

Replication Crisis from Previous Experiences to the Nursery

Del Olmo-De Dios MV¹ and Barrós-Loscertales A^{2*}

¹Rafalafena Health Center, Castellón, Spain

²Department of Basic, Clinical and Psychobiology Psychology, Jaume I, Castellón University, Spain

Perspective article

Received date: 04/11/2019;

Accepted date: 19/11/2019;

Published date: 25/11/2019

*For Correspondence:

Alfonso Barrós-Loscertales, Faculty of Health Sciences, Avd. Vicente Sos Baynat s/n. 12071. Castellón de la Plana, Spain.

Tel: +34964387661

E-mail: barrós@uji.es

Keywords: Reproducibility, Evidence-based treatment interventions, Replication crisis

ABSTRACT

Introduction: Application of evidence-based treatment intervention is expected to be based on a critical review of scientific literature.

Aim: Our aim is to describe the main issues raised by the so-called “reproducibility crisis” in the context of nursery research.

Background: Reproducibility is a key issue for science. When clinical practitioners look for an answer in the scientific literature for evidence-based interventions, they can be stuck in their aims. The fact of considering a single study as enough evidence for intervention, or the opposite, lack of effects, or agreement, in the effects between several studies, may discourage them.

Sources of evidence: In this manuscript, we review the main issues and resources related to the problem of replication in science for nursery research and practice contexts.

Discussion: The review begins with a description of the bias that may affect any research into the research community. Second, it describes the main issues related to the methodological consideration that has been the target of analyzing the replication problem.

Conclusion: Global policies and managing systems in nursery can improve their cost-effectiveness if they do an appropriate analysis of the issues related to the so call replication crisis in health science.

INTRODUCTION

Increasing the reliability and efficiency of research in health sciences will improve the quality of the results and their interpretation from the basic to the applied scientific literature by accelerating discovery and improving health interventions. The replicability crisis in the field of social psychology might be dated 2010-2012^[1], and quickly spread to other social and life science disciplines, including medicine and other research fields, and involved differences in reproducibility rates and effect sizes per discipline or journal^[2]. This manuscript summarizes those measures that may

help to optimize scientific progress in both the nursery and health science fields in general from notes taken from other fields, such as psychology, neuroscience, medicine and nursery. Particularly, those related to methods, evaluation, reporting, reproducibility and dissemination in order to increase information for researchers, practitioners and policymakers and agreements on the biases, limitations, potential improvements and current initiatives to improve scientific research in nursery. These changes are partly possible thanks to the evolution of information and communication technologies. We herein review a series of measures that others have suggested to improve research efficiency and the robustness of scientific findings^[3] by thinking particularly about their application in the nursery field.

Research strategies and practices can always improve to maximize the research community's efficiency and applicability. Replication of experiments is essential for science to advance. Meanwhile, we fight against our own thought full of subjectivity that promotes human creativity when faced with the threat of bias. Thus we tend to see an event as having been predictable only after it has occurred, i.e., hindsight bias; to focus on evidence that falls in line with our expectations, i.e., confirmation bias; or we tend to see patterns in random data, i.e., apophenia. All these biases involve the over interpretation of noise by understanding noise as the random facts that cannot be attributed to any cause of interest. Another bias may be more easily appraised, such as selection bias, information bias, performance bias, attrition bias, detection bias, and reporting bias or confounding. For example, it can be difficult to note the acceptance of outcomes that fit expectations as being appropriate during data analyses. Moreover, unexpected outcomes may be attributed to suboptimal designs or analyses or, otherwise, hypotheses may be reported with neither an indication nor recognition of their post hoc origin (so called Harking). Therefore, we need to prevent researchers from our own bias which, prompted by our scientific enthusiasm, may allow us to see patterns in noise^[3]. Several institutions offer formation, information that comes in the form of documents and checklists for identifying these biases and for guiding critical reading when reviewing the literature with the objective of translating research evidence into practice (Joanna Briggs Institute, <http://joannabriggs.org/>; Patient-Centred Outcomes Research Institute, <https://www.pcori.org/>). Other institutions directly oversee the integrity of research activities (e.g., Office of Research Integrity, ORI, <https://ori.hhs.gov/>; Standing Committee on Conflict of Interest, Scientific Misconduct and Ethical issues, ColME, <https://erc.europa.eu/erc-standing-committees/conflict-interests-scientific-misconduct-and-ethical-issues>; European Network of Research Integrity Offices, ENRIO, <http://www.enrio.eu/>).

