Optimal Pre-Analysis Plans:
Statistical Decisions Subject to Implementability

Maximilian Kasy   Jann Spiess

April 2024

Introduction

- Trial registration and pre-analysis plans (PAPs) have become a standard requirement for experimental research.
    - For clinical studies in medicine starting in the 1990s.
    - For experimental research in economics more recently.
- Standard justification: Guarantee validity of inference.
    - P-hacking, specification searching, and selective publication distort inference.
    - Tying researchers' hands prevents selective reporting.
    - Christensen and Miguel (2018); Miguel (2021).
- The widespread adoption of PAPs has not gone uncontested, however.
    - Coffman and Niederle (2015); Olken (2015); Duflo et al. (2020).

Open questions

1. Why do we need a commitment device?
   Standard decision theory has no time inconsistency!

2. How should the structure of PAPs look like?
   How can we derive optimal PAPs?

Key insight:

- Single-agent decision-theory cannot make sense of these debates.
- We need to consider multiple agents,
  conflicts of interest, and asymmetric information.

Our approach

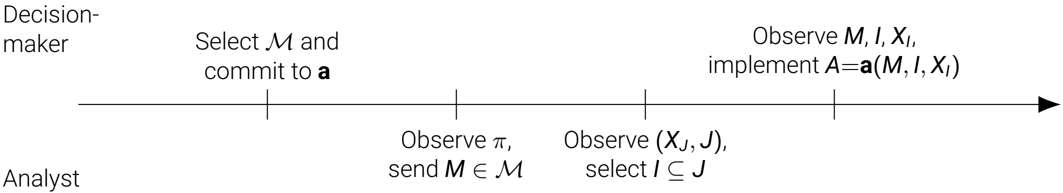
- Import insights from contract theory / mechanism design to statistics.
  - We consider (optimal) statistical decision rules
    subject to the constraint of implementability.
  - PAPs are generically necessary for implementation.
  - They allow to leverage researcher expertise
    while maintaining incentive compatibility of non-selective reporting.
- Our model:
  1. A decision-maker commits to a decision rule,
  2. then an analyst communicates a PAP,
  3. then observes the data,
     reports selected (!) statistics to the decision-maker,
  4. who then applies the decision rule.

Note: The model presented in this talk is different from that discussed in an earlier working paper on the same topic.

Setup: Notation

- Two parties, decision-maker and analyst.
- Message M ("pre-analysis plan") sent from analyst to decision-maker.
- Data $X = (X_1, \ldots, X_n) \sim P_\theta$.
  - Unknown parameter $\theta \in \Theta$.
- Index sets:
  - $K = \{1, \ldots, n\}$ fixed, finite, commonly known.
  - $J \subset K$ subset of data available to the analyst, privately known.
  - $I \subset J$ subset of available data reported to the decision-maker.
- Decision $A \in \mathcal{A} \subseteq \mathbb{R}$.

Setup: Timeline



Discussion

- The analyst can withhold information,
  but they cannot lie.
- The components of X might represent different
  - hypothesis tests,
  - estimates,
  - subgroups,
  - outcome variables, etc.
- Possible model interpretations:
  1. Drug approval (pharma company vs. FDA).
  2. Hypothesis testing (researcher vs. reader).
  3. Publication decision (researcher vs. journal).

Introduction

Setup

Motivating example: Normal testing

Implementable decision functions

Hypothesis testing

Conclusion and outlook

Motivating example: Normal testing

- $K = \{1, 2\}$.
- $X_1, X_2 \sim N(\theta, 1)$.
- The analyst knows J, but the decision-maker does not.
- Null hypothesis $H_0 : \theta \leq 0$.
- The analyst selectively reports, to get a rejection of the null.

Compare 5 testing rules

0. The optimal full data test (only available if $I = J = \{1, 2\}$).
1. The naive test (ignores selective reporting).
2. The conservative test (worst-case assumptions about unreported X_ι).
3. The optimal implementable test without a PAP.
4. The optimal implementable test with a PAP.

