# Rationalizing Pre-Analysis Plans: Statistical Decisions Subject to Implementability

Maximilian Kasy Jann Spiess

September 2021

# 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.
  - "PAPs are to frequentist inference what RCTs are to causality."
- Counter-arguments:
  - Pre-specification is costly.
  - Interesting findings are unexpected and flexibility is necessary.

# Open questions

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

2. Under what conditions are PAPs more or less useful?
   How do we trade off the benefits and costs of PAPs?

# Our approach

- Import insights from contract theory / mechanism design to statistics.
    - PAPs can be rationalized with multiple parties, conflicts of interest, and costly communication / asymmetric information.
    - We consider (optimal) statistical decision rules subject to the constraint of implementability.
- Our model:
    1. A decision-maker commits to a decision rule,
    2. then an analyst commits to a PAP,
    3. then observes the data, reports selected statistics to the decision-maker,
    4. who then applies the decision rule.
- PAPs are optimal when
    - there are many analyst degrees of freedom,
    - and/or communication costs are high.

# Alternative interpretations of our model

1. Publication decision:
   - A researcher wants to get published.
   - A journal wants to publish only studies for large enough true effects.
2. Drug approval:
   - A pharma company wants drug approval.
   - The public authority (FDA) wants to approve only effective drugs.
3. Hypothesis testing:
   - A researcher wants to always reject the null.
   - A reader wants to only reject when the null is false.

# Literature

- **P-hacking and publication bias**
  Ioannidis (2005); Gelman and Loken (2013); Andrews and Kasy (2019)
- **Contract theory and mechanism design**
  Hurwicz (1972); Glazer and Rubinstein (2004); Kamenica and Gentzkow (2011); Kamenica (2019)
- **Discussions of PAPs by empirical practitioners**
  Food and Drug Administration (1998); Coffman and Niederle (2015); Olken (2015); Christensen and Miguel (2018); Duflo et al. (2020)
- **Applied theory of research and the publication process**
  Chassang et al. (2012); Tetenov (2016); Ottaviani et al. (2017); Di Tillio et al. (2017); Spiess (2018); Henry and Ottaviani (2019); McCloskey and Michaillat (2020); Libgober (2020); Yoder (2020); Williams (2021); Abrams et al. (2021); Viviano et al. (2021); Banerjee et al. (2020); Frankel and Kasy (2021); Andrews and Shapiro (2020)

Introduction

Baseline model

- Assumptions
- Implementability and optimality

Analysis

- A minimal example: $\bar{n} = 3$
- Symmetric publication rules
- General solution

Model variations

- Frequentist testing
- Multiple parameters / hypotheses
- Analyst private information

Conclusion

# Setup

- Two agents: Decision-maker and analyst.
- The analyst observes a vector

$$X = (X_1, \ldots, X_{\bar{n}}),$$

where

$$X_i \overset{\text{iid}}{\sim} \text{Ber}(\theta).$$

- Analyst: Reports a subvector $X_I$ to the decision-maker, where

$$I \subset \{1, \ldots, \bar{n}\}.$$

- Decision-maker: Makes a decision

$$a \in \{0, 1\},$$

based on this report.

# Prior and objectives

- Common prior:
$$\theta \sim \text{Beta}(\alpha, \beta).$$
- Analyst's objective:
$$u^{\text{an}} = a - c \cdot |I|.$$
  $|I|$ is the size of the reported set,
  $c$ is the cost of communicating an additional component.
- Decision-maker's objective:
$$u^{\text{d-m}} = a \cdot (\theta - \underline{\theta}).$$
  $\underline{\theta}$ is a commonly known parameter.
  Minimum value of $\theta$ beyond which the decision-maker would like to choose $a = 1$.

# Timeline

1. The decision-maker commits to a decision rule
$$a = a(J, I, X_I).$$
2. The analyst reports a PAP
$$J \subseteq \{1, \ldots, \bar{n}\}.$$
3. The analyst next observes $X$, chooses $I \subseteq \{1, \ldots, \bar{n}\}$, and reports
$$(I, X_I).$$
4. The decision rule is applied and utilities are realized.

