INSTITUTE for REPLICATION

Email: instituteforreplication@gmail.com

X: https://twitter.com/I4Replication

Webpage: https://i4replication.org/

Making Reproducibility Research More Systematic

From the Social Sciences to Epidemiological and Medical Research?

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

Taking Stock: Health Sciences

- **Reproductions/replications in the health sciences:**
  - Similar issues
    - » Cobey et al. 2023, Lee and Hanage 2020, Peng et al. 2006, Wallach and Basu 2017
    - » Discussion on definitions and way forward, but lacking mass reproducibility efforts

- **Can we learn from the social sciences?**
  - Proposal at the end for public health/epidemiology

Definitions

- **Computational reproducibility (same code/data):**
  - Ability to duplicate the results of a prior study using the same data and procedures as were used by the original investigator.
- **Robustness reproducibility (sensitivity analysis):**
  - Ability to duplicate the results of a prior study using the same data but different procedures as were used by the original investigator.
- **Replicability (new data):**
  - Ability to duplicate the results of a prior study using new data.

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

Collaboration with Editors

- **Put together 3 special issues dedicated to replications**
  - Research & Politics, Canadian Journal of Economics and Economic Inquiry
  - Replication Section: Journal of Political Economy: Micro, World Development Perspectives, Spatial Economic Analysis, etc.

- **Collaboration with Psych Science and NHB**

- **Surveys of editors**
  - https://i4replication.org/publishing.html

Discipline	Journal	Editor	Q1 - Code	Q2 - Code	Long Answer
Economics	Journal of Economics	John Doe	Yes	No	In fact, this journal does not regularly publish commentary about prior publications...

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

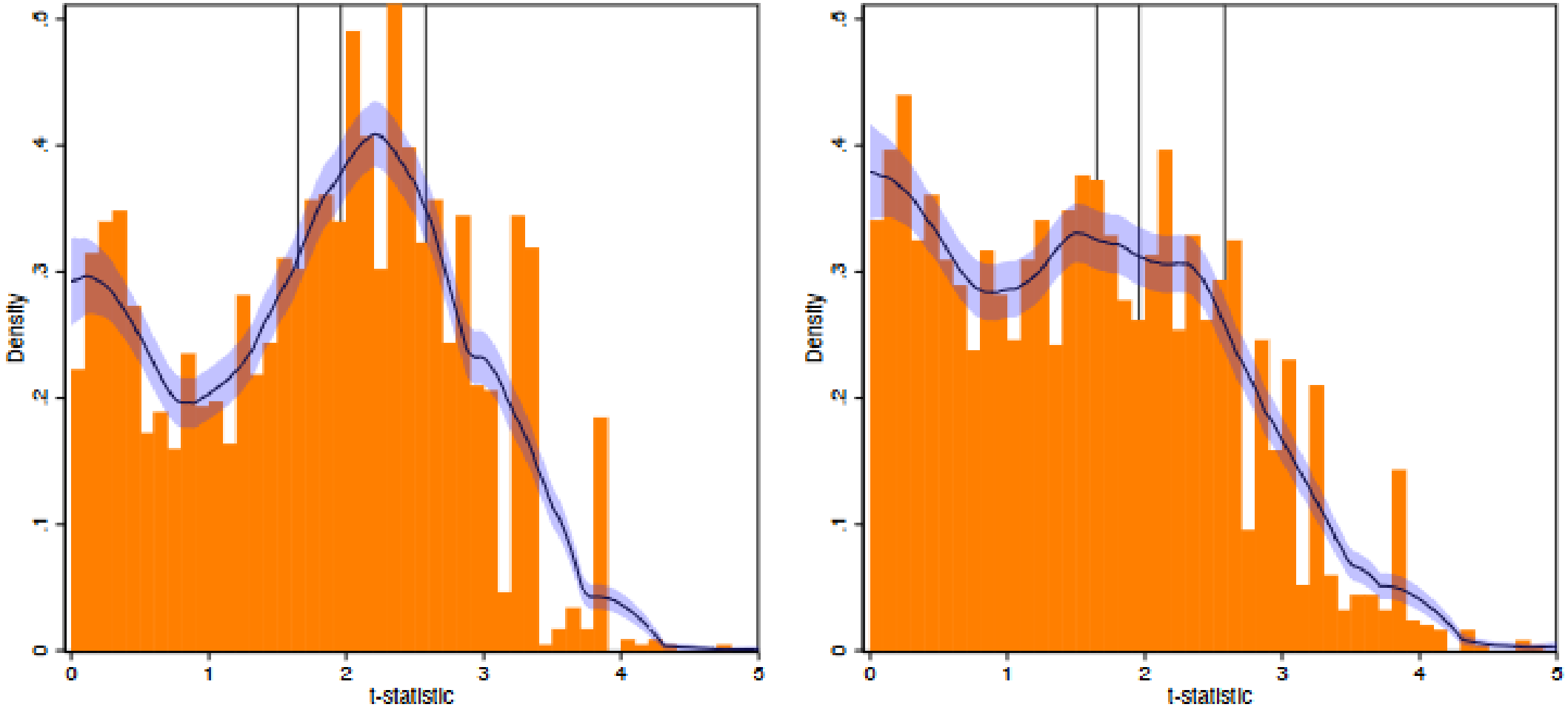
- **25% of studies have a coding error:**
  - Range from minor to MAJOR
    - » Ex. 75% of observations are duplicates
    - » Not cleaning raw data (e.g., St. Louis, St Louis, StLouis, …)
    - » Not fully interacting DID model
    - » Not specifying GMM function