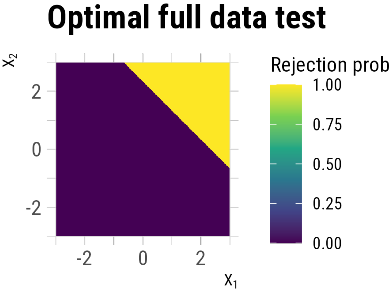
The optimal full data test

- Suppose availability and selective reporting were no concern.
- Then $X_1 + X_2$ is a sufficient statistic.
- By Neyman-Pearson, the uniformly most powerful test is given by

$$\mathbf{1}\left(X_1 + X_2 > \sqrt{2} \cdot z\right).$$

- Critical value:

$$z = \Phi^{-1}(1 - \alpha).$$



The naive test

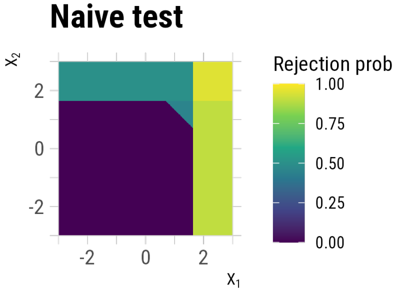
- Treat the reported data I as if there were no selective reporting.

$$\mathbf{a}_1(X_I, I) = \mathbf{1}\left(\sum_{\iota \in I} X_\iota > z \cdot \sqrt{|I|}\right).$$

- The analyst chooses $I \subset J$ to maximize rejection,

$$\bar{\mathbf{a}}_1(X_J, J) = \max_{I \subset J} \mathbf{a}(X_I, I).$$

- Such p-hacking violates size control!

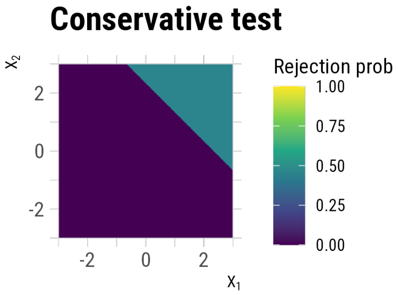


The conservative test

- Possible remedy:
  Worst-case assumptions about unreported components.

  $$\mathbf{a}_2(X_I, I) = \mathbf{1}\left(X_1 + X_2 > \sqrt{2}\cdot z \text{ and } I = K\right).$$

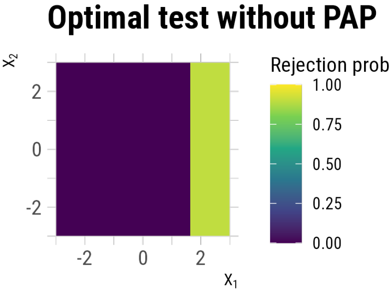
- This test controls size.
- But it has low power.



The optimal implementable test without PAP

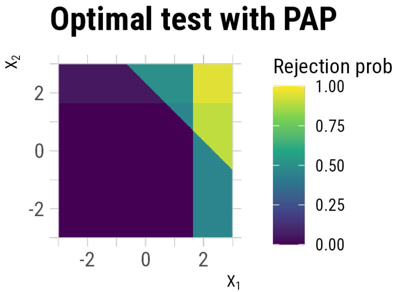
- Requirements:
  1. Size control.
  2. Incentive compatibility.
  3. Maximizes expected power.
- Solution without a PAP:
  1. Pick a full-data test,
  2. make worst-case assumptions about unreported components.
- Choose the full-data test to maximize expected power.
- Here:

