

Kursus Mandiri Metasains

Bagian 1:

Krisis Replikasi, Revolusi Kredibilitas, dan Pengantar Metasains

Rizqy Amelia Zein

- Dosen, Fakultas Psikologi, Universitas Airlangga
- Co-founder, #SainsTerbuka Airlangga 
- *Researcher-in-training*, Institute for Globally Distributed Open Research and Education (IGDORE)

Menghubungi saya?

✉ amelia.zein@psikologi.unair.ac.id

🐦 [@ameliazein](https://twitter.com/ameliazein)

🐙 [@rameliaz](https://github.com/rameliaz)

💻 <https://rameliaz.github.io>

Materi dalam paparan ini berlisensi © 4.0 (CC-BY-NC 4.0) dan tersedia di laman web kursus (<https://sainsterbukaua.github.io/meta-analysis/>). Kode tersedia secara terbuka di [repositori](#) 🐙 Sains Terbuka Airlangga



Krisis Replikasi

- Bayangkan anda sedang berjalan di Jalan Dharmawangsa dan sekilas melihat seseorang yang mirip dengan mantan anda.
- Pada pengamatan pertama, anda mungkin **tidak terlalu yakin** orang tersebut adalah mantan anda.
- Namun setelah **pengamatan berulang**, misalnya dengan memutar jalan dan melihatnya lagi yang **kedua, ketiga**, dan seterusnya, maka anda akan **lebih yakin** orang tersebut adalah/bukan mantan anda.
- Temuan riset akan **dinilai kredibel** apabila tim peneliti lain sukses mendapatkan temuan yang sama ketika **analisis diulang** atau **pengambilan data** diulang kembali (dengan protokol yang kurang lebih sama).
- Namun sayangnya, ada banyak sekali temuan penelitian yang lazim dipercayai, utamanya di Psikologi, gagal ditemukan kembali ketika dicoba-ulang oleh tim peneliti yang berbeda.
- ...sehingga menciptakan kondisi yang disebut juga dengan **krisis replikasi**.
- **Tidak hanya di Psikologi**, krisis ini juga ditemukan **di banyak disiplin ilmu lainnya**.

Mengapa hal ini bisa terjadi?

Gomez, Juristo, & Vegas (2010) menengarai ada setidaknya **lima penyebab** mengapa krisis ini bisa terjadi.

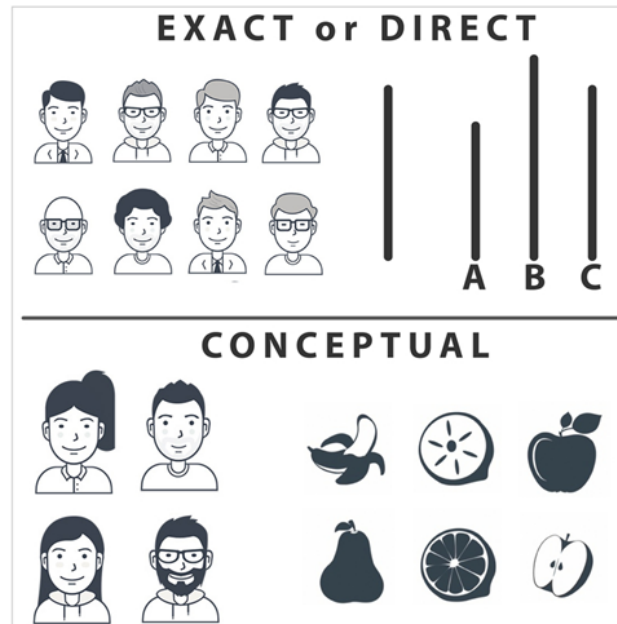
- **Lokasi penelitian** tempat dilakukannya penelitian coba-ulang (replikasi) berbeda dengan tempat dilakukannya penelitian pada penelitian asalnya.
- **Peneliti (eksperimenter)** yang melakukan penelitian bisa saja orang yang sama, orang yang samasekali berbeda, atau bahkan kombinasi antara peneliti yang sama dan berbeda dari penelitian asalnya.
- **Perangkat (aparatus)** penelitian yang digunakan, misalnya desain (protokol) penelitian, material studi, instrumen, dan perangkat lainnya yang digunakan di penelitian coba-ulang berbeda dengan penelitian asalnya.
- **Operasionalisasi** yang digunakan peneliti ketika melakukan coba-ulang berbeda dengan penelitian asalnya, dalam hal pengukuran variabel. Misalnya, penelitian coba-ulang dan penelitian asal menggunakan dua instrumen yang berbeda (PHQ-9 dan BDI) untuk mengukur variabel yang sama (depresi).
- Terakhir, perbedaan temuan penelitian mungkin dipengaruhi oleh **karakteristik populasi** yang berbeda.