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

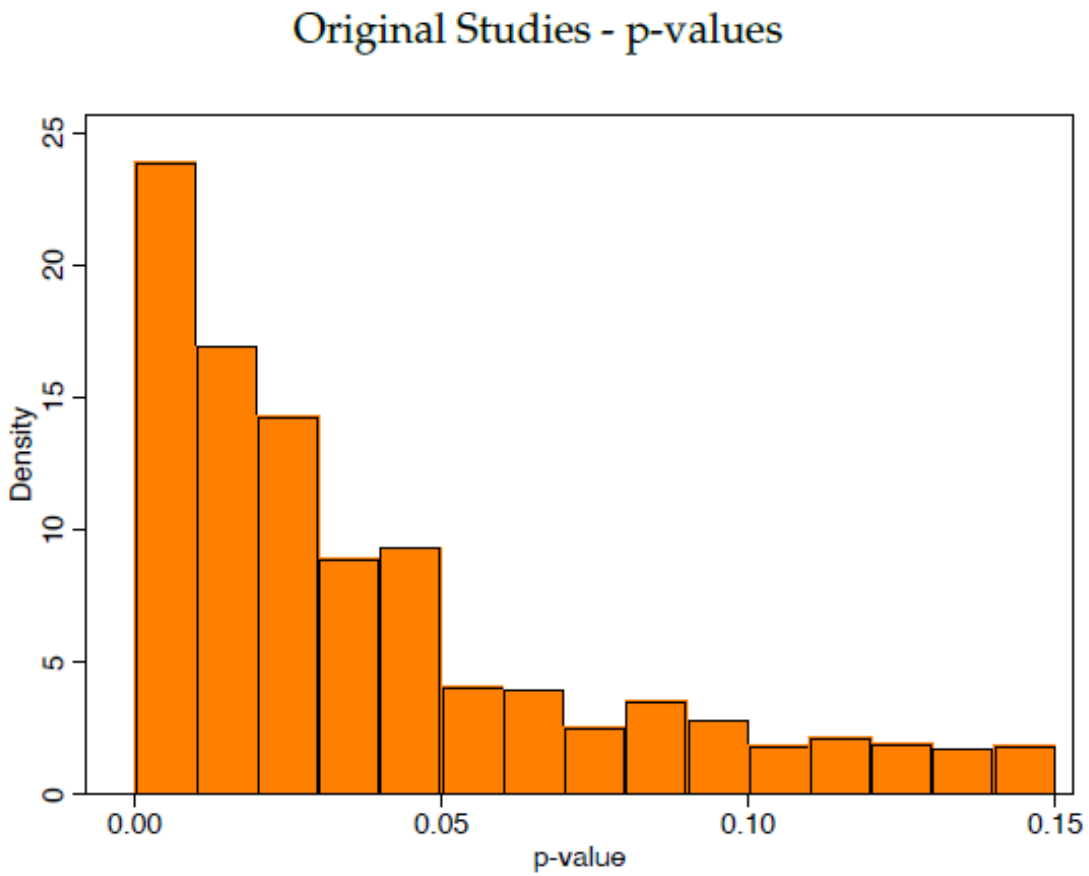
- **Important coding decisions buried in footnote or appendix**