# Implementability

- Let $x$ denote values that the random vector $X$ may take.
- Reduced form mapping (statistical decision rule)

$$x \mapsto \bar{a}(x).$$

- $\bar{a}(x)$ is implementable
  if there exist mappings $I(x)$ and $a(I, x_I)$
  such that for all $x$

$$\bar{a}(x) = a(I(x), x_{I(x)}),$$

  and

$$I(x) \in \underset{I}{\operatorname{argmax}}\ a(I, x_I) - c \cdot |I|.$$

# Optimal implementable publication rules

- The latter is the incentive compatibility constraint, which implies
  1. $$I(x) \in \underset{I}{\operatorname{argmin}} \{|I| : \; a(I, x_I) = 1\}$$
     whenever $\bar{a}(x) = 1$, and $I(x) = \emptyset$ else.
  2. $$|I(x)| \leq 1/c$$
     for all $x$.
- Our agenda:
  - Find implementable mappings (decision rules) $\bar{a}(x)$
  - that maximize the expected decision-maker utility $E[u^{\text{d-m}}]$.

# Notation

- Successes among all components: $s(X) = \sum_{i=1}^{\bar{n}} X_i$.
  Successes among the subset $I$: $s(X_I) = \sum_{i \in I} X_i$.
- Maximal number of components the analyst is willing to submit:

$$\bar{n}^{PC} = \max\{n:\ 1 - cn \geq 0\} = \lfloor 1/c \rfloor .$$

- First-best cutoff for the decision-maker:

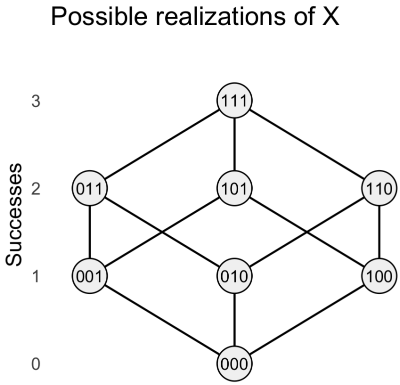
$$\underline{s}^*(n) = \min\left\{\underline{s}:\ E[\theta \big| s(X_{1,\ldots,n}){=}\underline{s}] \geq \underline{\theta}\right\} .$$

- Minimal cutoff for the decision-maker:

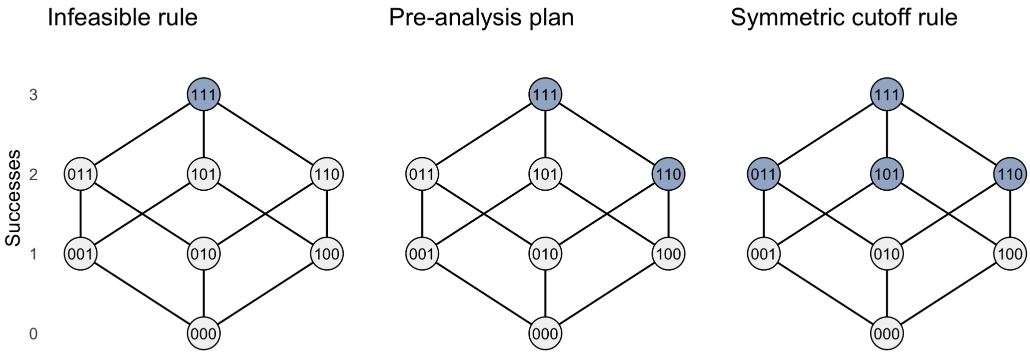
$$\underline{s}^{min}(n) = \min\left\{\underline{s}:\ E[\theta \big| s(X_{1,\ldots,n}){\geq}\underline{s}] \geq \underline{\theta}\right\} .$$

# A minimal example: $\bar{n} = 3$

- Suppose $\bar{n} = 3$.
  Possible realizations of $X$ form a cube.
- Vertical axis $=$
  number of successes $s(X)$.
- Suppose $\bar{n}^{PC} = 2$.
  Possible reports $(I, X_I) \cong$
  edges of the cube.
- Reduced form mappings $\bar{a}(x) \cong$
  set of nodes for which $a = 1$.



