Are Most Research Findings False?

. . . and what to do about it

Victor Gay

Toulouse School of Economics and IAST

Dinner Innovation December 2018

Plan

1. Basic Principles of Inference
2. Are Most Studies False?
3. The Need for Replications
4. Discussion

Plan

1. **Basic Principles of Inference**
   - **Statistical Significance**
   - Statistical Power
   - Prior Beliefs
2. Are Most Studies False?
3. The Need for Replications
4. Discussion

Basic Principles: Statistical Significance

- To test hypothesis H_1, suppose null hypothesis H_0 is true.
- E.g. $H_0 : \overline{X} = 0$, $H_1 : \overline{X} \neq 0$, where X is a random variable.
- Show data improbable under the null and reject H_0.

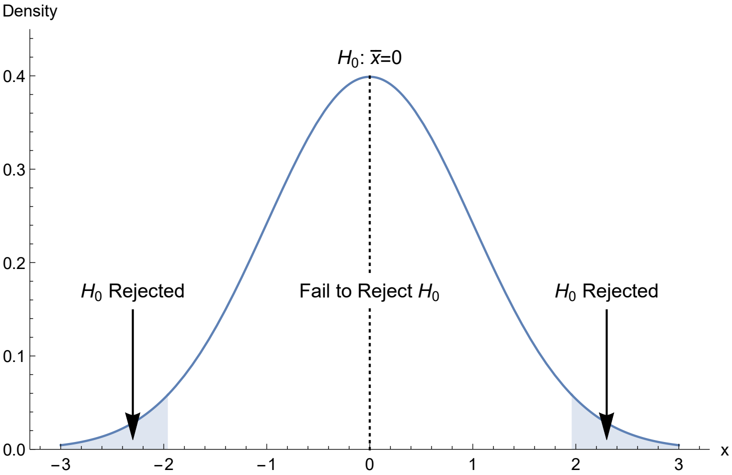
Basic Principles: Statistical Significance

- **p-value**:
  - Result is **statistically significant** if unprobable under H_0.
  - Probability of result equal or more extreme than observed under H_0.

Basic Principles: Statistical Significance

- **p-value**:
  - Result is **statistically significant** if unprobable under H_0.
  - Probability of result equal or more extreme than observed under H_0.
- **Significance level** α:
  - Threshold such that reject H_0 if $p \leq \alpha$.
  - Type I error (**false positive**): $\alpha = \Pr(p \leq \alpha | H_0)$.
  - Usually, $\alpha = 0.05$.

Basic Principles: Statistical Significance



Significance level: $\alpha = 0.05$.

Plan

1. **Basic Principles of Inference**
   - Statistical Significance
   - **Statistical Power**
   - Prior Beliefs
2. Are Most Studies False?
3. The Need for Replications
4. Discussion

Statistical Power

- Type II error (**false negative**): $\beta = \Pr(\texttt{fail reject } H_0 | H_0 \texttt{ false})$.
- Statistical power of a test: probability of rejecting H_0 when H_0 false.

  $\Longrightarrow$ Statistical power: $1 - \beta$.

Statistical Power

- Type II error (**false negative**): $\beta = \Pr(\texttt{fail reject } H_0 | H_0 \texttt{ false})$.
- Statistical power of a test: probability of rejecting H_0 when H_0 false.

  $\Longrightarrow$ Statistical power: $1 - \beta$.
- Factors affecting power of a test:
  - Effect size $\uparrow \Longrightarrow$ power $\uparrow$.
  - Sample size $\uparrow \Longrightarrow$ power $\uparrow$.

Statistical Power

- Type II error (**false negative**): $\beta = \Pr(\texttt{fail reject } H_0 | H_0 \texttt{ false})$.
- Statistical power of a test: probability of rejecting H_0 when H_0 false.

  $\Longrightarrow$ Statistical power: $1 - \beta$.
- Factors affecting power of a test:
  - Effect size $\uparrow \Longrightarrow$ power $\uparrow$.
  - Sample size $\uparrow \Longrightarrow$ power $\uparrow$.
- Rule of thumbs:
  - $1 - \beta = 0.80$.
  - Type I error 4-times as serious as type II error.