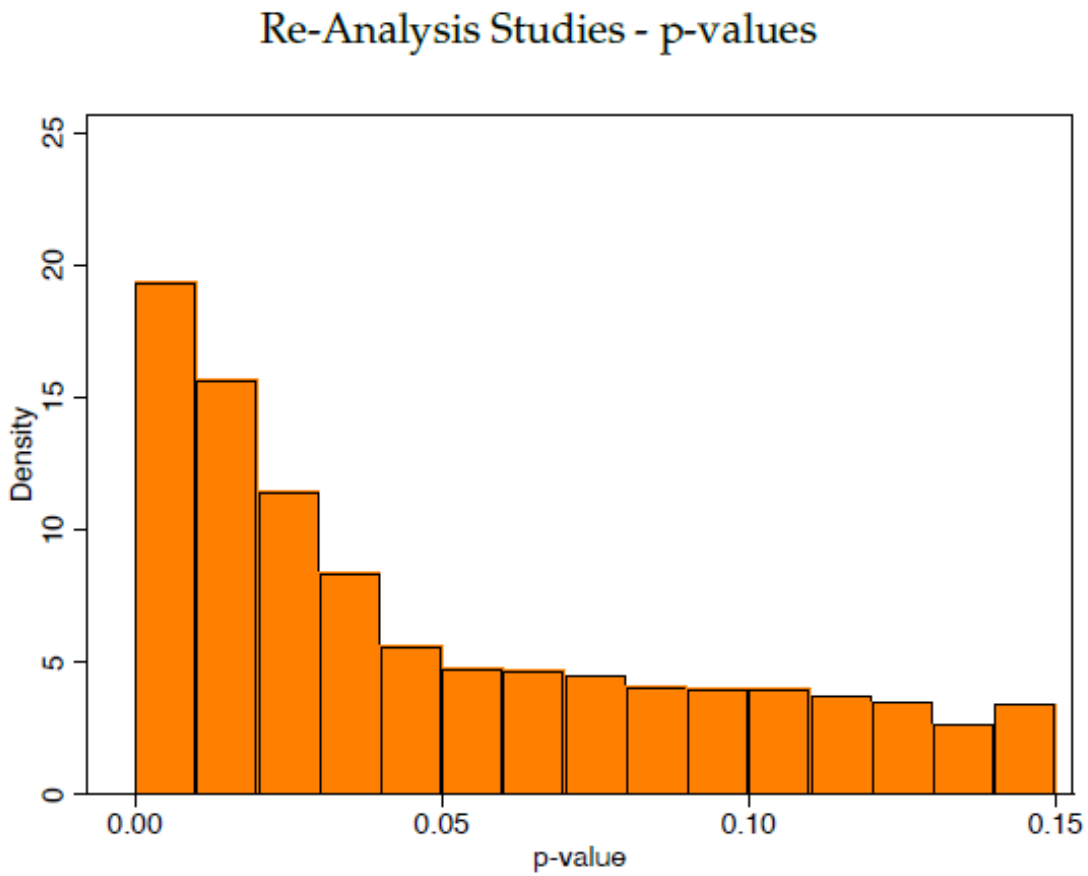
First Meta Paper: t-curves

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

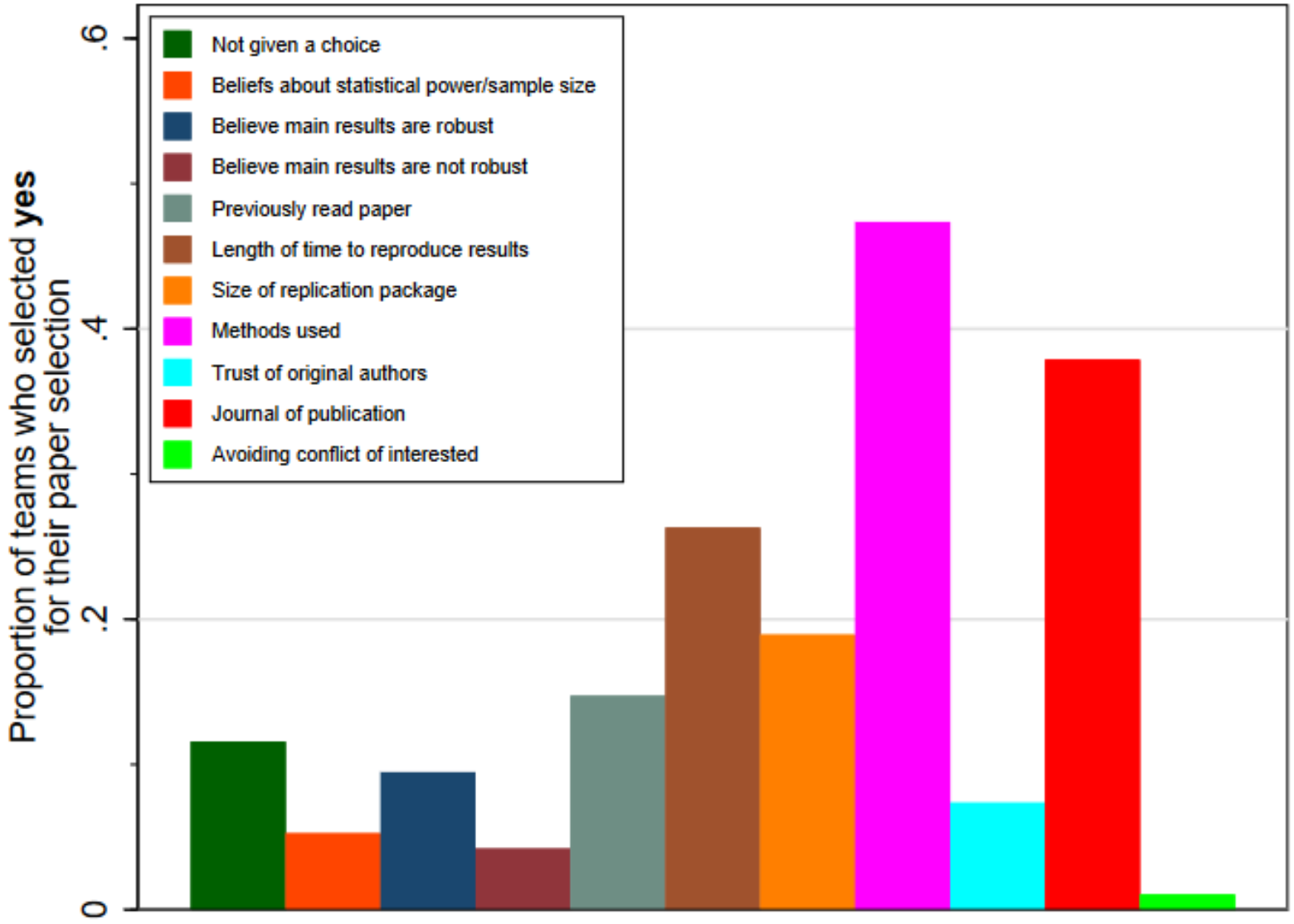
Proposal for Public Health

- **Major challenge is lack of data and code availability policies**
  - Working with editors
  - Need a full-time researcher/student to prepare replication package for replication games participants
- **Currently working on a grant proposal**
- **Need to build a board for public health/epidemiology**

