KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

KNAW REPORT
REPLICATION STUDIES

Eric-Jan Wagenmakers & Jean Philippe de Jong

Presentation KNAW report Replication Studies, 13 December 2017, Committee on Reproducibility and Replicability in Science, NAS (telco)

BACKGROUND

- Scientific progress requires research studies to be conducted rigorously, so that when they are repeated their results will be reasonably similar
- However, many replication studies have been unable to reproduce important results
- Studies with non-reproducible results can jeopardise scientific progress, waste resources, harm individuals and society, and erode public trust in science

GOALS OF KNAW REPORT (TO BE PUBLISHED ON JANUARY 15, 2018)

- Analysis of causes of non-reproducibility
- Assessment of desirability of replication studies
- Offer recommendations for improving reproducibility and for conducting replication studies

The report is based on:

- Analysis of scientific literature and reports
- Interviews with experts
- Workshop with invited experts
- Deliberations within the KNAW Replication Committee

SCOPE OF REPORT

- Analyses are (mainly) based on experiences within medical sciences, life sciences and psychology
- It could be argued that all scientific disciplines based on systematic observations should aim to generate reproducible results
- However, the importance of reproducibility and what can be replicated may differ significantly between disciplines
- KNAW invites other empirical disciplines to consider the relevance of this report's conclusions and recommendations for their field

DEFINITIONS OF REPLICATION STUDIES AND REPRODUCIBILITY

- A *replication study* is a study that is an independent repetition of an earlier, published study, using similar methods and conducted under similar circumstances
- *Reproducibility* concerns the extent to which the results of a replication study agree with those of the earlier study
- No consensus regarding definitions of replication, reproduction, replicability, reproducibility, nor their relation to terms such as robustness and reliability

Table 1. Occurrence of non-reproducibility		
Field	**Approach**	**Outcome**
Preclinical animal studies, general biology (I)	Researchers from Bayer HealthCare attempted to validate data on potential drug targets obtained in 67 projects by copying models exactly or by adapting them to internal needs	In 20% to 25% of cases, published data were completely in line with the results of the validation studies
Preclinical studies, oncology (II)	Amgen team attempted to reproduce the results of 53 'landmark' studies	Scientific results of 11% of the studies were confirmed
Preclinical studies, genetics (III)	Replication of data analyses provided in 18 articles on microarray-based gene expression studies	Two analyses (11%) were reproduced and six were partially reproduced or showed some discrepancies in results; ten could not be reproduced
Preclinical animal studies, neurology (IV)	Retesting of nine potential drugs in rigorous animal tests that had been reported to slow down disease in a mouse model for ALS	None (0%) of the drugs was found to slow down ALS
Observational and randomised studies in clinical medicine (V)	A retrospective analysis of the most highly-cited articles reporting on observational and randomised studies on postulated effective medical interventions	The conclusions of 16% of articles were contradicted by subsequent studies; in a further 16%, effects in subsequent research were weaker than initially found
Experimental psychology (VI)	Direct replications of 13 psychological phenomena across 36 independent samples	77% of phenomena were reproduced consistently
Experimental psychology (VII)	The Open Science Collaboration attempted to independently replicate selected results from 100 studies in psychology	36% of the replication studies produced significant results, compared to 97% of the original studies. The mean effect sizes were halved
Experimental economics (VIII)	Replication of 18 published studies in economics	A significant effect in the same direction as in the original study was found for 11 replications (61%); on average, the replicated effect size was 66% of the original

[i] Prinz, 2011, [ii] Begley 2012, [iii] Ioannidis 2009 , [iv] Perrin 2014, [v] Ioannidis 2005b, [vi] Klein 2014, [vii] OSC 2015, [viii] Camerer 2016

