Which findings should be published?

Alex Frankel      Maximilian Kasy

August 30, 2018
Introduction

• Not all empirical findings get published (prominently).
• Selection for publication might depend on findings.
  • Statistical significance,
  • surprisingness, or
  • confirmation of prior beliefs.
• This might be a problem.
  • Selective publication distorts statistical inference.
  • If only positive significant estimates get published, then published estimates are systematically upward-biased.
• Explanation of “replication crisis?”
• Ioannidis (2005), Christensen and Miguel (2016).
Introduction

Evidence on selective publication

• Data from Camerer et al. (2016), replications of 18 lab experiments in QJE and AER, 2011-2014.
  
  left  Histogram: Jump in density of z-stats at critical value.
  middle Original and replication estimates: More cases where original estimate is significant and replication not, than reversely.
  right Original estimate and standard error: Larger estimates for larger standard errors.

• Andrews and Kasy (2018): Can use replications (middle) or meta-studies (right) to identify selective publication.
Introduction
Reforming scientific publishing

• Publication bias motivates calls for reform: Publication should not select on findings.
  • De-emphasize statistical significance, ban “stars.”
  • Pre-analysis plans to avoid selective reporting of findings.
  • Registered reports reviewed and accepted prior to data collection.

• But: Is eliminating bias the right objective?
  How does it relate to informing decision makers?

• We characterize **optimal publication rules from an instrumental perspective**:
  • Study might inform the public about some state of the world.
  • Then the public chooses a policy action.
  • Take as given that not all findings get published (prominently).
Introduction
Key results

1. **Optimal** rules selectively **publish surprising findings.**
   In leading examples: Similar to two-sided or one sided tests.

2. But: Selective publication **always distorts inference.**
   There is a trade-off policy relevance vs. statistical credibility.

3. With **dynamics**: Additionally publish **precise null** results.

4. With **incentives**: Modify publication rule to **encourage more precise** studies.
Introduction
Example of relevance-credibility trade-off

• Suppose that there are many potential medical treatments tested in clinical trials.
• Most of them are ineffective.
• Doctors don’t have the time to read about all of them.
• Two possible publication policies:
  1. Publish only the most successful trials.
     • The published effects are systematically upward biased.
     • But doctors learn about the most promising treatments.
  2. Publish based on sample sizes and prior knowledge, but independent of findings.
     • Then the published effects are unbiased.
     • But doctors don’t learn about the most promising treatments.
Roadmap

1. **Baseline model.**
2. Optimal publication rules in the baseline model.
3. Selective publication and statistical inference.
4. Extension 1: Dynamic model.
5. Extension 2: Researcher incentives.
6. Conclusion.
Baseline model
Timeline and notation

<table>
<thead>
<tr>
<th>State of the world</th>
<th>( \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common prior</td>
<td>( \theta \sim \pi_0 )</td>
</tr>
</tbody>
</table>

**Study might be submitted**
- Exogenous submission probability: \( q \)
- Design (e.g., standard error): \( S \perp \theta \)
- Findings: \( X \sim f_{X|\theta,S} \)

**Journal decides whether to publish**
- Publication probability: \( D \in \{0, 1\} \)
- Publication cost: \( c \)
- Public updates beliefs: \( \pi_1 = \pi_1^{(X,S)} \) if \( D = 1 \)
  \( \pi_1 = \pi_1^0 \) if \( D = 0 \)

**Public chooses policy action**
- Utility: \( U(a, \theta) \)
- Social welfare: \( U(a, \theta) - Dc \)
Baseline model

Belief updating and policy decision

- Public belief when study is published: $\pi_1^{(X,S)}$.
  - Bayes posterior after observing $(X,S)$
  - Same as journal’s belief when study is submitted.
- Public belief when no study is published: $\pi_1^0$.
  Two alternative scenarios:
  1. Naive updating: $\pi_1^0 = \pi_0$.
  2. Bayes updating: $\pi_1^0$ is Bayes posterior given no publication.
- Public action $a = a^*(\pi_1)$ maximizes posterior expected welfare, $\mathbb{E}_{\theta \sim \pi_1}[U(a, \theta)]$.
  Default action $a^0 = a^*(\pi_1^0)$. 
Optimal publication rules

• Coming next: We show that ex-ante optimal rules, maximizing expected welfare, are those which ex-post publish findings that have a big impact on policy.

• Interim gross benefit $\Delta(\pi, a^0)$ of publishing equals
  • Expected welfare given publication, $E_{\theta \sim \pi}[U(a^*(\pi), \theta)]$,
  • minus expected welfare of default action, $E_{\theta \sim \pi}[U(a^0, \theta)]$.

• Interim optimal publication rule:
  Publish if interim benefit exceeds cost $c$.

