Suspect Research & Statistical Inferences

Jacob Schauer
jms@u.northwestern.edu

Northwestern University
IES Grant R305B140042

SREE 2018

Suspect research practices

- Suspect practices *can* lead to inaccurate findings (e.g., Hedges, 2017; Lindsay, 2012).
- How much? Depends on which suspect practice...
- Conditional data collection: failing significance, get more data!?
    1. Bias treatment effect estimates
    2. Bias can be large (~50%)
    3. Bias can arise even if no analyses are conducted

Conditional data collection

- John, Loewenstein, & Prelec (2012): $>$50% of respondents admitted they had collected more data based on a nonsignificant result.
- Fiedler & Schwartz (2016): $>$30% of respondents admitted to collecting more data in order to render a nonsignificant result significant.
- Simmons, Nelson, & Simonsohn (2011): Inflated type I error rates ($\sim$10-20%)
- Related to sequential trials in medicine (Nardini & Sprenger, 2012)

Conditional data collection

- Treatment effect $\theta \neq 0$
- Initial experiment ($\underline{n \text{ subjects in each of treatment \& control}}$)
  - Estimate T_0 (<u>mean difference</u>)
- Concomitant variable O correlated with T_0
- Based on O either:
  1. Report T_0
  2. Continue experiment
     - Recruit more subjects ($\underline{m \text{ subjects per arm}}$)
     - Report T_1

Conditional inferences

- We only observe an estimate conditional on O:
  1. $T_0|O$
  2. $T_1|O$

- Bias:
  1. $E[T_0|O] - \theta$
  2. $\underline{E[T_1|O] = \frac{n}{m+n}(E[T_0|O] - \theta)}$

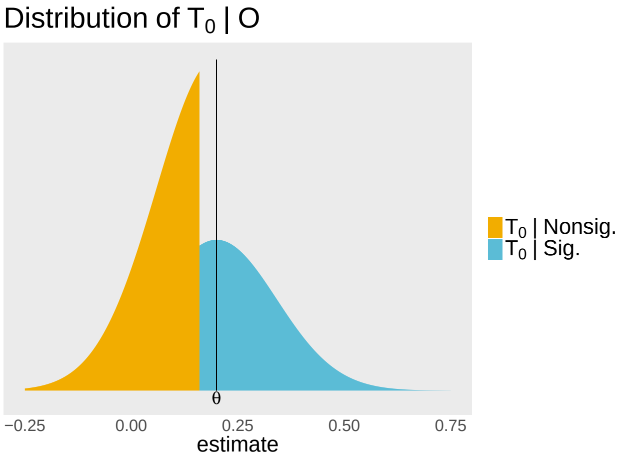
- **If O is correlated to T_0, the treatment effect can be biased**.

Model

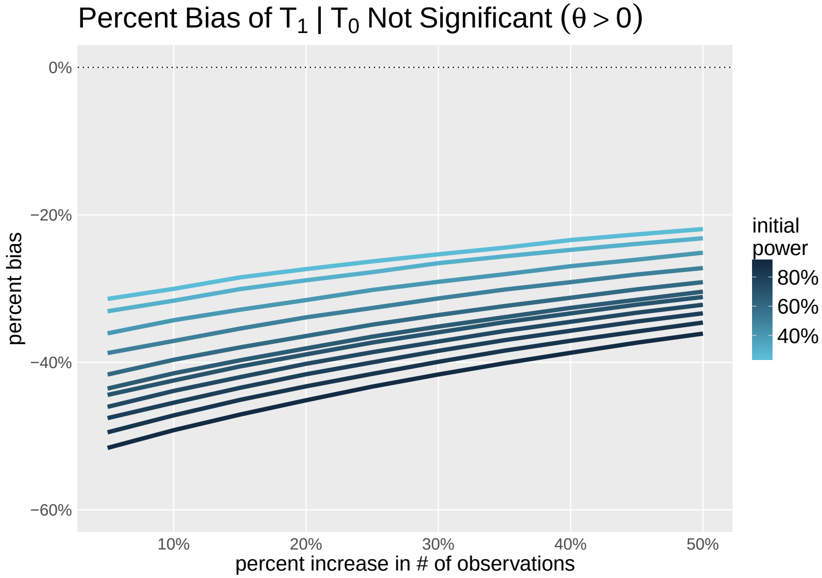
1. Data are normally distributed, with known variance.
2. n subjects per arm in initial experiment
   - $T_0 \sim N(\theta, 2\sigma^2/n)$
3. m subjects added per arm, whose responses are independent of past observations.