Appendix

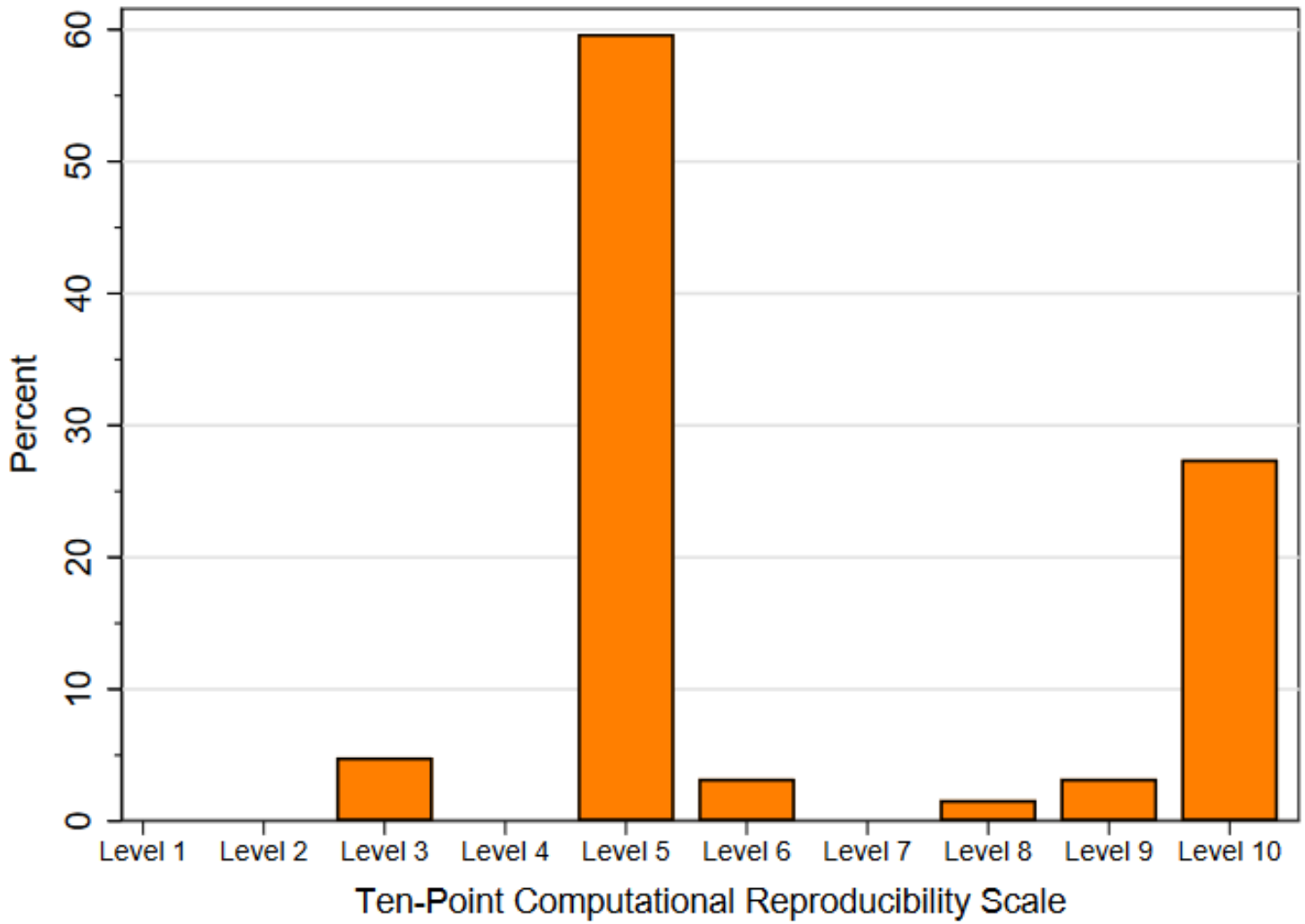
Paper Choice

Figure 13: For what reasons did you select your specific paper to reproduce and/or replicate from the list of papers provided? (Select all which apply)