The critical reading of any scientific report for clinicians interested in improving their practice based on evidence-based treatments is a personal and social responsibility. Their readings should depend on their interest in and concerns about the topic, as well as specific questions that may better guide them when looking for answers in the literature review. When they look, a first recommendation may be to look at the conflict of interest about the effectiveness of the intervention. A first approach, that of identifying existing primary research studies on the field of intervention, including systematic reviews and meta-analyses, may lead to a first informed read in our first approach. When usually involved in literature reading for the effectiveness of evidence-based interventions, other details considered in the experimental design are of interest. The following are also important: the specifics of diverse groups of patients studies (e.g., socio-demographic characteristics, severity of diseases, comorbidity or co-existing diseases), how the intervention of interest works and to what other interventions of interest can they be compared. Interestingly, identifying the relevance of the different outcomes from the tested intervention and its association with our practical interest and common clinical practices may lead to a careful reading of the applied procedures, approaches and measurement instruments. In this sense, a critical appraisal should include several issues raised in this review as to errors in the design, conduct and analysis of quantitative studies that may impact the reliability and validity of any study.

ISSUES AND BIAS

Bias

Ioannidis^[4] defined a bias to be the combination of design, data, analysis and presentation factors that may produce research findings when they should not be produced. As noted by Ioannidis^[4], this bias should not be confused with the chance probability of a finding that is false or true, but not correctly identified, which may take place even when previous factors are perfect. Bias is to be avoided by scrupulous design and its application, sophisticated techniques and procedures that reduce measurements errors.

Cognitive bias

A suggested solution for a cognitive bias is blinding participants, data collectors and data analyses from research aims. For example, during data acquisition, the identity of the experimental conditions should be blind to those research participants and researchers who apply treatments. The use of independent methodological support is established in areas like clinical trials given the well-understood financial conflicts of interest in conducting these studies. Methodological support from an independent single methodologist or a group of methodologists with no personal investment in a research topic may help in the design, monitoring, analysis or interpretation of research outcomes^[3].

These are suggested as an effective solution to mitigate self-deception and unwanted biases. Likewise, pre-registrations of studies, including study design, the analysis plan and the expected primary outcomes, are an effective way to avoid biases because data may not exist or, at least, outcomes may be unknown. Pre-registration protocols for randomized controlled trials in clinical medicine are a standard practice, but one that is not free of replication crisis effects^[5,6]. Interestingly, some journals favour manuscript publication if the study is already pre-registered in the journal section for this purpose (see <https://cos.io/rr/#journals> for a list of journals included in the Registered Report initiative of the Centre for Open Science). Therefore, pre-registration is advised for research plans before collecting data, or at least before analyzing data. There are many online services that support pre-registration, such as ClinicalTrials.gov, the AEA Registry, EGAP, the WHO Registry Network, and the well-known Open Science Framework (OSF).

Publication bias

Studies that obtain positive and novel results are more likely to be published than those that replicate previous results or obtain negative ones. This fact leads to changes in the outcomes of interest in a study depending on the observed results. Flexible analyses of multiple or multidimensional variables favour this outcome switching^[3] by leading to an intentional or unintentional selection of outcomes showing statistically significant results as those of interest. It should be taken into account that when we measure lots of variables (e.g., questionnaires, physiological variables, clinical scales, etc.) some of these variables are likely to give a positive result merely by random chance, which is known as a false-positive. Thus if a study plan and the analysis are pre-registered, the outcome to be considered is pre-specified and is much less likely to report a false-positive result. Moreover, if an outcome is not pre-specified, but reported, the authors must declare the timing and reasons for the change.