$$\mathbf{a}_3(X_I, I) = \mathbf{1}\left(X_1 > z \text{ and } 1 \in I\right).$$

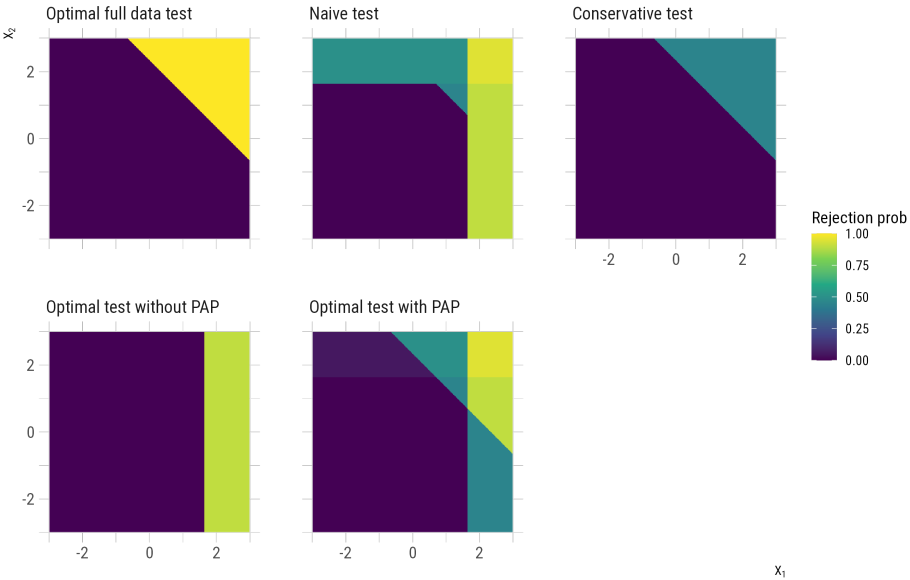


The optimal implementable test with PAP

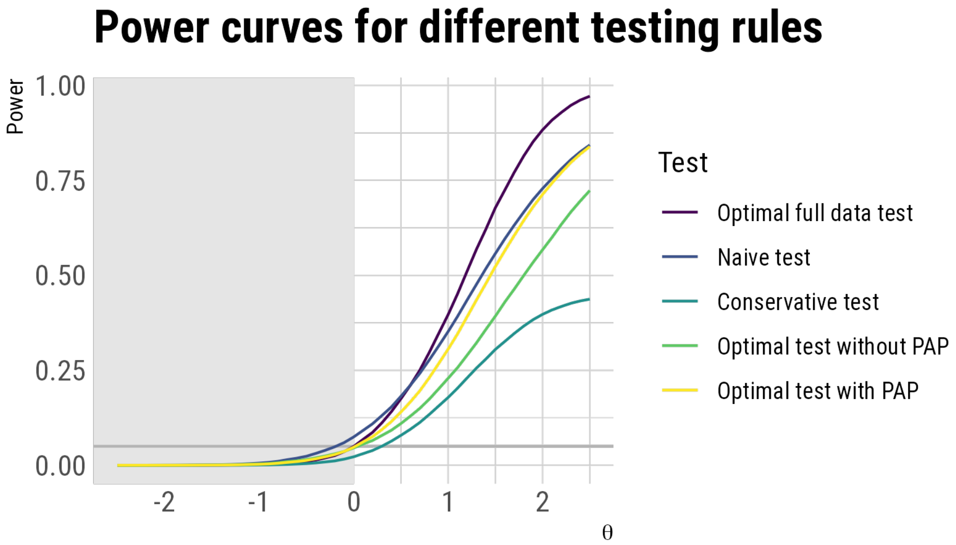
- Allow an analyst message before seeing data.
- Solution *with* a PAP :
  1. Let the *analyst* pick a full-data test,
  2. make worst-case assumptions about unreported components.
- The analyst knows J when choosing the full-data test.



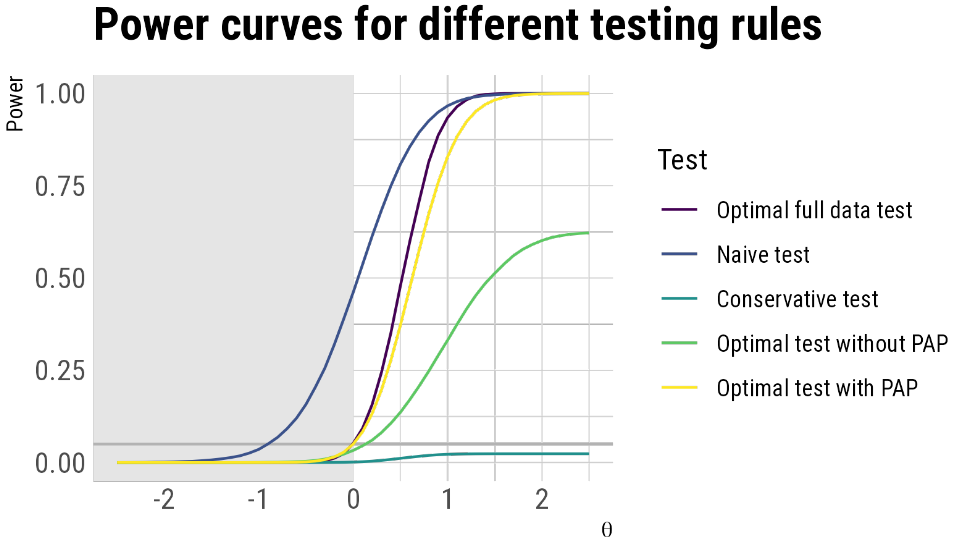
Rejection probabilities for different testing rules



Degrees of freedom $n = 2$



Degrees of freedom $n = 10$



Implementable decision functions

- A **reduced-form decision function** maps the full data into a decision **a**:

$$\bar{\mathbf{a}}(\pi, X_J, J)$$

- A reduced-form decision function $\bar{\mathbf{a}}$ is **implementable**
  - if there exist a decision function **a**
  - with best responses M^*, I^*
  - such that

$$\bar{\mathbf{a}}(\pi, X_J, J) = \mathbf{a}(M^*, X_{I^*}, I^*).$$

- **Assumption**:
  The analyst is an expected utility maximizer with utility

$$v(A)$$

  for a (strictly) monotonically increasing function v.

