contacted, or attempted to contact, the original authors for clarifications?
  - About 40% of replicators contacted (through I4R) the authors for clarifications
    - » Replication package was unclear, help to computationally reproduce the original authors' results; unable to access the original authors' data; verifying coding errors, etc.
  - About 66% mentioned that interacting with the original authors improved the quality of their report

First Meta Paper: About 350 Authors

- **110 robustness reproductions or replications:**
  - Very selected sample; most of these journals have a data editor

- **About 5,000 new point estimates from the following re-analyses:**
  - (i) alternative choice of control variables
  - (ii) changing the sample
  - (iii) changing the dependent variable
  - (iv) changing the main independent variable
  - (v) changing the estimation method/model
  - (vi) changing the method of inference
  - (vii) change weighting scheme
  - (viii) replication using new data

First Meta Paper

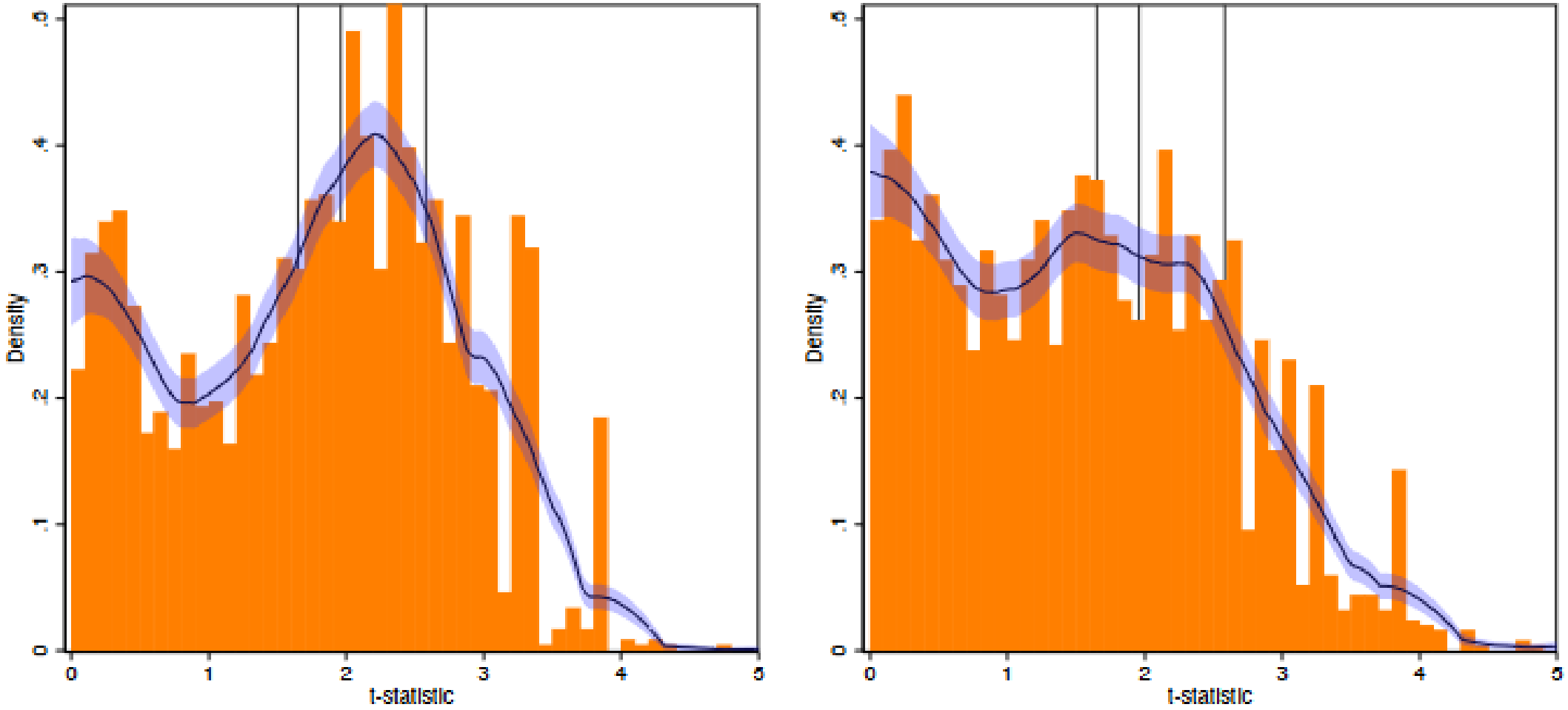
- **25% of studies have a coding error:**
  - Range from minor to MAJOR
    - » Ex. 75% of observations are duplicates
    - » Not cleaning raw data (e.g., St. Louis, St Louis, StLouis, …)
    - » Not fully interacting DID model
    - » Not specifying GMM function