Conditional on significance

- $O = \mathbf{1}\{|T_0| > 1.96\sqrt{2\sigma^2/n}\}$; ($\alpha = .05$, 2-tailed test)
- Stop if $O = 1$, continue if $O = 0$
- T_0 will be a truncated normal



Collecting data based on nonsignificance



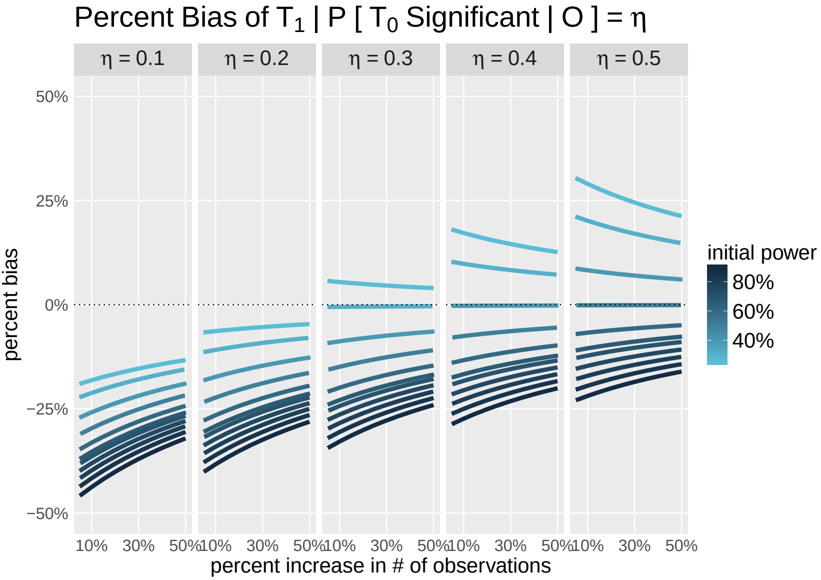
Other concomitant variables

- "If O is correlated to T_0, the treatment effect can be biased."
  - Me, three slides ago.
- Researchers may observe any number of variables correlated with T_0.
  - Casual observations may be correlated with T_0.
  - If more data are collected based on them, $T_1|O$ can be biased.
  - No analysis of intial data needed.
- How might these variables convey information about T_0?
  - How likely is it that T_0 will be significant given what was observed?

Information about possible significance

- Probabilistic model
- O provides information about how likely T_0 is to be significant:
  - $P[T_0 \text{ significant}|O] = \eta$
- For a given probability of significance (η), a researcher may
  - stop and report $T_0|\eta$
  - collect more data and report $T_1|\eta$
- Assume O conveys *only* information about the probability of significance.
  - $T_0|O$ has a reweighted normal distribution

Continuing due to improbable significance



Unknowable bias

- It may be impossible to determine exactly what information any observation conveys about T_0.
- If it carries *any* information, and the decision to collect more data depends on it, we know that T_1 can be biased.
- We may have no idea how biased a given result is.
- **Ad-hoc data collection can bias a treatment effect estimate even if no analysis of interim data is conducted. It may be impossible to know how much this bias is!**

Conclusions

- Bias from conditionally collected data can be substantial, even if a researcher does not actually run a significance test.
- Pre-registration can improve transparency, and help curtail more passive forms of CDC.
  - SREE!
- Blinding?
- Empirical replication of past results.

References

Fan, X. (Frank), DeMets, D. L., & Lan, K. K. G. (2004). Conditional bias of point estimates following a group sequential test. *Journal of Biopharmaceutical Statistics*, 14(2), 505–530.

Fiedler, K. & Schwarz, N. (2016). Questionable research practices revisited. *Social, Psychological, and Personality Science*, 7(1), 45–52.

Hedges, L. V. (2017). Challenges in building usable knowledge in education. *Journal of Research on Educational Effectiveness*, 11(1), 1-21.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532.

Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science*, 26(12), 1827–1832.

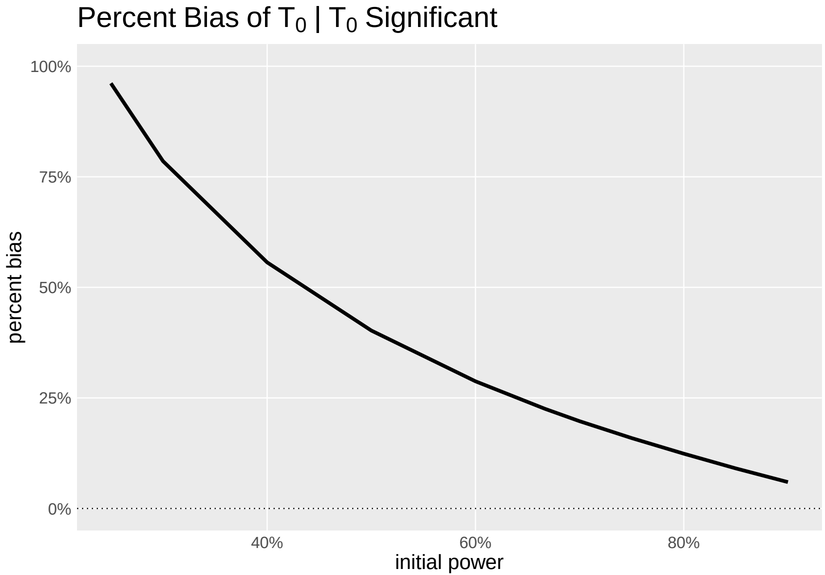
Nardini, C., & Sprenger, J. (2013). Bias and conditioning in sequential medical trials. *Philosophy of Science*, 80(5), 1053–1064.