OCCURRENCE OF NON-REPRODUCIBLE RESULTS

- Series of replication studies have been unable to reproduce many important results
- Disciplines as a whole may be subject to a substantial degree of non-reproducibility
- Studies with non-reproducible results carry a significant risk for both science and society and efforts should be made to achieve a high degree of reproducibility.
- The optimal degree of reproducibility is not expected to be the same in every discipline, and it may even depend on the specific type of research
- As a first step, empirical disciplines should assess the degree of non-reproducibility and its underlying causes

Table 2. Causes of non-reproducibility	
Area	**Cause**
Study methods	Weak experimental design/failure to control for biases
	Small sample size, increasing the risk of chance findings or inflated discoveries of otherwise true but weak signals (type I error)
	Low statistical power to detect the effect (type II error)
	Technical/human error in executing the study and poor quality control
	Fraud and fabrication of data
	Unknown variables that influence study outcomes
	Lack of rigour in statistical analyses
	Inappropriate statistical analyses
	Failure to conduct 'internal replication' (e.g. performing multiple measurements, cross-validation within a dataset, setting up control experiments)
Study reporting	Omitting null results (non-reporting and selective reporting) and selective analyses that make null results seem spuriously positive
	No sharing of data or methodology details
	Fishing (post hoc choices of dependent/independent variables based on results)
	Presenting post hoc hypotheses as tested hypotheses (HARKing)
	Outcome switching (a discrepancy between registered primary outcome and published primary outcome)
	Lack of adequate peer review
Incentive system	Rewarding many and 'high-impact' publications
	Rewarding positive and novel/'breakthrough' results
	Highly competitive funding systems
	Not rewarding open and reproducible practices
	Belief that a rigorous research process hampers discovery

CAUSES OF NON-REPRODUCIBLE RESULTS

- Many different factors can lead to non-reproducibility
- Even rigorously conducted studies may yield results that cannot be fully reproduced: a certain degree of non-reproducibility is inherent to the pursuit of science
- Nevertheless, avoidable factors that cause non-reproducibility should be addressed

Table 3. Strategies to improve reproducibility		
Area	**Strategy**	**Proposal**
Study methods	Improving study design	Comply with guidelines for designing and executing studies
	Improving methodological skills	Train future and current researchers in statistics and research methods
	Methodological support and oversight	Conduct an independent review of study protocols Involve methodologists in studies
	Collaboration	Multi-site studies Team-science consortia
	Standardisation	Standardise research activities with technologies/automation
	Quality control	Set up control mechanisms: checklists, audits
	Study preregistration	Facilitate study registries Require preregistration of hypothesis-testing studies
	Internal replication and validation	Conduct internal replication (e.g. repeat analyses in other datasets, repeat experiments) and other forms of internal validation (e.g. bootstrapping)
Study reporting	Improving reporting	Use journal guidelines and checklists Distinguish hypothesis-generating from hypothesis-testing studies Use mechanisms for correcting articles ('versioning')
	Transparency	Issue guidelines for storing and providing access to data and methods
	Diversifying peer review	Make peer review open Base peer review on study quality, not on study results and inflated claims
Incentive system	Rewarding null results	Publish 'negative' studies
	Less competition	Work through long-term contracts and funding Reward collaboration Reward mentoring and training
	Rewarding open practices	Reward sharing of methods and data
	Rewarding reproducible practices	Reward peer review Reward efforts to improve methods Reward high-quality, rigorous research rather than publications in 'high-impact' journals
	Stimulating research on research	Fund studies that monitor effects of proposed measures

STRATEGIES TO IMPROVE REPRODUCIBILITY

- Many of the factors that cause non-reproducibility can be addressed
- If these kinds of 'preventive' measures are implemented comprehensively and across the board, then the reproducibility of study results is likely to improve substantially
- However, replication studies *afterwards* are a normal and essential part of science too

THE DESIRABILITY OF REPLICATION STUDIES

- Replication of individual studies help to allay doubts about results. Important if results have a major impact on scientific progress, meeting societal goals, or waste of resources
- Systematic series of replication studies generate data on the occurrence of non-reproducibility, its causes and help monitor the effectiveness of measures aimed at improving non-reproducibility
- The desirability of replication studies depends on their benefits and costs compared to alternative strategies, such as conducting original study or other ways to invest research funds