Analyst best responses

- The optimal report I^* of the analyst satisfies

$$I^* \in \operatorname*{argmax}_{I \subseteq J} \mathbf{a}(M, X_I, I).$$

- The optimal message M^* satisfies

$$M^* \in \operatorname*{argmax}_{M} \mathrm{E}[v(\mathbf{a}(M, I^*, X_{I^*}))|\pi].$$

Preview of implementability results

- Without PAPs, implementability is equivalent to **monotonicity** in J: Reporting more can only increase the decision.
- With PAPs, implementability only requires monotonicity in J **conditional** on the analyst signal.
  $\Rightarrow$ Can leverage analyst expertise!
- Implementation can use different approaches:
  1. Truthful **revelation** of the analyst signal.
  2. **Delegation** to the analyst, letting them choose a decision function from a constrained set.
- For binary actions, the set of implementable decision functions is a **convex polytope**.
- Truthful revelation is closely related to **proper scoring**.

Implementability without PAPs

Proposition

If no pre-analysis messages $\boldsymbol{M}$ *are allowed,*
a reduced-form decision function $\bar{\mathbf{a}}(\pi, X_J, J)$ *is implementable iff*

1. $\bar{\mathbf{a}}$ *does not depend on* π*, and*

2. $\bar{\mathbf{a}}$ *is* ***monotonic*** *in* J*,*

$$\bar{\mathbf{a}}(X_I, I) \leq \bar{\mathbf{a}}(X_J, J)$$

for almost all $\boldsymbol{X}, \boldsymbol{J}$ *and all* $\boldsymbol{I} \subseteq \boldsymbol{J}$*.*

Proof

1. Suppose that both conditions hold.
    - Set $\mathbf{a}(X_I, I) = \bar{\mathbf{a}}(X_I, I)$.
    - Incentive compatibility of $I^* = J$ follows.
2. Consider a decision function $\bar{\mathbf{a}}$ that is implementable by $\mathbf{a}$.
    - Since I^* is an analyst best-response to this decision function $\mathbf{a}$,

$$\bar{\mathbf{a}}(\pi, X_J, J) = \max_{I \subseteq J} \mathbf{a}(X_I, I).$$

    - The maximum over subsets of J (weakly) increases in J. □

Note: The revelation principle does not directly apply here, due to partial verifiability!

Implementability with PAPs

Theorem

A reduced-form decision function $\bar{\mathbf{a}}$ *is implementable iff both of the following conditions hold:*

1. ***Truthful PAP***
   For almost all π *and all* π',
   $$\mathrm{E}[v(\bar{\mathbf{a}}(\pi', X_J, J))|\pi] \leq \mathrm{E}[v(\bar{\mathbf{a}}(\pi, X_J, J))|\pi].$$
2. ***Monotonicity***
   For almost all π, X, J, *and all* $I \subseteq J$
   $$\bar{\mathbf{a}}(\pi, X_I, I) \leq \bar{\mathbf{a}}(\pi, X_J, J)$$

Revelation and delegation

Proposition

A reduced-form decision rule $\bar{\mathbf{a}}$ *can be implemented iff:*

1. ***Implementation by truthful revelation***
   It can be implemented with a decision rule $\mathbf{a}$ *for which*

$$\mathbf{a}(\pi, X_J, J) = \bar{\mathbf{a}}(\pi, X_J, J).$$

2. ***Implementation by delegation***
   It can be implemented with a decision rule $\mathbf{a}$ *for which*

$$\mathbf{a}(b, X_J, J) = b(X_J, J),$$

where $\mathbf{b}$ *is restricted to lie in some set* $\mathcal{B}$.

Hypothesis testing

- Null hypothesis $\theta \in \Theta_0$.
- Rejection probability $A \in [0,1]$.

$\Rightarrow$ w.l.o.g. $v(A) = A$.

