INSTITUTE for REPLICATION

Email: instituteforreplication@gmail.com

Bsky: @i4replication.bsky.social

Webpage: https://i4replication.org/

Reproducibility, Replication Games, and The Institute for Replication

Journées du réseau français de la recherche reproductible

Abel Brodeur
University of Ottawa
Institute for Replication

Definitions

- **(1) Computational reproducibility:**
  - Same data and codes. Run the codes and check if results are similar

- **(2) Robustness/sensitivity analysis** Different codes

  If I say reproduce/reproducibility, I usually mean (1+2)

- **(3) Replication:**
  - New data

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

How do we fix this mess…

- **(Lack of) data and code sharing**
  - That's the biggest and most important hurdle. Fix this and we can mass reproduce
- **Editors will go out of their way to not publish comments/retract**
  - Mass reproducibility can fix this
- **$$$: Funders think replications are not "innovative"**
  - Don't need money for reproducibility
- **Original authors will be pissed at you for criticizing their work**
  - Need mediator + institutionalize the practice

Institute for Replication (I4R)

- **Launched in 2022**
  - Should be called Institute for Reproduction, but too late and replication sounds better

- **Initial focus on economics and political science:**
  - Now collaborations with *Nature Human Behaviour* and *Psychological Science*
  - Mass reproducing deforestation studies and air pollution
  - Soon leading public health journals

- **Objectives:**
  - ***Mass*** reproducibility and robustness (and selected replications with new data)
  - 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 9 economics and 4 political science journals (2022-)
  - Psychological Science (2024-)
  - Nature Human Behaviour (2023-)
  - Deforestation and air pollution (2012-2022)
    - » Several leading journals (e.g., Nature Ecology & Evolution, PNAS)
  - Public health (2025-)
    - » Journals to be determined. Emailing editors now to see if they care about reproducibility and enforce their data and code availability policy (they don't, but maybe they would…)

I4R's Strategies for Generating Reproductions/Replications

- (1) **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
  - » 29 events for 2024 with About 1,300 participants
    - Coming to a university near you in 2025…
      - Bordeaux (Aug 24th), **Paris FRRN** (Oct 3rd) and **Lyon ENS** (oct 9th)

I4R's Strategies for Generating Reproductions/Replications

- (2) **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

- (3) **Admin data, non-public data and lab experiments**
  - » Payments to replicators (up to USD 5,000)
    - Started this stream last 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 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