METHODOLOGICAL TRAINING

Appropriate and accurate experimental and statistical designs are essential for research validity and reliability. Therefore, knowing the importance of including control conditions (or groups); blinding, randomization and the counterbalancing of research treatments is basic; sample sizes big enough to replicate a true finding; the correct application of p-values, statistical power or effect size for conducting and interpreting research. Likewise, junior and senior researchers need continuous methodological education given the constant review of methodological good practices. Senior researchers are subject to biases introduced by adopting research strategies that worked in the past, but might be currently inappropriate. Afterward, information flows from their supervisor to their mentors, who learn these similar strategies and, if they are not encouraged to update and self-judge their application, they are prone to apply non updated designs to their research.

Hence it is important to be clear about the limitation of common research practices in clinical settings. The need of a control group is not usually considered in longitudinal studies. However, if we do not include a control group or a control period in the same research group, our conclusion from a pre-post comparison is merely descriptive. Even so, it may be recommendable to include an interim measure of our DV between the pre-treatment and the post-treatment measure, which suggest biases due to participants' desires to satisfy researchers or clinicians' expectations. Another example is to extract causal effects from group comparisons in cross-sectional studies. Causal effects cannot be drawn when comparing two different groups because we can never exclude a group effect that is independent of the main variable of interest. Thus if we compare two groups of patients that differ in the applied treatment and we find differences, we do not know if these effects are due to an uncontrolled variable. Nonetheless, a longitudinal study will allow causal inferences. In fact, when we write about group effects in between-subject designs, we explain them as "differences", while group effects in within-subject designs are described as "changes". Another common error is to use a covariance analysis to exclude the effect of a confounding variable which shows between- group differences or within-group baseline effects regarding our IV of interest. Using a covariance analysis in an analysis of variance is inappropriate or, at least, questionable, and it does not serve to exclude between-group differences in confounding or non interest variables because the subject assignments to research groups are randomly selected^[7]. However, covariation of observational variables in random assignments can be considered appropriate, but the randomization process should be clearly described and not be confused with non-controlled variables when assigning subjects to groups.

Scientific misconduct leads to non-replicable studies. Therefore, studies may not be replicated due to a questionable research practice rather than to a wrong hypothesis that has been falsified. Leslie et al.^[8] headed a research work with over 2000 psychologists about their involvement in questionable research practices, as several others have studied their prevalence by focusing on previous medical research^[9-14]. Beyond the interesting results of this research, we describe the research practices included in this research as "the steroids of scientific competition". Thus we try to allow readers to reflect on their own or others practices or to become aware of them during their career. Do we report all of our study dependent measures in a paper? Usually when we run an experiment, we collect several measures sometimes more than necessary or hypothesized or non-independent measures for the sake of economics, or sometimes we present "maybe", particularly in expensive procedures or technologies but we may not report them all, or even analyze them at all. Without

going into rationalization or self-defected arguments, to statistically include more dependent variables should be considered in the possibility of a type I error. Did you collect more data after looking to see whether the results were significant? To do a meantime analysis and deciding on increasing sample size without reporting a mean term “pilot” analysis involve a questionable practice. It is like briefly looking for our opponent’s cards and considering that our game will be independent of our quick glance. Similarly, stopping collecting data earlier than planned as we found the result that we were looking for is openly improper. However, there are proposal of analyses that contemplate sequential tests which may be justified when sample size is restricted and justified. Do we report all our study conditions? We can apply several conditions to our research, but reporting them all may go beyond our manuscript interests. This can be a rationalization of a questionable research practice that becomes blatantly improper when those conditions are not described in the methods section as if they were to be analyzed later in the analysis and results section. The inclusion of condition, albeit unanalyzed, may change the effects of those other conditions of interest under any condition without going into self-defection or rationalization. Did we round up a p-value to .05 that was, for example, .052 or close to it? Did we decide to exclude data after looking at the impact of doing so on the results? Did we claim that the results are unaffected by demographic variables (e.g., gender, age) when we were actually unsure or actually knew that they did? Did we falsify data? We consider that these last questions do not need further description. Likewise, these questions may derive from slightly different ones which have been shown to be “questionable”, or their applications are restricted or directly not advisable even when recognized as being post hoc and justified in the manuscript. For example, did we define as outliers some participants after looking at the impact of doing so on the results? Did we change lots of missing data to other values (e.g., mean sample values)? Did we apply unplanned analytical strategies? Did we regress out the effects of variables to obtain “significant” results?

