

BCB 731:

# *Defense Against the Dark Arts*



## Into the Garden of Forking Paths

*October 18th, 2023*



# Many Labs 2: Investigating Variation in Replicability Across Samples and Settings

GSAGE

Richard A. Klein<sup>1</sup>, Michelangelo Vianello<sup>2</sup>, Fred Hasselman<sup>3,4</sup>,  
 Byron G. Adams<sup>5,6</sup>, Reginald B. Adams, Jr.<sup>7</sup>, Simon Alper<sup>8</sup>,  
 Mark Aveyard<sup>9</sup>, Jordan R. Axelson<sup>10</sup>, Mayava T. Babalola<sup>11</sup>,  
 Stéphan Bahnik<sup>12</sup>, Rishabh Baratra<sup>13</sup>, Mihály Berkucsi<sup>14</sup>,  
 Michael J. Beaman<sup>15</sup>, Daniel J. Berntson<sup>16</sup>, Barbara Bozaleksza<sup>17</sup>,  
 Evan Dorn-Bianchi<sup>18</sup>, Daniel Bociurkiw<sup>19</sup>, Mark J. Brandt<sup>20</sup>, Robert Buschung<sup>20</sup>,  
 Anna Cabak Riedel<sup>21</sup>, Husjan Cal<sup>22</sup>, Fanny Cambier<sup>23,24</sup>,  
 Jesse Chandler<sup>25,26</sup>, Jen-Ho Chang<sup>27,28</sup>, Armand Chaudhary<sup>29,30</sup>,  
 Eva K. Chua<sup>31</sup>, William Cheung<sup>32</sup>, David C. Chudler<sup>33</sup>,  
 Jennifer A. Colman<sup>34</sup>, Kristian Colman<sup>35</sup>, Morgan A. Conroy<sup>36</sup>,  
 Katherine C. Cowan<sup>37</sup>, Cheryl C. Cromer<sup>38</sup>, Fiery Cushman<sup>39</sup>,  
 Zubairu K. Dagon<sup>40</sup>, Iker Dalgar<sup>41</sup>, Anna Dalla Rosa<sup>42</sup>,  
 William E. Davis<sup>43</sup>, Maikel de Brujin<sup>44</sup>, Leader De Schutter<sup>45</sup>,  
 Thibaut Devos<sup>46</sup>, Daniel J. Devos<sup>47</sup>, Carolyn Dickey<sup>48</sup>,  
 Nerea Domínguez<sup>49</sup>, Kristin Nicole Dukes<sup>50</sup>, Carsten Dunkel<sup>51</sup>,  
 Kevin Durheim<sup>52</sup>, Charles R. Eberle<sup>53</sup>, John E. Edlund<sup>54</sup>,  
 Anja Eller<sup>55</sup>, Alexander Scott English<sup>56</sup>, Carolyn Finck<sup>57</sup>,  
 Natalie Franklinowka<sup>58</sup>, Miguel-Angel Freyre<sup>59</sup>, Mike Friedman<sup>53,54</sup>,  
 Elisa Margarida Gama<sup>60</sup>, Daniel Gao<sup>61</sup>, Daniel Gao<sup>62</sup>,  
 Steffen R. Giesaen<sup>63</sup>, Tripti Gill<sup>64</sup>, Timo Gräfe<sup>65,66</sup>, Angel Gómez<sup>66</sup>,  
 Roberto González<sup>67</sup>, Jesse Graham<sup>68</sup>, Jon E. Grahe<sup>69</sup>, Ivan Gratch<sup>70</sup>,  
 Eva C. T. Green<sup>71</sup>, Kakul Ha<sup>72</sup>, Matthew Haigh<sup>73</sup>, Elizabeth L. Haines<sup>74</sup>,  
 Michael P. Hall<sup>75</sup>, Marie E. Han<sup>76</sup>, Jason W. Hicks<sup>77</sup>, Daniel Hinkeldey<sup>78</sup>,  
 Jeffrey H. Hines<sup>79</sup>, Brian Hines<sup>80</sup>, Alan Hirsch<sup>81</sup>, Yael Inbar<sup>82</sup>,  
 Áse H. Innes-Ker<sup>83</sup>, William Jiménez-Ledea<sup>84</sup>, Melisaa Sur John<sup>85</sup>,  
 Jennifer A. Joy-Gaba<sup>86</sup>, Roza C. Kamiloju<sup>87</sup>, Heather Barry Kappes<sup>88</sup>,  
 Serdar Karataş<sup>89</sup>, Harun Karick<sup>17,90</sup>, Victor N. Keller<sup>91</sup>, Anna Kendre<sup>92</sup>,  
 Norman Kippen<sup>93</sup>, Gergely Kovács<sup>94</sup>, Camilo Kouril<sup>95</sup>, Lucy E. Krueger<sup>96</sup>,  
 German Krumm<sup>97</sup>, Janice Kurzak<sup>98</sup>, Daniel L. Lazarev<sup>99</sup>,  
 Carmel A. Levitan<sup>100</sup>, Leah Lewellen<sup>101</sup>, Neil E. Lewis<sup>102</sup>, Samuel Lins<sup>103</sup>,  
 Nikolette P. Lipske<sup>104</sup>, Joy E. Luce<sup>104</sup>, Esther Maassen<sup>105</sup>,  
 Angela T. Mainous<sup>106</sup>, Winfried Malinga<sup>107</sup>, Robyn K. Mallett<sup>107</sup>,  
 Satia A. Mavrotis<sup>108</sup>, Janko Medović<sup>109</sup>, Fernando Mena-Pacheco<sup>104</sup>,  
 Tatjana L. Mihajlović<sup>110</sup>, Wendy L. Morris<sup>105</sup>, Sean C. Murphy<sup>107</sup>,  
 Andrey Myachykov<sup>109</sup>, Nick Neave<sup>111</sup>, Koen Neijenhuis<sup>104,109</sup>,

Many (most?)  
 “cognitive  
 biases” are  
 non-replicable

Klein et al.