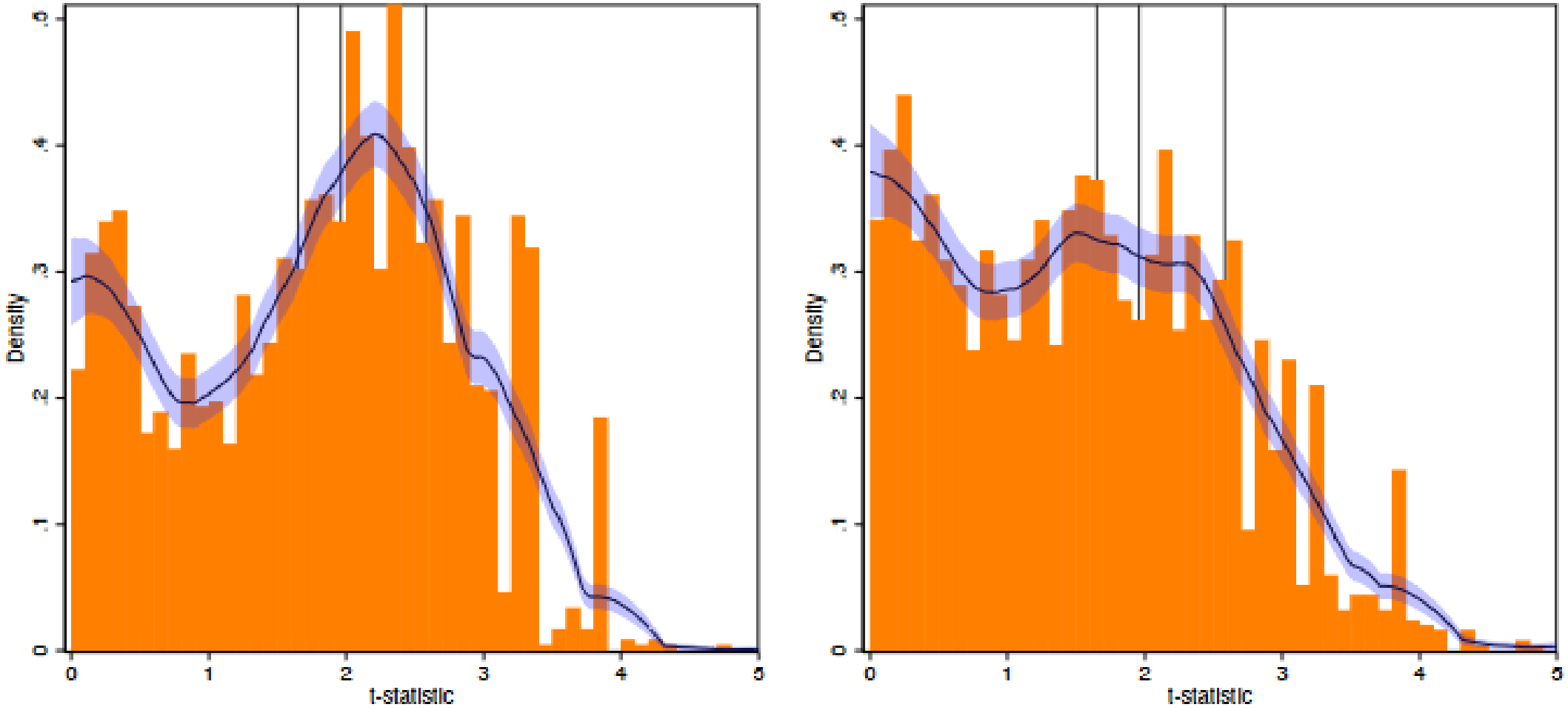
- **25% of studies have a coding error:**
  - Range from minor to MAJOR
    - » Ex. 75% of observations are duplicates
    - » Not cleaning raw data (e.g., St. Louis, St Louis, StLouis, …)
    - » Not fully interacting DID model
    - » Not specifying GMM function

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

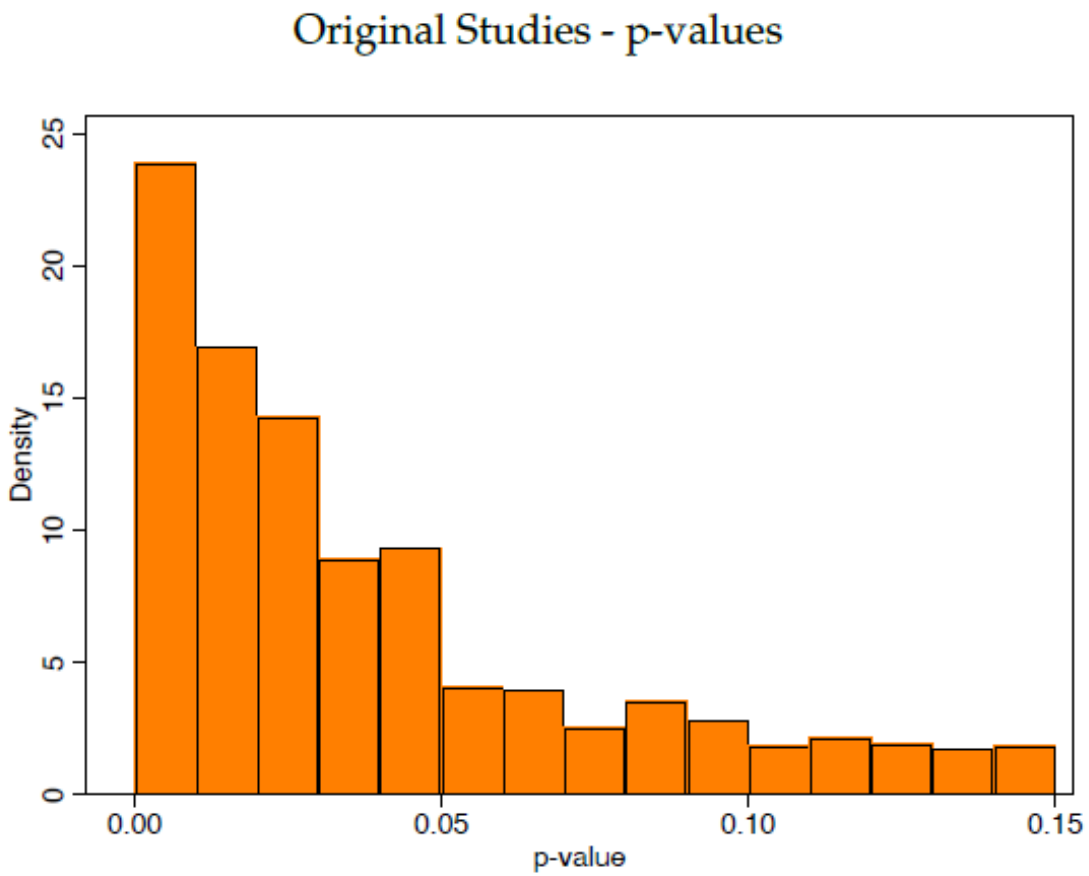
Important coding decisions buried in a footnote or appendix

