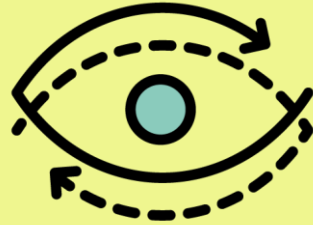


REPORT



TRUSTparency

RESULTS OF THE ASSESSMENT OF REPRODUCIBILITY INTERVENTIONS

Mariska Leeftang

Assessed reproducibility interventions

The assessment process led to a comprehensive list of assessed interventions, according to the stage of research, the intervention type and the specific intervention; and each intervention is accompanied by a row of assessment details and the overall judgment. See table below for an overview of all intervention types and their average judgement scores on evidence and acceptability. This score will eventually inform the decision on which intervention to choose of the end users of the toolbox.

In the coming months of the TRUSTparency project, we will test instances of some of these intervention types in pilot institutions and their feedback on the tools and interventions will help shape the description of the chosen intervention types and their implementability. The table below presents first draft.

Overview of all intervention types, the number of studies found and the judgment of the intervention types by both evidence and acceptability.

Intervention type	N	Evidence judgment [#]	Acceptability judgment ^{\$}
Journal policy on completeness of reporting: journal endorses or requires use of reporting checklists	18	Positive but weak effect; few (n=3) experimental designs; high heterogeneity. Low certainty	Burden lies with publishers; commonly used; high acceptability
Journal policy on data sharing: journal asks or demands sharing of data	30	Half positive effect; half no effect; no experimental designs; some heterogeneity.	Costs and effort depend on level of mandating; research may remain reluctant to share
Journal policy on registration: journal only publishes clinical trial when this was registered beforehand	7	Positive effect; no experimental designs; high heterogeneity. Low certainty.	Low costs and efforts; researchers may be reluctant
Journal awards open science badges: author team receives badge when they make their data, code or materials open and publicly accessible.	8	Positive effect; no experimental designs; high heterogeneity. Low certainty.	Low costs and efforts; effect of badges on reproducibility seems limited.
Results-free peer-review	1	Only 1 study; very low certainty of any effect.	Costs and efforts from the peer reviewers' side are low; for the journal it changes decision making
Outcome-agnostic acceptance by journal	1	Only 1 study; very low certainty of any effect.	Costs and efforts from the peer reviewers' side are low; for the journal it changes decision making
Funder policy on data management	3	Studies investigated NIH policy and found no effect. Low certainty.	Costs and effort depend on level of mandating; researchers may be reluctant
Research institution provides support services	2	Two very different non-comparative studies show a positive effect. Low certainty.	Requires infrastructure, staff and commitment. Potentially high costs.

Research institution provides training	3	Mixed effects; 2 comparative studies; heterogeneous types of training. Low certainty.	Acceptability depends on type of training; incentives may still be needed.
Research institution provides role models	1	Only 1 study; very low certainty of any effect.	Acceptable
Government policy on registration	2	Unclear effect in two very different studies. Before-after design. Low certainty.	Requires government to take action; no longer in the hands of the scientific community
Data repository	3	Three non-comparative studies mention a positive effect. Very low certainty.	Data repositories are already accepted measures to share data
Preregistration: researcher publishes a protocol of their study (before the results are known)	6	Weakly positive effect; no experimental designs; high heterogeneity. Low certainty.	Relatively cheap and effortless; training of referees may increase costs
Optimization of study design: A wide varying range of interventions: (1) standardization / adjustment of effect sizes; (2) best lab practices; (3) Python toolbox.	5	Positive or unclear effect; 1 experimental design; very variable. Too heterogeneous to make a statement.	Too heterogeneous to make a statement.
Statistical analysis plan	1	Only 1 study; very low certainty of any effect.	Acceptable, especially when not too much technical knowledge required.
Standardization of experimental protocol	2	Two articles of which one was a (descriptive) systematic review and the other an observational comparative study. The latter showed a positive effect. Low certainty.	Development of shared standards takes time, and may not be applicable in all contexts
Researcher trained on study protocol	3	Two studies introduced structured workflow next to training. Positive effect; low certainty.	Acceptability depends on type of training; incentives may still be needed.
Researcher uses standard method for data acquisition: standardized physical protocols, harmonized experimental procedures, and quality control frameworks	11	Positive effect; 1/11 experimental designs; high heterogeneity. Low certainty.	Diverse interventions, so varying acceptability.
Multi-site testing	1	Only 1 study; very low certainty of any effect.	We think this is not really an intervention
Standardized file system structure	3	Non-comparative studies; positive effect. Low certainty.	Acceptability may depend on type of structures and level of commitment
Data storage solution	3	Three non-comparative studies showing a positive effect. Low certainty	Acceptable, but may require funding and training
Direct data deposition	1	Only 1 study; unclear effect; very low certainty of any effect.	Based on one case study, it was difficult to judge acceptability.
Researcher uses standardized documentation framework	1	Only 1 study; very low certainty of any effect.	Overlaps with other standardizations