Disgust Sensitivity Predicts Homophobia (Inbar et al., 2009)  
 Assimilation & Contrast Effects (Schwarz et al., 1991)

Correspondence Bias (Miyamoto & Kitayama, 2002)

Perceived Intentionality for Side Effects (Knobe, 2003)

Trolley Dilemma 1 (Hauser et al., 2007)

False Consensus: Supermarket Scenario (Ross et al., 1977)

Moral Typcasting (Gray & Wegner, 2009)

False Consensus: Traffic-Ticket Scenario (Ross et al., 1977)

Preferences for Formal vs. Intuitive Reasoning (Norenzayan et al., 2002)

Less-Is-Better Effect (Hsee, 1998)

Effect of Framing (Tversky & Kahneman, 1981)

Cardinal Direction & SES (Huang et al., 2014)

Moral Foundations of Liberals vs. Conservatives (Graham et al., 2009)

Reluctance to Tempt Fate (Risen & Gilovich, 2008)

Trolley Dilemma 2 (Hauser et al., 2007)

Consumerism Undermines Trust (Bauer et al., 2012)

Influence of Incidental Anchors (Critcher & Gilovich, 2008)

SVO and Family Size (Van Lange et al., 1997)

Moral Violations & Cleansing (Zhong & Liljenquist, 2006)

Vertical Position & Power (Giessner & Schubert, 2007)

Directionality & Similarity (Tversky & Gati, 1978)

SMS & Well-Being (Anderson et al., 2012)

Priming “Heat” (Zaval et al., 2014)

Structure Promotes Goal Pursuit (Kay et al., 2014)

Disfluency Engages Analytic Processing (Alter et al., 2007)

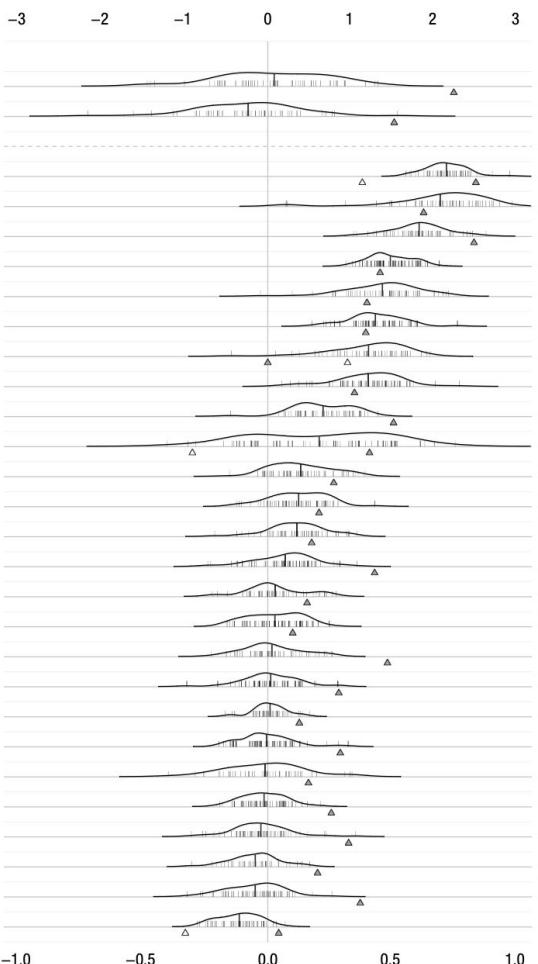
Effect of Choosing vs. Rejecting (Shafir, 1993)

Affect & Risk (Rottenstreich & Hsee, 2001)

Construing Actions as Choices (Savani et al., 2010)

Original Effect Size

Cohen's  $\eta^2$

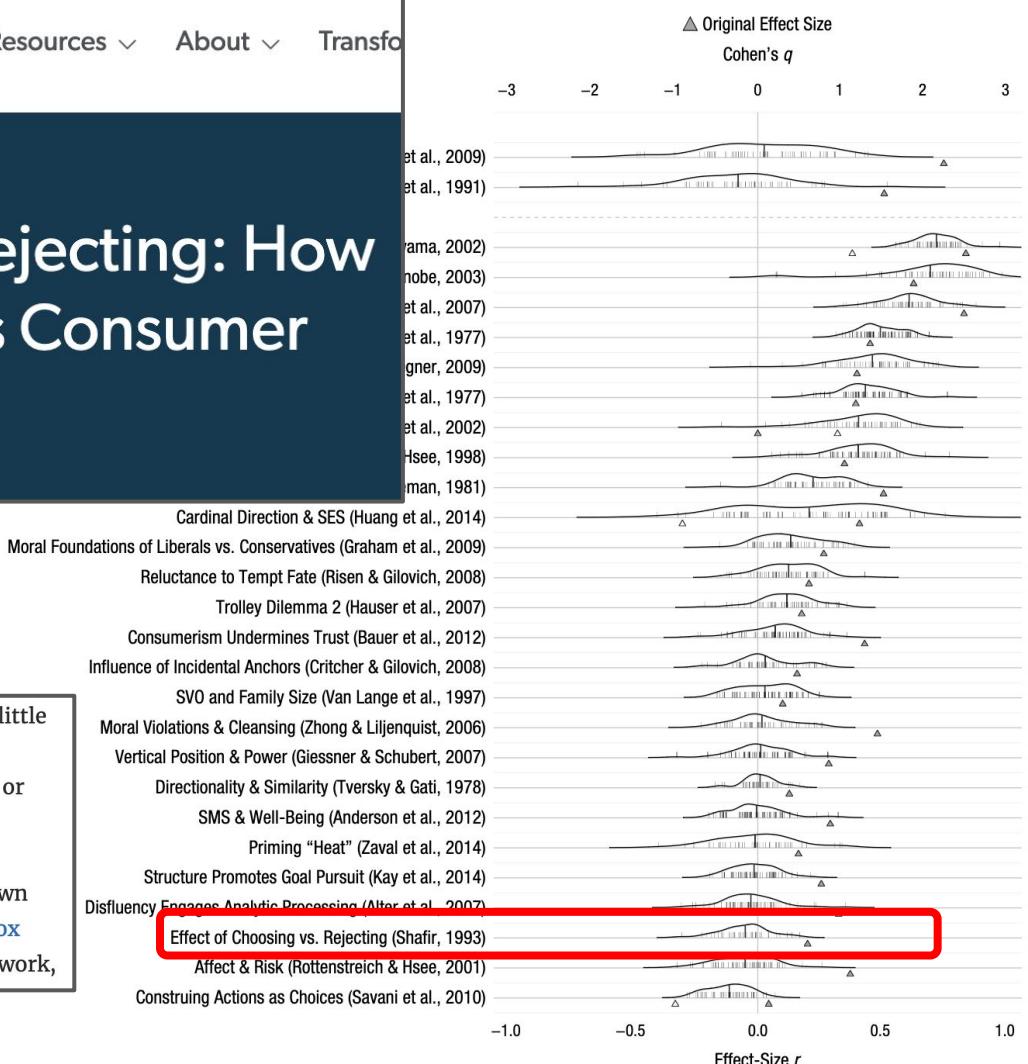


# Choosing By Selecting Or Rejecting: How Decision Strategy Influences Consumer Satisfaction

## Rejections Are More Contagious than Choices: How Another's Decisions Shape Our Own

Lana Xianglan Nan, Sang Kyu Park, Yang Yang 

Despite the prevalence of both choice and rejection behaviors in our daily lives, little is known about how others' choice and rejection behaviors influence our own preferences. Holding another person's decision outcome constant, are we more or less likely to conform to that decision if we perceive it as a choice versus as a rejection? Although the decision framing literature has greatly advanced our understanding of how our own act of choosing versus rejecting influences our own decision processes and outcomes (Dhar and Wertenbroch 2000; Laran and Wilcox 2011; Perfecto et al. 2017; Shafir 1993; Sokolova and Krishna 2016), none of this work,



