P-hacking in clinical trials and how incentives shape the distribution of results across phases

Jérôme Adda
Bocconi University

Christian Decker
University of Zurich

Marco Ottaviani
Bocconi University

June 2020

Motivation

- clinical research should conform to highest ethical standards
  $\Rightarrow$human lives may be at stake!
- economic incentives (large R&D costs, even larger potential profits) may generate conflicts of interest for investigators and pressure to withhold or "beautify" unfavorable results

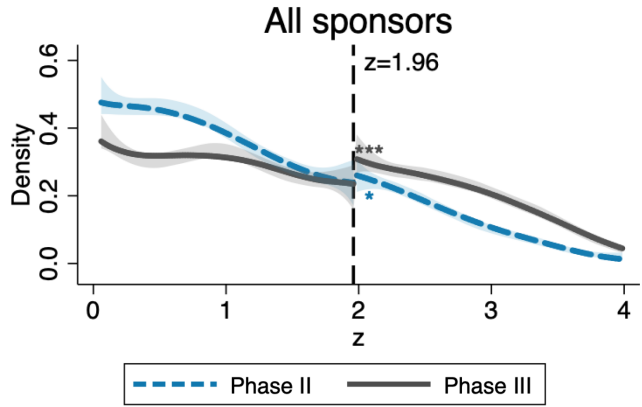
This Paper

- first systematic evaluation of distribution of p-values reported to *ClinicalTrials.gov*
- investigate "suspicious patterns" depending on economic 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)

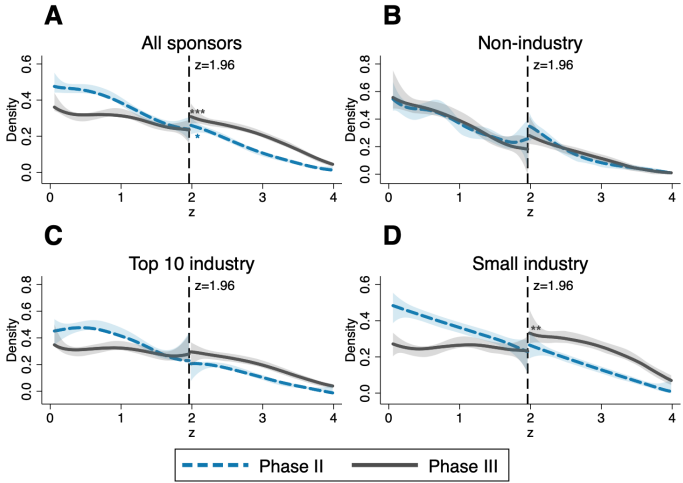
The Distribution of z-Scores on *ClinicalTrials.gov*

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



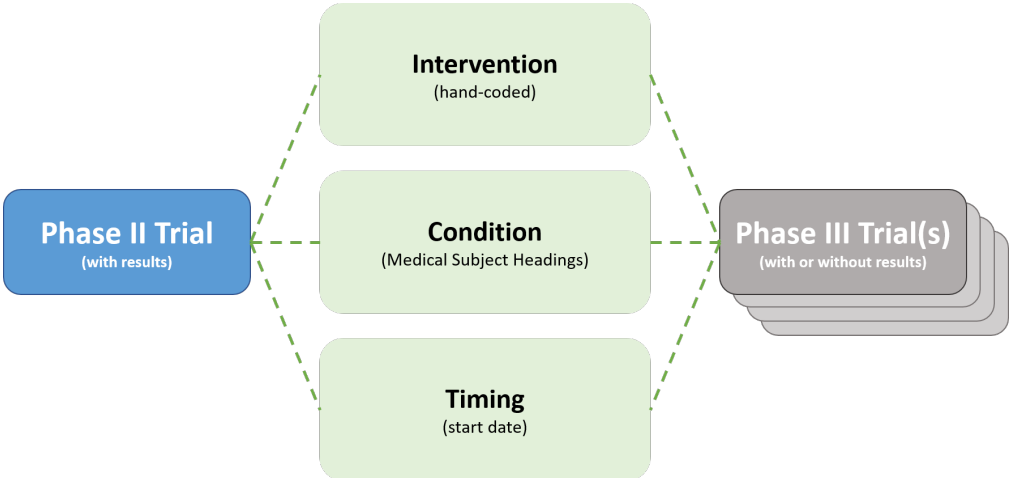
Breakdown by Affiliation of Lead Sponsor



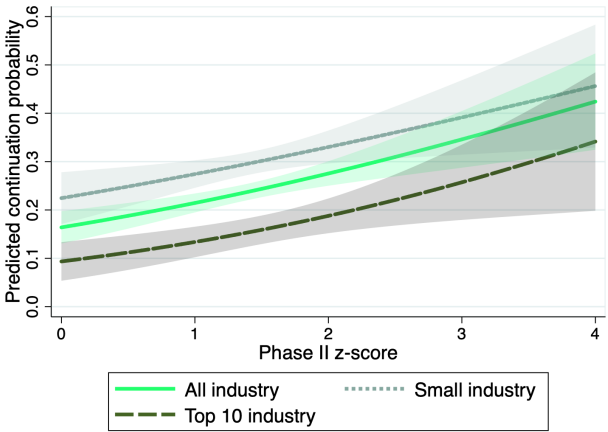
Takeaways

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



Takeaways

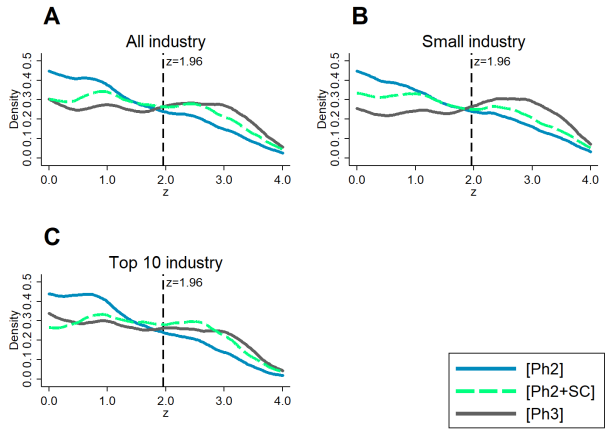
1. Higher phase II z-score significantly increases the probability of continuation to phase III.

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

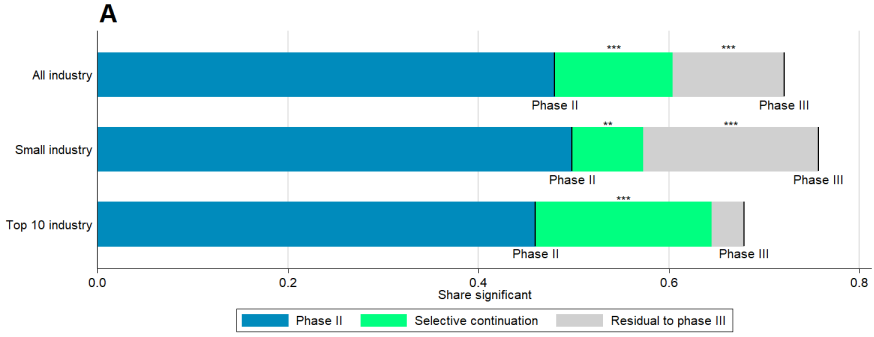
Controlling for *Selective Continuation*

- 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*

$\Rightarrow$ registries for pre-registration of RCTs and result databases help

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

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

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