Some Economics of Clinical Trials

Marco Ottaviani
Bocconi University

May 20, 2024

Economic Analyses of Science
Oxford University

Clinical Trials

- In clinical research, human lives are at stake
  $\Rightarrow$ we should demand highest standards
- Pressing economic incentives (large R&D costs, even larger potential profits)
  $\Rightarrow$ conflicts of interest for investigators
- Regulation: drug approval process (FDA, EMA) and ClinicalTrials.gov registry
- Opportunity for economics research:
  - Increasing availability of data
  - Institutional design questions

Cournot (1843) on P-Hacking

Exposition de la théorie des chances et des probabilités

"§101 ... A person not knowing how the data were analysed and whom the experimenter told the result of that analysis concerning the system ... but not **how many attempts he made** to achieve that result, is unable to judge with a determined chance of error whether the chances ... are equal or not..."

"... However, **unsuccessful tests usually leave no traces**; **the public only knows the results** which the experimenter thought to be deserving notice. It follows that a person alien to the testing is **absolutely unable** to regulate bets on **whether the result is, or is not attributable to anomalies of chance**."

Roadmap

1. Some theory
    a. Value of Selected Information
    b. Regulation of Optional Stopping

2. Clinical Trials
    a. Incentives Across Clinical Phases
    b. Preregistration
    c. Outcome changes

Roadmap

1. Some theory
    a. Value of Selected Information
    b. Regulation of Optional Stopping
2. Clinical Trials
    a. Incentives Across Clinical Phases
    b. Preregistration
    c. Outcome changes

Value of Selected Information?

Strategic Sample Selection (EMA, 2021), with Alfredo Di Tillio and Peter N. Sørensen

- Biased researchers in **observational studies**:
  - select sample non randomly from larger presample
  - choose specification
  - omit controls

Value of Selected Information?

Strategic Sample Selection (EMA, 2021), with Alfredo Di Tillio and Peter N. Sørensen

- Biased researchers in **observational studies**:
  - select sample non randomly from larger presample
  - choose specification
  - omit controls
- *Subversion of randomization to treatment* in **experiments**
  - When treatment is given to healthiest rather than random patients:

Value of Selected Information?

Strategic Sample Selection (EMA, 2021), with Alfredo Di Tillio and Peter N. Sørensen

- Biased researchers in **observational studies**:
  - select sample non randomly from larger presample
  - choose specification
  - omit controls

- *Subversion of randomization to treatment* in **experiments**
  - When treatment is given to healthiest rather than random patients:
    - favorable outcomes become *more likely*
    - but are also *less convincing* that treatment is effective

Value of Selected Information?

Strategic Sample Selection (EMA, 2021), with Alfredo Di Tillio and Peter N. Sørensen

- Biased researchers in **observational studies**:
  - select sample non randomly from larger presample
  - choose specification
  - omit controls
- *Subversion of randomization to treatment* in **experiments**
  - When treatment is given to healthiest rather than random patients:
    - favorable outcomes become *more likely*
    - but are also *less convincing* that treatment is effective
- **How does anticipated sample selection affect the value of information?**

Impact of ANTICIPATED Selection

- Compare information value of two experiments:
  - **Random** : $X = \theta + \varepsilon$ with $\varepsilon \sim F$ [BLUE]
  - **Selected** : max of k iid draws: $Y = \theta + \varepsilon_{(k)}$ with $\varepsilon_{(k)} \sim F^k$ [RED]

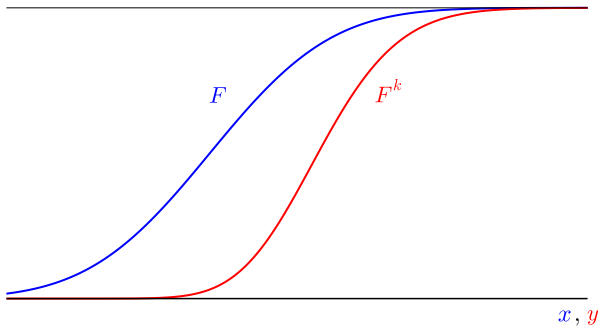


Illustration: Simple Hypothesis Testing

- Here we illustrate idea for **simple hypothesis testing:**

	θ_L	θ_H
reject	R	R
accept	θ_L	θ_H

$$\theta_L < R < \theta_H, \quad \text{prior } p = \Pr(\theta_H)$$

Illustration: Simple Hypothesis Testing

- Here we illustrate idea for **simple hypothesis testing:**

	θ_L	θ_H
reject	R	R
accept	θ_L	θ_H

$$\theta_L < R < \theta_H, \quad \text{prior } p = \Pr(\theta_H)$$

  - Generalizing Lehmann (1988), our results are valid for *general* Quah-Strulovici (2009) IDO preferences (including single crossing and Karlin-Rubin monotone decision problems)

Illustration: Simple Hypothesis Testing

- Here we illustrate idea for **simple hypothesis testing:**

	θ_L	θ_H
reject	R	R
accept	θ_L	θ_H

$$\theta_L < R < \theta_H, \quad \text{prior } p = \Pr(\theta_H)$$

  - Generalizing Lehmann (1988), our results are valid for *general* Quah-Strulovici (2009) IDO preferences (including single crossing and Karlin-Rubin monotone decision problems)

- Location experiment $x = \theta + \varepsilon$, with $\varepsilon \sim F$ independent from θ
  - Assume **logconcave** density $f \Leftrightarrow$ monotone likelihood ratio property

Illustration: Simple Hypothesis Testing

- Here we illustrate idea for **simple hypothesis testing:**

	θ_L	θ_H
reject	R	R
accept	θ_L	θ_H

$$\theta_L < R < \theta_H, \quad \text{prior } p = \Pr(\theta_H)$$

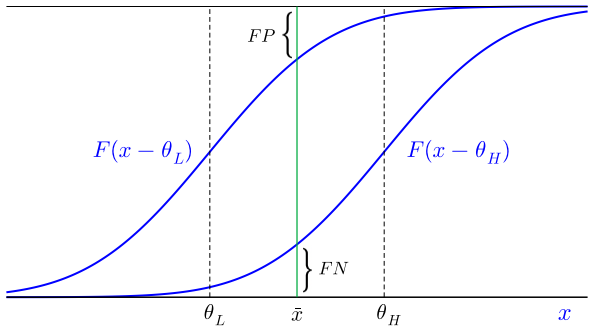
  - Generalizing Lehmann (1988), our results are valid for *general* Quah-Strulovici (2009) IDO preferences (including single crossing and Karlin-Rubin monotone decision problems)

- Location experiment $x = \theta + \varepsilon$, with $\varepsilon \sim F$ independent from θ
  - Assume **logconcave** density $f \Leftrightarrow$ monotone likelihood ratio property

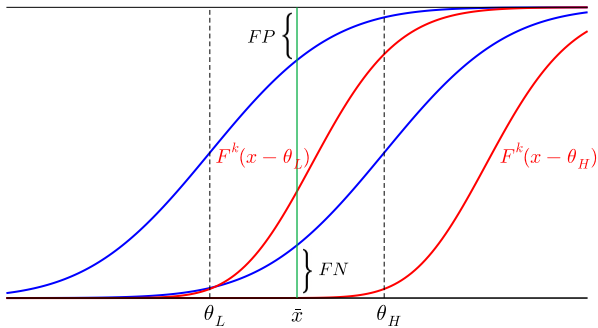
- With a single draw, cutoff rule optimal: accept iff

$$\frac{f(x-\theta_H)}{f(x-\theta_L)} \geq \frac{1-p}{p}\frac{R-\theta_L}{\theta_H - R} \Leftrightarrow x \geq \bar{x}$$

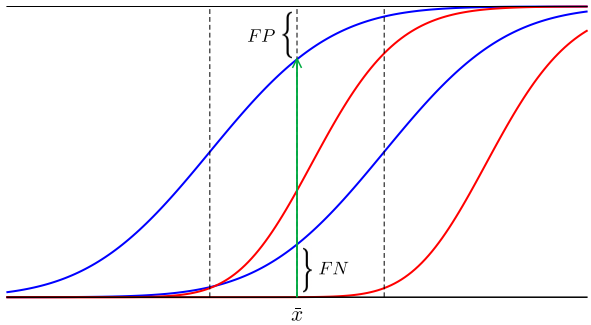
Random Experiment



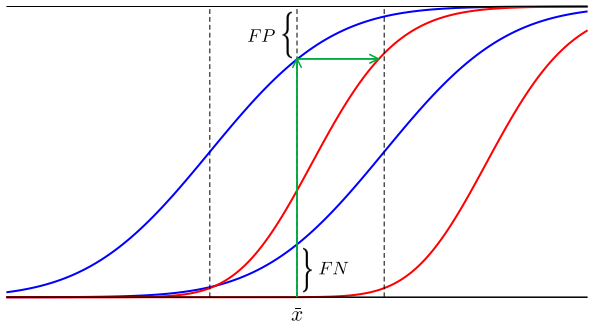
Selected Experiment



Random v. Selected



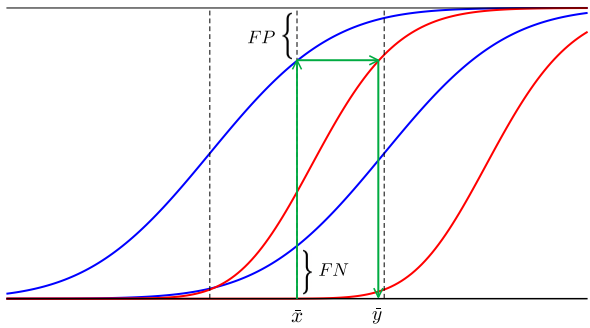
Random v. Selected



Random v. Selected

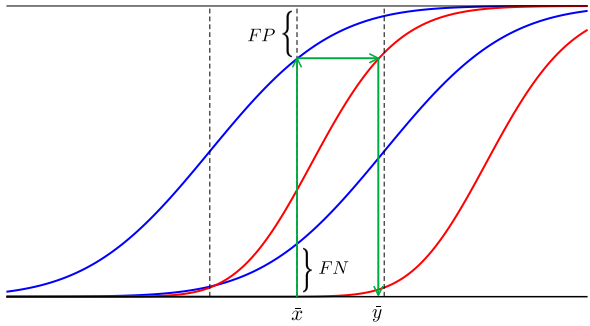
$$\underbrace{1 - F^k(\bar{y} - \theta_L) = 1 - F(\bar{x} - \theta_L)} \quad \Rightarrow \bar{y} = (F^k)^{-1} F(\bar{x} - \theta_L) + \theta_L$$

Using cutoff $\bar{y}$ in Y that matches False Positives



Random v. Selected

$$\underbrace{1 - F^k(\bar{y} - \theta_L) = 1 - F(\bar{x} - \theta_L)}_{\text{Using cutoff } \bar{y} \text{ in } Y \text{ that matches False Positives}} \quad \Rightarrow \bar{y} = (F^k)^{-1} F(\bar{x} - \theta_L) + \theta_L$$

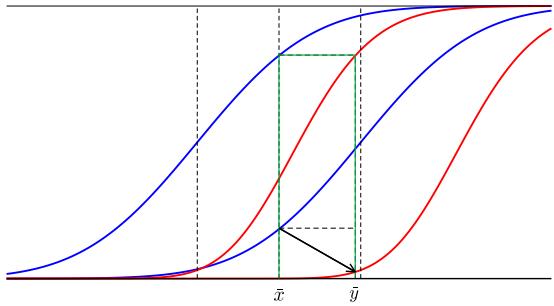


$$\underbrace{F^k(\bar{y} - \theta_H) \overset{?}{\leq} F(\bar{x} - \theta_H)}_{\text{Are False Negatives reduced? Yes, with } F \text{ normal! More generally?}}$$

Random v. Selected

$$\underbrace{1 - F^k(\bar{y} - \theta_L) = 1 - F(\bar{x} - \theta_L)} \quad \Rightarrow \bar{y} = (F^k)^{-1} F(\bar{x} - \theta_L) + \theta_L$$

Using cutoff $\bar{y}$ in Y that matches False Positives

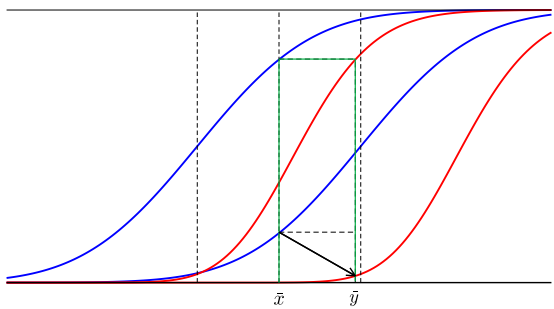


$$\underbrace{F^k(\bar{y} - \theta_H) \overset{?}{\leq} F(\bar{x} - \theta_H)}$$

Are False Negatives reduced? Yes, with F normal! More generally?