# The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time\*

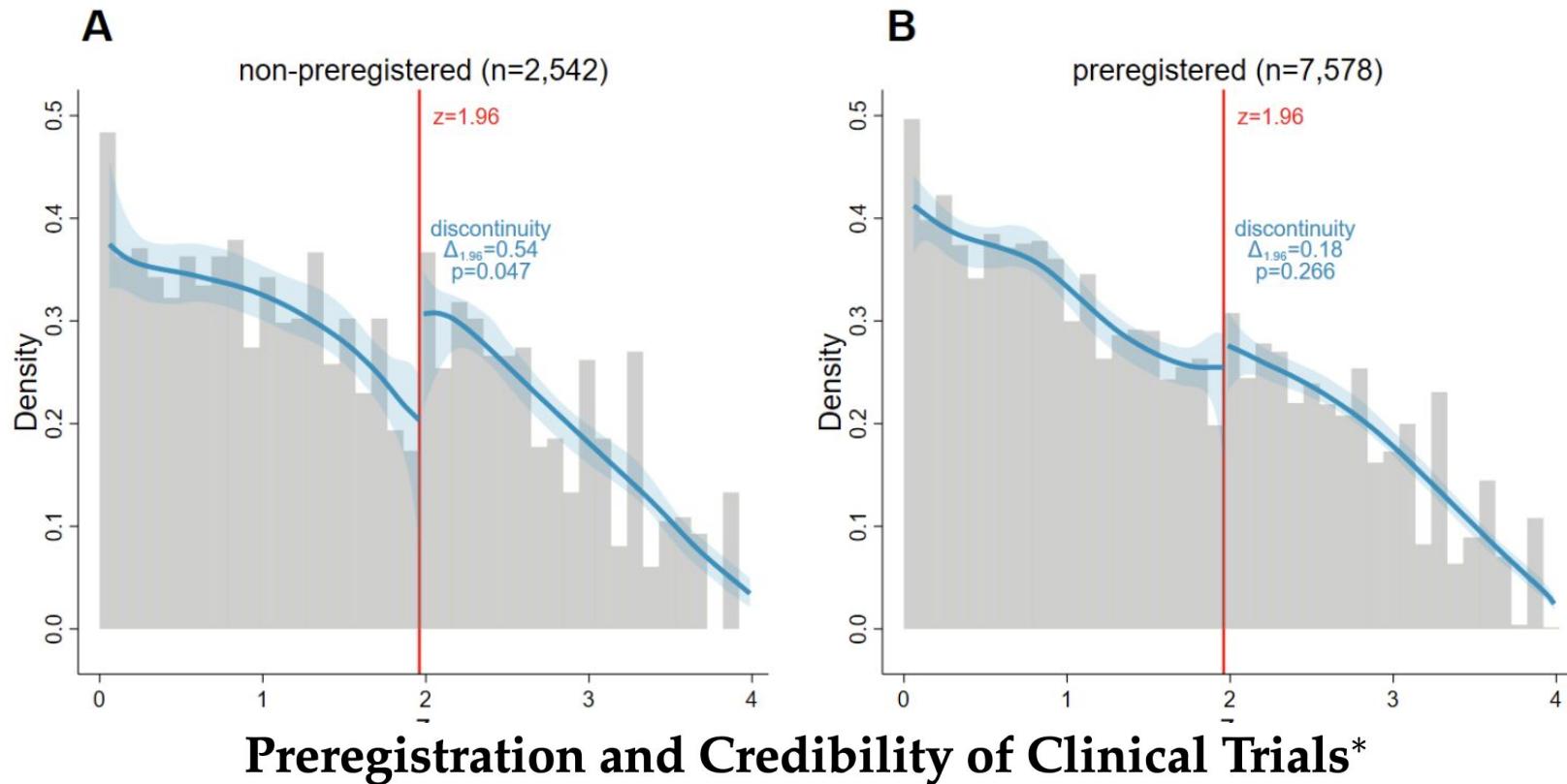
Andrew Gelman<sup>†</sup> and Eric Loken<sup>‡</sup>

14 Nov 2013

*“I thought of a labyrinth of labyrinths, of one sinuous spreading labyrinth that would encompass the past and the future . . . I felt myself to be, for an unknown period of time, an abstract perceiver of the world.” — Borges (1941)*

1. Simple classical test based on a unique test statistic,  $T$ , which when applied to the observed data yields  $T(y)$ .
2. Classical test pre-chosen from a set of possible tests: thus,  $T(y; \phi)$ , with preregistered  $\phi$ . For example,  $\phi$  might correspond to choices of control variables in a regression, transformations, and data coding and excluding rules, as well as the decision of which main effect or interaction to focus on.
3. Researcher degrees of freedom without fishing: computing a single test based on the data, but in an environment where a different test would have been performed given different data; thus  $T(y; \phi(y))$ , where the function  $\phi(\cdot)$  is observed in the observed case.
4. “Fishing”: computing  $T(y; \phi_j)$  for  $j = 1, \dots, J$ : that is, performing  $J$  tests and then reporting the best result given the data, thus  $T(y; \phi^{\text{best}}(y))$ .

# Preregistration reduces R-DOF



# Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials

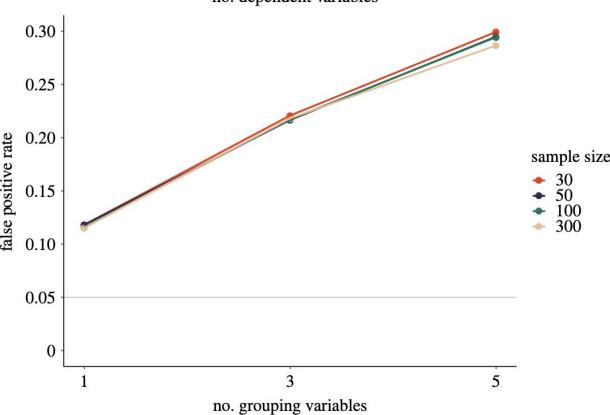
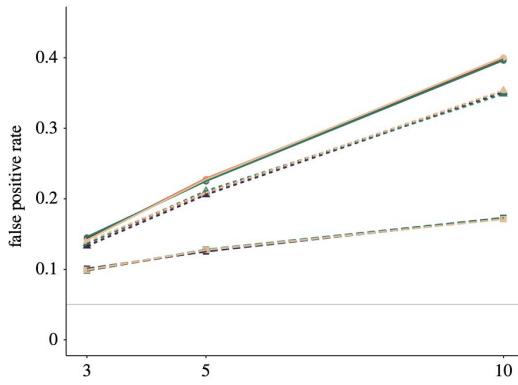
## ...but not a panacea