Coba-ulang (*Replicable*) \neq Reka-ulang (*Reproducible*)

- ...meskipun kedua istilah ini sering digunakan bergantian.
- **Coba-ulang (*Replicable*)** ➡ desain (protokol) penelitian sama, pengambilan data dilakukan di tempat yang berbeda, dilakukan oleh tim peneliti yang berbeda sehingga menghasilkan data yang berbeda (independen) dari penelitian asalnya.
 - Data berbeda, peneliti berbeda.
- **Reka-ulang (*Reproducible*)** ➡ desain (protokol) penelitian sama, data penelitian dari studi asal dianalisis kembali oleh tim penelitian yang berbeda.
 - Data sama, tim peneliti berbeda.

Penelitian Replikasi

- Umumnya dapat dilakukan dengan **dua cara**¹ yang berbeda, yaitu:
 - *Direct replication*
 - *Conceptual replication*



¹ Diener, E. & Biswas-Diener, R. (2020). *The replication crisis in psychology*. In R. Biswas-Diener & E. Diener (Eds), Noba textbook series: Psychology. Champaign, IL: DEF publishers.

Replication Continuum					
	← Highly Similar			Highly Dissimilar →	
	Direct Replication			Conceptual Replication	
Design Facet	Exact Replication (All facets under researcher control are the same)	Very Close Replication (Procedure or physical setting is different)	Close Replication (IV or DV stimuli are different)	Far Replication (IV or DV operationalization or population is different)	Very Far Replication (IV or DV constructs are different)
Effect, Hypothesis	Same	Same	Same	Same	Same
IV Construct	Same	Same	Same	Same	Different
DV Construct	Same	Same	Same	Same	Different
IV Operationalization	Same	Same	Same	Different	
DV Operationalization	Same	Same	Same	Different	
Population (e.g., age)	Same	Same	Same	Different	
IV Stimuli	Same	Same	Different		
DV Stimuli	Same	Same	Different		
Procedural Details	Same	Different			
Physical Setting	Same	Different			
Contextual Variables	Different				
⋮	⋮				

Fig. 1. Taxonomy for classifying a replication study's methodological similarity to an original study. "Same" indicates that the design facet in question is the same as in the original study, and "different" indicates that it is different. IV = independent variable; DV = dependent variable. "Population" refers to major population characteristics, such as age and whether the sample is drawn from the community or a special clinical population. Procedural details are minor experimental particulars (e.g., task instructions, font, font size). Contextual variables are design facets beyond a researcher's control (e.g., history, culture, language).

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). [A unified framework to quantify the credibility of scientific findings](#). *Advances in Methods and Practices in Psychological Science*, 1(3), 389-402.

Krisis Replikasi di Psikologi

Journal	% Findings Replicated
Journal of Personality and Social Psychology: Social	23
Journal of Experimental Psychology: Learning, Memory, and Cognition	48
Psychological Science, social articles	29
Psychological Science, cognitive articles	53
Overall	36

Diener, E. & Biswas-Diener, R. (2020). [The replication crisis in psychology](#). In R. Biswas-Diener & E. Diener (Eds), Noba textbook series: Psychology. Champaign, IL: DEF publishers.