Random v. Selected

$$\underbrace{1 - F^k(\bar{y} - \theta_L) = 1 - F(\bar{x} - \theta_L)}_{\text{Using cutoff } \bar{y} \text{ in } Y \text{ that matches False Positives}} \quad \Rightarrow \bar{y} = (F^k)^{-1} F(\bar{x} - \theta_L) + \theta_L$$



$$\underbrace{F^k(\bar{y} - \theta_H) \overset{?}{\leq} F(\bar{x} - \theta_H)}_{\text{Are False Negatives reduced? Yes, with } F \text{ normal! More generally?}}$$

$\Leftrightarrow F^k$ is **less dispersed** than F ie $\left(F^k\right)^{-1}(q) - F^{-1}(q) \searrow q$

Dispersion of Selected Experiment

When is F^k steeper than F at same quantile?

Dispersion of Selected Experiment

When is F^k steeper than F at same quantile?

- Theorem (general sample size n): *Fixing sample size n,* ***as pre-sample size*** k ***increases,*** *experiment becomes* ***more*** *(less)* ***informative*** *in every monotone problem* ***if reverse hazard rate RHR*** $f(x|\theta)/F(x|\theta)$ ***is log-supermodular*** *(log-submodular, w/ support unbounded above)*

$$\frac{f(x|\theta')/F(x|\theta')}{f(x|\theta)/F(x|\theta)}$$ *increasing (decreasing) in x, for all $\theta' > \theta$*

Dispersion of Selected Experiment

When is F^k steeper than F at same quantile?

- Theorem (general sample size n): *Fixing sample size n,* ***as pre-sample size*** k ***increases,*** *experiment becomes* ***more*** *(less)* ***informative*** *in every monotone problem* ***if reverse hazard rate RHR*** $f(x|\theta)/F(x|\theta)$ ***is log-supermodular*** *(log-submodular, w/ support unbounded above)*

$$\frac{f(x|\theta')/F(x|\theta')}{f(x|\theta)/F(x|\theta)} \quad \text{increasing (decreasing) in } x, \text{ for all } \theta' > \theta$$

- For location $F_\theta(x) = F(x-\theta)$ experiment:

Dispersion of Selected Experiment

When is F^k steeper than F at same quantile?

- Theorem (general sample size n): *Fixing sample size n,* ***as pre-sample size*** k ***increases,*** *experiment becomes* ***more*** *(less)* ***informative*** *in every monotone problem* ***if reverse hazard rate RHR*** $f(x|\theta)/F(x|\theta)$ ***is log-supermodular*** *(log-submodular, w/ support unbounded above)*

$$\frac{f(x|\theta')/F(x|\theta')}{f(x|\theta)/F(x|\theta)} \quad \text{increasing (decreasing) in } x, \text{ for all } \theta' > \theta$$

- For location $F_\theta(x) = F(x-\theta)$ experiment:
  - **Beneficial** selection: logconcave RHR $\frac{f(x)}{F(x)}$ (e.g., Normal)

Dispersion of Selected Experiment

When is F^k steeper than F at same quantile?

- Theorem (general sample size n): *Fixing sample size n,* ***as pre-sample size*** k ***increases,*** *experiment becomes* ***more*** *(less)* ***informative*** *in every monotone problem* ***if reverse hazard rate RHR*** $f(x|\theta)/F(x|\theta)$ ***is log-supermodular*** *(log-submodular, w/ support unbounded above)*

$$\frac{f(x|\theta')/F(x|\theta')}{f(x|\theta)/F(x|\theta)} \quad \textit{increasing (decreasing) in } x\textit{, for all } \theta' > \theta$$

- For location $F_\theta(x) = F(x-\theta)$ experiment:
  - **Beneficial** selection: logconcave RHR $\frac{f(x)}{F(x)}$ (e.g., Normal)
  - **Neutral** selection: loglinear RHR $\frac{f(x)}{F(x)}$ (Gumbel)

Dispersion of Selected Experiment

When is F^k steeper than F at same quantile?

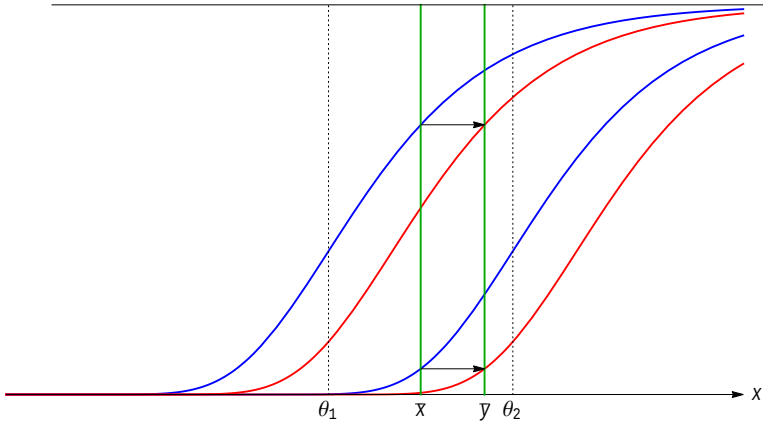
- Theorem (general sample size n): *Fixing sample size n,* ***as pre-sample size*** k ***increases,*** *experiment becomes* ***more*** *(less)* ***informative*** *in every monotone problem* ***if reverse hazard rate RHR*** $f(x|\theta)/F(x|\theta)$ ***is log-supermodular*** *(log-submodular, w/ support unbounded above)*

$$\frac{f(x|\theta')/F(x|\theta')}{f(x|\theta)/F(x|\theta)} \quad \textit{increasing (decreasing) in } x\textit{, for all } \theta' > \theta$$

- For location $F_\theta(x) = F(x-\theta)$ experiment:
  - **Beneficial** selection: logconcave RHR $\frac{f(x)}{F(x)}$ (e.g., Normal)
  - **Neutral** selection: loglinear RHR $\frac{f(x)}{F(x)}$ (Gumbel)
  - **Harmful** selection: logconvex RHR $\frac{f(x)}{F(x)}$ (e.g., Exponential)

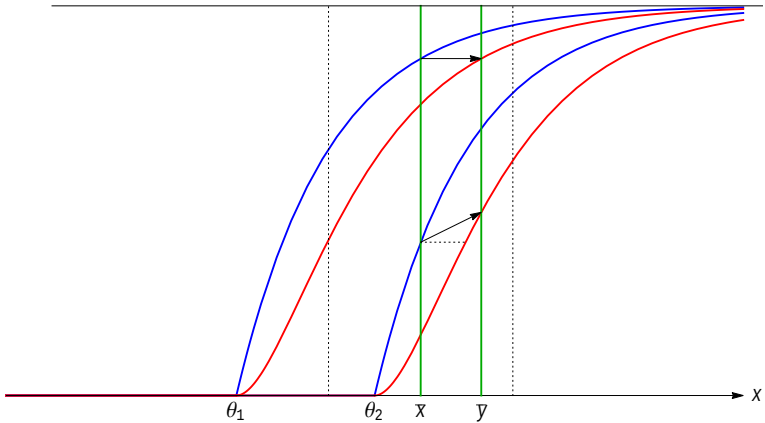
NEUTRAL Selection: Loglinear f/F

Gumbel noise: $F(\varepsilon) = e^{-e^{-\varepsilon}}$



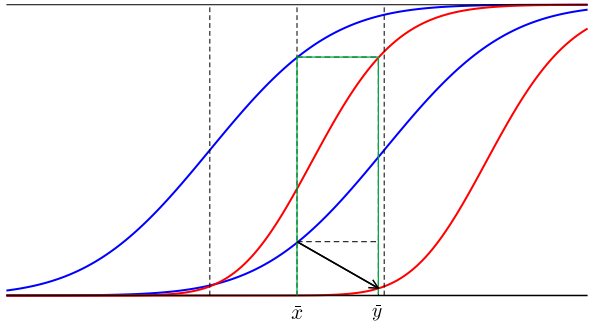
HARMFUL Selection: f LESS Logconcave than F

Exponential noise: $F(\varepsilon) = 1 - e^{-\varepsilon}$



BENEFICIAL Selection: f MORE Logconcave than F

Normal noise: $\varepsilon \sim \mathcal{N}$



Roadmap

1. Some theory
    a. Value of Selected Information
    b. Regulation of Optional Stopping
2. Clinical Trials
    a. Incentives Across Clinical Phases
    b. Preregistration
    c. Outcome changes

Regulation of Optional Stopping

Research and the Approval Process: The Organization of Persuasion (AER, 2019), with Emeric Henry

- Sender (pharma co) benefits from **approval** of drug with uncertain efficacy
  - Evaluator = FDA regulator

$$\theta^H > 0 \text{ in state } H \text{ and } \theta^L < 0 \text{ in state } L$$

Regulation of Optional Stopping

Research and the Approval Process: The Organization of Persuasion (AER, 2019), with Emeric Henry

- Sender (pharma co) benefits from **approval** of drug with uncertain efficacy
  - Evaluator = FDA regulator

$$\theta^H > 0 \text{ in state } H \text{ and } \theta^L < 0 \text{ in state } L$$

  - Researcher = pharma company benefits from approval

$$v > 0$$

Regulation of Optional Stopping

Research and the Approval Process: The Organization of Persuasion (AER, 2019), with Emeric Henry

- Sender (pharma co) benefits from **approval** of drug with uncertain efficacy
  - Evaluator = FDA regulator

$$\theta^H > 0 \text{ in state } H \text{ and } \theta^L < 0 \text{ in state } L$$

  - Researcher = pharma company benefits from approval

$$v > 0$$

- Researcher **sequentially** acquires & diffuses **costly information**
  - instantaneous trial result from state-dependent Brownian motion

Regulation of Optional Stopping

Research and the Approval Process: The Organization of Persuasion (AER, 2019), with Emeric Henry

- Sender (pharma co) benefits from **approval** of drug with uncertain efficacy
  - Evaluator = FDA regulator

$$\theta^H > 0 \text{ in state } H \text{ and } \theta^L < 0 \text{ in state } L$$

  - Researcher = pharma company benefits from approval

$$v > 0$$

- Researcher **sequentially** acquires & diffuses **costly information**
  - instantaneous trial result from state-dependent Brownian motion
- Evaluator has **coarse instruments** for regulation
  - approve/reject, ask for additional evidence [impose liability]

Organizational Deconstruction of Wald

TWO players:

1. **Researcher**

a. directly **controls information** acquisition & pays info cost

b. but **always wants approval** and does not directly value info

Organizational Deconstruction of Wald

TWO players:

1. **Researcher**

a. directly **controls information** acquisition & pays info cost

b. but **always wants approval** and does not directly value info

2. **Evaluator**

a. directly controls **approval decision** with uncertain payoff

b. **benefits from info**, but can only obtain it indirectly from researcher

Organizational Deconstruction of Wald

TWO players:

1. **Researcher**

a. directly **controls information** acquisition & pays info cost

b. but **always wants approval** and does not directly value info

2. **Evaluator**

a. directly controls **approval decision** with uncertain payoff

b. **benefits from info**, but can only obtain it indirectly from researcher

Wald's **social planner**

a. controls **all decisions** (rejection/approval) & info acquisition

b. obtains **total payoff** θ+ v (evaluator+researcher) & pays info cost

Equilibrium Persuasion

- Researcher acquires costly information to persuade evaluator

Equilibrium Persuasion

- Researcher acquires costly information to persuade evaluator
- Researcher stops acquiring costly information when cumulative evidence is
  - either favorable enough to persuade evaluator to approve
  - or so discouraging to make it economical to abandon research

Equilibrium Persuasion

- Researcher acquires costly information to persuade evaluator
- Researcher stops acquiring costly information when cumulative evidence is
  - either favorable enough to persuade evaluator to approve
  - or so discouraging to make it economical to abandon research
- Evaluator does not internalize research cost
  - optimal to commit to softening approval standard to encourage research

Roadmap

1. Some theory
   a. Value of Selected Information
   b. Regulation of Optional Stopping

2. Clinical Trials
   a. Incentives Across Clinical Phases
   b. Preregistration
   c. Outcome changes

P-Hacking and Incentives Across Clinical Phases

P-Hacking in Clinical Trials and How Incentives Shape the Distribution of Results Across Phases (PNAS, 2020), with Jérôme Adda and Christian Decker

- Does pressure to withhold or “beautify” unfavorable results lead to p-hacking?