**Table 2.** Differences Between Primary Outcomes in Trial Registration and in Published Article for Studies With a Clear Description of the Primary Outcome in the Registry and Discrepancies Favoring Statistically Significant Results

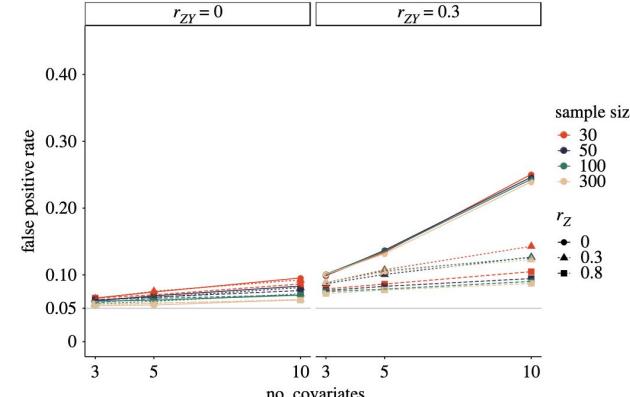
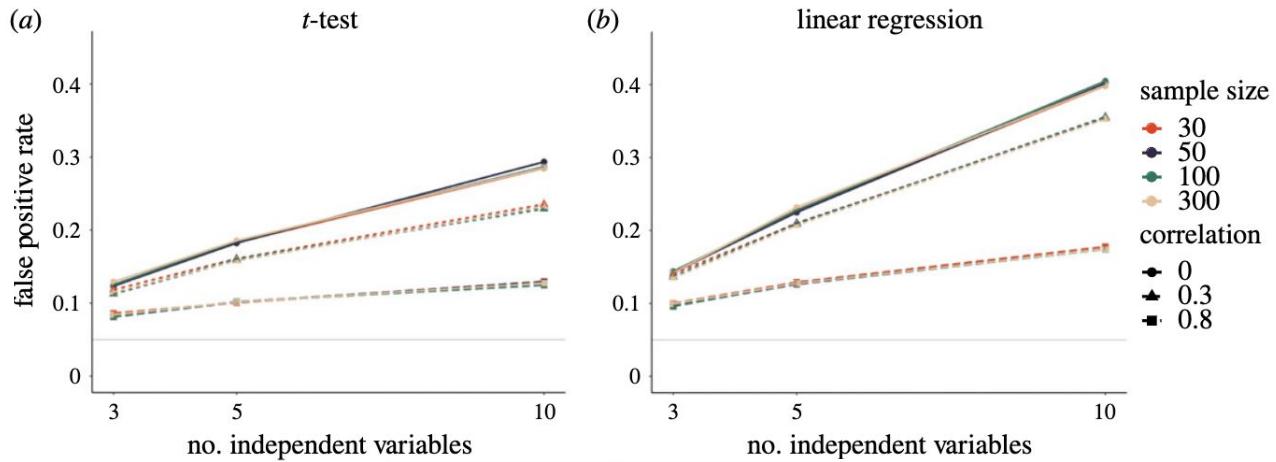
	No. (%) of Articles		
	All (n = 147)	General Medical Journals (n = 75)	Specialty Journals (n = 72)
Articles with different primary outcomes in trial registration and in published article	46 (31.3) <sup>a</sup>	22 (29.3) <sup>b</sup>	24 (33.3) <sup>c</sup>
Registered primary outcome omitted in text	15 (10.2)	8 (10.7)	7 (9.7)
New primary outcome introduced in text	22 (15.0)	11 (14.7)	11 (15.3)
Different timing of assessment of primary outcome	4 (2.7)	1 (1.3)	3 (4.2)
Published primary outcome described as secondary outcome in registry	8 (5.4)	5 (6.7)	3 (4.2)
Registered primary outcome reported as secondary outcome in text	6 (4.0)	4 (5.3)	2 (2.8)
Discrepancies in primary outcome favoring statistically significant results, No. <sup>d</sup>	46	22	24
Yes	19 (41.3)	9 (40.9)	10 (41.7) <sup>e</sup>
No	4 (8.7)	1 (4.5)	3 (12.5)
Impossible to conclude	23 (50.0)	12 (45.5)	11 (45.8)

# Big little lies: a compendium and simulation of *p*-hacking strategies

Angelika M. Stefan<sup>1,2</sup> and Felix D. Schönbrodt<sup>3</sup>



## How quickly do we get total noise with increasing researcher degrees of freedom?



# Does immigration reduce support for social services? (79 teams)

PNAS

RESEARCH ARTICLE

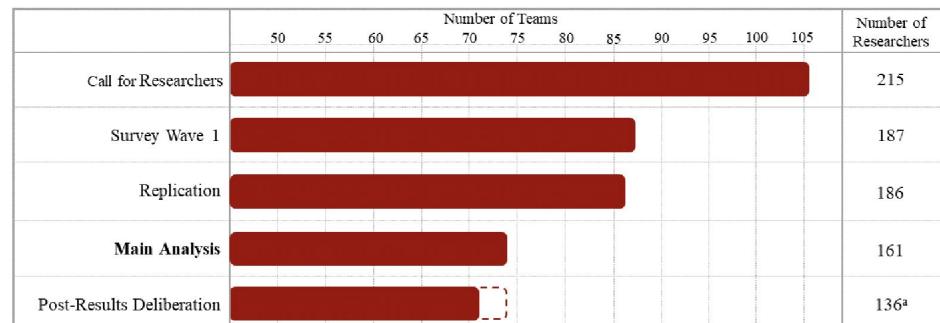
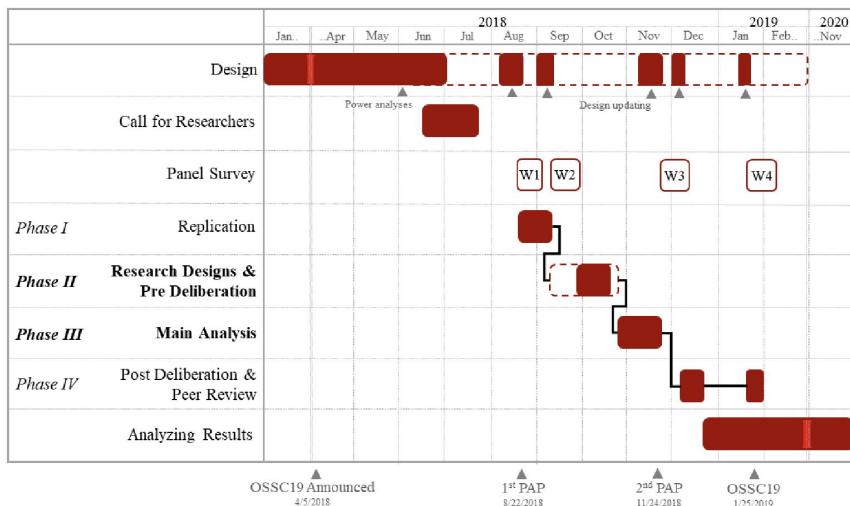
SOCIAL SCIENCES