First Meta Paper: Computational Reproducibility

Figure 1: 10-Point Computationally Reproducibility Sore

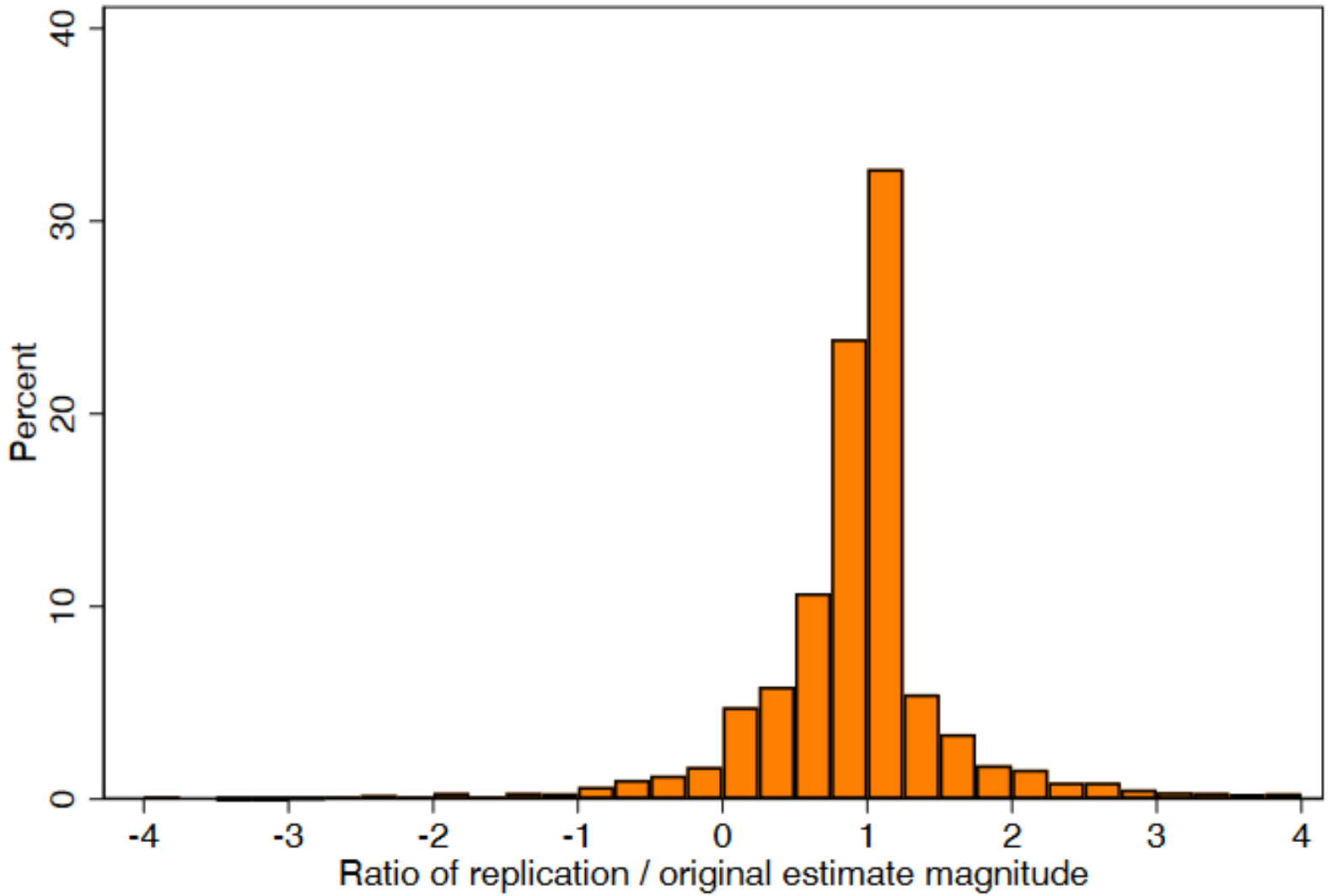


Many-Analysts: Results Next Week!

- 1-2. “Does reproducibility/replicability rate depend on **replicators’ experience coding**?” or “**academic experience**?”
- 3. “Does reproducibility/replicability rate depend on the **authors’ experience**?”
- 4. “Does reproducibility/replicability rate depend on the **interaction of the authors’ experience and replicators’ experience**?”
- 5. “Does reproducibility/replicability rate depend on the **interaction of the authors’ prestige and replicators’ prestige**?”
- 6-7. “Does reproducibility/replicability rate depend on the original authors providing **raw data**?” or “**raw or intermediate data**?”
- 8. “Does reproducibility/replicability rate depend on the original authors providing **cleaning code**?

First Meta Paper: Relative Effect Size

Figure 5: Relative Effect Size



How Can you Contribute?

- **This is your Institute! Help us make your discipline more open and credible!**

- **Editors:**
  - Contact us if you want to put together a special issue, section, or data and codes policy

- **Researchers:**
  - Contact us if you'd like to reproduce/replicate a study
  - Or participate in Replication Games
  - Or if you have already replicated a study and want to disseminate it