Reproducibility

Reading a scientific article can be just as exciting as reading a catching book. General population, clinicians, or even scientists, may be seduced by a good introduction and discussion of the manuscript, but we easily skim the methodology, even when we are working on that research field, but it depends on our research stage (e.g., when we are writing the discussion, we may not pay attention to study the methodology, but discussion issues and conclusions instead). However, this is the main work of our research. As good researchers or clinicians, we should be able to evaluate whether the methodology is appropriate to test the hypothesis by assuming that the hypothesis is exploratory or the data are driven. In this sense, we need to consider experimental and statistical designs, the acquired data, the data analysis and the procedure to implement the methodology. We must not “believe” a scientific claim as being true without been able to evaluate the evidence supporting it.

Open Science Initiative

OSI is a social enterprise that supports the transparency of the scientific process to support evidence. Open science refers to the process of making the contents and processes of scientific research transparent and accessible to others^[3]. Historically, there have been very few opportunities to make the research process accessible but, with the arrival of the Internet, most barriers are open, even though we find others, such as financial interest and few incentives to open science. The transparency and openness promotion (TOP), as proposed by Nosek and colleagues^[15], involve author guidelines for journals and funder policies to improve transparency and reproducibility in science. TOP has involved a widely available data from published reports to reproduce or extend analyses in the Science, PLOS or Springer Nature journals. Other proposals suggest that journals assign a badge to articles with open data as an incentive for this initiative, and point out that journals value these practices. For example, the journal Psychological Science remember that the psychology field was the origin of the so-called reproducibility crisis – has adopted these badges by increasing data sharing by more than 10-fold^[3]. Similarly, funding agencies such as Research Councils in the UK, the National Institutes of Health (NIH) and the National Science Foundation (NSF) in the USA, are incentivizing and increasing pressure to make access to their data public. Likewise, the Open Science initiative extends to every step in the scientific process, including peer reviews^[16]. For example, the F1000 Research publishing platform for life scientists (<https://f1000research.com>) offers transparent refereeing and the inclusion of all source data. This platform offers the immediate publication of articles (e.g., including outcomes with negative results and replications lacking novelty) and other scientific documents (e.g., posters, slides) after passing an in-house quality check. The peer review process by experts suggested by the author is open after publication. Authors replies are encouraged to be open to the referee reports. The articles that pass the peer review are indexed in Pubmed, among other bibliographic databases. The role of the publishing evaluation is very important but, conventionally, this process is done privately and anonymously. Non anonymous reviews provide empirical evidence for the quality of the received review^[17]. Similarly, the Wellcome Open Research (<https://wellcomeopenresearch.org>), promoted by the Wellcome Trust, is a platform powered by F1000 for the rapid publication of any scientific result with transparent post-publication peer reviews. Other forms to rapidly disseminate research are based on preprint services, such as “arXIV” for physics, but include an archive for data analysis, statistics and probability (<https://arxiv.org/list/physics.data-an>); “bioRxiv” for biology (<https://>

www.biorxiv.org/) accepts some clinical papers; “PsyArXiv” for psychology (<https://psyarxiv.com/>); JMIR Preprints (<https://preprints.jmir.org/>), while MedArXiv is for medicine and health sciences, and (<http://yoda.yale.edu/medrxiv>) is a work in progress project with some reservations. Some health professionals are cautious about giving open access to unreviewed health research and suggest the potential risks for patients’ health, among other arguments. Some journals may consider that a preprint does not preclude publication, but others may not. Promoters of MedArXiv are looking for feedback on this initiative (medarxiv@yale.edu) at the time of this publication.