- **Mentioning something in the paper, but doing something else in the code**
  - Rare, but happened twice for inference

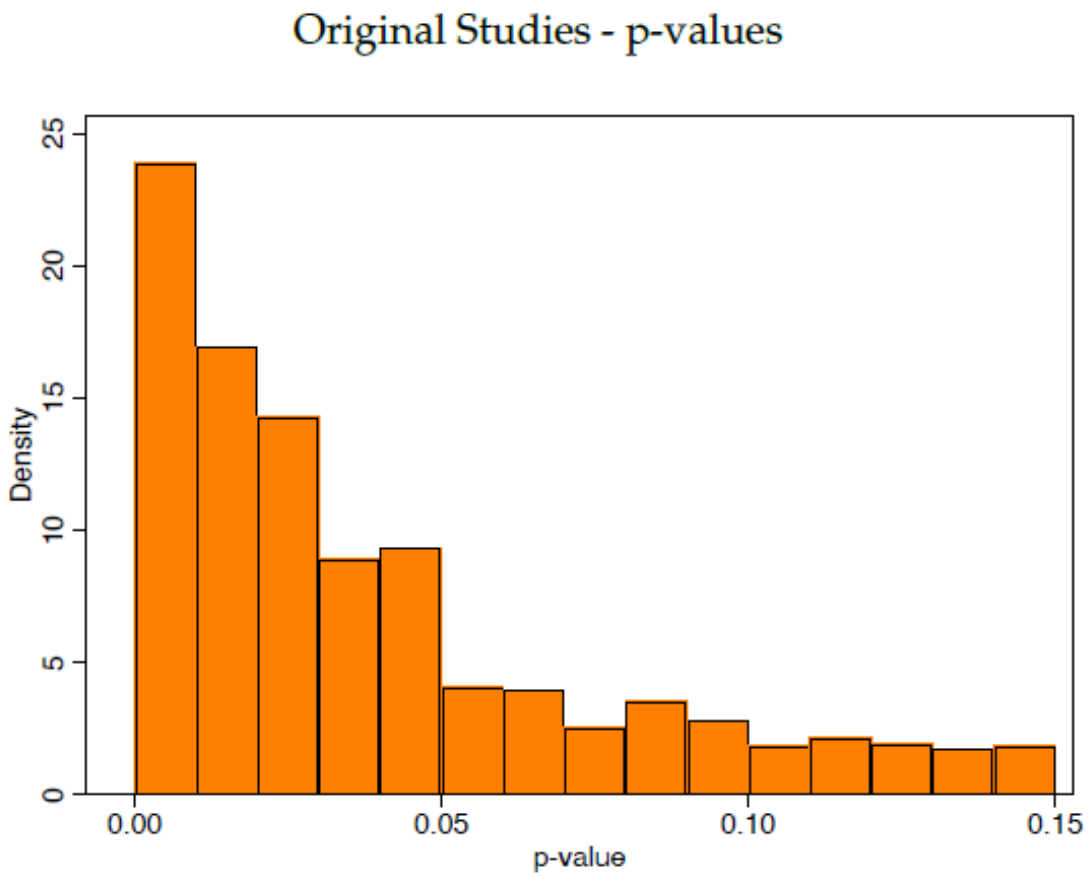
- **Important coding decisions buried in footnote or appendix**