Simmons, J. P., Nelson, L. D. & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11), 1359–1366.

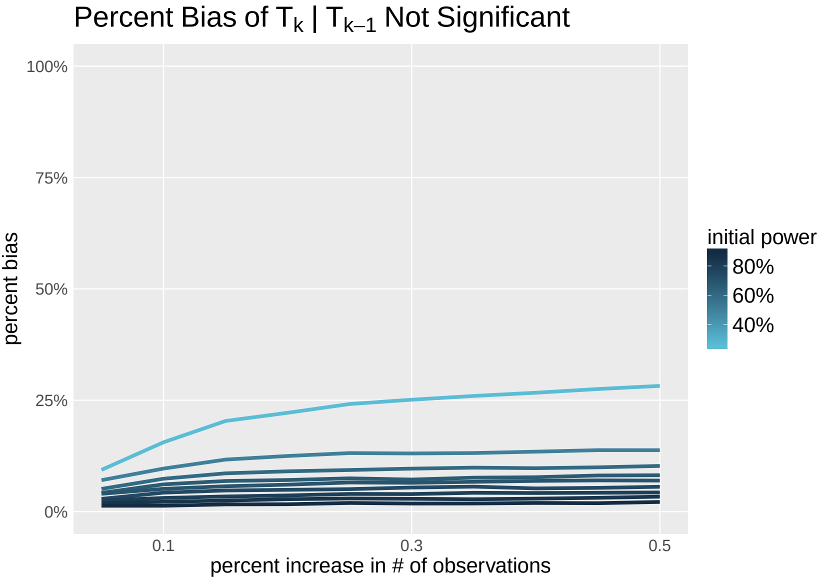
Thank You!

Jacob Schauer
PhD Student
Department of Statistics
Northwestern University
jms@u.northwestern.edu

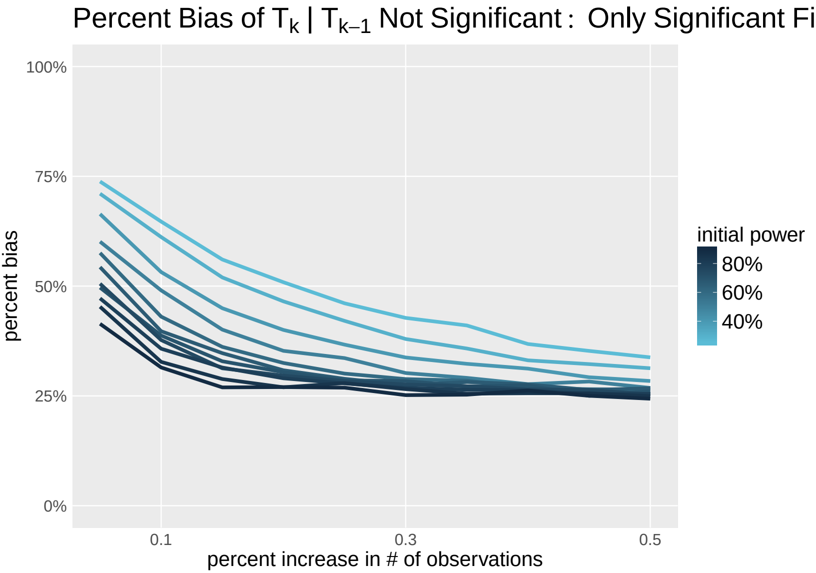
Stopping for significance



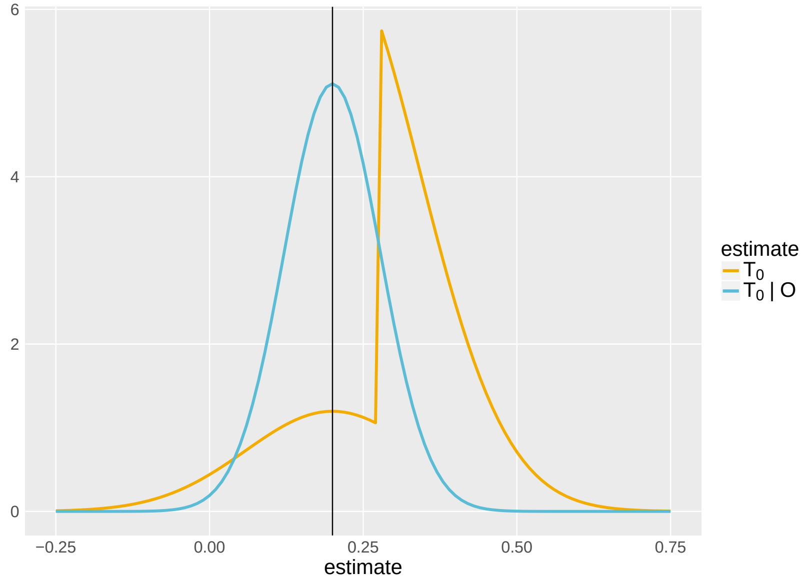
Repeated waves



Repeated waves, reporting only significant results



Distributions: probable significance



Stopping for probable significance