Software tool that documents data retrieval procedures	2	To very different studies show positive effect	Too heterogeneous to assess acceptability.
Real-time procedure documentation	1	Only 1 study; very low certainty of any effect.	Based on one case study, it was difficult to judge acceptability.
Standardized data preparation procedure: A wide varying range of interventions: (1) harmonisation of data; (2) automated preparation or reporting of data; (3) standardization -omics data and analyses	5	Low certainty; although it seems logical that standardized methods improve reproducibility, these studies do not prove that	Diverse interventions, so varying acceptability.
Data quality control	1	Only 1 study; very low certainty of any effect.	Depends a lot on exact control measurements and infrastructure
Researcher uses common method for data analysis: Coordinating research centres and software developers mandating shared standardized analytical algorithms to process data	19	Low certainty; although it seems logical that standardized methods improve reproducibility, these studies do not prove that. No experimental designs.	Depends on level of standardization and mandates; it asks from all collaborators to adhere to a fixed structure.
Tool to standardize data analysis procedures	2	Comparative before-after studies; positive effect; low certainty.	Unclear
Multiverse analysis	1	Only 1 study; very low certainty of any effect.	Is this really an intervention?
Authors use effect-size filter	3	Positive effects, but not based on comparative analyses. Low certainty.	Is this really an intervention; unclear acceptability.
Researcher uses standardized documentation framework	2	Positive effect, based on anecdotal evidence. Very low certainty.	Acceptable, as these frameworks rely on codebooks and existing repositories. But may require commitment
Open workflow platform	3	Positive effect in three very different (non-comparative) studies on very different interventions. Low certainty.	Using cloud computing can lead to dependence on third-party providers.
Researchers use tool to document analysis procedures	4	Positive effect in three very different (non-comparative) studies on very different interventions. Low certainty.	Low-cost, low-to-moderate-effort tools
Use of reporting checklist: use of a reporting checklist or writing aid by individual researchers	15	Overall effective, but weak effects; 3 (quasi) experimental designs; relatively homogeneous. Moderate certainty.	Passive guideline existence alone is cheap and easy, but also less effective than active or tool-supported interventions
Researchers share data	4	Retrospective non-comparative studies showing positive effect. Low certainty.	Acceptability depends on privacy issues and effort needed to make shared data really reusable.
Data management tool: variable tools that accommodate or support data analysis, data storage, FAIR-ization, etc	5	Positive effect; no experimental designs; high heterogeneity. Low certainty.	Too heterogeneous to make a statement.
Peer-reviewer checks reporting	4	Positive effect; three randomized experiments.	Especially a dedicated peer reviewer may cost

		Moderate-high certainty that this may have an effect.	time and effort; but overall it is feasible
Post-publication data requests	2	Comparative studies, but effect is unclear. This is a post-hoc intervention.	Low costs, low efforts; but takes action after publication
Discrepancy review	1	Positive effect in only 1 case-study. Very low certainty.	May cause discussions about study afterwards.
Replication	3	Unclear effect of a post-hoc intervention; non-comparative studies. Low certainty.	This does not improve the quality of the research prospectively.

Evidence judgment is a judgment based on the quality of the studies that reported on this intervention. The question was “How certain are we that this intervention improves reproducibility?”

\$ Acceptability judgment is based on a subjective indication of the costs, effort and required knowledge and infrastructure of the interventions. These may vary between settings and countries.