Error Types

	Null hypothesis H_0	
	True	False
Reject H_0	Type I error *false positive* α	Correct *true positive* $1-\beta$
Fail to reject H_0	Correct *true negative* $1-\alpha$	Type II error *false negative* β

Plan

1. Basic Principles of Inference
   - Statistical Significance
   - Statistical Power
   - **Prior Beliefs**
2. Are Most Studies False?
3. The Need for Replications
4. Discussion

Prior Beliefs

- Prior probability that the relationship is true i.e. H_0 should be rejected: π.
- Depends on:
  - Field
  - Scientific relationship
  - Prior research

Plan

Post-Study Probability

- Mechanics of statistical inference $\Longrightarrow$ many findings are false.
- Reliance on p-values $\Longrightarrow$ excessive number of **false positives**.

Post-Study Probability

- Notations:
  - n: number of scientific associations to be investigated.
  - π: fraction of true associations (or prior).
  - $1 - \beta$: power of test.
  - α: statistical significance, or false positive probability.

Post-Study Probability

- Notations:
  - n: number of scientific associations to be investigated.
  - π: fraction of true associations (or prior).
  - $1 - \beta$: power of test.
  - α: statistical significance, or false positive probability.
- Quantities of interest:
  - True associations: $\pi \cdot n$
  - False associations: $(1 - \pi) \cdot n$

Post-Study Probability

- Notations:
  - n: number of scientific associations to be investigated.
  - π: fraction of true associations (or prior).
  - $1-\beta$: power of test.
  - α: statistical significance, or false positive probability.
- Quantities of interest:
  - True associations: $\pi \cdot n$
  - False associations: $(1-\pi) \cdot n$
  - **True associations declared true:** $(1-\beta) \cdot \pi \cdot n$

Post-Study Probability

- Notations:
  - n: number of scientific associations to be investigated.
  - π: fraction of true associations (or prior).
  - $1 - \beta$: power of test.
  - α: statistical significance, or false positive probability.
- Quantities of interest:
  - True associations: $\pi \cdot n$
  - False associations: $(1 - \pi) \cdot n$
  - **True associations declared true:** $(1 - \beta) \cdot \pi \cdot n$
  - **False associations declared true:** $\alpha \cdot (1 - \pi) \cdot n$

Post-Study Probability

- Post-study probability (PSP):
  - PSP: probability that an association declared true is actually true.
  - PSP$= \Pr(H_0 \texttt{ false} | \texttt{reject } H_0)$.

Post-Study Probability

- Post-study probability (PSP):
  - PSP: probability that an association declared true is actually true.
  - PSP$= \Pr(H_0 \texttt{ false} | \texttt{reject } H_0)$.
- Using Baysian updating (Bayes rule).

$$\text{PSP} = \frac{\texttt{number of true associations declared true}}{\texttt{number of associations declared true}}$$

Post-Study Probability

- Post-study probability (PSP):
  - PSP: probability that an association declared true is actually true.
  - PSP$= \Pr\left(H_0 \texttt{ false} \middle| \texttt{reject } H_0\right)$.
- Using Baysian updating (Bayes rule).

$$\text{PSP} = \frac{\texttt{number of true associations declared true}}{\texttt{number of associations declared true}}$$

$$\text{PSP} = \frac{(1-\beta)\,\pi}{\underbrace{(1-\beta)\,\pi}_{\texttt{true positives}} + \underbrace{\alpha\,(1-\pi)}_{\texttt{false positives}}}$$

Post-Study Probability

PSP: probability that an association declared true is actually true.

$$\texttt{Post-Study Probability} = \frac{(1-\beta)\,\pi}{\underbrace{(1-\beta)\,\pi}_{\texttt{true positives}} + \underbrace{\alpha\,(1-\pi)}_{\texttt{false positives}}}$$

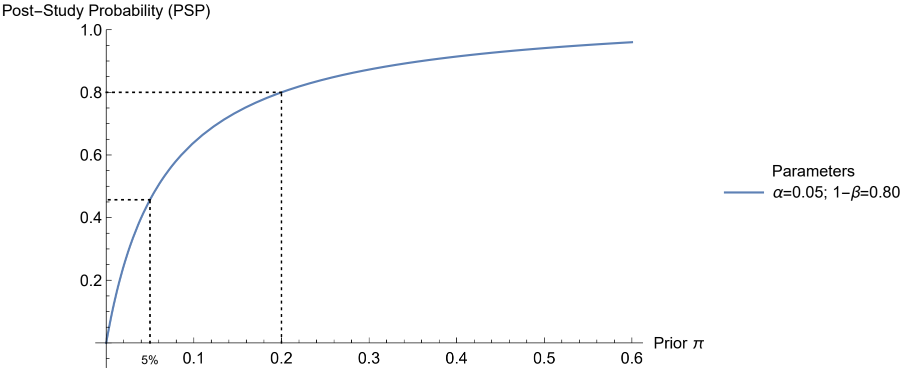
Comparative statics:

- Prior $\pi \uparrow \implies$ PSP $\uparrow$.
- Statistical significance $\alpha \downarrow \implies$ PSP $\uparrow$.
- Power $1 - \beta \uparrow \implies$ PSP $\uparrow$.