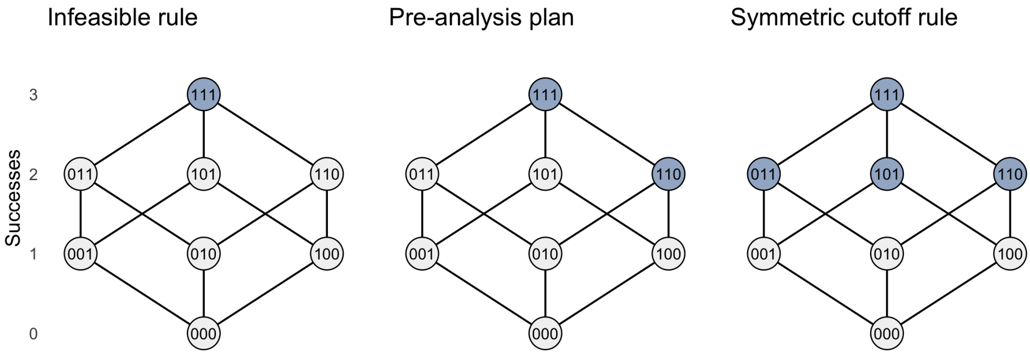
# A minimal example: $\bar{n} = 3$



## A minimal example: $\bar{n} = 3$
## Case I: Symmetric cutoff rule is optimal

- Suppose $\bar{n} = 3$, $\bar{n}^{PC} = 2$, and $\underline{s}^*(3) = 2$.
- The **unconstrained efficient** solution is given by
$$\bar{a}(X) = \mathbf{1}(s(X) \geq 2).$$
- This solution **can be implemented** by
$$a(I, X_I) = \mathbf{1}(s(X_I) \geq 2).$$
- **No PAP** is needed to implement this solution.

# A minimal example: $\bar{n} = 3$



# A minimal example: $\bar{n} = 3$
# Case II: PAP is optimal

- Suppose again that $\bar{n} = 3$, and $\bar{n}^{PC} = 2$. Suppose now

$$\underline{s}^*(3) = 3, \qquad \underline{s}^*(2) = 2$$

- The **unconstrained efficient** solution is given by

$$\bar{a}(X) = \mathbf{1}(s(X) = 3).$$

  There is **no** incentive compatible **implementation** of this solution.

- The **PAP** solution for $J = \{1, 2\}$,

$$a(J, I, X_I) = \mathbf{1}(I = \{1, 2\}, s(X_I) = 2),$$

  yields $E[u^{\text{d-m}}] > 0$, and is **constrained optimal**.

# Symmetric publication rules

- Denote $t(X_I) = |I| - s(X_I)$.
- Consider now, for general $\bar{n}$, symmetric rules of the form

$$a(s(X_I), t(X_I)),$$

**Lemma (Symmetrically implementable rules)**

*$\bar{a}(\cdot)$ is a reduced form publication rule that is implementable by such a symmetric rule iff it is of the form*

$$\bar{a}(X) = \mathbf{1}(s(X) \in \mathscr{S}),$$

*where $\mathscr{S}$ is a union of intervals of length at least $\bar{n} - \bar{n}^{PC}$.*

# Optimal symmetric rules

## Proposition (Optimal symmetric publication rule)

*The optimal reduced-form publication rule that is symmetrically implementable takes the form*