P-Hacking and Incentives Across Clinical Phases

P-Hacking in Clinical Trials and How Incentives Shape the Distribution of Results Across Phases (PNAS, 2020), with Jérôme Adda and Christian Decker

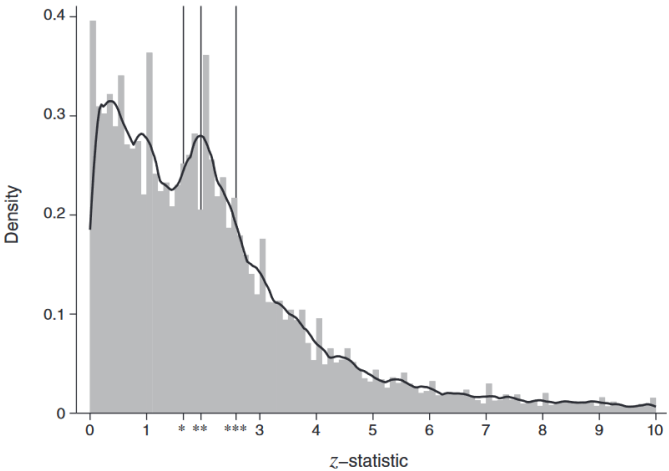
- Does pressure to withhold or “beautify” unfavorable results lead to p-hacking?
- Initial evaluation of distribution of p-values reported to *ClinicalTrials.gov*

P-Hacking and Incentives Across Clinical Phases

P-Hacking in Clinical Trials and How Incentives Shape the Distribution of Results Across Phases (PNAS, 2020), with Jérôme Adda and Christian Decker

- Does pressure to withhold or "beautify" unfavorable results lead to p-hacking?
- Initial evaluation of distribution of p-values reported to *ClinicalTrials.gov*
- Investigate "suspicious patterns" depending on incentives resulting from
  - affiliation of lead sponsor (non-industry, small industry, large industry)
  - phase of clinical research (high-stake phase III, lower-stake phase II)

Evidence for P-Hacking in Economics



Brodeur, Cook, and Heyes (2020)

- P-hacking: intentionally or unconsciously exploring various ways of analyzing data and selectively reporting the ones that yield best results
- "Spike"/excess mass right above significance threshold commonly interpreted as evidence for p-hacking and/or selective reporting
- Similar findings for results in academic publications in Political Science, Psychology, Life Sciences

The *ClinicalTrials.gov* Registry

- US online registry of clinical research studies in human volunteers, maintained by the *National Institute of Health* and the *Food and Drug Administration* FDA
- Established in 2000 with the objective to increase transparency in clinical research by collecting information and results of ALL trials (including unpublished ones)

The *ClinicalTrials.gov* Registry

- US online registry of clinical research studies in human volunteers, maintained by the *National Institute of Health* and the *Food and Drug Administration* FDA
- Established in 2000 with the objective to increase transparency in clinical research by collecting information and results of ALL trials (including unpublished ones)

- *FDA Amendments Act* (2007/2017) prescribes, if certain (often arguable) criteria are met,
  - to register the trial no later than 21 days after enrollment of first participant
  - to submit results of the trial no later than twelve months after completion
- In theory, fines for non-compliance, but never enforced for long time (now first cases under way)
  $\Rightarrow$ compliance still quite poor, though slightly improving over time

Related Literature – P-Hacking

- Evidence for P-Hacking in Academic Publications
  - Economics: Brodeur et al. (2016); Brodeur, Cook, and Heyes (2020)
  - Political Science: Gerber and Malhotra (2008); Gerber at al. (2010)
  - Psychology: Simonsohn, Nelson, and Simmons (2014); Hartgerink et al. (2016)
  - Life Sciences: Holman et al. (2015)

- Methodology to Detect P-Hacking: Elliott, Kudrin, and Wüthrich (2022)

p-Values and z-Scores

- Focus on p-values as comparable outcome measure for all kinds of different trials
- Only consider p-values that are reported exactly (not "$p < 0.01$" or "$p < 0.001$")

p-Values and z-Scores

- Focus on p-values as comparable outcome measure for all kinds of different trials
- Only consider p-values that are reported exactly (not "$p < 0.01$" or "$p < 0.001$")
- Apply one-to-one mapping to transform p-values to z-scores

$$z = -\Phi^{-1}\left(\frac{p}{2}\right)$$

$\Rightarrow$ allows to investigate both overall shape of distribution and region around significance threshold more easily

The Distribution of z-Scores on *ClinicalTrials.gov*

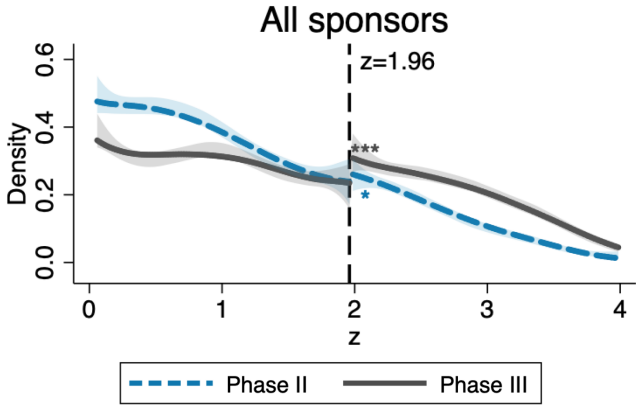
12,621 p-values from tests performed on primary outcomes of 4,977 trials

- Pre-approval interventional superiority studies on drugs (phase II and phase III)
- Conducted mainly between 2007 and 2019
- p-values transformed to z-statistics

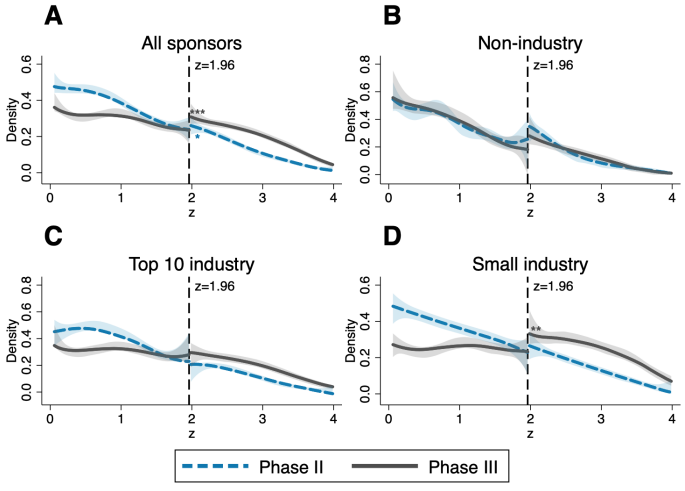
The Distribution of z-Scores on *ClinicalTrials.gov*

12,621 p-values from tests performed on primary outcomes of 4,977 trials

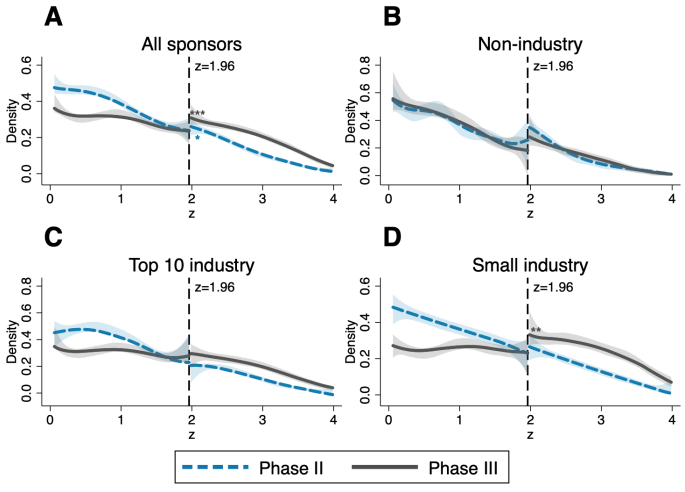
- Pre-approval interventional superiority studies on drugs (phase II and phase III)
- Conducted mainly between 2007 and 2019
- p-values transformed to z-statistics



Breakdown by Affiliation of Lead Sponsor



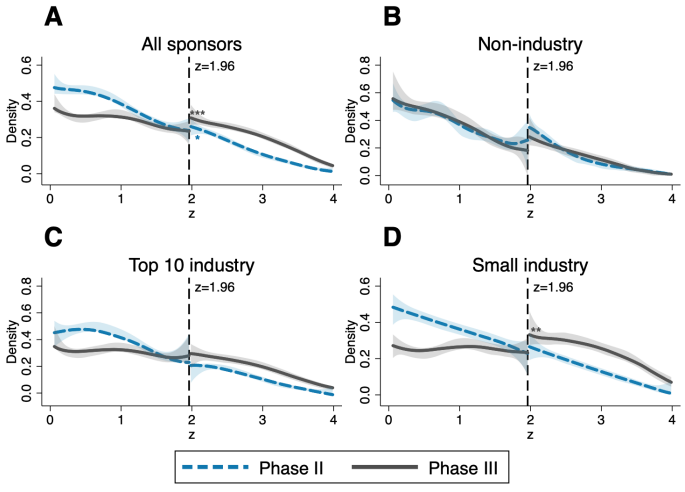
Breakdown by Affiliation of Lead Sponsor



Takeaways

1. No spike in density functions right above 1.96.
   $\Rightarrow$ good news!

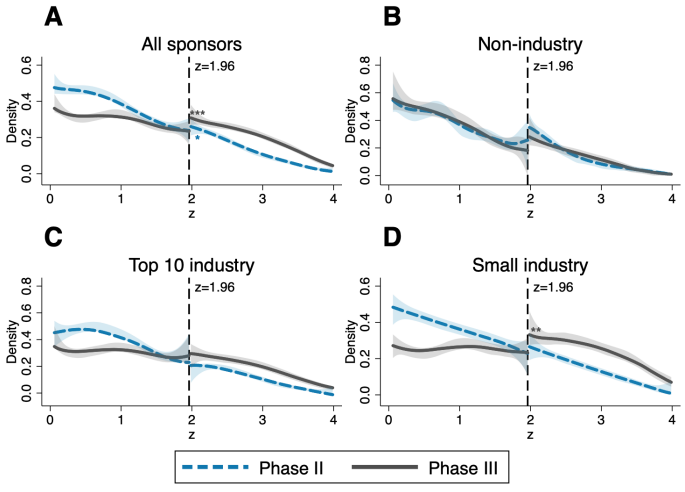
Breakdown by Affiliation of Lead Sponsor