OPEN ACCESS



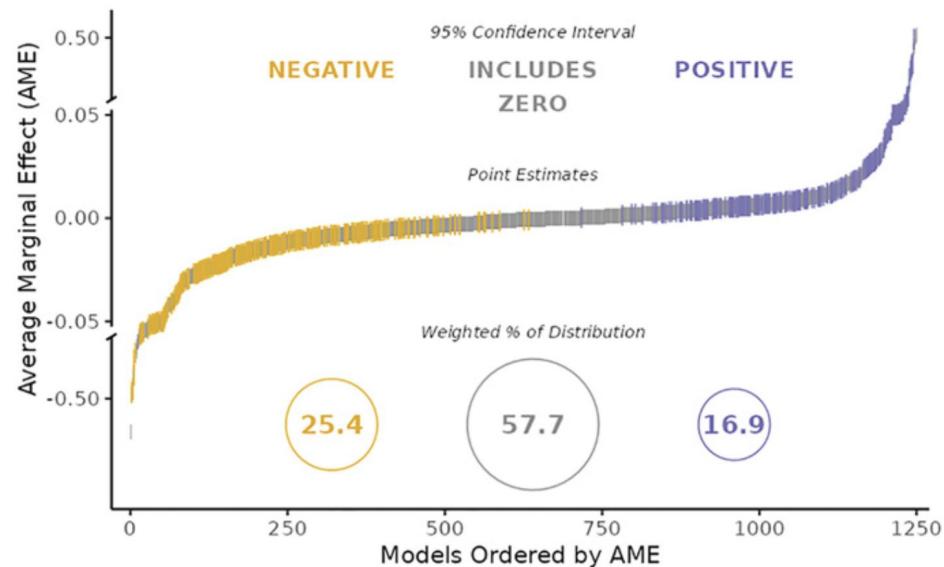
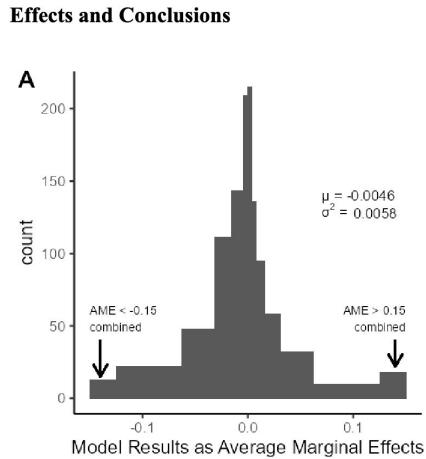
Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty

Edited by Douglas Massey, Princeton University, Princeton, NJ; received March 6, 2022; accepted August 22, 2022



<sup>a</sup> "Full Participation" required completion of the Replication and Main Analysis, not necessarily the Deliberation.

# Does immigration reduce support for social services? (79 teams)



Same data, different conclusions: Radical dispersion in empirical results  
when independent analysts operationalize and test the same hypothesis<sup>☆</sup>

Martin Schweinsberg <sup>a,1,\*</sup>, Michael Feldman <sup>b,1,\*</sup>, Nicola Staub <sup>b,1</sup>, Olmo R. van den Akker <sup>c</sup>,  
Robbie C.M. van Aert <sup>c</sup>, Marcel A.L.M. van Assen <sup>d</sup>, Yang Liu <sup>e</sup>, Tim Althoff <sup>e</sup>, Jeffrey Heer <sup>e</sup>,  
Alex Kale <sup>e</sup>, Zainab Mohamed <sup>f</sup>, Hashem Amireh <sup>g</sup>, Vaishali Venkatesh Prasad <sup>a</sup>,

Abraham Bernstein <sup>b,\*</sup>, Emily Robinson, Kaisa Snel

David Robinson, Nikhil Madan <sup>k</sup>, Raphael Silberz

Toshio Murase <sup>o</sup>, Benjamin Mandl <sup>p</sup>, Domenico Vi

C. Schaumans <sup>q</sup>, Stijn Kelchtermans <sup>r</sup>, Chan Nasee

S. Richard Chan <sup>v</sup>, Prestone Adie <sup>w</sup>, Paulius Alabu

Jeff Alstott <sup>z</sup>, Andrew A. Nelson <sup>aa</sup>, Eduardo Ariñ

Jason Baik, Laura Winther Balling <sup>ad</sup>, Sachin Ban

Brenda Barros-Rivera <sup>ah</sup>, Matt Bauer <sup>ai</sup>, Enuh Blais

Robert A. Briers <sup>am</sup>, Oliver Burkhard, Miguel-Ang

Olivia Chen, Michael Clark <sup>ao</sup>, Brent Cohn, Alex C

Paul G. Curran <sup>ar</sup>, Wilson Cyrus-Lai <sup>h</sup>, David Dai <sup>a</sup>

Henrik Danielsson <sup>au</sup>, Rosaria de F.S.M. Russo <sup>av</sup>,

Frank Dondelinger <sup>ax</sup>, Carolina Duarte de Souza <sup>ay</sup>

Ben Mark Dunn <sup>ag</sup>, Peter Adriaan Edelsbrunner <sup>ba</sup>

Yuanyuan Gong <sup>be</sup>, Erin Grand, Brandon Greenaw

Antony B. Hong <sup>h</sup>, David Hood, Justin Hsueh, Lil

Keith A. Hultman <sup>bj</sup>, Azka Javaid <sup>bk</sup>, Lily Ji Jiang

David Kane <sup>bn</sup>, Gregor Kappler <sup>bo</sup>, Erikson Kaszub

Madian Khabsa, Bennett Kleinberg <sup>bp</sup>, Jens Kouro

Angelos-Miltiadis Krypotos <sup>br</sup>, Dejan Lavbić <sup>do</sup>, Ru

Wei Yang Lim <sup>bt</sup>, Silvia Liverani <sup>bu</sup>, Bianca Loh <sup>h</sup>, I