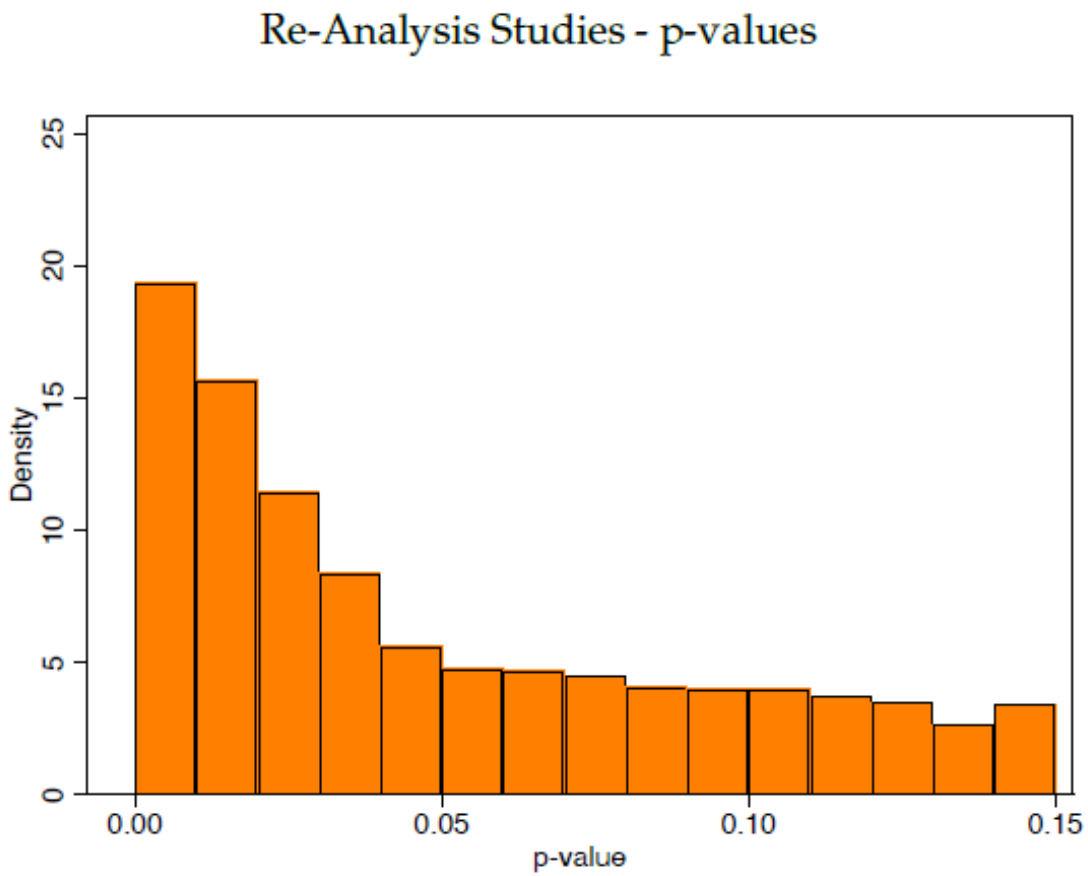
First Meta Paper: t-curves

Figure 3: Distributions of t-Statistics for Original Studies and Re-Analyses



First Meta Paper: p-curves





Robustness Reproducibility Rate

- **About 70% of re-analyses remain significant at 5% and same sign**

Table 4: Shifts in Statistical Significance Regions

	Re-Analysis Significance Level					
Original Significance Level	Sign Change	Not Sig.	Sig. at 10%	Sig. at 5%	Sig. at 1%	Total
Not Significant	12.83	77.32	4.54	2.77	2.54	100.00
Significant at 10%	6.49	45.89	27.27	13.42	6.93	100.00
Significant at 5%	3.45	26.91	10.00	44.36	15.27	100.00
Significant at 1%	5.08	11.24	3.91	6.99	72.77	100.00
Total	7.31	37.70	7.14	13.31	34.55	100.00

Robustness Reproducibility Rate

- **Barriers to sensitivity analysis:**
  - Self-report: by far the main barrier is the lack of raw data

- **Re-analyses by type:**
  - Lowest robustness reproducibility rates for: (i) changing the dependent variable, (ii) sample and (iii) weights
  - Highest for: (iv) changing independent variable, (v) inference method
  - Middle-range: (vi) new data, (vii) change estimation, (viii) change controls

Conclusion

- **High computational reproducibility rates**
- **Severe issues with only a small number of studies**
- **Potential robustness/sensitivity issues for some studies**
- **Positive impact on views of the discipline:**
  - 40% of replicators report that the quality of the replication package led them to have a more optimistic view of the discipline
  - Another 40% reported no impact on their views