• Want to maximize ex-ante expected welfare:

$$ EW(p, a^0) = E[U(a^0, \theta)] $$

$$ + q \cdot E[p(X, S) \cdot (\Delta(\pi_1^{(X, S)}, a^0) - c)] . $$

• Immediate consequence:
  Optimal policy is interim optimal given $a^0$. 
Optimal publication rules

Optimality and interim optimality

- **Under naive updating:**
  - Default action $a^0 = a^*(\pi_0)$ does not depend on $p$.
  - **Interim optimal** rule given $a^0$ is optimal.

- **Under Bayes updating:**
  - $a^0$ maximizes $EW(p, a^0)$ given $p$.
  - $p$ maximizes $EW(p, a^0)$ given $a^0$, when interim optimal.
  - These conditions are **necessary but not sufficient** for joint optimality.

- **Commitment does not matter** in our model.
  - Ex-ante optimal is interim optimal.
  - This changes once we consider researcher incentives (endogenous study submission).
Leading examples

- **Normal prior and signal**, normal posterior: 

\[ \theta \sim \pi_0 = \mathcal{N}(\mu_0, \sigma_0^2) \]

\[ X|\theta, S \sim \mathcal{N}(\theta, S^2) \]

- **Canonical utility functions:**
  1. Quadratic loss utility, \( \mathcal{A} = \mathbb{R} \):

\[ U(a, \theta) = -(a - \theta)^2 \]

  Optimal policy action: \( a = \) posterior mean.

  2. Binary action utility, \( \mathcal{A} = \{0, 1\} \):

\[ U(a, \theta) = a \cdot \theta \]

  Optimal policy action: \( a = 1 \) iff posterior mean is positive.
Leading examples

Interim optimal rules

- Quadratic loss utility: "Two-sided test." Publish if
  \[ |\mu_1^{(X,S)} - a^0| \geq \sqrt{c}. \]

- Binary action utility: "One-sided test." Publish if
  \[ a^0 = 0 \text{ and } \mu_1^{(X,S)} \geq c, \quad \text{or} \]
  \[ a^0 = 1 \text{ and } \mu_1^{(X,S)} \leq -c. \]

- Normal prior and signals:
  \[ \mu_1^{(X,S)} = \frac{\sigma_0^2}{s^2 + \sigma_0^2} X + \frac{s^2}{s^2 + \sigma_0^2} \mu_0. \]
Leading examples
Quadratic loss utility, normal prior, normal signals

• Optimal publication region (shaded).
  left  Axes are standard error $S$, estimate $X$.
  right Axes are standard error $S$, “t-statistic” $(X - \mu_0)/S$.

• Note:
  • Given $S$, publish outside symmetric interval around $\mu_0$.
  • Critical value for t-statistic is non-monotonic in $S$. 
Leading examples
Binary action utility, normal prior, normal signals

- Optimal publication region (shaded).
  - **left** Axes are standard error $S$, estimate $X$.
  - **right** Axes are standard error $S$, “t-statistic” $(X - \mu_0)/S$.

- **Note:**
  - When prior mean is negative, optimal rule publishes for large enough positive $X$. 
Generalizing beyond these examples

Two key results that generalize:

- **Don’t publish null results:**
  A finding that induces $a^*(\pi^I) = a^0 = a^*(\pi^0_1)$ always has 0 interim benefit and should never get published.

- **Publish findings outside interval:**
  Suppose
  - $U$ is supermodular.
  - $f_{X|\theta,S}$ satisfies monotone likelihood ratio property given $S = s$.
  - Updating is either naive or Bayes.

  Then there exists an interval $I^s \subseteq \mathbb{R}$ such that $(X, s)$ is published under the optimal rule if and only if $X \notin I^s$. 

Roadmap

1. Baseline model.
2. Optimal publication rules in the baseline model.
3. Selective publication and statistical inference.
4. Extension 1: Dynamic model.
5. Extension 2: Researcher incentives.
6. Conclusion.
Selective publication and inference

- Just showed:
  Optimal publication rules select on findings.
- But: Selective publication rules can distort inference.
- We show a stronger result:
  Any selective publication rule distorts inference.
- Put differently:
  **If we desire that standard inference be valid, then the publication rule must not select on findings at all.**
- Next two slides:
  1. Bias and size distortions,
  2. distortions of likelihood and of naive posterior,
  when publication is based on statistical significance.
Selective publication and inference
Distortions of frequentist inference.

- $X \mid \theta \sim \mathcal{N}(\theta, 1)$; publish iff $X > 1.96$.
  
  **Left** Bias of $X$ as an estimator of $\theta$, conditional on publication.
  
  **Right** Coverage probability of $[X - 1.96, X + 1.96]$ as a confidence set for $\theta$, conditional on publication.
Selective publication and inference
Distortions of likelihood and Bayesian inference.

- Same model.
  - Left: **Probability of publication** conditional on $\theta$.
  - Right: **Bayesian default belief** and naive default belief, for prior $\theta \sim \mathcal{N}(0, 4)$.
Selective publication and inference
Validity of inference is equivalent to no selection

For normal signals and prior support with non-empty interior, the following statements are equivalent:

1. Non-selective publication.  
   \( p(x, s) \) is constant in \( x \) for each \( s \).

2. Publication probability constant in state.  
   \( \mathbb{E}[p(X, S)|\theta, S = s] \) is constant over \( \theta \in \Theta_0 \) for each \( s \).