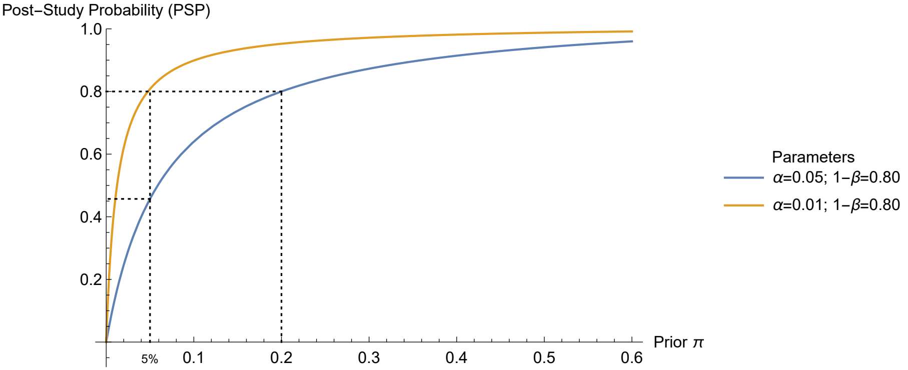
Plan

1. Basic Principles of Inference
2. **Are Most Studies False?**
   - Post-Study Probability
   - **What Can We Learn From One Study?**
   - What Can We Learn From Many Studies?
   - Aggravating Problems
3. The Need for Replications
4. Discussion

What Can We Learn From One Study?

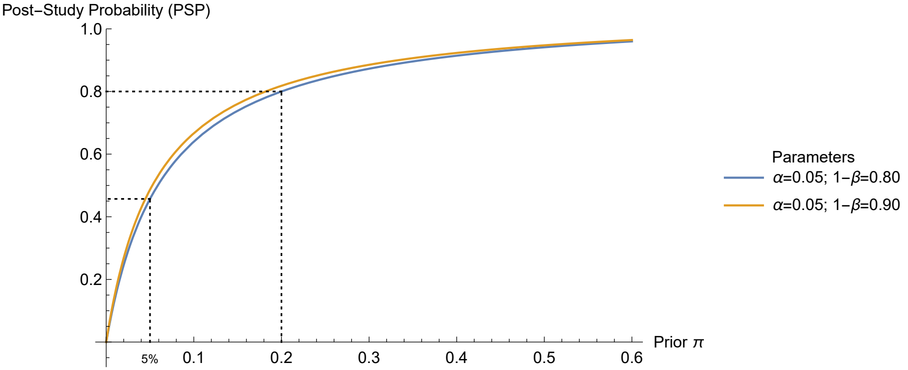


What Can We Learn From One Study?



Statistical significance $\alpha \downarrow \Longrightarrow$ PSP $\uparrow$

What Can We Learn From One Study?



Statistical power $1-\beta\uparrow \implies \text{PSP}\uparrow$

Plan

What Can We Learn From Many Studies?

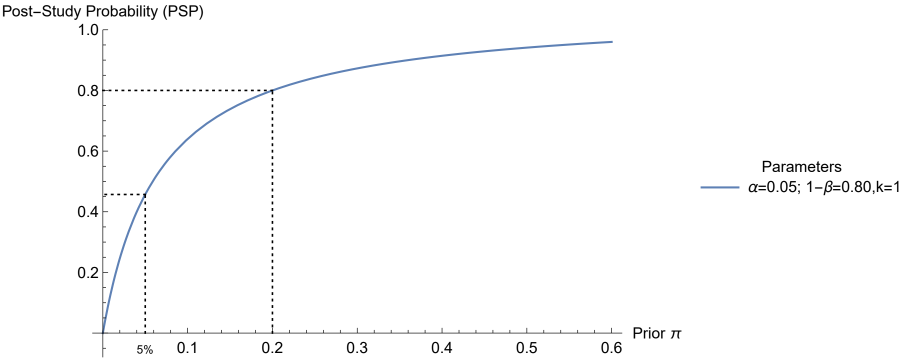
- Usually number of researchers simultaneously working $k > 1$.
- Two implications:
  - Number of true positives $\uparrow$.
  - Number of false positives $\uparrow$.
- But false positives $\uparrow$ faster than true positives. Formula

What Can We Learn From Many Studies?