Reproducibility Project: Psychology (RP:P)¹

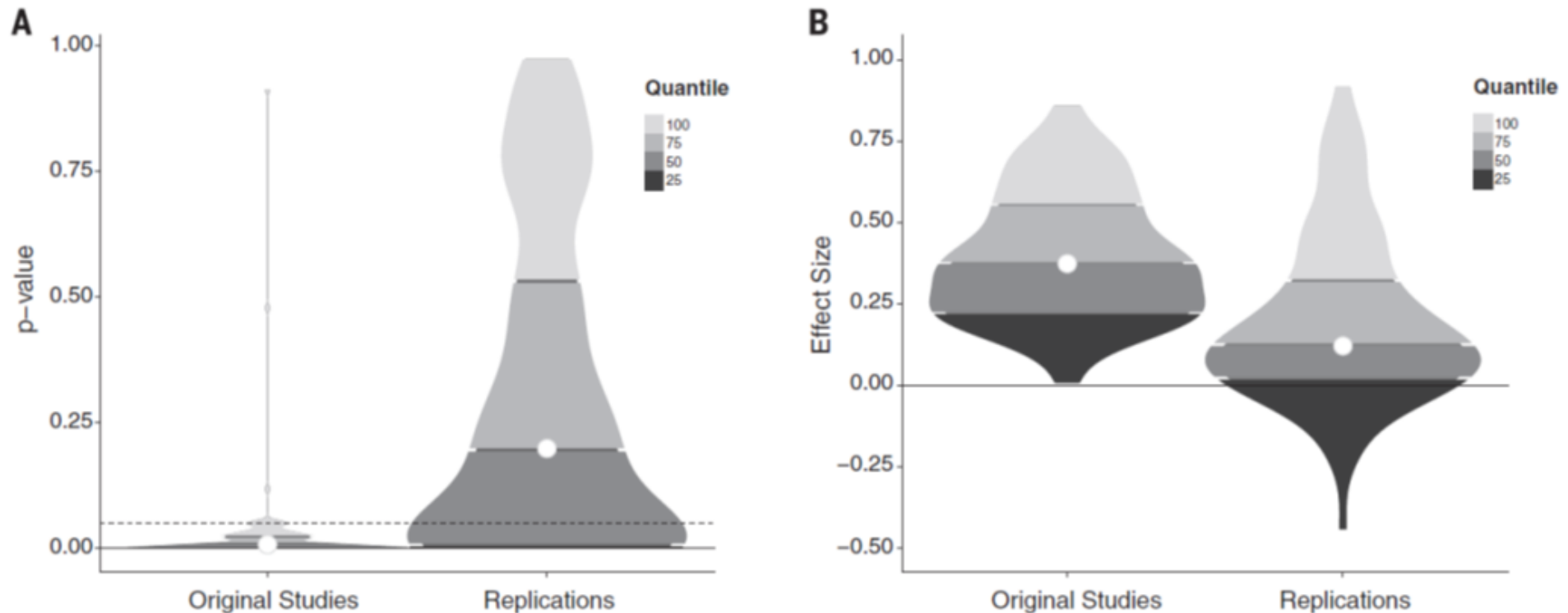


Fig. 1. Density plots of original and replication *P* values and effect sizes. (A) *P* values. (B) Effect sizes (correlation coefficients). Lowest quantiles for *P* values are not visible because they are clustered near zero.

¹ Open Science Collaboration. (2015). [Estimating the reproducibility of psychological science](#). *Science*, 349(6251), aac4716.

Kasus Fluoxetine



Kasus Fluoxetine

- Fluoxetine adalah obat antidepresan yang sangat populer dan menghasilkan penjualan hingga miliaran dolar US hanya di Amerika Serikat saja.
- Namun, sebuah **studi meta-analisis** cenderung amat kecil bahkan mengarah ke *placebo*.
- Para ahli **sepakat menyimpulkan** bahwa Prozac tidak bermanfaat untuk merawat pasien dengan gejala depresi.
- Mengapa ini bisa terjadi? Salah satu penyebabnya adalah **bias publikasi**, yaitu peneliti hanya **menerbitkan temuan penelitian yang positif** saja sehingga menghasilkan kesimpulan yang keliru.

Kasus Fluoxetine

Ioannidis (2008) menyimpulkan bahwa riset klinis obat-obat antidepresan cenderung menyuguhkan hasil yang keliru karena:

- Kesalahan peneliti dalam menginterpretasi **signifikansi statistik (p -value)**
- Protokol (desain) penelitian yang terdistorsi atau mungkin dimanipulasi
- *Sampling bias*
- Peneliti melakukan studi *follow-up* tidak lama setelah intervensi diberikan
- *Selective* atau *distortive report*, yaitu kondisi ketika peneliti melaporkan **hanya hasil positif** dan mengabaikan informasi atau temuan lainnya yang kontradiktif dengan hipotesis penelitian yang diajukan ➔ **bias konfirmasi**

