Replicability of predictions across studies: challenges and opportunities

Giovanni Parmigiani

Discussant: Gianmarco Altoè

Department of Developmental and Social Psychology
University of Padova

22 September 2022, Padova

Statistical methods and models for complex data

800 years of research to understand a complex world

Avoiding confusion between reproducibility and replicability

Reproducibility:
Obtaining *consistent* results using **the same input data**; computational steps, methods, and code. This definition is synonymous with "computational reproducibility."

Replicability:
Obtaining *consistent* results across studies aimed at answering the same scientific question, **each of which has obtained its own data**.

PLEASE NOTE:
As G. P. pointed out, much effort has been spent on studying the **replicability of results** *(usually coming from hypothesis testing), while much less work has addressed the relevant issue of the* **replicability of predictions** *(usually coming from machine learning).*

Source: *National Academies of Sciences, Engineering, and Medicine. (2019). Reproducibility and replicability in science.*

The crisis of replicability of results

Ioannidis, J. P. (**2005**).
Why most published research findings are false. *PLoS medicine*

Open Science Collaboration (2015). Estimating the reproducibility (**?!**) of psychological science. *Science*

- 100 published effects
- mean effects size in the replications was 50% smaller than in the original studies
- success rate of replication was 36%

Errington, T. M., et al (2021). Investigating the replicability of preclinical cancer biology. *Elife*

- 158 published effects
- median effect size in the replications was 85% smaller than in the original studies
- success rate of replication was 46%

PLEASE NOTE:

- Tools for assessing replicability and their interpretation are still hotly debated
- Studying the reasons of non-replicability is fundamental for science

How to increase replicability

Among other proposals, a good suggestion is to:

"Increase Meta-Analytic reasoning among researchers"
or, in the words of G. P.
"Use multiple studies (for training)"

In some sense, also **Galileo Galilei** who worked at our University around 1600 claimed that *"to evaluate your hypotheses multiple experiments are needed"*

Replicability of complex results

- **Cross-Study Replicability in Cluster Analysis**
  - A general and flexible framework to quantify replicability, *globally* (for the whole sample) and *locally* (for individual clusters)
  - Masoero, L., Thomas, E., Parmigiani, G., et al (preprint, 2022)
- **Multi-study factor analysis**
  - A novel class of [exploratory] factor analysis methodologies for the *joint* analysis of multiple studies (MSFA)
  - The goal is to separately identify and estimate:

    - common [*replicable*] factors shared across studies

    - study-specific [*unreplicable*] factors
  - De Vito, R., Bellio, R., Trippa, L., & Parmigiani, G. (2019)

Replicability of predictions

- **Shift from:**
  Specific algorithms trained and validated within a single study

- **to:**
  Generalist (and replicable) algorithms (i.e., Cross-Study Learners) trained and validated using multiple studies

Bernau, C., Riester, M., Boulesteix, A. L., Parmigiani, G., et al (2014).
Patil, P., & Parmigiani, G. (2018).

... and finally, 2 questions ;)

- Considering your work on the replicability of clusters and factors,
  what do you think about evaluating replicability of *factor-mixture analysis* (where factors and clusters are simultaneously estimated)?

- Regarding Cross-Study Learners,
  what are the most important things to take into account to define an efficient *training studies subsetting strategy*?

, gianmarco.altoe@unipd.it