- Usually number of researchers simultaneously working $k > 1$.
- Two implications:
  - Number of true positives $\uparrow$.
  - Number of false positives $\uparrow$.
- But false positives $\uparrow$ faster than true positives. Formula

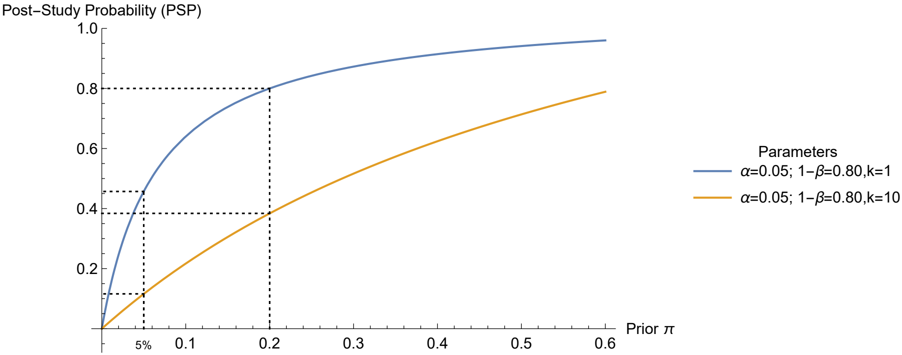
$\Longrightarrow$ Post-study probability $\downarrow$ when number of researchers $\uparrow$

What Can We Learn From Many Studies?



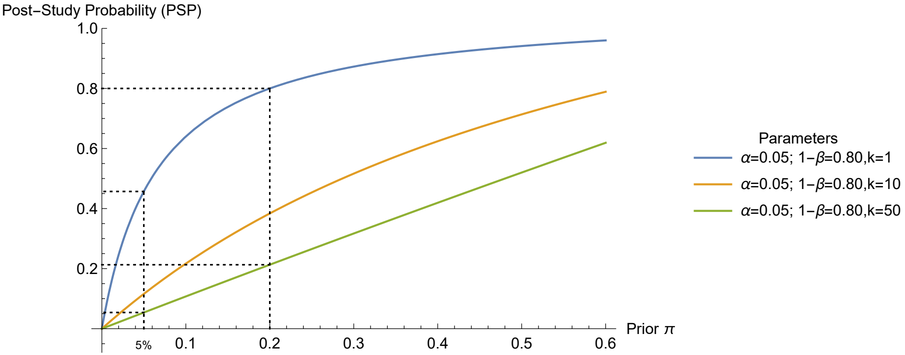
$k = 1$ researcher

What Can We Learn From Many Studies?



$k = 10$ researchers

What Can We Learn From Many Studies?



$k = 50$ researchers

Plan

1. Basic Principles of Inference
2. **Are Most Studies False?**
   - Post-Study Probability
   - What Can We Learn From One Study?
   - What Can We Learn From Many Studies?
   - **Aggravating Problems**
3. The Need for Replications
4. Discussion

Aggravating Problems

- Problems increasing the probability of false positives:
  - p-hacking.
  - Multiple testing.
  - Publication bias.
- Other problems:
  - Generalizability of lab behavior.
  - Scaling-up problem (population heterogeneity, compliance...).

Plan

1. Basic Principles of Inference
2. Are Most Studies False?
3. **The Need for Replications**
    - **What is a Replication?**
    - Benefits of Replications
    - Incentives for Replications
4. Discussion

What is a Replication?

- Replication tests:
  - Verification tests:
    - Original data, same sample, identical method.
    - Solves measurement errors, coding errors, data construction errors.

What is a Replication?