Kyle MacDonald <sup>bx</sup>, Christopher R. Madan <sup>by</sup>, Lass

Alexandra Mangold, Adrienne Marshall <sup>ca</sup>, Helena

Katherine L. McLain <sup>a</sup>, Amelia A. McNamara <sup>cc</sup>, M

Ben Moore <sup>ce</sup>, Andrew Moore, Eric Nantz <sup>cf</sup>, Ziaud

Colleen S Nell <sup>ch</sup>, Andrew Arthur Nelson <sup>aa</sup>, Gusta

Christopher E. O'Brien, Patrick O'Neill <sup>ck</sup>, Kieran

Diana Palsetia <sup>al</sup>, Bianca Pereira, Ivan Pozdniakov <sup>g</sup>,

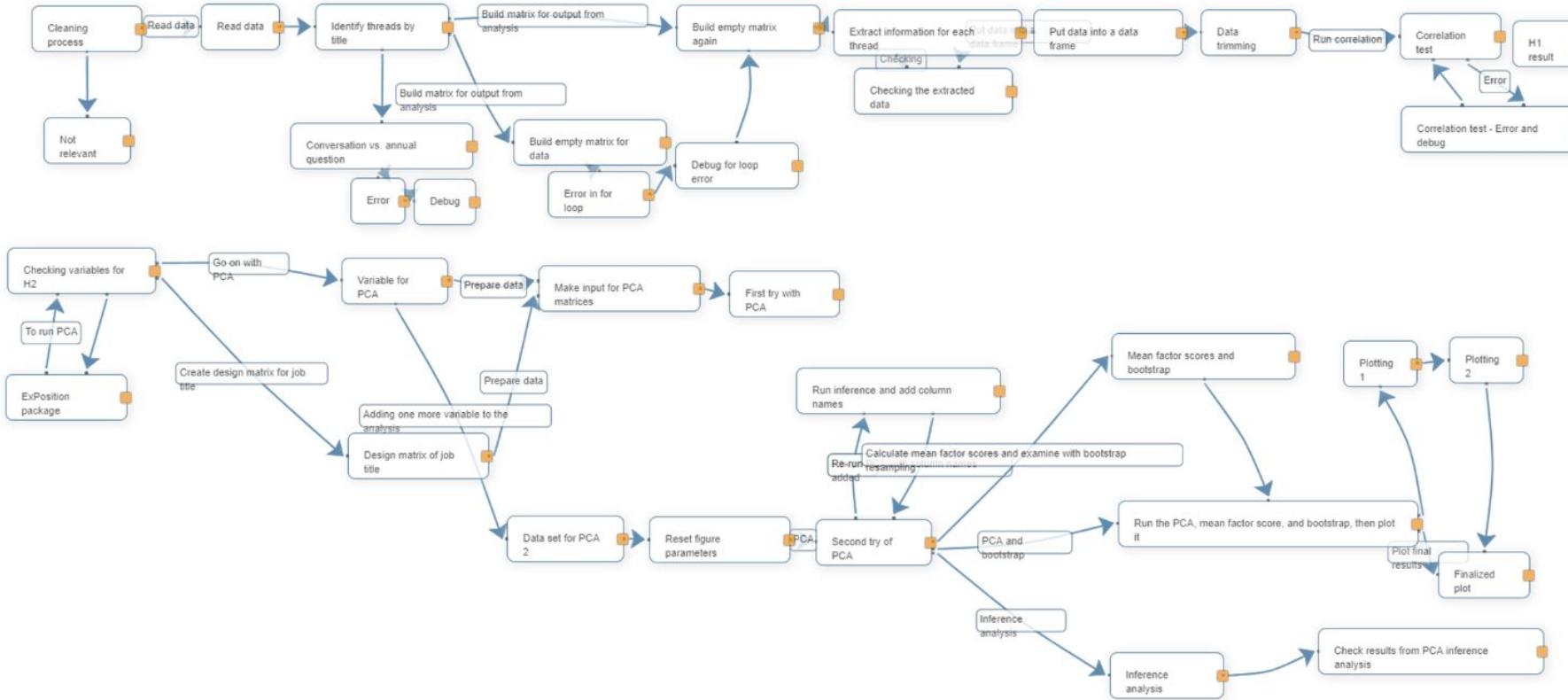
Travis Riddle <sup>cd</sup>, Amal (Akmal) Ridhwan Omar Ali <sup>cr</sup>,

Ivan Ropovik <sup>cs</sup>, Joshua M. Rosenberg <sup>ct</sup>,

Stephane Rothen, Michael Schulte-Mecklenbeck <sup>cu</sup>, Nirek Sharma <sup>cv</sup>, Gordon Shotwell <sup>cw</sup>,

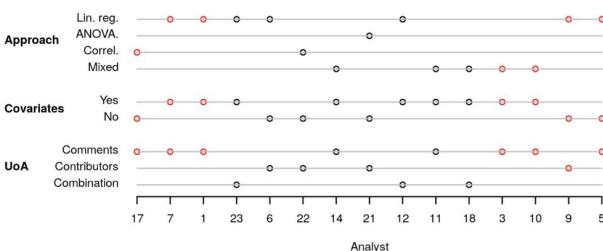
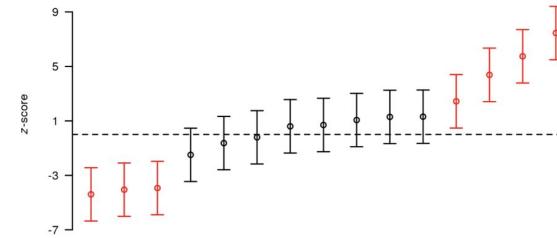
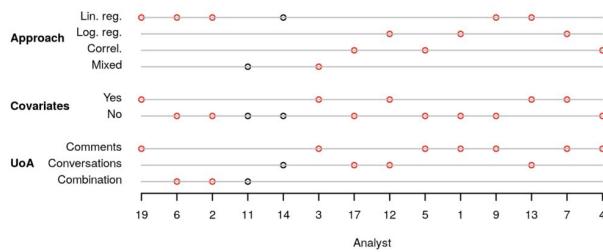
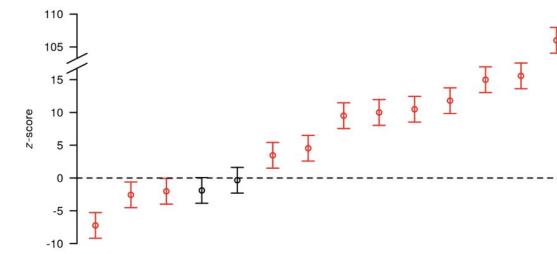
Martin Skarzynski, William Stedden, Victoria Stodden <sup>cx</sup>, Martin A. Stoffel <sup>cy</sup>, Scott Stoltzman <sup>cz</sup>,