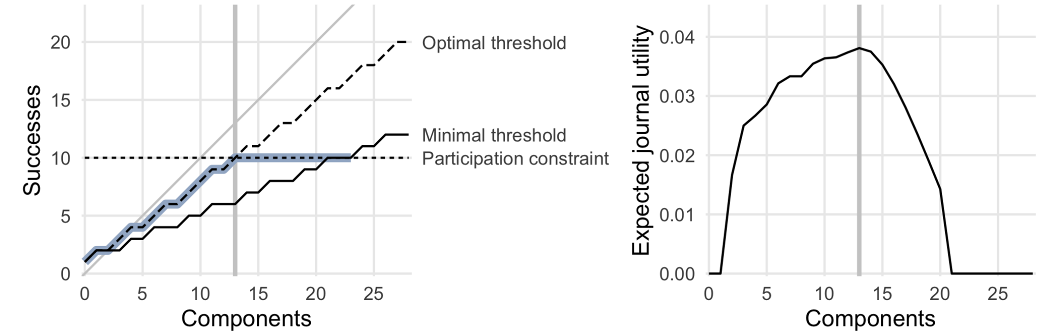
$$\bar{a} = \mathbf{1}(s(X) \geq \min(\underline{s}^*, \bar{n}^{PC})),$$

*if* $\bar{n}^{PC} \geq \underline{s}^{min}$*, and can be implemented by*

$$a = \mathbf{1}(s(X_I) \geq \min(\underline{s}^*, \bar{n}^{PC})).$$

*Otherwise the optimal publication rule is given by* $a \equiv 0$.

# Symmetric cutoff without PAP, uniform prior




If the number of components $\bar{n}$ is to the right of the maximum $\bar{n}^*$:

- PAPs increase decision-maker welfare
- by forcing the analyst to ignore all components $i > \bar{n}^*$.

# General implementable rules

### Lemma

*The implementable publication functions* $\bar{a}(x)$ *are exactly those that are of the form*

$$\bar{a}(x) = \mathbf{1}(x \in \cup_j C_{I_j, w_j}),$$

*for some set of* $\{(I_j, w_j)\}$*, where* $C_{I,w}$ *are the cylinder sets*

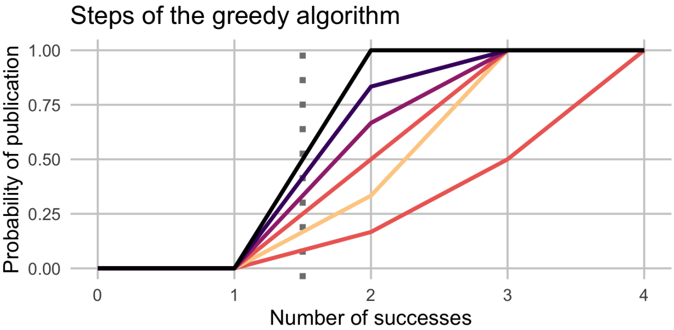
$$C_{I,w} = \{x : \ x_I = w\},$$

*and* $|I_j| = \bar{n}^{PC}$ *for all* $j$*.*

# (Approximately) optimal implementable rules

- Conceptually:
  - Optimal solution is given by the maximizer of $E[u^{\text{d-m}}]$
  - among the implementable reduced form rules $\bar{a}(x) = \mathbf{1}(x \in \cup_j C_{I_j, w_j})$.
- This is a hard combinatorial optimization problem!
  - Large number of possible unions of cylinder set.
  - No simplifying properties such as super-modularity.
- Alternatives:
  1. Restricted rules (e.g. cut-off rules with PAPs).
  2. Heuristic optimization algorithms (e.g. greedy optimization).

# Greedy algorithm for $\bar{n} = 4$, $\bar{n}^{PC} = 2$, $\underline{\theta} = 0.6$




- Each step increases the probability of publication.
- The first step is the PAP solution. The last step is the cutoff solution.
- Hue codes expected decision-maker utility. Step 2 yields the highest utility.

# Model variation I: Frequentist testing

- Setup same as in the baseline model, except for the decision-maker objective:
  - Consider the **null hypothesis** $\theta \leq \underline{\theta}$.
  - $\Rightarrow X_i$ is a valid test for the **significance level** $\underline{\theta}$.

