Reproducibility of Results in Medical and Biomedical Research

Statement of the Working Group “Quality in Clinical Research” of the DFG Senate Commission on Key Questions in Clinical Research

The DFG Senate Commission on Animal Protection and Experimentation supports this statement.

March 2018

Dr. Katja Hartig
Life Sciences 3: Medicine
Tel.: +49 (0)228 885-2359
E-mail: katja.hartig@dfg.de

Dr. Britta Mädge
Life Sciences 1: Molecular and Organismic Biology
Tel.: +49 (0)228 885-2453
E-mail: britta.maedge@dfg.de

In April 2017, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) published a general position on the reproducibility of research results[1] and encouraged the different scientific disciplines to participate in and continue the discussion by reflecting on the quality of the research process on a discipline-specific basis. In medical and biomedical research, successes and knowledge gain are generally highly visible and often of direct human and economic relevance. The discussion of issues relating to the scientific quality and reproducibility of research results, triggered by the series of articles in *The Lancet* on "Increasing value, reducing waste"[2], is therefore quite rightly considered particularly important in these fields.

This statement is a response from the fields of medicine and biomedicine – where research questions are directly or indirectly oriented towards the understanding of diseases and approaches to patient treatment – to requests for a deeper subject-specific engagement. In medicine and biomedicine there has been intensive discussion for several years on the reproducibility[3] of scientific results[4]. In what follows, the specific challenges facing these two fields of research are explained and solutions are proposed to increase reproducibility. The recommendations are aimed at researchers in the areas of medicine and biomedicine, but may also be of interest to researchers and other stakeholders in other areas of the life sciences.

---

[1] http://www.dfg.de/en/service/press/press_releases/2017/press_release_no_13.

[2] Increasing value, reducing waste. *The Lancet*. 2014; 383.

[3] A definition of the key terms used in the statement can be found at the end of the text.

[4] DFG roundtable discussions:

http://www.dfg.de/en/dfg_profile/statutory_bodies/senate/clinical_research/events/workshop_0315/index.html.

http://www.dfg.de/en/dfg_profile/statutory_bodies/senate/clinical_research/events/workshop_1215/index.html.

1 Putting the discussion about reproducibility in a discipline-specific context

In medicine and biomedicine, as in other scientific disciplines, researchers study what happens and obtain findings that cannot be directly reproduced due to their singularity and/or contingency[5]. **However, for the majority of research questions and approaches in medical and biomedical research, reproducibility should be ensured as much as possible**. This aim is built into the obligation to comply with good scientific practice[6] and to engage in self-critical examination of research work, which is equally incumbent upon all areas of science.

This is associated with the following **discipline-specific challenges**:

- The **research process** in medicine and biomedicine consists of **recurring cycles** of exploratory, hypothesis-generating phases in which new observations are made and unknown connections are identified, followed by consolidation phases in which these observations are deepened and connections are verified. To make new findings usable for patients, it is usually necessary to transfer knowledge from *in vitro* to *in vivo* models and then gradually generalise them as far as possible through different model organisms.
- In addition to experimental approaches, researchers are increasingly developing theoretical approaches that describe dynamic or more complex processes with the aid of **modelling and simulation** using existing research data. New knowledge can only be gained when researchers can build on their or others' previous results. The validity of published results is an essential requirement for this continuity in research.
- Biological and medical research seeks to achieve a deeper understanding of life processes and often uses **living organisms** with limited capacity for **standardisation**[7]. A high degree of standardisation also limits transferability of findings. In all investigations, an appropriate balance must therefore be found between the standardisation required for quality assurance and the taking into account of the genetic and phenotypic heterogeneity of living organisms.
- For research on living organisms and systems, the **accessibility of cells, tissues and materials** also plays an important role in the reuse of scientific findings. Here too, there is a need to create and improve the necessary infrastructures and technical standards.
- In medicine and biomedicine, experimental fields in particular are characterised by the **consistent development of methods and technologies**. However, each new method must be established, well tested and standardised before it goes into general use.
- The transfer of findings from one model to another and ultimately to humans is broadly referred to as **translation to the clinics**. It is based on a continual consideration of the

[5] Examples include clinical reports of individual cases and the EHEC outbreak.

[6] http://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html.

[7] J. Lithgo, M. Driscoll and P. Phillips: A long journey to reproducible results. *Nature*. 2017; 548: 387-388.

suitability of models, especially the need for animal and human studies. Findings obtained in one model organism may be valid and reproducible in themselves but not transferable to other model systems or to humans. Insufficient ability to transfer from one system to another *per se* does not mean low quality research, but is an unavoidable challenge associated with the **use of different model systems**.

- In recent decades a large number of high-throughput technologies has been developed, bringing both methodological requirements and entirely **new requirements in terms of description, documentation and archiving of data and the access to it**[8]. At the present time, IT applications and suitable infrastructures for professional data management are limited.

[8] The FAIR principles of data management and data stewardship (http://www.nature.com/articles/sdata201618).