- Size control at level $\alpha \in (0,1)$:

$$\sup_{\pi,\theta\in\Theta_0,J\subseteq\{1,\ldots,n\}} \mathrm{E}[\bar{\mathbf{a}}(\pi, X_J, J)|\theta, \pi, J] \leq \alpha.$$

- Expected power:

$$\mathrm{E}[\bar{\mathbf{a}}(\pi, X_J, J)].$$

Preview of optimal implementable tests

- Implementable tests are montonic,
  so that size control only binds for the full data.
- The **optimal test**
  - maximizes expected power,
  - subject to size control
  - and implementability.
- This test can be implemented as follows:
  - Ask the **analyst** to **choose a full-data test** that controls size.
  - For any report, **assume the worst** about the **unreported components**.
- The **analyst** problem of choosing the optimal full data test
  is a (simple) **linear program**.

Implementing the optimal test by delegation

Theorem

- *The test with maximal expected power*
- *subject to implementability and size control*
- *can be implemented by requiring the analyst to communicate a full-data test* $\mathbf{t}$ *which satisfies, for all* $\theta \in \Theta_0$,

$$E[t(X)|\theta] \leq \alpha$$

- *and then implementing the test*

$$b(X_I, I) = \min_{X';\, X'_I = X_I} t(X').$$

Sketch of proof

- Anything that can be implemented can be implemented by delegation.
- Implementable rules are monotonic.
- Monotonic rules satisfy size control iff they satisfy full-data size control.
- Subject to this constraint, analyst and decision-maker are aligned.
- Expected power given full-data size control and monotonicity is maximized by

$$b(X_I, I) = \min_{X';\, X'_I = X_I} t(X').$$

$\square$

The analyst's problem as a linear program

$$\max_{b} \int b(X_J, J) d\,\mathrm{P}_\pi(X, J), \qquad \text{(Interim expected power)}$$

$$\text{s.t.} \int b(X, K) d\,\mathrm{P}_{\theta_0}(X) \leq \alpha, \qquad \text{(Size control)}$$

$$b(X_J, J) \in [0, 1] \qquad \forall J, X, \qquad \text{(Support)}$$

$$b(X_J, J) \leq b(X, K) \qquad \forall J, X. \qquad \text{(Monotonicity)}$$

The optimal test when the analyst knows J

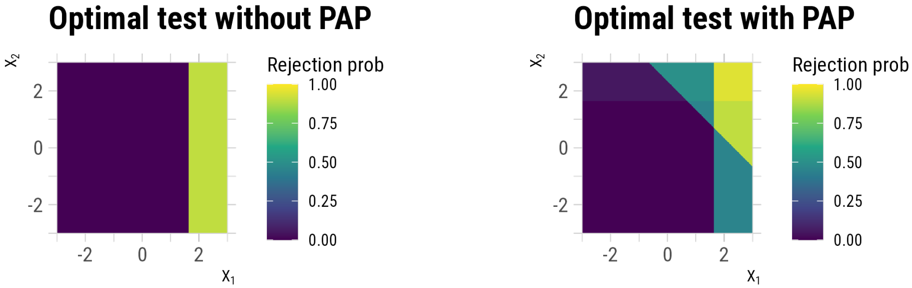
Proposition

- *Suppose that the analyst observes J before specifying the PAP.*
- *Then there exists a solution b to the analyst's problem such that $b(X_K, K) = b(X_J, J)$ for all values of X.*
- *Any solution of the analyst's problem that is of this form furthermore satisfies that*

$$b(X_K, K) = \begin{cases} 1 & \textit{when } d\,\mathrm{P}_\pi(X_J, J) > \kappa_J \cdot d\,\mathrm{P}_{\theta_0}(X_J, J) \\ 0 & \textit{when } d\,\mathrm{P}_\pi(X_J, J) < \kappa_J \cdot d\,\mathrm{P}_{\theta_0}(X_J, J) \end{cases}.$$

for some critical value κ.

Example revisited



Discussion

- Conflicts of interest, private information.
  $\Rightarrow$ Not all decision rules are implementable.
- Mechanism design: Optimal implementable rules.
- Statistical reporting: Partial verifiability.
  1. No lying about reported statistics.
  2. Private information about which statistics were available.
- Pre-analysis plans:
  - No role in single-agent decision-theory.
  - But increase the set of implementable rules in multi-agent settings.
- We characterize
  1. implementable rules,
  2. optimal implementable hypothesis tests.

Thank you!