Kasus *Psi (precognition)*

- [1] Bayangkan apabila saya meminta anda untuk mengingat sebuah daftar kata yang cukup panjang
- [2] Kemudian saya meminta anda untuk menyebutkan kata-kata tersebut
- [3] Setelah itu, saya minta anda untuk mengetik sejumlah kata, yang saya pilih secara acak dari daftar kata yang sebelumnya saya tunjukkan kepada anda
- [4] Hal menarik mulai terjadi, ternyata anda (secara signifikan) mengingat lebih baik (di tahap [2]) kata-kata yang saya minta untuk anda ketik (di tahap [3])
- Wow saya dapat membuktikan bahwa anda punya kemampuan **memprediksi apa yang terjadi di masa depan!** ✨

Hah, apa?



Kasus *Psi (precognition)*

- Eksperimen ini sungguh ada dan berdasarkan **penelitian Psikologi** yang dilakukan seorang ilmuwan Psikologi Sosial yang sangat dihormati (dulunya), yaitu Daryl Bem dari Stanford University.
- Dari penelitian tersebut, 9 dari 10 studi menunjukkan *p-value* yang signifikan ($p < .05$) dengan rerata besaran efek yang cenderung moderat (Cohen's $d=0.22$).
- Beberapa peneliti penasaran mencoba-ulang eksperimen tersebut, misalnya **Ritchie, dkk. (2012)** dan **Galak, dkk. (2012)**, bahkan Bem sendiri mencoba-ulang bersama timnya. Namun hasilnya, **nihil**, temuan penelitian yang sama tidak ditemukan kembali.

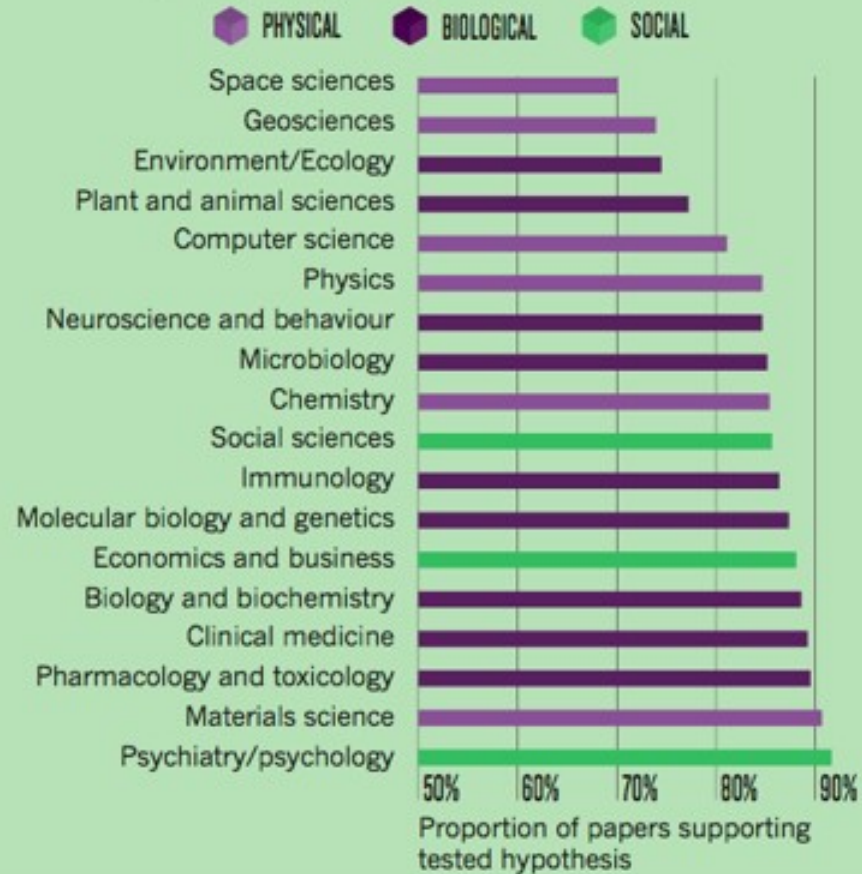
Kok bisa sih? 🤖

- *Selective reporting* ➡ Bem mungkin melakukan eksperimen berkali-kali, namun hanya melaporkan temuan penelitian yang sesuai dengan keinginannya saja.
- Tidak melaporkan semua kondisi dalam eksperimennya.
- Mengukur beberapa variabel dependen, namun hanya melaporkan yang hasilnya signifikan saja.
- Terlalu murah hati ketika membulatkan *p-value*.
- Bisa jadi, Bem memodifikasi hipotesis penelitiannya setelah melihat data ➡ *hypothesizing after the results are known*
- Membuang data *outlier*.
- Menghentikan pengambilan data ketika *p-value* sudah < 0.05 (*optional stopping*).
- ...bisa jadi Bem hanya sedang beruntung 😊