INSTITUTE for REPLICATION

Email: instituteforreplication@gmail.com

Twitter: https://twitter.com/I4Replication

Webpage: https://i4replication.org/

Figure 11: Code Availability

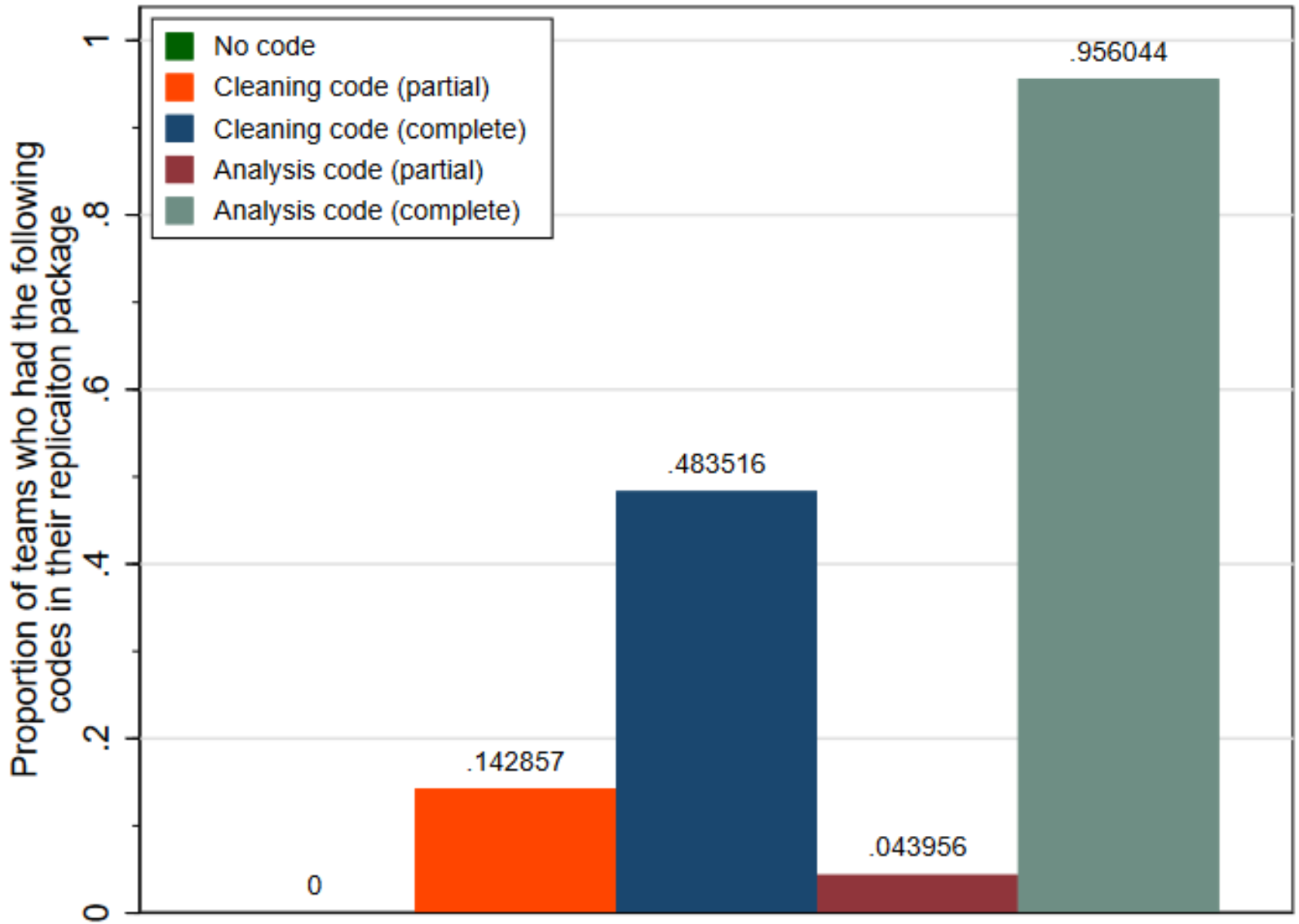
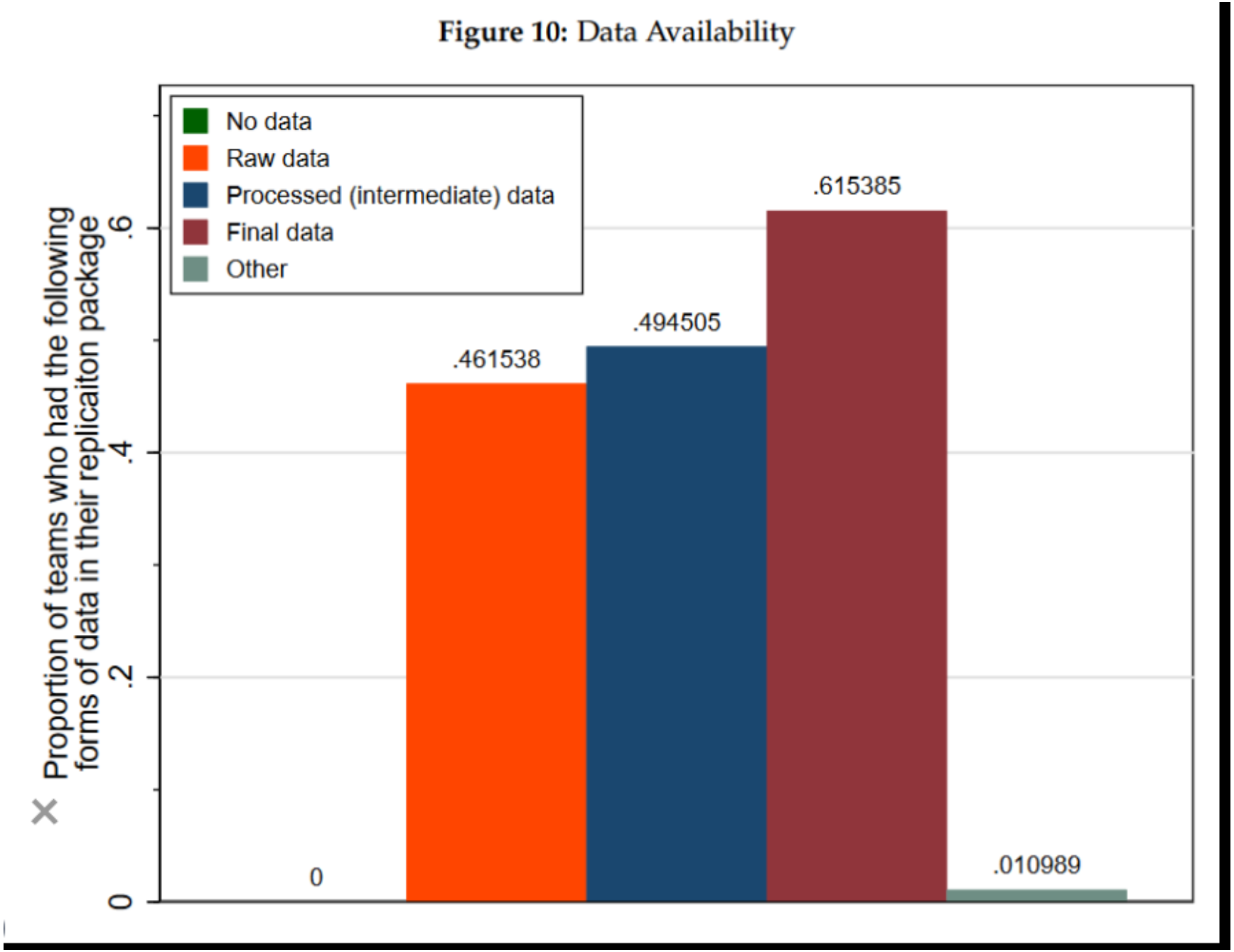
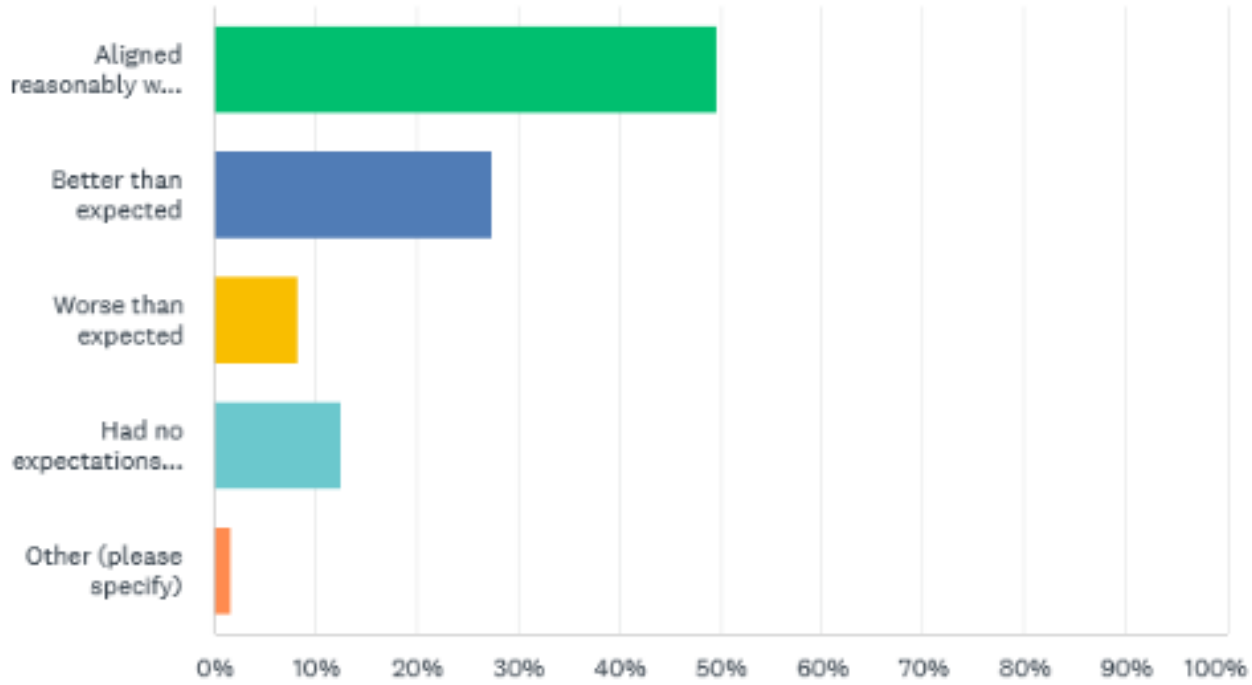