2 Putting the reproducibility debate in a broader context

Important methodological prerequisites and higher-level framework conditions both play a key role in ensuring reproducibility.

- **Misguided incentive systems result in more false-positive findings and make replication studies unattractive**. Curiosity, as well as the competition between ideas and new approaches, are essential drivers of the production of new scientific knowledge. Researchers also compete for available resources. Since research achievements are primarily assessed on the basis of publication output, there is an incentive to publish as much as possible in the most prominent journals. A 'publication bias' can be observed in favour of the first description of findings and of results with statistically significant effects. This leads to the following problems: The use of inferential statistics inevitably causes the occurrence of false-positive findings. If significant results are more likely to be published, there will be a higher incidence of positive findings in the literature even though the studies were all performed correctly[9]. This makes it difficult to form a comprehensive and realistic picture of the state of knowledge. When preference is given to original as opposed to consolidating results, this also hampers the carrying out of replication studies and makes this type of qualitative examination unattractive.

  Time pressures on research projects, e.g. resulting from short funding periods, rapid evaluation cycles and short contract periods for both research and support staff, promote the practice of producing and publishing results quickly and in small chunks.
- **Legal and ethical regulations**: There is a fundamental conflict of aims between carrying out direct replications and the legal requirements of animal welfare and the ethical considerations involved in human studies. The need for repetition must be balanced against the avoidance of superfluous work.
- **Insufficient or unsuitable infrastructures for the storage of research data and materials** make it more difficult to document and reuse research results. There is also a lack of suitable basic and advanced training to ensure that researchers have the necessary methodological knowledge to make use of available infrastructures.

[9] J. P. A. Ioannidis: Why most published research findings are false. *PLoS Medicine*. 2005; 2: 696-701.

3 Prerequisites for reproducibility

Several essential prerequisites must be in place if research results aimed directly or derivatively at the understanding of disease and patient treatments are to be reproduced. The importance of these prerequisites increases as the research process proceeds and is particularly relevant in the case of consolidating approaches.

- **Validity of models and standardisation of methods**: The choice of a model system requires basic considerations as to the advantages and limitations of a particular model. The models and methods used must be sufficiently well established. Applicable standards must be taken into account.
- **Adequate statistical planning**: Before an investigation gets underway, proper consideration must be given to statistics, for example planning the number of cases or random samples and multiple stages of correction. Blind or double-blind studies should be used wherever possible.
- **Careful management of research data and materials**: The materials used and research data generated must be fully described and made available in databases and tissue banks for repeat studies, verification and reuse by other researchers.
- **Complete description of methods and analyses**: This includes the careful, complete and accurate description of hypotheses and statistical methods used for evaluation in reports and publications. Selectively presented results ('p-hacking') and the post-hoc adaptation of the hypothesis to the results ('HARKing') significantly impair the meaningfulness of the results.

These requirements must of course also be applied to replication studies, which seek to repeat and verify research results.

In view of the dynamic development of methodologies and the fundamental challenges posed by the translation of results, there is a particular need for an ongoing and self-critical examination of the research process. Identifying the reasons for unsuccessful replication can make a significant contribution to the overall improvement of quality.

4 Self-imposed obligations and specific recommendations

The reasons for insufficient reproducibility in medicine and biomedicine are complex. They are influenced by structures in the research landscape and by stakeholders outside research institutions. Creating the necessary environment for improving the reproducibility of results therefore requires a shared commitment by multiple stakeholders.

Below, a number of recommendations to different stakeholders are listed which, in the view of the working group, could help to improve the basis for the reproducibility of research results.

- **Research institutions**: Research institutions should provide researchers with the necessary resources, basic infrastructure and scope for the adequate management of research data. The continuity and stability of research teams and the creation of attractive career paths for service-oriented research units, suitable basic and advanced training, and the establishment of advisory services with low-threshold access are all important components.
- **Animal welfare and ethics committees, regulatory authorities**: The high relevance of the repetition and validation of scientific results should be adequately taken into account in the approval process and the ongoing development of legal standards (e.g. animal welfare or data protection).
- **Publishers and journals**: A growing number of scientific journals is publishing replication studies[10] and results with few or no effects. Other publishers should be encouraged to follow this positive example. The development and provision of checklists and guidelines[11] which can be used to verify the validity or documentation of results is much to be welcomed.
- **Scientific organisations and the general public**: The potential of research results and knowledge gain should be considered in a more realistic way in relation to the treatment of disease – overreaching expectations produce pressure to succeed and are thus counterproductive to careful research.
- **Funding organisations:** For research funding organisations, including the DFG, there are various potential starting points for further discussion:
  - In the submission, review and evaluation of research proposals, it is recommended that the prerequisites listed in section 3 should be adequately taken into account.
  - In the review and evaluation of research proposals, greater value should be placed on an individual's research performance above and beyond publication output. Participation in quality-reflecting or quality-promoting activities could be given greater emphasis than is currently the case.

[10] https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology.

[11] http://www.equator-network.org/ or http://circres.ahajournals.org/content/121/5/472.

