Data Verification and Reproducibility

Research verification involves testing claims of prior research to determine if the research was performed and produced accurate results. Research verification techniques include:

- Research replication: Testing the same claim from a prior study with different data.
- Research reproduction: Testing the same claim from a prior study using the same data and same model or analysis strategy.
- Research robustness**:** Testing the same claim from a prior study using the same data but a different model or analysis strategy.

Sometimes data replication and reproduction are seen as interchangeable, but they are 2 different concepts, replication is results orientated, whilst reproduction is method orientated.

Stages of Verification

- Formulating a hypothesis
  Documenting the central thesis and creating a Data Management Plan
- Designing the study
  Using version control in a study plan or protocol, calculate power or sample size as underpowered studies are prone to irreproducibility. Register the study on a platform such as Protocol.oi
- Running the study
  When collecting anonymised data try and share it in a repository. Share materials and use an electronic lab notebook. Share code and software.
- Analysing the data
  Use software such as R or Python and attempt to automate the process as much as possible. Turn scripts into reports and code into functions.
- Reporting the Study
  Publishing in an Open journal from PLOS or BMJ, which links to original study. Use preprints.

Does it make better research?

Research Verification is a key driver for open research and changing research culture. But does it actually make research better?

It confronts established scientific publishing which has tended to select only positive results or novel ideas without the replicability to back it up.

Both reproducibility and replicability look to verify the reliability of an effect, forming a solid foundation on which further scientific knowledge can be built.

Data verification brings transparency to scientific endeavour and research is more robust and reduces Data Hedging. This, also known as P-haking , is repeatedly searching a dataset or trying alternative analysis until the ‘right’ result is found.

There are lots of resources to learn more, including the Equator Network, UKRN and the Open Science training handbook and Centre for Open Science

Much is made of the reproducibility crisis and data verification is a way of making reproduction easier. But research even if the same data is used, adopting the same methods may not lead to the exact same result being produced. This can be due to several aspects, including different computational environments, differences in the software versions and implicit biases that were not clearly stated.

However, some reproducibility and replication is better than none.