3. Frequentist unbiasedness.  
   \( \mathbb{E}[X|\theta, S = s, D = 1] = \theta \) for \( \theta \in \Theta_0 \) and for all \( s \).

4. Bayesian validity of naive updating.  
   For all distributions \( F_S \), the Bayesian default belief \( \pi^0_1 \) is equal to the prior \( \pi_0 \).
Selective publication and inference

Intuition and implications

• Sketch of proof:
  • Non-selective publication ⇒ the other conditions: immediate.
  • Constant publication probability ⇒ non-selective publication:
    Completeness of the normal location family.
  • Unbiasedness ⇒ constant publication probability:
    “Tweedie’s formula” and integration.

• Optimal publication if we require non-selectivity?

• Suppose
  • There are normal signals.
  • Updating is either naive or Bayesian.
  • The publication rule is restricted to not select on $X$.

Then there exists $\bar{s} \geq 0$ for which the optimal rule publishes a study if and only if $S \leq \bar{s}$.
Selective publication and inference

Optimal non-selective publication region

- For quadratic loss utility, normal prior, normal signals.
- Subject to the constraint that \( p(x, s) \) is restricted to not depend on \( x \).
Roadmap

1. Baseline model.
2. Optimal publication rules in the baseline model.
3. Selective publication and statistical inference.
4. **Extension 1: Dynamic model.**
5. Extension 2: Researcher incentives.
6. Conclusion.
A dynamic two-period model

- Period 1 as before, with study \((X_1, S_1)\), action \(a_1 = a^*(\pi_1)\).
- Now additionally: Period 2 study, always published.
- Independent estimate
  \[X_2|\theta, X_1, S_1 \sim F_{X_2|\theta}.\]
- Period 2 action \(a_2 = a^*(\pi_2)\).
- Social welfare
  \[\alpha U(a_1, \theta) - Dc + (1 - \alpha) U(a_2, \theta).\]
A dynamic two-period model
Quadratic loss utility, normal prior, normal signals, naive updating

- Optimal publication region (shaded).
- Note:
  - For $S$ small enough, publish even when $X = \mu_0$. 
A dynamic two-period model

General implications

- Publishing a precise (null) result in period 1 can help reduce mistakes in period 2.
- Holds under more general conditions, for normal signals:
  1. The benefit of publication is strictly positive whenever \( \pi_1^l \neq \pi_1^0 \).
  2. The benefit goes to 0 as either \( s_2 \to 0 \) or \( s_2 \to \infty \).
- Put differently:
  1. Even **null results that improve precision** are valuable to prevent future mistakes.
  2. This value disappears for
     a) very precise future information
        (won't make any mistakes either way), and
     b) very imprecise future information
        (effectively back to one-period case).
Researchers Incentives

- Thus far: study submission and design exogenous, random.
- Assume now that a researcher
  1. decides whether or not to submit a study,
  2. and picks a design $S$.
- Normal signals with standard error $S$.
- Researcher utility:
  1. Utility 1 from getting published,
  2. cost $\kappa(S)$ depending on design $S$.
- Expected researcher utility

  $$E_{\theta \sim \pi_0, X \sim N(\theta, S^2)}[p(X, S)] - \kappa(S).$$

- Outside option with utility 0.
- Journal faces
  1. participation constraint (PC) and
  2. incentive compatibility constraint (ICC).
• Journal objective as before, $U(a, \theta) - Dc$.
• Journal commits to publication rule $p(x, s)$ ex-ante. Commitment matters in this extension!
• Optimal publication rule subject to (PC) and (ICC)?
• Solution: Relative to baseline model, journal **distorts publication rule** in two ways
  • Reject imprecise studies (large $S$) – even if valuable ex post.
  • For low enough $S$, set interim benefit threshold for acceptance below $c$. 
Conclusion
Summary

- Eliminating selection on findings has costs as well as benefits. Important for reform debates!
- Key results:
  1. Optimal rules selectively publish surprising findings. In leading examples: Similar to two-sided or one sided tests.
  2. But: Selective publication always distorts inference. There is a trade-off policy relevance vs. statistical credibility.
  3. With dynamics: Additionally publish precise null results.
  4. With incentives: Modify publication rule to encourage more precise studies.
Conclusion

Outlook

Different ways of thinking about statistics / econometrics:

1. Making decisions based on data.
   - Objective function?
   - Set of feasible actions?
   - Prior information?

2. Statistics as (optimal) communication.
   - Not just “you and the data.”
   - What do we communicate to whom?
   - Subject to what costs and benefits?
     Why not publish everything? Attention?

3. Statistics / research as a social process.
   - Researchers, editors and referees, policymakers.
   - Incentives, information, strategic behavior.
   - Social learning, paradigm changes.

Much to be done!
Thank you!