Takeaways

1. No spike in density functions right above 1.96.
   $\Rightarrow$ good news!
2. Discontinuity in phase III density function at 1.96 (driven by small industry).
   $\Rightarrow$ suggestive of some selective reporting

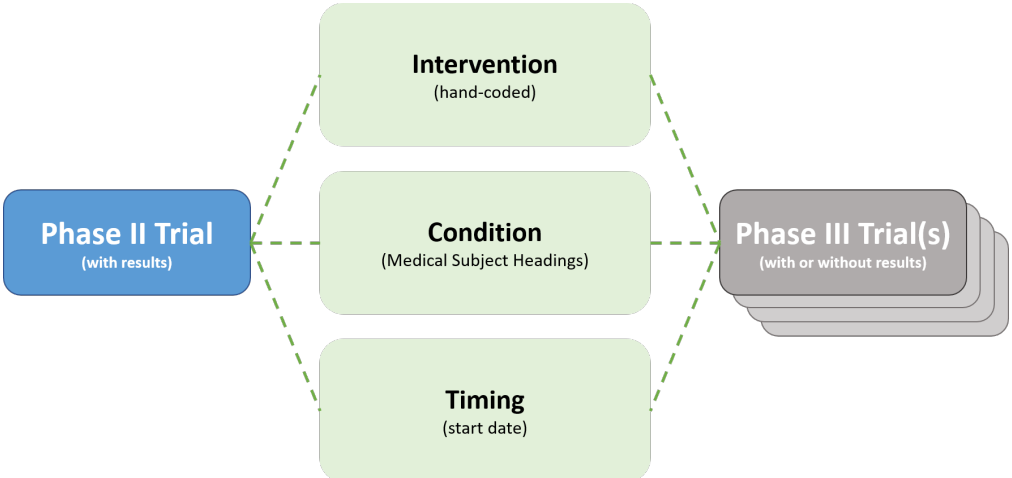
Breakdown by Affiliation of Lead Sponsor



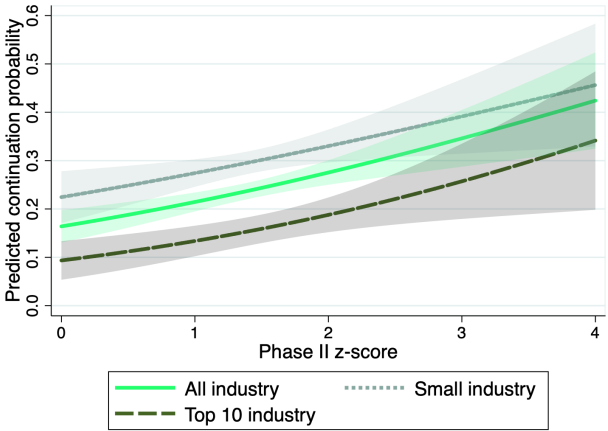
Takeaways

1. No spike in density functions right above 1.96.
   ⇒ good news!
2. Discontinuity in phase III density function at 1.96 (driven by small industry).
   ⇒ suggestive of some selective reporting
3. Excess mass of significant results in phase III compared to phase II for industry sponsored trials.
   ⇒ selective reporting or *selective continuation*?

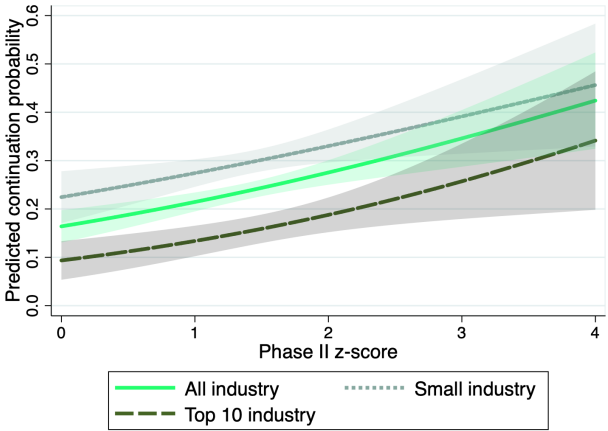
Linking Trials across Phases



Selective Continuation from Phase II to Phase III



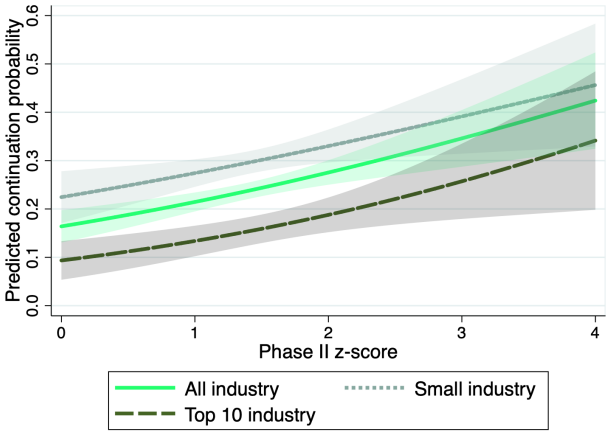
Selective Continuation from Phase II to Phase III



Takeaways

1. Higher phase II z-score significantly increases the probability of continuation to phase III.

Selective Continuation from Phase II to Phase III



Takeaways

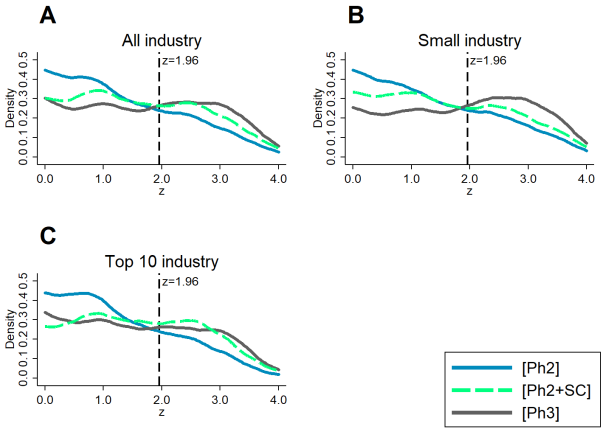
1. Higher phase II z-score significantly increases the probability of continuation to phase III.

2. Larger companies continue research projects more selectively.
   ⇒ higher opportunity costs?
   ⇒ more efficient managerial decisions?

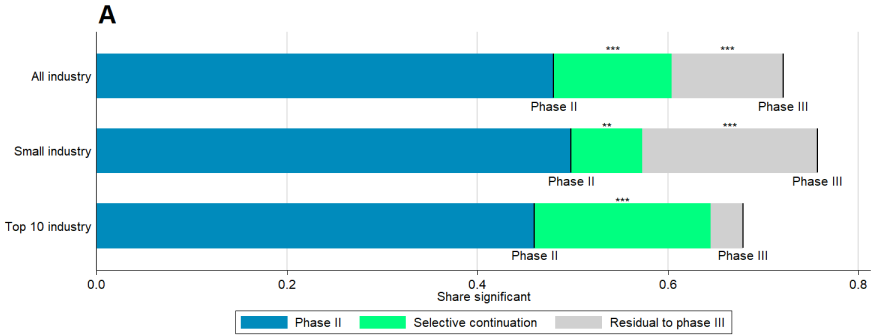
Controlling for *Selective Continuation*

- Estimate phase II density reweighting each observation by continuation probability predicted by selection function
  $\Rightarrow$ predicted phase III density

- Selection function increasing in phase II z-score
  $\Rightarrow$ counterfactual z-density rotates counter-clockwise, increasing share of significant results



Decomposition of the Difference in Significant Results between Phase II and Phase III



Takeaways

1. Large sponsors: *selective continuation* can explain excess share of significant results in phase III almost entirely
2. Small sponsors: *selective continuation* less pronounced, can only account for less than one third of excess share

Conclusion

- No indication of widespread manipulation of results reported to *ClinicalTrials.gov*

$\Rightarrow$ research registries and result databases seem to help

- Two different methodologies identify suspicious reporting patterns only for phase III trials by smaller industry sponsors (robust to definition of large vs. small)

$\Rightarrow$ economic incentives matter!

$\Rightarrow$ discipline of reputational concerns stronger for large companies?

$\Rightarrow$ disclosure regulations should focus particularly (but not exclusively) on smaller industry sponsors

Roadmap

1. Some theory
   a. Value of Selected Information
   b. Regulation of Optional Stopping
2. Clinical Trials
   a. Incentives Across Clinical Phases
   b. Preregistration
   c. Outcome changes

“Any attempt to analyze one’s data without first deciding exactly how one is going to conduct the key analysis will almost inevitably end in p-hacking. And so, for researchers who collect new data, the solution to this problem is straight-forward: Researchers must decide exactly how they will conduct their key analysis before they collect their data. And then they must commit to it. This is called pre-registration.”

Simmons, Nelson, and Simonsohn: Pre-registration: Why and How
Journal of Consumer Psychology, 2021

“Any attempt to analyze one’s data without first deciding exactly how one is going to conduct the key analysis will almost inevitably end in p-hacking. And so, for researchers who collect new data, the solution to this problem is straight-forward: Researchers must decide exactly how they will conduct their key analysis before they collect their data. And then they must commit to it. This is called pre-registration.”

Simmons, Nelson, and Simonsohn: Pre-registration: Why and How
Journal of Consumer Psychology, 2021

“A major shortcoming of preregistration as a normative standard is that the increased transparency it provides may be more illusory than real. Under present systems of preregistration, there is still substantial room for selective reporting and researchers’ degrees of freedom.”

Pham and Oh: Preregistration Is Neither Sufficient nor Necessary for Good Science
Journal of Consumer Psychology, 2021

Preregistration

Preregistration and Credibility of Clinical Trials (medRxiv, 2023)

- Preregistration commonly seen as solution to p-hacking
  - research registries like *ClinicalTrials.gov* and the AEA RCT Registry have been established (Abrams, Libgober, and List, 2020)
  - idea formalized in economic theory: preregistration as means for commitment/signaling in a persuasion game between researcher and evaluator (Williams, 2021; Felgenhauer 2021)

Preregistration

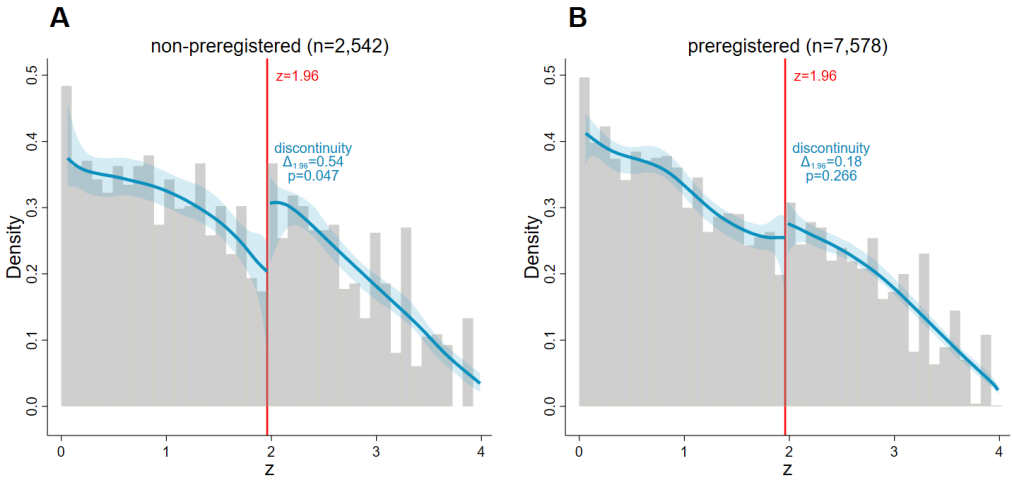
Preregistration and Credibility of Clinical Trials (medRxiv, 2023)

- Preregistration commonly seen as solution to p-hacking
  - research registries like *ClinicalTrials.gov* and the AEA RCT Registry have been established (Abrams, Libgober, and List, 2020)
  - idea formalized in economic theory: preregistration as means for commitment/signaling in a persuasion game between researcher and evaluator (Williams, 2021; Felgenhauer 2021)

- "Preregistration prevents p-hacking" sounds plausible, but there are some critical voices and so far empirical evidence for this claim is limited

$\Rightarrow$ **Our Research Question:**
Is preregistration a reliable signal for credibility of clinical trials?