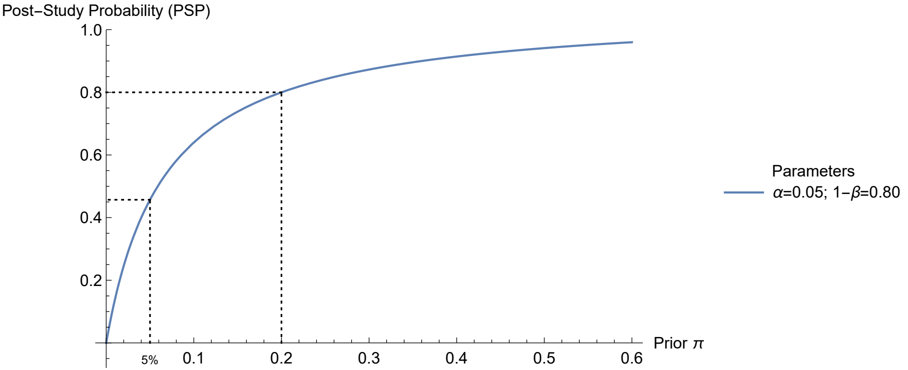
- Replication tests:
  - Verification tests:
    - Original data, same sample, identical method.
    - Solves measurement errors, coding errors, data construction errors.
  - Reproduction tests:
    - Resampling same population, identical method.
    - Solves sampling errors, low power.

What is a Replication?

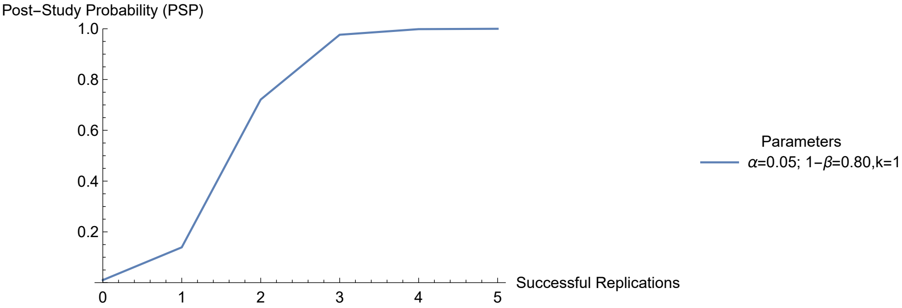
- Replication tests:
  - Verification tests:
    - Original data, same sample, identical method.
    - Solves measurement errors, coding errors, data construction errors.
  - Reproduction tests:
    - Resampling same population, identical method.
    - Solves sampling errors, low power.
- Robustness tests:
  - Reanalysis: same population, different method.
  - Extension: resampling, different method.

Plan

Benefits of Replications

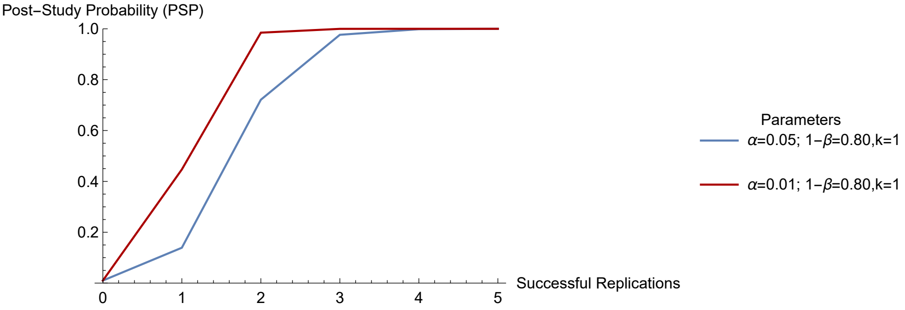


Benefits of Replications



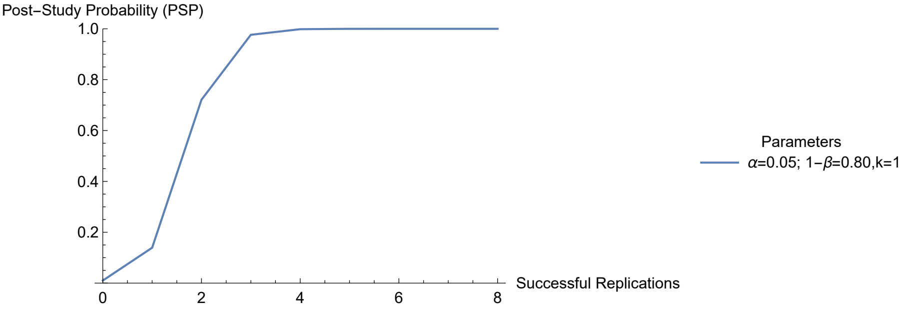
Starting prior: $\pi = 1\%$.

