Metrics for Reproducibility
A call for contributions

Contact: Rachel Heyard and Samuel Pawel (Center for Reproducible Science, University of Zurich)

Background and aim of the study

One of the objectives of the Horizon Europe project iRISE (improving Reproducibility In SciencE) is a review of the metrics to quantify reproducibility that are currently used or were suggested. Hereafter we use reproducibility as an overarching term for aspects including computational reproducibility, replicability, translatability, and generalizability. A detailed protocol of our review has been uploaded to the Open Science Framework (osf.io/j65wb). The first part of our review will investigate which metrics have been **used** to quantify reproducibility. A systematic search for all papers using a certain metric or a set of metric to quantify a specific type of reproducibility was judged unfeasible. Therefore, we decided to collect large-scale effort to quantify the reproducibility of, for example, a whole field of studies. Such large-scale reproducibility projects are specifically interesting as many of them used a whole set of metrics. We now need help by the larger research community to identify these projects.

Large-scale reproducibility projects - Definition

We are not interested in not single efforts to reproduce or repeat part or all of an "original" study or finding. Instead, we want to collect larger projects where a group or consortium of researchers attempts to reproduce or repeat a set of original studies, or the same original study several times. These include the large-scale replication projects in psychology, experimental economics, and other fields, but also so-called many-lab projects [REFS]. As meta-researchers experienced with the design and the evaluation of replication studies, we are specifically interested in projects concerned with other types of reproducibility which we are not aware of. Such projects could include, but are not limited to, efforts to reproduce the coding of qualitative data, to translate or generalise the effects of an intervention in another population, to reproduce the analysis code solely based on the methods description, and XXX.

To qualify as a large-scale reproducibility project, the project team should on top of conducting the set of reproducibility studies and attempts additionnally seek summarising the results of the set of studies. The summarising procedure will be of special interest to us.

Note that, eventhough they are very valuable, databases with a collection of single-study attempts at testing the reproducibiltiy of one finding do not qualify as large-scale reproducibility projects.

Submit reproducibility projects

We need your help to point us to large-scale reproducibility projects (as defined above) that we do not yet know about. A list with the names, description and links of projects already collected can be found here [link to public list]. To submit a project that is not on our list yet, please fill out this survey [link]. To submit several projects you can fill in the survey multiple times. If you want to, you can provide your name and contact details at the end of the survey to help us trace back the origin of the collected data. Providing this information is completely optional and we will not share it at any time.

Get involved in data extraction, and screening for second part of our study

Once the projects were collected our team will start extracting the information on the reproducibility metrics used by the projects. For this task, we would benefit from additional help. Further, the second part of our review involves systematic search for methodological papers suggesting reproducibility metrics. Also for the screening and the data extraction of this part of the review we would appreciate

help. Substantial contributions to screening an data extraction will be credited with authorship on the final paper summarising our results. If you want to contribute, please let us know via email (let us know if you have prior experience with reviews and/or the topic).