Guidelines

There is a group of documents and reports that offer standards to improve the quality of research reporting. Thus the Transparency and Openness Promotion (TOP) guidelines offer standards for journals and funders to incentivize or require greater transparency in planning and reporting research [15]. The Consolidated Standards of Reporting Trials [18] provide guidelines for the transparent, complete and accurate reporting of randomized controlled trials in biological sciences [19,20], as endorsed by 600 journals and prominent editorial groups. Similarly, other extended versions provide guidelines for psychological and social interventions [21] by extending nine items from CONSORT 2010 [18], adding a new item related to stakeholder involvement in trials, and modifying the original CONSORT 2010 [18] flow diagram [21]. The Preferred Reporting Items of Systematic Reviews and Meta-analyses (PRISMA) present a guideline for reporting systematic reviews and meta-analyses [22]. Similarly, PRISMA-P is a guideline for protocols of systematic reviews [23]. In 2017, and as part of the 2017 Peer Review Congress, a working group of journal editors and experts supported an overall effort to develop a minimal set of expectations and standards that journals could agree to ask their authors in order to meet in the life sciences context.

There are over 300 guidelines for observational studies, prognostic studies, predictive models, diagnostic tests, systematic reviews and meta-analyses in humans and laboratory methods for humans and animals. These guidelines are aggregated in the Equator Network (<http://www.equator-network.org/>) [24]. However, the use of reporting guidelines is what makes them effective and successful, rather than only for publication.

Triangularization

Munafr and Davey Smith [25] suggest applying multiple approaches to address a question when looking for an agreement across different methodologies, techniques and analytical strategies depending on different assumptions. These authors advocate that consistent findings through replication may reflect systematic failings in study methods and analyses, besides other previously mentioned biases and dark forces that lead us all to direct our sight in the same direction. Thus a cube is a cube if we have a three dimensional view of the volume, and all its faces give as a different view of the same phenomena. Otherwise, we may all think that we are looking at a cube when it is actually the frontal view of the base of a pyramid. So we consider taking some cautions before reducing errors in triangularization following the checklist for triangularization suggested by Munafr and Davey-Smith [25]. Triangularization involves changes in the research groups composition and coordination –or even science community changes–, including members specialized in different disciplines. These changes involve experts who think in parallel and others who coordinate the leaving aside of the classical hierarchy, recognizing each contributor and its contribution from subjects recruitment, and the data preparation of “junior” researchers to methodological guidance or a critical review at any level of “seniors”. Munafr and Davey-Smith suggest a long list of individuals specifying their contributions both fully and specifically. Likewise, the peer review process should be partitioned according to sub studies or specialized subsections. In fact we consider that this suggestion can be currently applied to research belonging to one discipline, but which needs multidisciplinary knowledge to conduct it, particularly those involving state-of-the-art technologies and methodologies involving novel mathematical, physical or chemical advances. Similarly, this suggestion is applicable to funders, as pointed out by Munafr and Davey-Smith [25,26].

CONCLUSIONS AND IMPLICATIONS FOR NURSING AND HEALTH POLICY

The field of metascience –the scientific study of science itself– is flourishing and has generated substantial empirical evidence for the existence and prevalence of threats to efficiency in knowledge accumulation. Similarly, research initiatives may encourage governments and institutions to develop metapolicies based on empirical evidence in the data mining age. Likewise, we consider that the topics described in this manuscript are just as important for experimental and applied researchers as they are for clinical practitioners.