Preview: It seems yes!



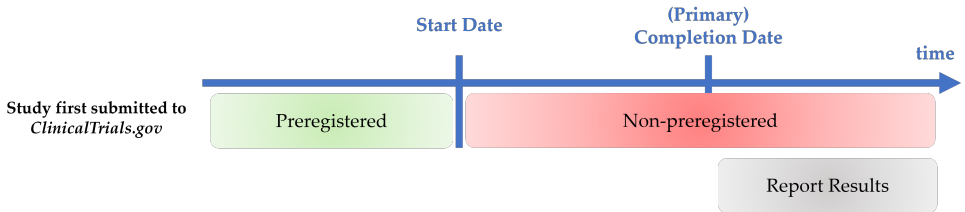
Sample

- *ClinicalTrials.gov* database as of Feb 18, 2023
- Focus on preapproval (Phase II & III), interventional (¬observational), superiority (¬non-inferiority) studies on drugs (¬devices or others)

Sample

- *ClinicalTrials.gov* database as of Feb 18, 2023
- Focus on preapproval (Phase II & III), interventional (¬observational), superiority (¬non-inferiority) studies on drugs (¬devices or others)

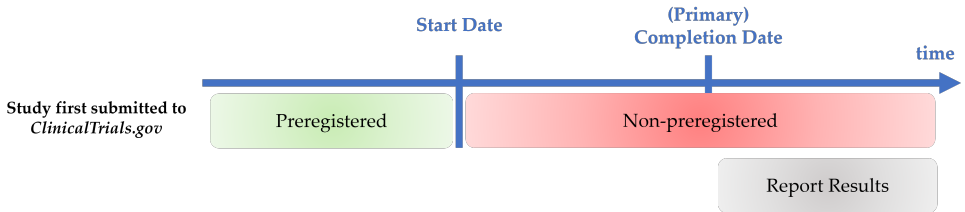
- Define trial as preregistered if first submitted to registry before start date Trend



Sample

- *ClinicalTrials.gov* database as of Feb 18, 2023
- Focus on preapproval (Phase II & III), interventional (¬observational), superiority (¬non-inferiority) studies on drugs (¬devices or others)

- Define trial as preregistered if first submitted to registry before start date ▸ Trend



⇒ out of 50,730 trials, 22,065 (43%) report any results, and 4,810 (9.5%) provide statistical analysis with at least one exact p-value ▸ Summary Stats ▸ Balance

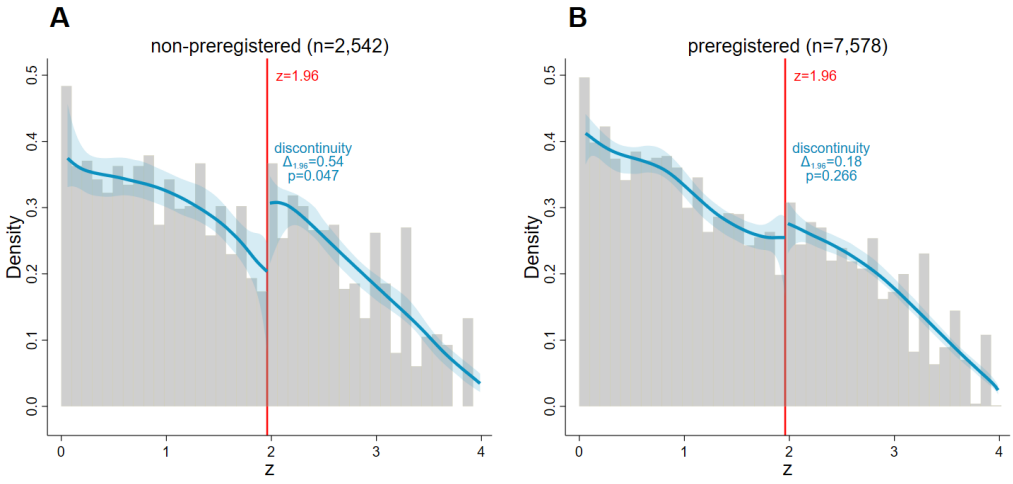
Methodology I: Density Discontinuity Tests

- In the absence of p-hacking, density of p-values/z-statistics is continuous (Andrews and Kasy, 2019; Elliott, Kudrin, and Wüthrich, 2022)

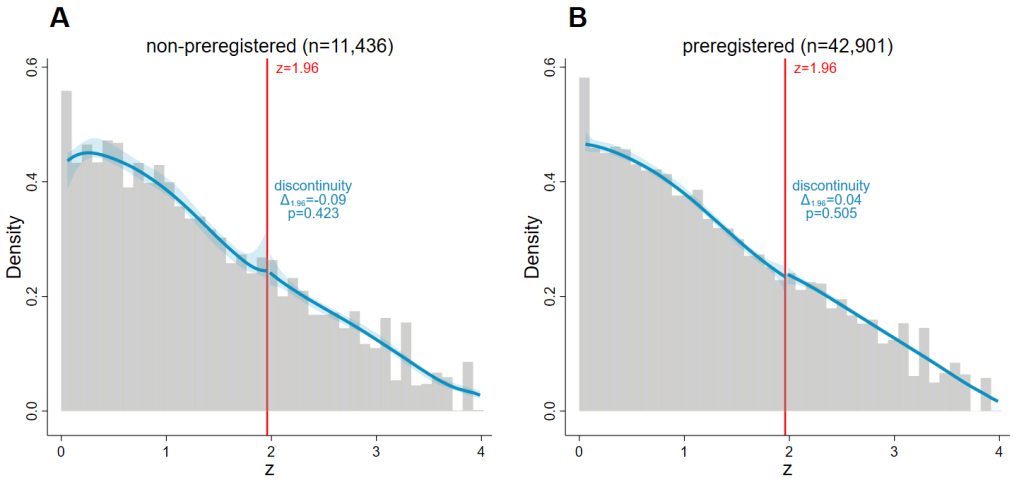
$$\Rightarrow \text{test } H_0 : \lim_{z \nearrow c} f(z) = \lim_{z \searrow c} f(z) \quad \text{vs.} \quad H_1 : \lim_{z \nearrow c} f(z) \neq \lim_{z \searrow c} f(z)$$

- "Improved" version of McCrary-Test based on local polynominal density estimators with bias correction at the boundary (Cattaneo, Jansson, and Ma, 2018, 2020)
- Advantages over other methods to detect p-hacking (caliper test, p-curve,. . .)
  - no pre-binning
  - fully data-driven bandwidth selection
  - (potentially) exploits entire distribution

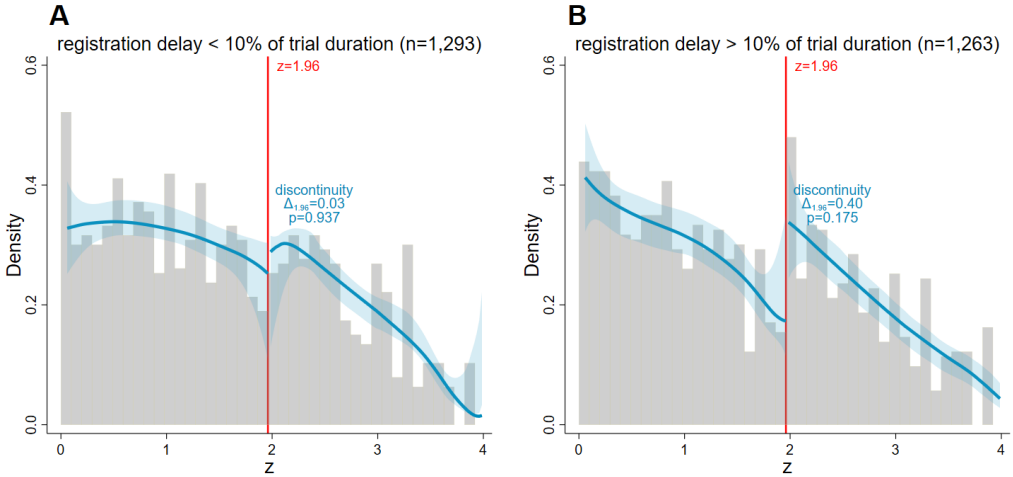
Main Result: Density Discontinuity Tests for Primary Outcomes



Placebo: Non-Primary Outcomes



Larger Registration Delay ⇒ Larger Discontinuity (Primary Outcomes)



Selection on Observables/Unobservables?

	(1) non-preregistered	(2) preregistered	(3) difference (2)–(1)
data monitoring committee	0.464	0.575	0.111***
	(0.499)	(0.494)	(0.018)
subject masked	0.711	0.768	0.057***
	(0.453)	(0.422)	(0.014)
caregiver masked	0.420	0.444	0.024
	(0.494)	(0.497)	(0.017)
investigator masked	0.697	0.751	0.053***
	(0.460)	(0.433)	(0.015)
outcomes assessor masked	0.398	0.431	0.033**
	(0.490)	(0.495)	(0.016)
mask folds	2.226	2.394	0.168***
	(1.576)	(1.505)	(0.051)
PI no employee of sponsor	0.878	0.914	0.036***
	(0.327)	(0.281)	(0.010)
Observations	1,206	3,604	4,810

*** p<0.01, ** p<0.05, * p<0.1

- Preregistered trials tend to have other design characteristics that are considered superior in terms of integrity.
- Selection on unobservable researcher characteristics?

$\Rightarrow$ Is it really lack of preregistration that opens the door for p-hacking?

Methodology II: Caliper Test

- Compare number of z-statistics in a narrow window above and below the significance threshold (Gerber and Malhotra, 2008)

- For all z-scores in the window $[c-h, c+h]$ estimate regression model (Brodeur, Cook, and Heyes, 2020)

$$significant_{ij} = \alpha + \beta preregistered_j + \boldsymbol{x}'_j\boldsymbol{\gamma} + \varepsilon_{ij},$$

$\Rightarrow$ advantage: regression framework allows to control for other trial characteristics $\boldsymbol{x}$ and/or fixed effects

- Caution: need to choose window size h and interpretation less "clean" than density discontinuity tests

Caliper Tests for $z \in 1.96 \pm 0.2$ (Primary Outcomes, LPM)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
preregistered	-0.0862**	-0.0740*	-0.0965**	-0.0927**	-0.0991*	-0.116*	-0.121*
	(0.0350)	(0.0396)	(0.0377)	(0.0426)	(0.0508)	(0.0614)	(0.0704)
data monitoring committee			0.0253	0.000886			-0.0144
			(0.0332)	(0.0362)			(0.0567)
subject masked			0.0104	-0.0200			-0.0757
			(0.0868)	(0.0873)			(0.151)
caregiver masked			0.00782	0.00686			0.00861
			(0.0443)	(0.0438)			(0.0658)
investigator masked			-0.0787	-0.0425			0.0724
			(0.0835)	(0.0774)			(0.135)
outcomes assessor masked			0.0416	0.0598			0.0643
			(0.0427)	(0.0424)			(0.0645)
PI not employee of sponsor			-0.0169	0.0677			0.149
			(0.0599)	(0.0741)			(0.153)
Non-Prereg. Sig. Rate	0.633	0.633	0.638	0.638	0.610	0.610	0.602
Observations	1,033	1,033	912	912	844	844	727
No. of trials	850	850	747	747	661	661	562
R-squared	0.006	0.065	0.011	0.082	0.178	0.222	0.251
Controls		✔		✔		✔	✔
Start Year FE		✔		✔		✔	✔
MeSH Condition FE		✔		✔		✔	✔
Sponsor FE					✔	✔	✔
Window	1.96±0.2	1.96±0.2	1.96±0.2	1.96±0.2	1.96±0.2	1.96±0.2	1.96±0.2

SEs clustered at trial level; *** p<0.01, ** p<0.05, * p<0.1. Controls: Phase Dummies, Sponsor Group Dummies, Placebo Control, sqrt enrollment.