Hypothesis	Significant in predicted (+) direction	Not significant in predicted (+) direction	Not significant in opposite (-) direction	Significant in opposite (-) direction
<i>H1: A woman's tendency to participate actively in the conversation correlates positively with the number of females in the discussion</i>	64.3% (n=9)	0% (n=0)	14.2% (n=2)	21.4% (n=3)
<i>H2: Higher status participants are more verbose than lower status participants</i>	28.6% (n=4)	21.4% (n=3)	28.6% (n=4)	21.4% (n=3)



## Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis<sup>☆</sup>

Martin Schweinsberg <sup>a,1,\*</sup>, Michael Feldman <sup>b,1,\*</sup>, Nicola Staub <sup>b,1</sup>, Olmo R. van den Akker <sup>c</sup>, Robbie C.M. van Aert <sup>c</sup>, Marcel A.L.M. van Assen <sup>d</sup>, Yang Liu <sup>e</sup>, Tim Althoff <sup>e</sup>, Jeffrey Heer <sup>e</sup>, Alex Kale <sup>e</sup>, Zainab Mohamed <sup>f</sup>, Hashem Amireh <sup>g</sup>, Vaishali Venkatesh Prasad <sup>a</sup>, Abraham Bernstein <sup>b,\*</sup>, Emily Robinson, Kaisa Snellman <sup>h</sup>, S. Amy Sommer <sup>i</sup>, Sarah M.G. Otner <sup>j</sup>, David Robinson, Nikhil Madan <sup>k</sup>, Raphael Silberzahn <sup>l</sup>, Pavel Goldstein <sup>m</sup>, Warren Tierney <sup>n</sup>, Toshio Murase <sup>o</sup>, Benjamin Mandl <sup>p</sup>, Domenico Viganola <sup>p</sup>, Carolin Strobl <sup>b</sup>, Catherine B. C. Schaumans <sup>q</sup>, Stijn Kelchtermans <sup>r</sup>, Chan Naseeb <sup>s</sup>, S. Mason Garrison <sup>t</sup>, Tal Yarkoni <sup>u</sup>, C. S. Richard Chan <sup>v</sup>, Prestone Adie <sup>w</sup>, Paulius Alaburda, Casper Albers <sup>x</sup>, Sara Alspaugh <sup>y</sup>, Jeff Alstott <sup>z</sup>, Andrew A. Nelson <sup>aa</sup>, Eduardo Ariño de la Rubia <sup>ab</sup>, Adbi Arzi <sup>h</sup>, Štěpán Bahník <sup>ac</sup>, Jason Baik, Laura Winther Balling <sup>ad</sup>, Sachin Bunker <sup>ae</sup>, David AA Baranger <sup>af</sup>, Dale J. Barr <sup>ag</sup>, Brenda Barros-Rivera <sup>ah</sup>, Matt Bauer <sup>ai</sup>, Enuh Blaise <sup>aj</sup>, Lisa Boelen <sup>ak</sup>, Katerina Bohle Carbonell <sup>al</sup>, Robert A. Briers <sup>am</sup>, Oliver Burkhard, Miguel-Angel Canela <sup>an</sup>, Laura Castrillo, Timothy Catlett, Olivia Chen, Michael Clark <sup>ao</sup>, Brent Cohn, Alex Coppock <sup>ap</sup>, Natália Cugueró-Escofet <sup>aq</sup>, Paul G. Curran <sup>ar</sup>, Wilson Cyrus-Lai <sup>h</sup>, David Dai <sup>as</sup>, Giulio Valentino Dalla Riva <sup>at</sup>, Henrik Danielsson <sup>au</sup>, Rosaria de F.S.M. Russo <sup>av</sup>, Niko de Silva <sup>a</sup>, Curdin Derungs <sup>aw</sup>, Frank Dondelingier <sup>ax</sup>, Carolina Duarte de Souza <sup>ay</sup>, B. Tyson Dube, Marina Dubova <sup>az</sup>, Ben Mark Dunn <sup>ag</sup>, Peter Adriaan Edelsbrunner <sup>ba</sup>, Sara Finley <sup>bb</sup>, Nick Fox <sup>bc</sup>, Timo Gnambs <sup>bd</sup>, Yuanyuan Gong <sup>be</sup>, Erin Grand, Brandon Greenawalt <sup>bf</sup>, Dan Han, Paul H.P. Hanel <sup>bg</sup>, Antony B. Hong <sup>h</sup>, David Hood, Justin Hsueh, Lilian Huang <sup>bh</sup>, Kent N. Hui <sup>bi</sup>, Keith A. Hultman <sup>bj</sup>, Azka Javaid <sup>bk</sup>, Lily Ji Jiang <sup>bl</sup>, Jonathan Jong <sup>bm</sup>, Jash Kamdar, David Kane <sup>bn</sup>, Gregor Kappler <sup>bo</sup>, Erikson Kaszubowski <sup>ay</sup>, Christopher M. Kavanagh, Madian Khabsa, Bennett Kleinberg <sup>bp</sup>, Jens Kouros, Heather Krause <sup>bq</sup>, Angelos-Miltiadis Kryptotos <sup>br</sup>, Dejan Lavbić <sup>do</sup>, Rui Ling Lee <sup>bs</sup>, Timothy Leffel <sup>bh</sup>, Wei Yang Lim <sup>bt</sup>, Silvia Liverani <sup>bu</sup>, Bianca Loh <sup>h</sup>, Dorte Lønsmann <sup>bv</sup>, Jia Wei Low <sup>bw</sup>, Alton Lu <sup>e</sup>, Kyle MacDonald <sup>bx</sup>, Christopher R. Madan <sup>by</sup>, Lasse Hjorth Madsen <sup>bz</sup>, Christina Maimone <sup>al</sup>, Alexandra Mangold, Adrienne Marshall <sup>ca</sup>, Helena Ester Matskevich <sup>e</sup>, Kimia Mavon <sup>bn</sup>, Katherine L. McLain <sup>a</sup>, Amelia A. McNamara <sup>cc</sup>, Mhairi McNeill, Ulf Mertens <sup>cd</sup>, David Miller <sup>al</sup>, Ben Moore <sup>ce</sup>, Andrew Moore, Eric Nantz <sup>cf</sup>, Ziauddin Nasrullah <sup>a</sup>, Valentina Nejkovic <sup>cg</sup>, Colleen S Nell <sup>ch</sup>, Andrew Arthur Nelson <sup>aa</sup>, Gustav Nilsonne <sup>ci</sup>, Rory Nolan <sup>cj</sup>, Christopher E. O'Brien, Patrick O'Neill <sup>ck</sup>, Kieran O'Shea <sup>ag</sup>, Toto Olita <sup>cl</sup>, Jahna Otterbacher <sup>cm</sup>, Diana Palsetia <sup>al</sup>, Bianca Pereira, Ivan Pozdniakov <sup>cn</sup>, John Protzko <sup>co</sup>, Jean-Nicolas Reyt <sup>cp</sup>, Travis Riddle <sup>cq</sup>, Amal (Akmal) Ridhwan Omar Ali <sup>cr</sup>, Ivan Ropovik <sup>cs</sup>, Joshua M. Rosenberg <sup>ct</sup>, Stephane Rothen, Michael Schulte-Mecklenbeck <sup>cu</sup>, Nirek Sharma <sup>cv</sup>, Gordon Shotwell <sup>cw</sup>, Martin Skarzynski, William Stedden, Victoria Stodden <sup>cx</sup>, Martin A. Stoffel <sup>cy</sup>, Scott Stoltzman <sup>cz</sup>,