Nursing global policies and managing systems aware of a spread replication crisis across several health sciences can be harmful or beneficial. Harmful or detrimental in the absence of metascience in nursing research. Economical global policies will reduce the number of available nursing screening test or treatments, professional continuous career development and training and patients satisfactions, whether the replications crisis is used as reason for cost restrictions justified on variability of nursing practices which might negatively impact treatment efficiency. On the other

hand, beneficial effects of policies and managing systems based on the knowledge extracted from the metascience emerging from the replication crisis can improve cost-effectiveness of treatments across health centers adapted to their own managing system and patients characteristics. Thus, appropriate policies and managing initiatives serving evidence based treatments based on the appropriate analysis of the so called replication crisis may facilitate the transference of knowledge from nursing research to practice and vice versa, encouraging nurses to lead the management of relevant protocols, guidelines and care bundles based on scientific evidence.

As written by Munafo "Publication is the currency of academic science". This manuscript has avoided going into the incentives that motivate the publication of novel and positive results. The concerns set out in this manuscript about scientific work should not be treated in a sensationalist way, and should not be used as a topic of political contention linked to an attempt of diminish regulations or discredit scientific research, but to responsibly improve research practices. The replication of any given scientific finding can range within a continuum from replication to total failure .

REFERENCES

1. Pashler H and Wagenmakers EJ. Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science* 2012;7:528-530.
2. Collaboration, Open Science. Estimating the reproducibility of psychological science. *Science* 2015;349: 4716.
3. Munafo MR, et al. A manifesto for reproducible science. *Nature Human Behavior*, 2017.
4. Ioannidis JPA. Why most published research findings are false. *Plos Medicine* 2005;2:e124.
5. Lenzer J, et al. Ensuring the integrity of clinical practice guidelines: A tool for protecting patients. *BMJ* 2013;347:f5535.
6. Krauss A. Why all randomised controlled trials produce biased results? *Annals of Medicine* 2018;50:312-322.
7. Miller GA and Chapman JP. Misunderstanding analysis of variance. *Journal of Abnormal Psychology* 2001;110:40-48.
8. John LK, et al. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science* 2012;23:524-532.
9. Gardner W, et al. Authors' reports about research integrity problems in clinical trials. *Contemporary Clinical Trial* 2005;26:244-251.
10. Geggie D. A survey of newly appointed consultants' attitudes towards research fraud. *Journal of Medical Ethics* 2001;27:344-346.
11. Henry DA, et al. Medical specialists and pharmaceutical industry-sponsored research: A survey of the Australian experience. *Medical Journal of Australia* 2005;182:557-560.
12. List J, et al. Academic economists behaving badly? A survey on three areas of unethical behavior. *Economic Inquiry* 2001;39:162-170.
13. Martinson BC, et al. Scientists behaving badly. *Nature* 2005;435:737-738.
14. Swazey JP, et al. Ethical problems in academic research. *American Scientist* 1993;81:542-553.
15. Nosek BA, et al. Promoting an open research culture. *Science* 2015;348:1422-1425.
16. Morey RD, et al. The peer reviewers' openness initiative: Incentivizing open research practices through peer review. *The Royal Society Open Science* 2016;3:150547.
17. Walsh E, et al. Open peer review: A randomized trial. *Brit J Psychiat* 2000;176:47-51.
18. Schulz KF, et al. Consort 2010 statement: Updated guidelines for reporting parallel group randomized trials. *BMJ* 2010;340:c332.
19. Begg C, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;276:637-639.
20. Moher D, et al. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;272:122-124.
21. Grant S, et al. Consort-SPI 2018 Explanation and elaboration: Guidance for reporting social and psychological intervention trials. *BMC Trials* 2018;19:406.
22. Liberati A, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6:e1000100.
23. Shamseer L, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ* 2016;349:g7647.
24. Grant S, et al. Developing a reporting guideline for social and psychological intervention trials. *Res Social Work Prac* 2013;23:595-602.
25. Munafo MR and Davey-Smith G. Robust research needs many lines of evidence. *Comment in Nature* 2018;553:399-401.

26. <https://respectfulinsolence.com/2017/01/30/is-there-a-reproducibility-crisis-in-biomedical-research-2017-edition/>.