▸ Other Bandwidths

Conclusion and Future Directions

- We find a discontinuity in the density at the 5% significance threshold for non-preregistered trials, but no discontinuity for preregistered trials
- Differences robust in caliper tests controlling for other trial characteristics and sponsor FE
- No evidence of p-hacking for lower-stake secondary outcomes independent of preregistration status

- Potential future directions
  - explicit selection and signaling model of preregistration
  - weighing advantages and disadvantages of preregistration (optimal policy?)
  - mechanisms: track outcome changes

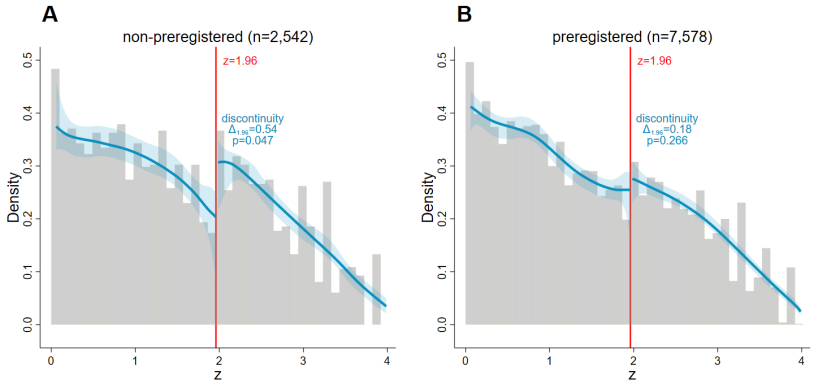
Roadmap

1. Some theory
    a. Value of Selected Information
    b. Regulation of Optional Stopping
2. Clinical Trials
    a. Incentives Across Clinical Phases
    b. Preregistration
    c. Outcome changes

Outcome Changes and Distribution of Results

Outcome Changes and Distribution of Results in Registered Clinical Trials (2024, in progress), with Christian Decker and Marta Maxia

Preregistration signals credibility in clinical trials $\Rightarrow$ **underlying mechanism**?
$\Rightarrow$ Do changes in outcome variables affect distributions of results?



Which Outcome Changes Matter?

Sample $\Rightarrow$ same as in Decker and Ottaviani (2023) ▸ sample

Classification task:

1. Compare initial *vs* final version of primary outcome measure
2. Set variable `major_change = True` if:
   - Originally pre-specified outcome **omitted**
   - Introduction of a **new outcome**
   - Switch from **primary to secondary**
   - Switch from **secondary to primary**
3. Exclude typo corrections, redundant measurement details

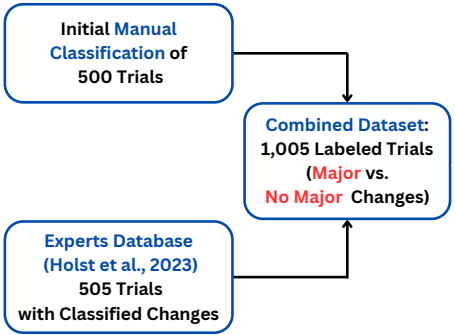
Web Scraping Outcome Descriptions

Primary (Original) (Submitted: 2008-09-11)	ICMJE	• Liebowitz Social Anxiety Scale [Time Frame: post, 6-months, and 12-months follow-up]

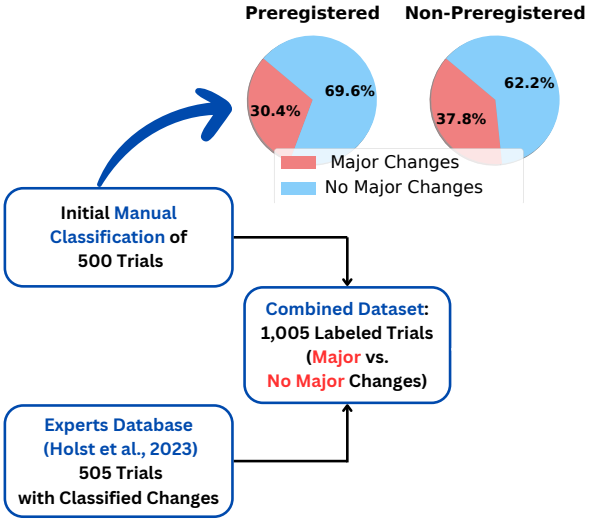
Primary (Current) (Submitted: 2011-01-06)	ICMJE	• Brief Social Phobia Scale [Time Frame: pre, within, post, 6-months, and 12-months follow-up] ◦ Brief Social Phobia Scale BSPS is an interview assessing fear and avoidance in seven social situations and severity of four bodily symptoms.

- Identify two levels of information detail:
  1. **Full** descriptions
  2. **Short** descriptions (**blue rectangles**)

Classification: Manual plus Machine Learning

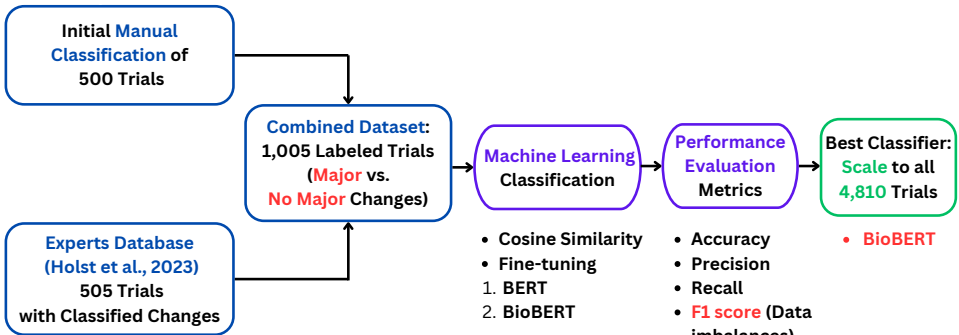


Classification: Manual plus Machine Learning

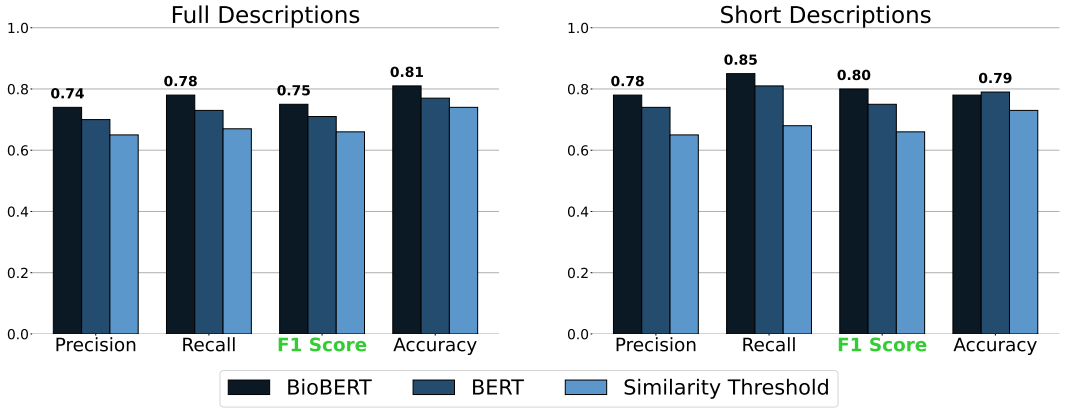


Classification: Manual plus Machine Learning

cosine fine-tuning

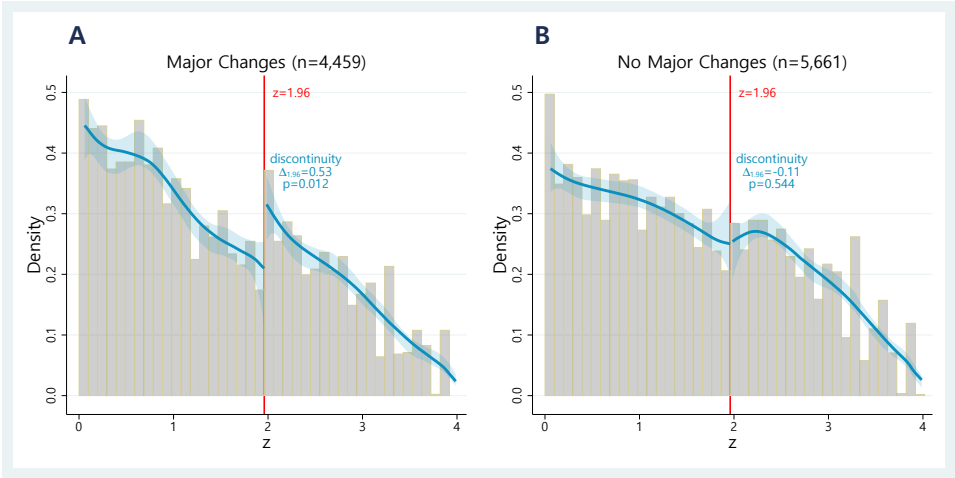


Performance Comparison: Average Metric Scores



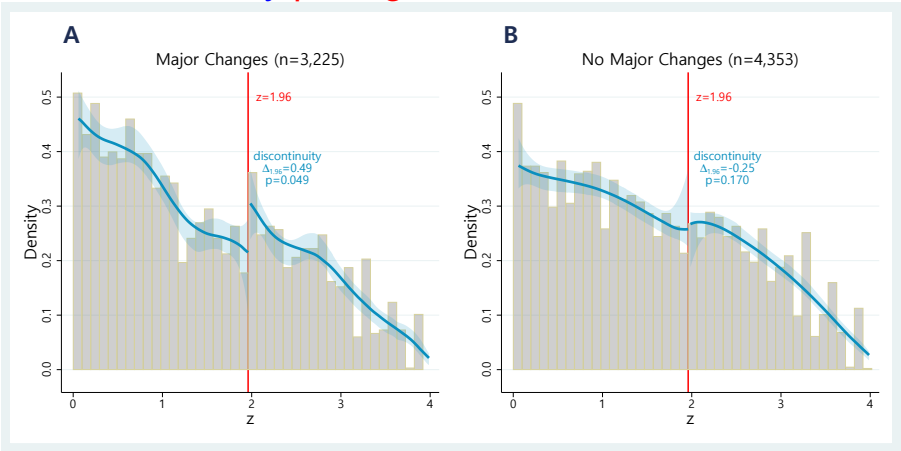
- Averages calculated across the two classes: major and no major changes
- Highest performance $\Rightarrow$ BioBERT on Short Descriptions

Is there a jump at the significance threshold for all trials?



- **Significant discontinuity** for trials at high risk of major change

What if we consider only preregistered trials?