- **First best** rule (uniformly most powerful test):
  Critical value $\underline{s}^{test}(\bar{n})$, $U \sim Uniform([0,1])$,

$$\bar{a}(X) = \mathbf{1}(s(X) + U \geq \underline{s}^{test}(\bar{n})).$$

- When $\underline{s}^{test}(\bar{n}) > \bar{n}^{PC}$, the first best is **not implementable**.
  In this case no cutoff rule exists that
  1. controls size, and
  2. has non-trivial power.

- **Second best**:
  Use PAP to restrict $\bar{n}$ to the largest value
  such that $\bar{a}(X)$ is implementable.

# Model variation II: Multiple parameters / hypotheses

- Setup same as in the baseline model, except for the decision-maker objective:

$$u^{\text{d-m}}(a) = a \cdot \sum_{i \in I} (\theta_i - \underline{\theta}),$$

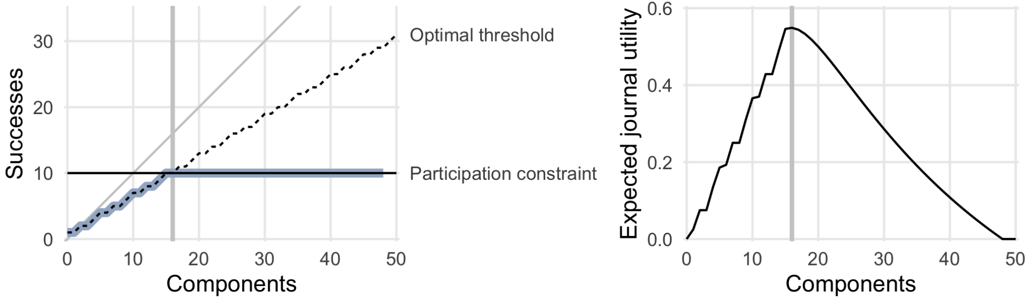
  where there are parameters $\theta_i$ for every $i$.

- Joint distribution of data and parameters:

$$\begin{aligned} X_i | \theta_1, \ldots \theta_{\bar{n}}, \bar{\theta} &\sim Ber(\theta_i) \\ \theta_i | \bar{\theta} &\sim Beta(m\bar{\theta}, m(1-\bar{\theta})) \\ \bar{\theta} &\sim \pi, \end{aligned}$$

- Selective reporting distorts inference.
  - For large $\bar{n}$ or $c$, the first best is not implementable,
  - but a PAP allows to implement the second best.

# Model variation II: Multiple parameters / hypotheses



# Model variation III: Analyst private information about signal validity

- Setup same as baseline model,
  except observability is determined by $W = (W_1, \ldots, W_{\bar{n}})$.

- Before choosing $J$, the decision-maker observes $W$.
  After choosing $J$, she observes the vector $X' = (W_1 X_1, \ldots, W_{\bar{n}} X_{\bar{n}})$,
  and reports a subvector of $X'$ to the decision-maker.

- $\bar{n}' = |W|$, is common knowledge.
  The decision-maker's prior over $W$ given $\bar{n}'$ is uniform
  over all permutations of the components $i$.

- Solutions are exactly the same as in the baseline model.
  except we need the decision-maker (not the decision-maker) to choose the PAP.

# Summary

- Single agent (statistical) decision theory can not rationalize PAPs.
- Mechanism design allows us to study implementable statistical decision rules.
- In our model, PAPs are optimal when
  1. there are many decision-maker degrees of freedom
  2. and communication costs are high.
- Variations of the baseline model: Qualitative conclusions are robust.
  1. Replacing the decision-maker objective by size and power of a statistical test.
  2. Multiple parameters or hypotheses.
  3. Analyst private information about signal validity.
  4. No decision-maker commitment.
  5. Ex-ante uncertainty about the available number of components $\bar{n}$.
  6. The decision-maker bears the communication cost.

Thank you!