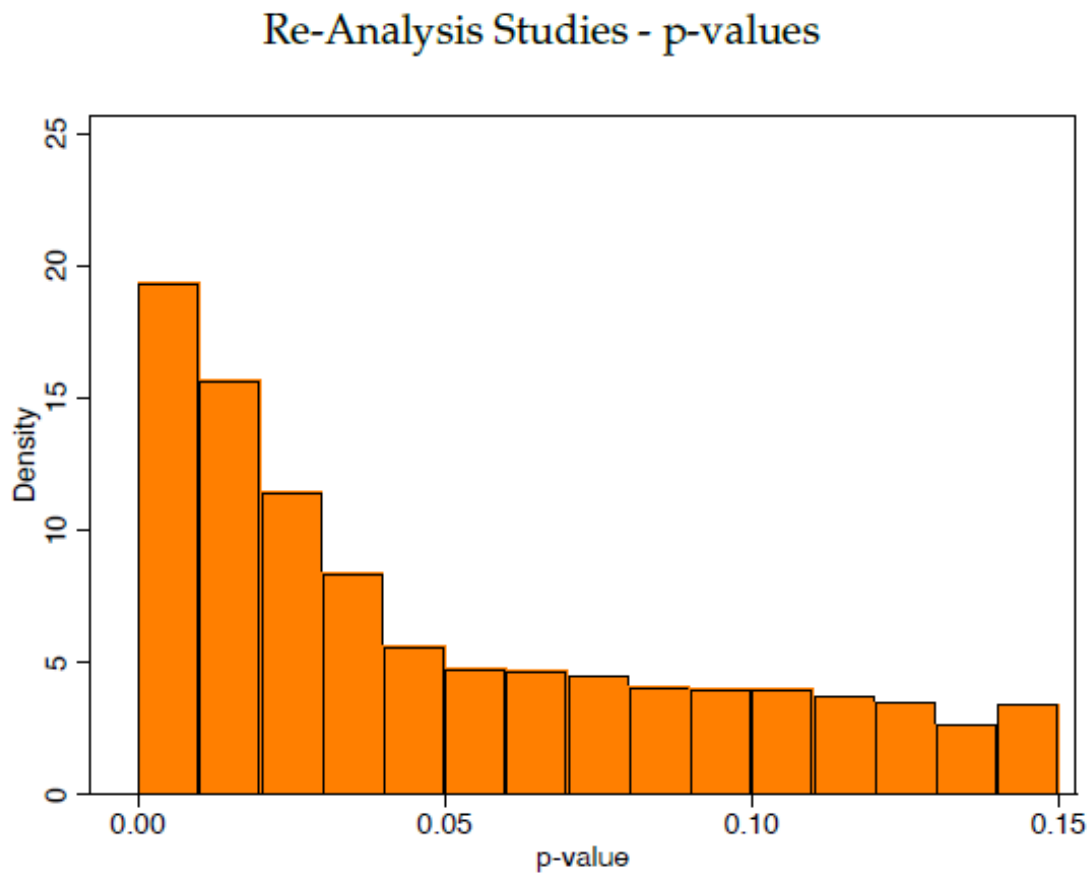
First Meta Paper: t-curves

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

Q&A: Hot Takes

- **Replication with new data without reproduction first is non-sensical**
- **Lack of reproduction/replication is your (researchers') fault, not funders/editors**
- **A study not reproducing/replicating is actually fine by me**
  - Not everything should reproduce/replicate