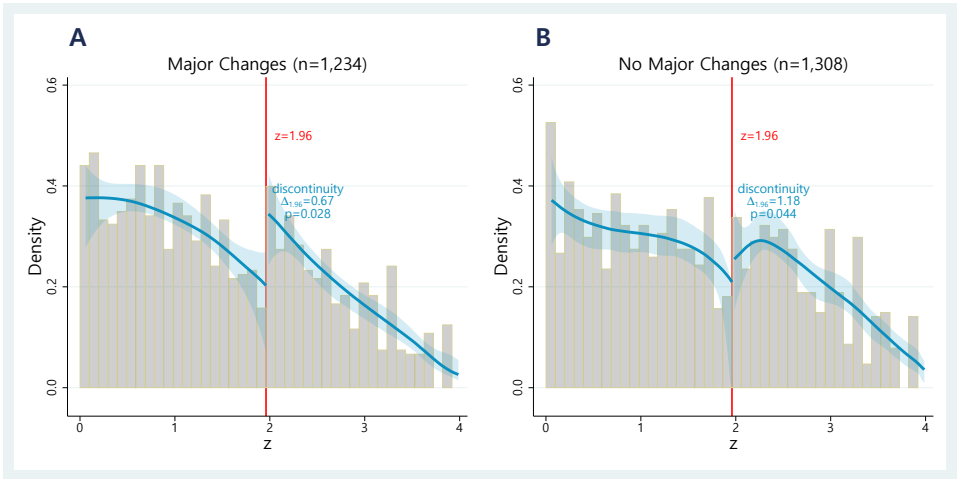


- **Significant discontinuity** for trials at high risk of major change

⇒ In contrast to **no discontinuity** with pooled data (Decker and Ottaviani, 2023)

preregistration

What if we consider only non-preregistered trials?



- **Discontinuity** in both groups: undisclosed changes?

Adding Controls? Caliper Tests for $z \in 1.96 \pm 0.1$ - Preregistered Trials

	Preregistered			
	(1)	(2)	(3)	(4)
major change	**0.108****	**0.0862***	**0.106***	**0.111****
	(0.0500)	(0.0512)	(0.0569)	(0.0563)
data monitoring committee			0.0304	0.0240
			(0.0566)	(0.0573)
subject masked			0.0337	0.0288
			(0.128)	(0.137)
caregiver masked			0.0224	0.0276
			(0.0767)	(0.0757)
investigator masked			-0.202*	-0.118
			(0.119)	(0.131)
outcomes assessor masked			0.0259	0.0211
			(0.0690)	(0.0688)
PI no employee			0.163*	0.219**
			(0.0936)	(0.0979)
Observations	378	378	338	338
R-squared	0.012	0.127	0.153	0.179
No. of trials	336	336	299	299
Controls		✔		✔
Start Year FE		✔	✔	✔
Mesh Condition FE		✔	✔	✔
Window	1.96 ± 0.1	1.96 ± 0.1	1.96 ± 0.1	1.96 ± 0.1
Major Changes Sig. Rate	0.671	0.671	0.684	0.684

Other Bandwidths All Trials

$$significant_{ij} = \alpha + \beta change_j + x_j'\gamma + \varepsilon_{ij}, \quad z \in 1.96 \pm 0.1$$

- **Other design features**: superior research integrity
- Additional **controls** as in Decker and Ottaviani (2023)

Summary Stats

- Effect of outcome changes: seems to remain
- **Robust** in probit and logit

Probit Logit

Conclusion

1. **BioBERT** outperforms **BERT** and **cosine similarity** in classification task
2. Significant impact of outcome changes on trial results: potential manipulation
3. Non-preregistered trials: discontinuity in both groups
   $\Rightarrow$ **preregistration** matters

Future Research Directions

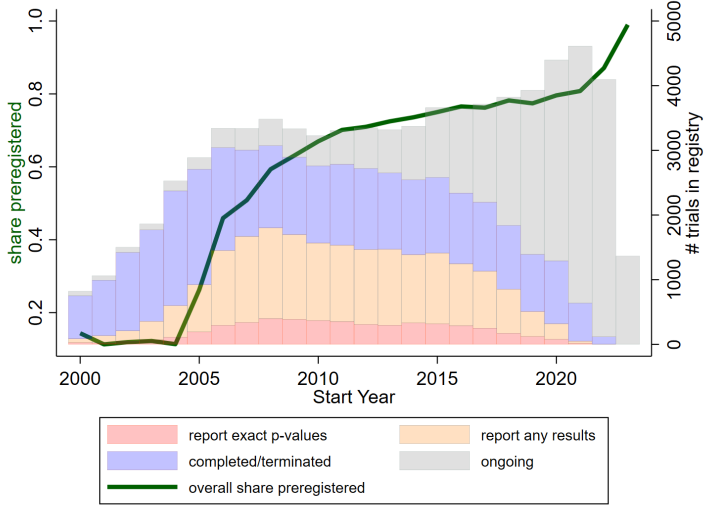
Open questions:

$\Rightarrow$ What about the **dark side** of preregistration?

$\Rightarrow$ Is there a cost in terms of novelty?

Thank you!

Time Trend



back

Summary Statistics

	mean	sd	min	median	max
duration [months]	32.05	25.73	0	25	261
# exact p-values from primary outcomes	2.10	2.26	1	1	30
# exact p-values from secondary outcomes	11.30	50.42	0	2	2,632
enrollment	569.93	1,623.86	4	225	27,395
placebo-controlled	0.694	0.461			
preregistered	0.749	0.433			
phase II	0.443	0.497			
phase III	0.519	0.500			
phase II/III combined	0.038	0.192			
non-industry sponsor	0.257	0.437			
top 10 industry sponsor	0.233	0.423			
small industry sponsor	0.510	0.500			
data monitoring committee	0.549	0.498			
subject masked	0.754	0.431			
caregiver masked	0.438	0.496			
investigator masked	0.737	0.440			
outcomes assessor masked	0.423	0.494			
mask folds	2.35	1.52	0	2	4
PI not employee of sponsor	0.905	0.294			

back

	(1) exact p-values provided	(2) results provided, but no exact p-values	(3) no results provided	(4) difference (1)–(2)	(5) difference (1)–(3)	(6) difference (2)–(3)
preregistered	0.749	0.669	0.435	0.080***	0.314***	0.234***
	(0.433)	(0.471)	(0.496)	(0.008)	(0.008)	(0.005)
data monitoring committee	0.549	0.495	0.469	0.053***	0.079***	0.026***
	(0.498)	(0.500)	(0.499)	(0.009)	(0.008)	(0.005)
subject masked	0.754	0.421	0.388	0.333***	0.366***	0.034***
	(0.431)	(0.494)	(0.487)	(0.008)	(0.007)	(0.005)
caregiver masked	0.438	0.238	0.229	0.199***	0.208***	0.009**
	(0.496)	(0.426)	(0.420)	(0.007)	(0.007)	(0.004)
investigator masked	0.737	0.405	0.363	0.332***	0.374***	0.043***
	(0.440)	(0.491)	(0.481)	(0.008)	(0.007)	(0.005)
outcome assessor masked	0.423	0.237	0.224	0.187***	0.199***	0.013***
	(0.494)	(0.425)	(0.417)	(0.007)	(0.007)	(0.004)
mask folds	2.352	1.301	1.203	1.051***	1.149***	0.098***
	(1.524)	(1.604)	(1.583)	(0.026)	(0.025)	(0.015)
PI no employee of sponsor	0.905	0.781		0.124***		
	(0.294)	(0.414)		(0.006)		
enrollment	569.9	242.1	245.0	327.9***	324.9***	-2.9
	(1,623.9)	(999.2)	(978.0)	(19.0)	(17.2)	(9.6)
placebo-controlled	0.694	0.350	0.289	0.344***	0.404***	0.060***
	(0.461)	(0.477)	(0.453)	(0.008)	(0.007)	(0.004)
industry-sponsored	0.743	0.572	0.466	0.172***	0.278***	0.106***
	(0.437)	(0.495)	(0.499)	(0.008)	(0.008)	(0.005)
phase III	0.519	0.368	0.389	0.151***	0.130***	-0.021***
	(0.500)	(0.482)	(0.488)	(0.008)	(0.008)	(0.005)
Observations	4,810	17,255	28,665	22,065	33,475	45,920

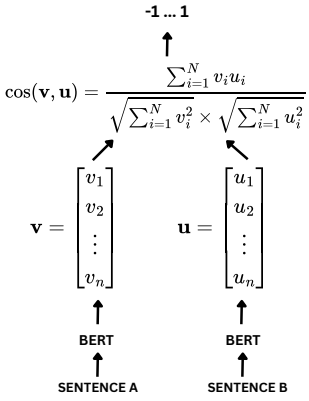
back

Caliper Test Regressions with Alternative Bandwidths

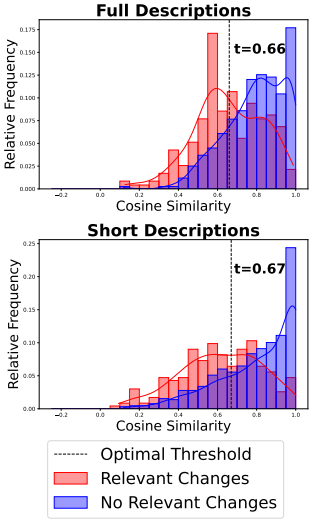
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$z \in 1.96 \pm 0.05$	-0.0874	-0.0681	-0.120*	-0.126	-0.0603	-0.101	-0.162
$[N = 254]$	(0.0633)	(0.0849)	(0.0687)	(0.0955)	(0.112)	(0.153)	(0.251)
$z \in 1.96 \pm 0.10$	-0.0680	-0.0558	-0.0826*	-0.0871	-0.00184	0.0688	0.0452
$[N = 512]$	(0.0464)	(0.0517)	(0.0495)	(0.0566)	(0.0766)	(0.0887)	(0.114)
$z \in 1.96 \pm 0.15$	-0.0895**	-0.0793*	-0.115***	-0.117**	-0.103*	-0.0976	-0.136*
$[N = 786]$	(0.0391)	(0.0440)	(0.0418)	(0.0463)	(0.0573)	(0.0706)	(0.0804)
$z \in 1.96 \pm 0.20$	**-0.0862****	**-0.0740***	**-0.0965****	**-0.0927****	**-0.0991***	**-0.116***	**-0.121***
$[N = 1,033]$	**(0.0350)**	**(0.0396)**	**(0.0377)**	**(0.0426)**	**(0.0508)**	**(0.0614)**	**(0.0704)**
$z \in 1.96 \pm 0.25$	-0.0677**	-0.0505	-0.0754**	-0.0658*	-0.0592	-0.0523	-0.0538
$[N = 1,327]$	(0.0316)	(0.0349)	(0.0332)	(0.0365)	(0.0442)	(0.0523)	(0.0572)
$z \in 1.96 \pm 0.30$	-0.0455	-0.0331	-0.0554*	-0.0500	-0.0348	-0.0168	-0.0191
$[N = 1,585]$	(0.0287)	(0.0323)	(0.0304)	(0.0341)	(0.0402)	(0.0476)	(0.0518)
$z \in 1.96 \pm 0.35$	-0.0458*	-0.0424	-0.0511*	-0.0550*	-0.0250	-0.0224	-0.0146
$[N = 1,826]$	(0.0274)	(0.0310)	(0.0294)	(0.0326)	(0.0373)	(0.0445)	(0.0480)
$z \in 1.96 \pm 0.40$	-0.0417	-0.0250	-0.0471*	-0.0318	-0.0226	-0.00454	0.00475
$[N = 2,102]$	(0.0256)	(0.0290)	(0.0272)	(0.0305)	(0.0344)	(0.0411)	(0.0436)
$z \in 1.96 \pm 0.50$	-0.0463**	-0.0415	-0.0510**	-0.0504*	-0.0312	-0.0169	-0.0199
$[N = 2,645]$	(0.0235)	(0.0267)	(0.0249)	(0.0278)	(0.0305)	(0.0363)	(0.0384)
$z \in 1.96 \pm 0.60$	-0.0413*	-0.0525**	-0.0372	-0.0527*	-0.0233	-0.0230	-0.0232
$[N = 3,135]$	(0.0229)	(0.0257)	(0.0242)	(0.0269)	(0.0287)	(0.0334)	(0.0350)
Controls	no	yes	no	yes	no	yes	yes
Start Year FE	no	yes	no	yes	no	yes	yes
Mesh Condition FE	no	yes	no	yes	no	yes	yes
Other Design Features	no	no	yes	yes	no	no	yes
Sponsor FE	no	no	no	no	yes	yes	yes