Figure 10: Data Availability



Which of the following best describes how the replication package aligned with your expectations:

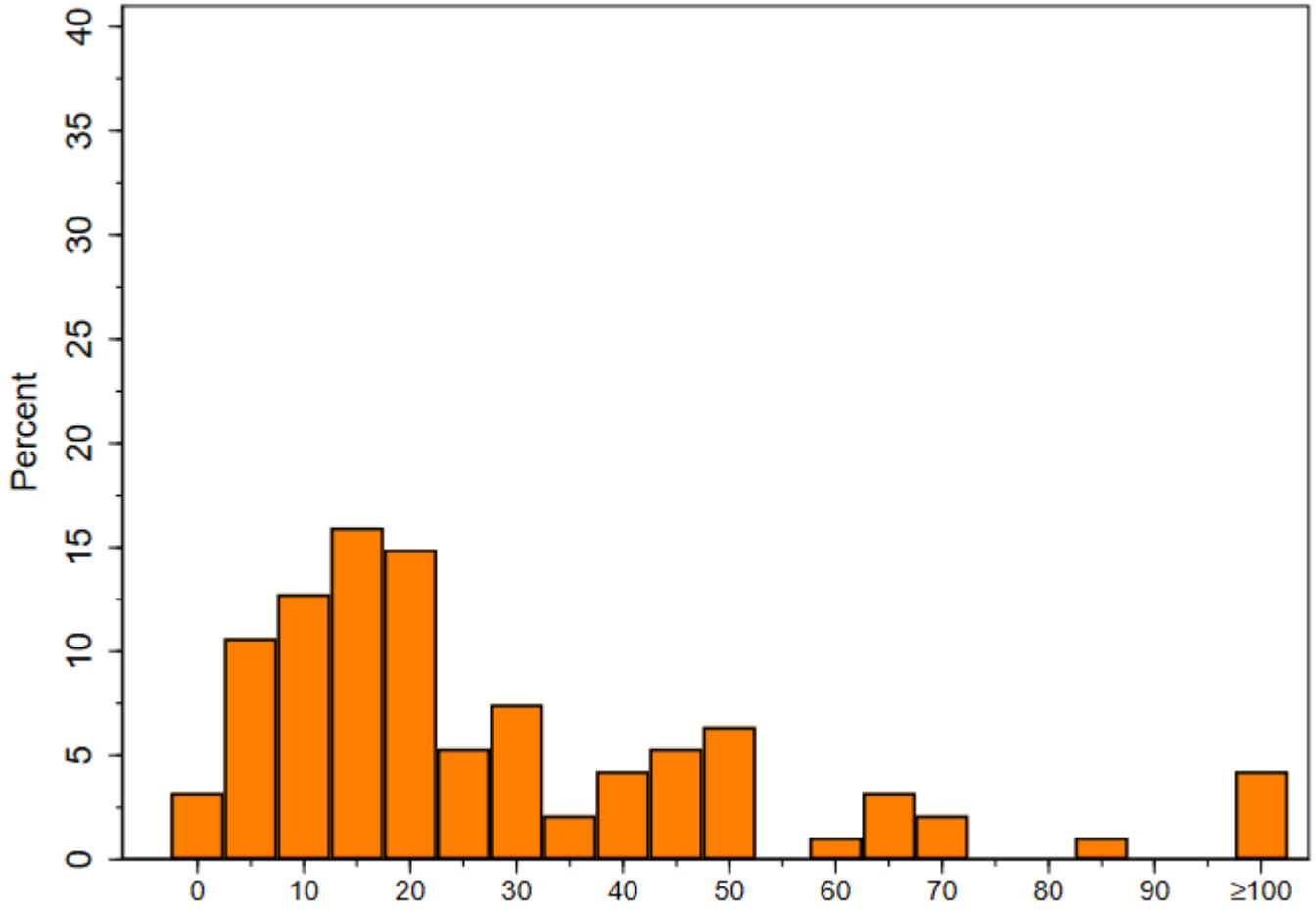
Answered: 276 Skipped: 37



Notes: This Figure illustrates the responses to the question: "For what reasons did you select your specific paper to reproduce and/or replicate from the list of papers provided?".

Active Work Days

Figure 15: Number of Active Work Days



Publishing Replications

- **Special issues dedicated to replications:**
  - Politics: Research & Politics
  - Economics: (i) Canadian Journal of Economics and (ii) Economic Inquiry
- **Journals with replication section**
  - Journal of Political Economy: Micro, World Development Perspectives, Spatial Economic Analysis, etc.
- **Survey of editors and more information:**
  - https://i4replication.org/publishing.html

Discipline	Journal	Editor	Q1 - Code	Q2 - Code	Long Answer
Economics	Journal of Economics	John Doe	Yes	No	In fact, this journal does not regularly publish commentary about prior publications... Read more

Incentives for Replicators

- **Help with publication and dissemination of their replication**
  - Put in contact with other replicator(s) replicating same study
  - Special issue at selected journals
  - Submit session proposals to conferences including AEA P&P
  - Replicators are automatically coauthors for meta-paper
  - Collaborating with instructors who have their graduate students replicate studies

Platform and Guidelines for Replications

- Social Science Reproduction Platform
  - https://www.socialsciencereproduction.org/

- Guide for Accelerating Computational Reproducibility in the Social Sciences (ACRe): https://bitss.github.io/ACRE/

- Template (word document) for replicators

Proprietary data

- **Most studies include multiple data types:**
  - Prepare a file with the name of the data set and how it can be accessed
    - » Data editors already do part of this work
  - This information will then be shared with the editorial board
  - Role of editors is to identify potential replicators who have access to the data and have excellent knowledge of the literature

- **Collaboration with *cascad*:** https://www.cascad.tech/

Instructions given to replicators

- **What are sensible robustness checks?**
  - Set of possible specifications is very large
    - » Possibility of adding variables to the analysis
- **What should replicators focus on:**
  - Identify the main or preferred specification and main results
    - » Not always described by the original authors
    - » Different replicators may focus on different results depending on data availability or skills (similar to a referee report)

Instructions given to replicators

- **Using same sample: Examples of robustness checks:**
  - Coding of dep var, main indep variable and controls
  - Weight
  - Standard errors
  - Outliers
  - Choice of parameters
  - Compare PAP to non-PAP
  - Anything else? Need to talk to macroeconomists…

Instructions given to replicators

- **Changing sample: Examples of robustness checks:**
  - Look at the raw data and check data restrictions made (e.g., dropping individuals outside and age range)
  - Do the restrictions make sense? Are the results robust to changing those restrictions?

- **Adding new variables: Examples of robustness checks:**
  - Adding key missing control variables