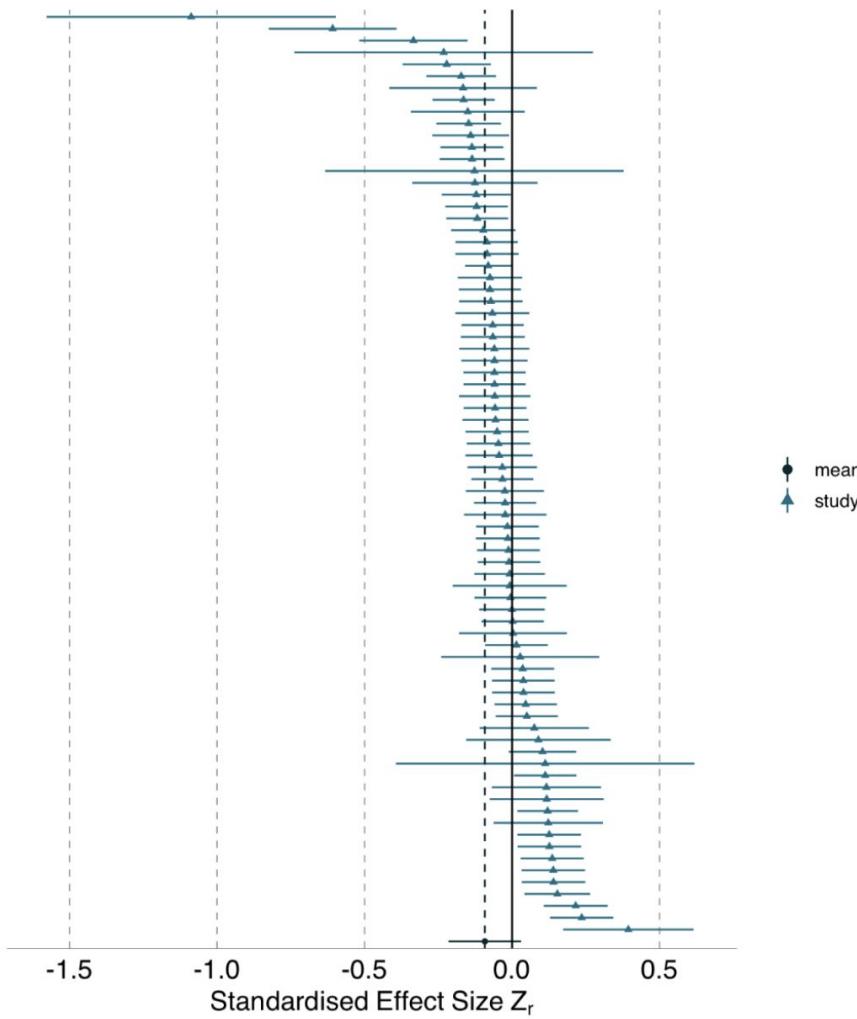


# EvoBio (bird nest competition) & ecology (plant competition): 174 teams

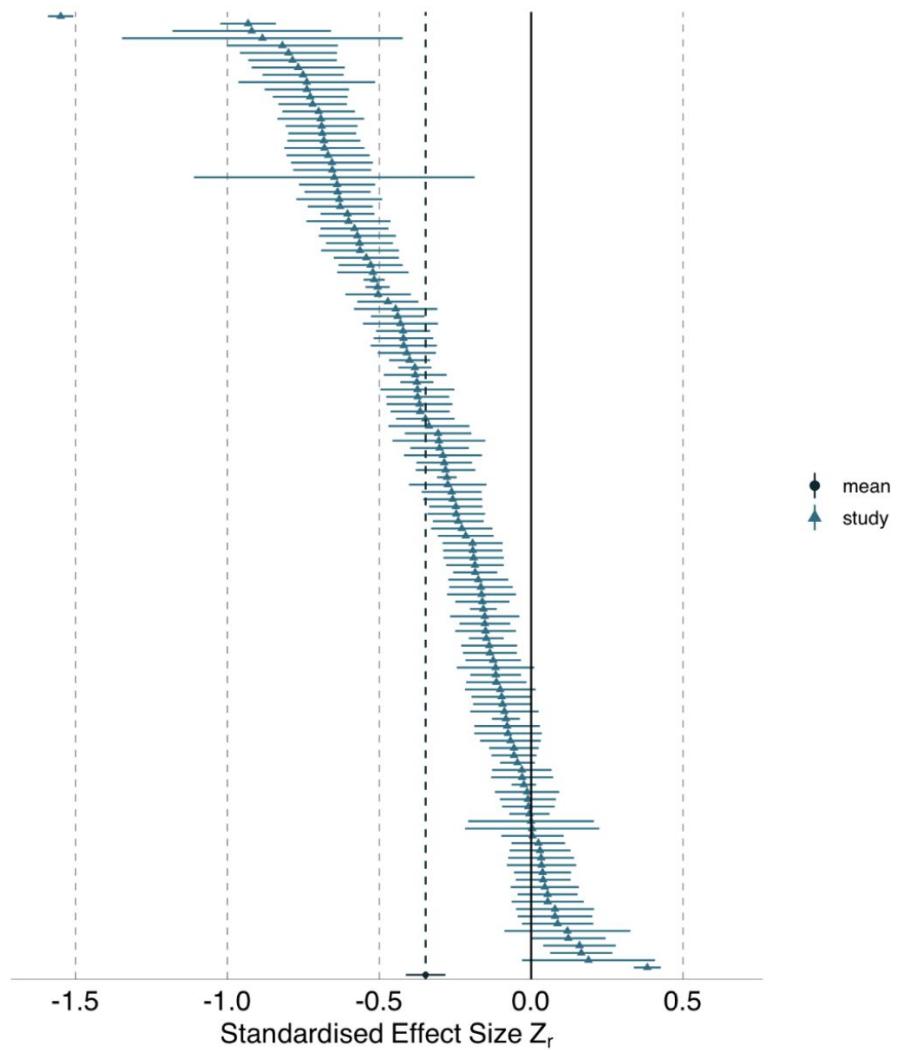
# Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology.



# No effect



# Real effect with small number of informative variables



 *Fin*