ACCENTUATE THE POSITIVE

A literature analysis across disciplines reveals a tendency to publish only 'positive' studies — those that support the tested hypothesis. Psychiatry and psychology are the worst offenders.



Yang menjadi masalah... ⚡

- Apakah kita dapat mempercayai informasi atau temuan penelitian yang disajikan dalam buku-buku teks?
- Yang mana informasi yang reliabel? Mana yang tidak? Bagaimana menentukannya?
- Lebih buruk lagi, apakah ilmu yang kita pelajari **benar-benar saintifik**?
- Kalau **sebagian besar temuan penelitian adalah *false-positive***, apa yang membuatnya **berbeda dengan berita palsu**?

Sekarang, saatnya Revolusi Kredibilitas!

- Dengan mengadopsi transparansi, kecermatan ilmiah mudah diperoleh dan dapat meningkatkan kredibilitas temuan penelitian.
- Memberikan ruang yang cukup bagi proses koreksi diri (*self-correction*), tidak hanya penemuan (*discovery*).
- Hasil negatif (*null findings*) **sama pentingnya** dengan hasil positif.
- Membuka akses pada material dan luaran riset akan menghasilkan **dampak yang lebih besar**.
- Mulai populer disiplin ilmu **Metasains** (*research-on-research*) yang memungkinkan peneliti untuk menyelidiki apa yang perlu dikoreksi dari agenda dan proses riset.

Metasains (Ioannidis, dkk., 2015)

- ..merupakan "disiplin ilmu yang bertujuan untuk **mengevaluasi dan meningkatkan kualitas praktik meneliti** yang termasuk mencakup area tematik seperti **metode, pelaporan, reka-ulang (*reproducibility*), evaluasi, dan insentif**. Singkatnya, bagaimana **melakukan, melaporkan, memverifikasi, mengoreksi, dan memberikan penghargaan** atas kerja saintifik.."
- Bisa menggunakan desain penelitian yang umum digunakan, bisa juga dengan pendekatan *evidence synthesis* ➡ yang kedua akan lebih banyak dibahas dalam kursus ini.

Meta-research area	Specific interests (nonexhaustive list)
Methods: "performing research"—study design, methods, statistics, research synthesis, collaboration, and ethics	Biases and questionable practices in conducting research, methods to reduce such biases, meta-analysis, research synthesis, integration of evidence, crossdesign synthesis, collaborative team science and consortia, research integrity and ethics
Reporting: "communicating research"—reporting standards, study registration, disclosing conflicts of interest, information to patients, public, and policy-makers	Biases and questionable practices in reporting, explaining, disseminating and popularizing research, conflicts of interest disclosure and management, study registration and other bias-prevention measures, and methods to monitor and reduce such issues
Reproducibility: "verifying research"—sharing data and methods, repeatability, replicability, reproducibility, and self-correction	Obstacles to sharing data and methods, replication studies, replicability and reproducibility of published research, methods to improve them, effectiveness of correction and self-correction of the literature, and methods to improve them
Evaluation: "evaluating research"—prepublication peer review, postpublication peer review, research funding criteria, and other means of evaluating scientific quality	Effectiveness, costs, and benefits of old and new approaches to peer review and other science assessment methods, and methods to improve them
Incentives: "rewarding research": promotion criteria, rewards, and penalties in research evaluation for individuals, teams, and institutions	Accuracy, effectiveness, costs, and benefits of old and new approaches to ranking and evaluating the performance, quality, value of research, individuals, teams, and institutions

doi:10.1371/journal.pbio.1002264.t001

Terima kasih banyak! 😊



Paparan disusun dengan menggunakan  *package* **xaringan** dengan *template* dan *fonts* dari **R-Ladies**.

Chakra dibuat dengan **remark.js**, **knitr**, dan **R Markdown**.