Statement

Replicability of Research Results

A Statement by the German Research Foundation

April 2017

Contact:
Dr. Katja Hartig
Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)
Katja.Hartig@dfg.de

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)
Kennedyallee 40 · 53175 Bonn · Germany
Phone: + 49 228 885-1 · Fax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de

Replication is a very important method for testing empirical knowledge claims based on experimental and quantitative research in medicine, the natural, life, engineering, social and behavioural sciences, as well as the humanities. Since a series of articles on the replicability of research results (in biomedicine) appeared in the *Lancet* journal in 2014[1], a lively and public debate developed around the catchphrase "replication crisis". In essence, the debate concerns the question of the quality of research, and affects science as a whole.

However, replication is not the only test method. There are others, including, for example, theoretical-conceptual discussion and criticism, modelling, mathematical modelling, simulation and more. Thus, when discussing the question of replication and replicability, it needs to be borne in mind that:

- Scientific results can be replicable, but they need not be. **Replicability is not a universal criterion for scientific knowledge.** The expectation that all scientific findings must be replicable cannot be satisfied, if only because numerous research areas investigate unique events such as climate change, supernovas, volcanic eruptions or past events. Other research areas focus on the observation and analysis of contingent phenomena (e.g. in the earth system sciences or in astrophysics) or investigate phenomena that cannot be observed repeatedly for other reasons (e.g., ethical, financial or technical reasons). Furthermore, there are forms of research that have reached such a degree of complexity in their experimental methodology that replicative repetition can be difficult.
- **Ascertaining the replicability or non-replicability of a scientific result is itself a scientific result.** As such, it is not final; rather, like all scientific knowledge, it is subject to methodological scepticism and further investigation.
- Non-replicability of an empirically-based scientific knowledge claim can, but does not necessarily, indicate its falsification. **Non-replicability is not a universal proof by falsification.** On the contrary, under certain circumstances non-replicability can also be understood as currently-not-yet-replicability, or as a signal for the existence of a new, previously unknown scientific context, and thus an indicator of speculative research for the advancement of scientific knowledge.
- Non-replicability of an empirically-based scientific knowledge claim can, but does not necessarily, result from poor scientific practice or scientific misconduct. **Non-replicability is not a universal indicator of poor science.** For the purpose of ex-

[1] Chalmers et al. *Lancet*, Vol. 383 (2014); Ioannidis et al. *Lancet*, Vol. 383 (2014); Salman et al. *Lancet*, Vol. 383 (2014); Chan et al. *Lancet*, Vol. 383 (2014); Glaszui et al. *Lancet*, Vol. 383 (2014)

amining whether poor science in this sense was involved in the preparation, execution, evaluation, description or publication of an experiment, we refer to the *Recommendations for Safeguarding Good Scientific Practice* published by the DFG and the *Process Guideline*[2] developed for this purpose.

- The fact that robust, although not replicable, scientific knowledge exists must on no account be misappropriated as an excuse or apology for non-replicability in cases where the replicability of a scientific knowledge claim must be expected for methodological reasons.

In light of these principles, a differentiated assessment should be made of the papers that have appeared since the *Lancet* series in 2014 and the public debate that has followed them. The DFG takes this discussion as an occasion to state its position as follows:

- Debate in science policy and in the public arena often simplistically equates replicability with good science, and non-replicability with poor science, and thus draws the incorrect conclusions indicated above.
- The discussion triggered by the *Lancet* series, metastudies[3] on replicability, and related initiatives[4] are themselves evidence of functioning mechanisms for scientific self-regulation.
- In their totality, the reported cases of non-replicability are a cause for concern, even if it is assumed that they are attributable to a variety of causes and by no means all indicate bad scientific practice or scientific misconduct. To this extent, the discussion about the so-called replication crisis indicates a **quality problem in research**, even though its scope cannot be precisely determined. Researchers, research institutions and research organisations must take this quality issue very seriously. It endangers the performance of science as well as society's confidence in it.
- Insofar as elements of this quality problem, in the form of poor scientific practices or **scientific misconduct**, are attributable to individuals, this is addressed within the research community by the control mechanisms of the ombudsman for German research, the local ombudsman bodies, the research institutions' own procedures for investigating and penalising cases of scientific misconduct, and, in the case of DFG-

[2] Recommendations for Safeguarding Good Scientific Practice and Process Guideline for Good Scientific Practice: www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html
[3] Open Science Collaboration (2015); Gilbert, King, Pettigrew & Wilson (2016)
[4] NIH Rigor and Reproducibility Initiative; Center for Open Science; ARRIVE; Force 11

funded research, particularly also the Committee of Inquiry on Allegations of Scientific Misconduct.

- Along with individual misconduct, there are also **structural reasons** for the quality problem in research. The weight of quantitatively-parametrising control, evaluation and gratification systems prevailing in research today has the effect of creating increased (and continually increasing) pressure to compete and to accelerate results. This is manifested in decisions (and underlying decision criteria) about career moves, financial support, location of publication, and institutional structural trends. The scrupulous care required for the preparation, execution, evaluation, description and publication of experimental and empirical-quantitative research requires time, opportunity, funding and staff. Such care is often threatened, rather than supported, by competitive and time pressures. The same applies to the appraisal and publication of so-called neutral or negative or redundant results.

Given a structural framework that can all too easily be misunderstood as an invitation to quick-and-dirty research practices, the **DFG also acknowledges its own responsibility**. In exercising this responsibility, the DFG must be mindful of various aspects of its activity as an organisation promoting research and scientific self-governance. The DFG

- will focus on the specific insights to be expected from a research project when it comes to evaluating project proposals;
- will ensure that, in the ongoing development of its funding portfolio and in the review, evaluation and decision-making processes for which it is responsible, the main criterion for scientific judgement will be the quality of publications rather than their quantity or location;
- will also take into account that replication as a method for testing experimental and empirical quantitative research results must be systematically strengthened;
- will therefore facilitate and support processes of subject-specific investigation of questions concerning the replicability of research results; this also includes the development of subject-specific criteria for funding replication studies[5] as well as the funding itself;
- will continue to pay particular attention to questions of research data management and current challenges that emerge from digitalisation;

[5] Recommendations of the Psychology Review Board: Psychologische Rundschau 67/3: pages 163ff. (2016)

- will promote the development of infrastructure and methodological tools as well as their use for this purpose;
- will remain fully committed to its wide-ranging efforts to promote good scientific practice, which set standards in the German science system;
- calls on academic publishers, scientific institutions and ethics commissions, as well as lawmakers and scientific policymakers, to do everything in their power to combat the structural reasons for replication difficulties.