Benefits of Replications



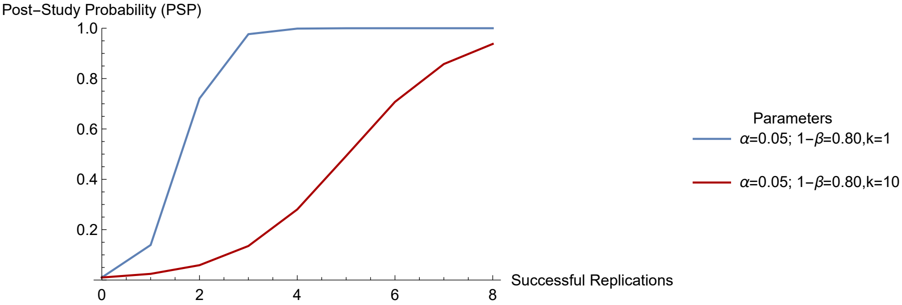
Starting prior: $\pi = 1\%$. $\downarrow$ statistical significance $\alpha \Longrightarrow \uparrow$ PSP

Benefits of Replications



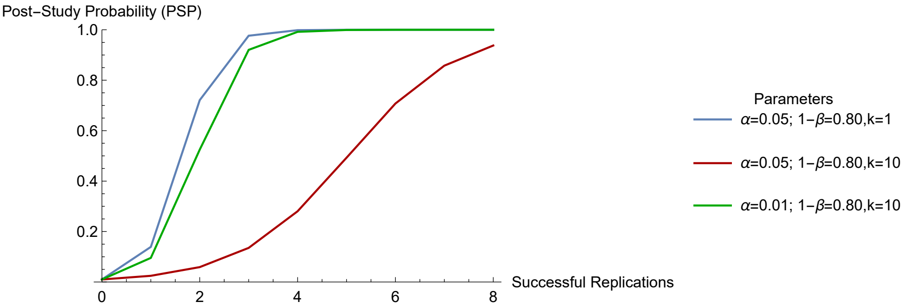
Starting prior: $\pi = 1\%$.

Benefits of Replications



Starting prior: $\pi = 1\%$. $\uparrow$ researchers $k \Longrightarrow \downarrow$ PSP

Benefits of Replications



Starting prior: $\pi = 1\%$. $\uparrow$ researchers $k \Longrightarrow \downarrow$ PSP

Plan

1. Basic Principles of Inference
2. Are Most Studies False?
3. **The Need for Replications**
    - What is a Replication?
    - Benefits of Replications
    - **Incentives for Replications**
4. Discussion

Incentives for Replications

- Very weak incentives for replications, except in specific fields.
    - Publication bias.
    - Career concerns...

Incentives for Replications

- Very weak incentives for replications, except in specific fields.
    - Publication bias.
    - Career concerns...
- Solutions?
    - Dedicated journals
    - Financial incentives
    - Co-authorship

Plan

1. Basic Principles of Inference
2. Are Most Studies False?
3. The Need for Replications
4. **Discussion**

Practitioners' Perspective

- Interpretation of new findings?
- Investment following findings?
- Incentives for replication?

References

- Benjamin et al. 2017. "Redefine Statistical Significance." *Nature Human Behavior*
- Ioannidis. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2(8):1418–22.
- Maniadis, Tufano, and List. 2014. "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects." *American Economic Review* 104(1):277–90.
- Moonesinghe, Khoury, and Janssen. 2007 "Most Published Research Findings are False, But a Little Replication Goes a Long Way." *PLoS Medicine* 4(2): 218–21.

What Can We Learn From Many Studies?

- Numerator:
  - Probability true positive: $1 - \beta$.
  - Probability no-one declares true positive: $[1 - (1 - \beta)]^k = \beta^k$.
  - Probability at least one declares true positive: $1 - \beta^k$.
- Denominator:
  - Probability false positive: α.
  - Probability no-one declares false positive: $(1 - \alpha)^k$.
  - Probability at least on declares false positive: $1 - (1 - \alpha)^k$.

$$\text{PSP}^{\texttt{comp}} = \frac{(1 - \beta^k)\ \pi}{(1 - \beta^k)\ \pi + [1 - (1 - \alpha)^k]\ (1 - \pi)}$$

Back