Table 4. Assessment of the desirability of replication studies	
Criteria	**The desirability of a replication study:**
Knowledge	- is higher when results from a previous study seem more implausible - is higher when there are more doubts about the validity of the methods or the proper execution of a previous study - is higher when its results may have a major impact on scientific knowledge - is higher when it may help improve research methods
Impact	- is higher when its results may have a major societal impact - is higher when it may help avoid wasting research resources on a scientific dead end - is higher when it may improve the functioning of a whole discipline (replication series)
Cost	- is lower when it requires more resources and time investment by researchers - is lower when it places a heavier burden on human and animal test subjects
Alternatives	- must be weighed against performing innovative studies - must be weighed against taking other measures to improve reproducibility

REPLICATION STUDIES IN PRACTICE

- Replication studies appear to account for a small fraction of all published literature (reliable data are lacking)
- Proper replication studies can be designed in various ways (Who will carry out the study; What parts of the original study are being replicated; How similar are those replicated parts to the original study?)
- Investigators should offer sound arguments for their replication approach
- Replication practices vary considerably across disciplines and are still evolving
- Several disciplines have taken important steps towards developing good replication practices

LESSONS FROM REPLICATION PRACTICES IN DIFFERENT DISCIPLINES

(Preclinical animal research, clinical research, experimental psychology, genetic epidemiology and biochemistry)

- A series of replication studies can cause a discipline to reflect on its research practices
- The development of guidelines for study design and reporting is an important starting point for improving reproducibility
- Improving reproducibility requires a significant effort on the part of the research community with proper incentives from stakeholders such as scientific journals, institutions and funding agencies

BARRIERS TO CONDUCTING MORE REPLICATION STUDIES

- Studies are often not reported in sufficient detail, making it impossible for other researchers to design a proper replication study
- Researchers may be unsure about the right design for a replication study and the interpretation of its results
- Researchers do not always appreciate the value of replication studies and may find it difficult to get them funded and published

Table 5. Strategies to stimulate replication studies*

Area	Strategy	Proposal	Researcher	Funding agencies	Journals	Institutions
Information-sharing	Preregistration	Register study protocols in a database with relevant details about the methodology and analysis	X			
		Make preregistration a requirement for funding		X		
		Include preregistration in guidelines for proper research conduct				X
		Require preregistration for publication			X	
	Quality of reporting	Develop reporting guidelines and checklists for a structured reporting format with relevant details of the methodology	X		X	
		Lift length restrictions for methods sections			X	
		Make adherence to reporting guidelines part of the review process	x		X	
	Repositories	Store detailed information about study data and methods in repositories that are accessible to other researchers	X			
		Make storage of study information in a repository mandatory		X	X	X
Know-how	Data	Conduct replication series to generate data on reproducibility and its causes and monitor the effectiveness of measures to improve reproducibility	X	x		x
	Assessment of benefits and costs	Assess the overall desirability of replication studies in a discipline based on expected costs and benefits compared to alternatives	X	x		
		Make a careful assessment of the desirability of a replication study in individual cases	X			
	Education	Educate researchers in the history and philosophy of science and the essential role of replications	x			X
		Train researchers in assessing (replication) study designs and interpreting results, especially with respect to reproducibility	x			X
		Familiarise young researchers with replication studies through on-the-job learning, for example by conducting a replication study as part of their PhD training	X			x
	Best practices	Share best practices with other disciplines by publishing about them and discussing them, e.g. within scientific societies	X		x	x
Incentives	Positive culture	Commend replication activities through awards, prizes, and editorial comments in journals	X	X	X	X
		Report regularly on replication attempts, making replication studies a mainstream activity	X		X	
		Frame willingness to help with replication studies when needed as a good research practice	X			
		Have researchers share any specific materials, instruments, skills and other resources required to conduct a replication study	X			x
	Funding opportunities	See that replication studies comprise a sizable fraction of funding		X		
		Set up programmes that allocate money specifically to proposals for replication studies		X		
		Require replication to be part of individual research proposals		X		
		Monitor what proportion of funding goes towards replication efforts		X		
	Publication	Amend editorial policies to state that replication studies have a fair chance of being published			X	
		Encourage replication studies explicitly, e.g. through a tailored format			X	
		Publish at least one replication study of a previous report in that same journal			X	
	Career evaluations	Credit the efforts of researchers to improve reproducibility of research and conduct replication studies		x		X