back

Cosine Similarity Optimal Threshold

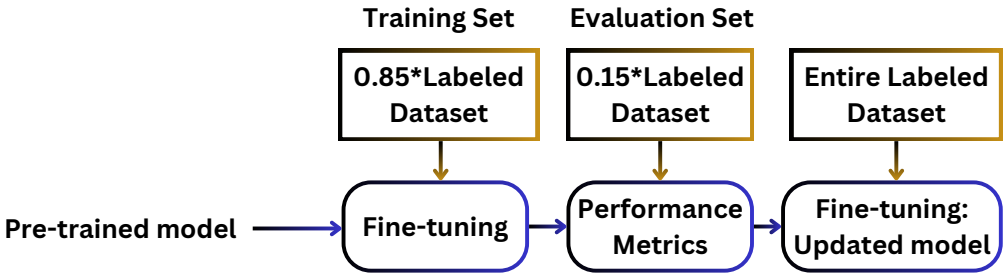


- **Sentence-Transformers : assigns similarity scores.**
- **Optimal threshold maximizes F1 score: best separates "Relevant Changes" from "No Relevant Changes".**



back

Fine-Tuning Large Language Models



- Fine-tuning of two large language models
  1. BERT: Bidirectional Encoder Representations from Transformers, developed by Google (Devlin et al., 2018)
  2. BioBERT: BERT for Biomedical Text Mining, domain-specific (Lee et al., 2019)

back

Summary Statistics of Trials at a High Risk *vs* at a Low Risk of Change

	(1) **no major** change	(2) **major** change	(3) difference (2)-(1)
preregistered	0.775	0.702	-0.073***
	(0.418)	(0.458)	(0.013)
data monitoring committee	0.521	0.600	0.079***
	(0.500)	(0.490)	(0.016)
subject masked	0.780	0.707	-0.073***
	(0.414)	(0.455)	(0.013)
caregiver masked	0.467	0.384	-0.083***
	(0.499)	(0.487)	(0.015)
investigator masked	0.765	0.685	-0.080***
	(0.424)	(0.465)	(0.014)
outcomes assessor masked	0.446	0.381	-0.066***
	(0.497)	(0.486)	(0.015)
PI no employee	0.913	0.890	-0.022***
	(0.282)	(0.312)	(0.009)
industry	0.778	0.679	-0.099***
	(0.416)	(0.467)	(0.014)
top10rev	0.243	0.216	-0.027**
	(0.429)	(0.411)	(0.013)
phase III	0.521	0.514	-0.007
	(0.500)	(0.500)	(0.015)
placebo	0.720	0.645	-0.075***
	(0.449)	(0.479)	(0.014)
enrollment	507.2	685	177.7***
	(1,408.8)	(1,953.1)	(53.7)
Observations	3,113	1,697	4,810

back

Caliper Tests Probit - Preregistered Trials

	Preregistered			
	(1)	(2)	(3)	(4)
$z \in 1.96 \pm 0.10$	.11**	.089*	.11**	.11**
$[N = 378]$	(0.049)	(0.048)	(0.052)	(0.051)
$z \in 1.96 \pm 0.15$	.096**	.081**	.095**	.097**
$[N = 591]$	(0.041)	(0.04)	(0.042)	(0.042)
$z \in 1.96 \pm 0.20$	.058	.053	.061	.064*
$[N = 782]$	(0.036)	(0.035)	(0.038)	(0.038)
$z \in 1.96 \pm 0.25$	.073**	.072**	.084**	.085**
$[N = 995]$	(0.032)	(0.031)	(0.034)	(0.034)
$z \in 1.96 \pm 0.30$	.059**	.063**	.075**	.076**
$[N = 1,177]$	(0.03)	(0.03)	(0.032)	(0.032)
$z \in 1.96 \pm 0.35$	.038	.043	.057*	.057*
$[N = 1,364]$	(0.029)	(0.029)	(0.031)	(0.031)
$z \in 1.96 \pm 0.40$	.044	.045*	.049*	.051*
$[N = 1,564]$	(0.027)	(0.027)	(0.029)	(0.029)
Controls		✔		✔
Other Design Features			✔	✔
Start Year FE		✔	✔	✔
Mesh Condition FE		✔	✔	✔

back

Caliper Tests Logit - Preregistered Trials

	Preregistered			
	(1)	(2)	(3)	(4)
$z \in 1.96 \pm 0.10$	.11**	.09*	.11**	.12**
$[N = 378]$	(0.049)	(0.049)	(0.054)	(0.052)
$z \in 1.96 \pm 0.15$	.095**	.081**	.095**	.099**
$[N = 591]$	(0.04)	(0.04)	(0.043)	(0.042)
$z \in 1.96 \pm 0.20$	.058	.053	.06	.064*
$[N = 782]$	(0.036)	(0.035)	(0.038)	(0.038)
$z \in 1.96 \pm 0.25$	.073**	.071**	.083**	.085**
$[N = 995]$	(0.032)	(0.031)	(0.034)	(0.034)
$z \in 1.96 \pm 0.30$	.059**	.063**	.075**	.076**
$[N = 1,177]$	(0.03)	(0.03)	(0.032)	(0.032)
$z \in 1.96 \pm 0.35$	.038	.042	.057*	.057*
$[N = 1,364]$	(0.029)	(0.029)	(0.031)	(0.031)
$z \in 1.96 \pm 0.40$	.044	.045*	.049*	.051*
$[N = 1,564]$	(0.027)	(0.027)	(0.029)	(0.029)
Controls		✔		✔
Other Design Features			✔	✔
Start Year FE		✔	✔	✔
Mesh Condition FE		✔	✔	✔

back

Adding Controls? Caliper Tests - All Trials

	(1)	(2)	(3)	(4)	(5)	(6)
$z \in 1.96 \pm 0.05$	0.041	0.043	0.11	-0.025	-0.014	0.027
$[N = 254]$	(0.059)	(0.059)	(0.1)	(0.059)	(0.068)	(0.064)
$z \in 1.96 \pm 0.10$	0.097**	0.094**	0.054	0.049	0.064	0.062
$[N = 512]$	(0.042)	(0.042)	(0.077)	(0.043)	(0.049)	(0.048)
$z \in 1.96 \pm 0.15$	0.085**	0.079**	0.027	0.055	0.07*	0.069*
$[N = 786]$	(0.035)	(0.035)	(0.066)	(0.035)	(0.039)	(0.038)
$z \in 1.96 \pm 0.20$	0.053*	0.047	0.012	0.03	0.045	0.041
$[N = 1{,}033]$	(0.031)	(0.031)	(0.06)	(0.031)	(0.034)	(0.034)
$z \in 1.96 \pm 0.25$	0.071**	0.068**	0.053	0.055**	0.072**	0.069**
$[N = 1{,}327]$	(0.028)	(0.028)	(0.054)	(0.028)	(0.03)	(0.03)
$z \in 1.96 \pm 0.30$	0.066***	0.063**	0.074	0.055**	0.067**	0.066**
$[N = 1{,}585]$	(0.025)	(0.026)	(0.049)	(0.026)	(0.029)	(0.028)
$z \in 1.96 \pm 0.35$	0.05**	0.047*	0.073	0.043*	0.054**	0.054**
$[N = 1{,}826]$	(0.024)	(0.024)	(0.047)	(0.025)	(0.027)	(0.027)
$z \in 1.96 \pm 0.40$	0.048**	0.046**	0.052	0.043*	0.046*	0.046
$[N = 2{,}102]$	(0.023)	(0.023)	(0.044)	(0.023)	(0.025)	(0.025)
Preregistration Status		✔	✔	✔	✔	✔
Controls				✔		✔
Other Design Features					✔	✔
Start Year FE				✔	✔	✔
Mesh Condition FE				✔	✔	✔
Interaction effect			✔			

- Bandwidth values vary between 0.05 and 0.4
- Other design features: superior research integrity
- Same **controls** as in Decker and Ottaviani (2023)

back

Caliper Test with Alternative Bandwidths - Preregistered Trials

	Preregistered			
	(1)	(2)	(3)	(4)
$z \in 1.96 \pm 0.05$	0.021	-0.027	0.00072	0.042
$[N = 188]$	(0.071)	(0.073)	(0.085)	(0.082)
$z \in 1.96 \pm 0.10$	**0.11****	**0.086***	**0.11***	**0.11****
$[N = 378]$	(0.05)	(0.051)	(0.057)	(0.056)
$z \in 1.96 \pm 0.15$	**0.096****	**0.08***	**0.095****	**0.095****
$[N = 591]$	(0.041)	(0.041)	(0.044)	(0.044)
$z \in 1.96 \pm 0.20$	0.059	0.052	0.061	0.061
$[N = 782]$	(0.036)	(0.036)	(0.039)	(0.039)
$z \in 1.96 \pm 0.25$	0.073**	0.071**	0.084**	0.085**
$[N = 995]$	(0.032)	(0.032)	(0.035)	(0.035)
$z \in 1.96 \pm 0.30$	0.059**	0.062**	0.075**	0.076**
$[N = 1,177]$	(0.03)	(0.03)	(0.033)	(0.033)
$z \in 1.96 \pm 0.35$	0.038	0.042	0.057*	0.057*
$[N = 1,364]$	(0.029)	(0.029)	(0.032)	(0.031)
$z \in 1.96 \pm 0.40$	0.044	0.044	0.049	0.051*
$[N = 1,564]$	(0.027)	(0.027)	(0.03)	(0.029)
Controls		✔		✔
Other Design Features			✔	✔
Start Year FE		✔	✔	✔
Mesh Condition FE		✔	✔	✔

- **Bandwidth** values vary between 0.05 and 0.4
- Other design features: superior research integrity
- Same **controls** as in Decker and Ottaviani (2023)
- Effect of outcome changes: **remains** for preregistered trials

back

Sponsor Fixed effects ? Preregistered Trials

	Preregistered			
	(1)	(2)	(3)	(4)
$z \in 1.96 \pm 0.05$	-.14	-.24	-.24	-.39*
$[N = 114]$	(0.14)	(0.17)	(0.21)	(0.23)
$z \in 1.96 \pm 0.10$	.0061	.015	.088	.09
$[N = 263]$	(0.08)	(0.091)	(0.12)	(0.12)
$z \in 1.96 \pm 0.15$	.011	.025	.068	.052
$[N = 453]$	(0.061)	(0.065)	(0.075)	(0.076)
$z \in 1.96 \pm 0.20$	.03	.058	.085	.09
$[N = 630]$	(0.051)	(0.053)	(0.061)	(0.061)
$z \in 1.96 \pm 0.25$	.051	.06	.078	.079
$[N = 814]$	(0.044)	(0.045)	(0.052)	(0.052)
$z \in 1.96 \pm 0.30$	.033	.036	.055	.053
$[N = 999]$	(0.04)	(0.042)	(0.047)	(0.047)
$z \in 1.96 \pm 0.35$	.011	.014	.031	.024
$[N = 1,177]$	(0.038)	(0.039)	(0.044)	(0.044)
$z \in 1.96 \pm 0.40$	.016	.013	.026	.018
$[N = 1,369]$	(0.035)	(0.036)	(0.04)	(0.04)
Controls		✔		✔
Other Design Features			✔	✔
Start Year FE		✔	✔	✔
Mesh Condition FE		✔	✔	✔
Sponsor FE	✔	✔	✔	✔

- Effect of outcome changes: **absorbed**
- Unobserved characteristics of researchers or their sponsoring organizations?

back