- Replication studies contribute to knowledge gain. For medicine and biomedicine, proposals for replication studies must be evaluated according to a set of review criteria, which should be developed, in order to be competitive against all other proposals. The subject-specific criteria already developed by the DFG's Psychology review board for the funding of replication studies[12] provide some useful starting points.
- Longer project durations and/or interim evaluations which focus less on quantitative parameters may help to reduce the pressure to produce results.
- The final reports for research projects should also describe unforeseeable, low-effect or no-effect results. Reports should be publicly accessible and citable.
- All efforts designed to create broad and open access to scientific information and results should be continued.
- Many funding organisations are currently engaging with the implementation of measures to increase the reproducibility of scientific results. A joint procedure or dialogue on experiences would be desirable.

Although some of the structural conditions in which research takes place is difficult for individual researchers to influence directly, nevertheless it is researchers themselves who conduct research projects, evaluate them and publish the data. The task and responsibility therefore falls on them to initiate crucial changes. Scientists and their research societies and professional associations should therefore commit to the following principles:

- The **prerequisites** listed in section 3 **must be satisfied more consistently than is currently the case** in the planning, implementation and description of research projects.
- Researchers must be willing to publish results that contradict their initial hypotheses or that show few or no effects. The scientific community itself must promote the **broad and complete visibility of results**. This should be achieved in part by means of existing publication options[13].
- **Basic and advanced training for researchers** should focus much more on self-reflection, openness and a culture of error acceptance. Learning methods for quality improvement should be regarded as an essential feature of a researcher's **continuing professional development**.
- Researchers should **support the replication** of their investigations by others in every respect. The categorisation of non-reproducible results and the identification of the **reasons for this should be discussed openly and in cooperation** with the authors of the original study.

[12] http://www.dfg.de/foerderung/faq/geistes_sozialwissenschaften/index.html.

[13] Examples of publishers which publish results with few or no effects: BioMedCentral, PloSOne; for preprint servers: arXiv, bioRxiv; for preregistration (registered reports): Open Science Framework, Royal Society Open Science.

Reliability, clarity and transparency are essential pillars of scientific progress and lay the foundations for public trust in science. This is especially true when scientific findings contribute to the understanding and treatment of disease and scientists are seeking to convert findings into therapies. In addition to the ethical responsibility borne by science, this also gives research an economic relevance. Each of these objectives rightly demands a high degree of quality and diligence. This makes it vitally important for the scientific community to fulfil this requirement. However, the research process in the area of medicine and biomedicine is also based on tentative exploration, failure and branching out into unknown areas in which a complete reproducibility of results may not be ensured. An appropriate balance between new and exploratory research and scientific approaches that verifies and validates results and their robustness is essential to dynamic fields of research such as medicine and biomedicine. Many steps are required before findings can be applied to patients, and these steps must be carefully planned and documented. The specific implementation of the requirements defined here demands cooperation across the entire scientific community in the area medicine and biomedicine, which holds key starting points for change in its own hands.

The terms ‘replication’ and ‘replicability’ as well as the often-used terms ‘reproduction’ and ‘reproducibility’, and the differences between them, have yet to be clearly defined. These terms are not used consistently in either German or English.

In this paper only the terms ‘replication’ and ‘reproducibility’ are used, with the following meaning: **Replication** is the repetition of an investigation, experiment or study which is claimed to be repeatable, but also refers generally to the possibility, regardless of results, of repeating something or performing an experiment again.

Reproducibility means, correspondingly, the ability or capacity to confirm results on a repeat basis within the framework for error.

Working group “Quality in Clinical Research”

Members of the working group

Professor Dr. Bernd Fleischmann, Bonn (chair)
Professor Dr. Christopher Baum, Hanover
Professor Dr. Klaus-Michael Debatin, Ulm
Professor Dr. Simone Fulda, Frankfurt
Professor Dr. Wolfgang Herr, Regensburg
Professor Dr. Andreas Meyer-Lindenberg, Mannheim
Professor Dr. Georg Peters, Münster
Professor Dr. Esther von Stebut-Borschitz, Cologne
Dr. Frank Wissing, Berlin

Guests of the working group

Professor Dr. Julia Fischer, Göttingen
Professor Dr. Michael Sendtner, Würzburg
Professor Dr. Brunna Tuschen-Caffier, Freiburg
Professor Dr. Onur Güntürkün, Bochum
Professor Dr. Roland Lill, Marburg

DFG Head Office

Dr. Katja Hartig, Bonn
Dr. Britta Mädge, Bonn
Dr. Tanja Kollei, Bonn

The SCCR’s Scientific Secretariat

Corina Schnabel M.A., Freiburg

Deutsche Forschungsgemeinschaft (DFG)
German Research Foundation
Kennedyallee 40, 53175 Bonn
Postal address: 53170 Bonn, Germany
Tel.: + 49 (0)228 885-1
Fax: + 49 (0)228 885-2777
postmaster@dfg.de
www.dfg.de/en