*** An uppercase X indicates the main stakeholder for carrying out a proposal and a lowercase x indicates the involvement of additional stakeholders**

RECOMMENDATION 1

Improve study methods. Researchers should conduct research more rigorously by strengthening standardisation, quality control, evidence-based guidelines and checklists, validation studies and internal replications. Institutions should provide researchers with more training and support for rigorous study design, research practices that improve reproducibility, and the appropriate analysis and interpretation of the results of studies.

RECOMMENDATION 2

Improve study reporting. Funding agencies and journals should require preregistration of hypothesis-testing studies. Journals should issue detailed evidence-based guidelines and checklists for reporting studies and ensure compliance with them. Journals and funding agencies should require storage of study data and methods in accessible repositories.

RECOMMENDATION 3

Create proper incentives for reproducible research. Journals should be more open to publishing studies with null results and incentivise researchers to report such results. Rather than reward researchers mainly for 'high-impact' publications, 'innovative' studies and inflated claims, institutions, funding agencies and journals should also offer them incentives for conducting rigorous studies and producing reproducible research results.

RECOMMENDATION 4

Improve know-how about replication studies. Researchers should share best replication practices and the resources (e.g. methods, software, materials, samples, detailed analysis plans) required to conduct a particular replication study. Institutions should teach researchers how to design replication studies and assess reproducibility.

RECOMMENDATION 5

Create better incentives for replication studies. Funding agencies should increase funding for replication studies (e.g. by setting up programmes that allocate money specifically to replication studies and by requiring researchers to include replication in their individual proposals). Journals should encourage the submission of replication studies. Institutions should properly credit replication studies in career evaluations.

INTERNATIONAL PANELISTS HAVE BEEN ASKED TO FOCUS ON THE FOLLOWING TOPICS:

1. What is the level of awareness, interest, concern, and involvement in reproducibility and replicability (R&R) of research results within your national scientific societies?
2. Are there specific areas of science that are more likely to have issues with reproducing scientific results?
3. What R&R issues exist for cross disciplinary research?

KNAW COMMITTEE ON REPLICATION STUDIES

Prof. Johan Mackenbach (Professor of Public Health, Erasmus MC), *Chair*

Prof. Cock van Duijn (Professor of Genetic Epidemiology, Erasmus MC)

Prof. Harry Büller (Professor of Vascular Medicine, Academic Medical Center Amsterdam)

Prof. Aad van der Vaart (Professor of Stochastics, Leiden University)

Prof. Eric-Jan Wagenmakers (Professor of Neurocognitive Modelling, University of Amsterdam)

Dr Patricia Dankers (Associate Professor of Biomaterials, Eindhoven University of Technology)

Prof. Lex Bouter (Professor of Methodology and Integrity, VU University Amsterdam)

Dr Jean Philippe de Jong (KNAW), *Secretary*

CONTACT KNAW

ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES
Dr. Jean Philippe de Jong

- Senior policy officer medical sciences and scientific integrity
- Secretary to Council for Medical Sciences
- Secretary to Medical, Biomedical and Health Sciences Domain

- Het Trippenhuis, Kloveniersburgwal 29
- Postbus 19121, 1000 GC Amsterdam, The Netherlands
- +31 (0) 20 551 0762
- jean.philippe.de.jong @knaw.nl
- www.knaw.nl