BIH QUEST
Center for Responsible Research

DLR Projektträger
Deutsches Zentrum für Luft- und Raumfahrt e.V.

Preclinical Confirmatory Studies and Systematic Reviews

Glossary

Term	Definition	Reference
Accessibility of results	The degree to which results are accessible e.g. for other researchers, see also Open Access and Open Data	-
Between-lab replication	The process of performing a replication in a different laboratory setting from the original experiment. Even if the experimental design is unchanged, unknown sources of variation will be added to the replication and may contribute to generalizability of the findings (e.g., animal husbandry).	
Bias	Systematic error in an estimate (e.g., variance, effect size) caused by violations of assumptions, inadequacies in the design, conduct, or analysis of an experiment.	https://catalogofbias.org/
Blinding	Procedure to prevent performance bias. Will ensure that experimenters are unaware of the intervention condition during outcome assessment and analysis	https://catalogofbias.org/biases/performance-bias/
Comparator	Comparator group acts as a control with an effective intervention (as opposed to a placebo group). Allows comparison of effect sizes between two effective treatments.	Active comparator arm: https://clinicaltrials.gov/ct2/about-studies/glossary

Term	Definition	Reference
Confirmatory study	A study aimed at corroborating empirically specific relationships between defined factors, based on previous (exploratory) observations. Results from confirmatory studies may verify a hypothesis.	
Construct Validity	Extent to which a measured outcome is representative of the phenomenon studied	https://doi.org/10.1371/journal.pmed.1001489
Control (negative, positive control)	A baseline comparison group that receives no or placebo treatment (negative) or an effective comparison treatment (positive; see comparator).	-
Convergent evidence	Convergent evidence refers to the ability of different measuring instruments to show similar effects in the same direction. (For example, measuring upregulation of mRNA should also result in a protein expression as measured by PCR and AQUA.)	
Data management	Protocol describing data storage, data sharing, data preservation, and data description	https://www.dcc.ac.uk/dmps
Data security	Models to ensure management of threats to data privacy, storage, longevity	-

Term	Definition	Reference
Data sharing	Making data accessible and reusable for others to different degrees by providing direct links to repositories or describing procedures through which data can be obtained.	-
Data synthesis strategy	Strategy to combine data from various sources into one estimate of an effect (e.g., strategy for a meta-analysis)	-
Design of a study	The setup of a study, specifically what interventions take place, what measurements are taken, what procedures are implemented to reduce bias	https://www.nc3rs.org.uk/experimental-design-assistant-eda
Discriminant evidence	Discriminant evidence refers to measurements that rule out alternative explanations for the current evidence. That is asking the question of what kind of data would be necessary to disprove my hypothesized cause-effect relationship.	
Dissemination of results	The way results are communicated and shared with peers and a broader public. See also Open Access	https://research.uq.edu.au/research-support/ethics-integrity-and-compliance/research-integrity/publication-and-dissemination-research https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007704

Term	Definition	Reference
Effect and Effect Size	By the clinician (or biomedical scientist), a relevant effect may be postulated. From this an effect size is derived considering a variation when observing this effect. This effect size refers to a statistic which estimates the magnitude of an effect. In this context, the Smallest Effect Size of Interest (SESOI) is the smallest effect size that is considered theoretically and/or practically interesting and can be taken to justify the sample size for a given experiment. To determine the SESOI, previous evidence as well as practical aspects (e.g., feasibility) may be considered. The SESOI might be informed by a biologically relevant effect (size) that can be defined as an effect considered by expert judgment as important and meaningful for human, animal, plant or environmental health. It therefore implies a change that may alter how decisions for a specific problem are taken. In that line, a clinically relevant effect size denotes the minimal average improvement in outcome of interest to clinical decision-makers.	https://doi.org/10.1111/j.1469-185X.2007.00027.x doi:10.31234/osf.io/9d3yf https://doi.org/10.1101/2022.01.17.476585 https://onlinelibrary.wiley.com/doi/10.1002/sim.4780050103
Efficacy of the intervention	Outcome measure of the effect of interest. See also Effect Size.	-
Experimental Unit	The experimental unit (EU) is the entity that is randomly and independently assigned to experimental conditions. This is equal to the sample size (N).	https://doi.org/10.1371/journal.pbio.2005282
Exploratory study	Studies that search for novel biological associations and phenomena without having previous assumptions or hypotheses, often employing less rigorous experimental designs, low sample sizes. An exploratory study may also be used to identify confounders (e.g., physiological parameters relevant to the	https://doi.org/10.1371/journal.pbio.1001863

Term	Definition	Reference
	research question). The result of an exploratory study may allow for the generation of a hypothesis that can later be tested in a confirmatory study.	
External validity	External validity refers to the generalizability of a study and whether its findings are applicable or transferable to a broader context. External validity is potentially increased via replications between labs, tests in additional (animal) models, and consideration of both sexes. Whereas the introduction of heterogeneity or multimodal studies can increase external validity, this can at the same time be a trade-off to internal validity. Vice versa, idiosyncrasies of a given experimental setting might reduce the generalizability of the study.	
Extraction of data from primary studies	Process of extracting the relevant information to a synthesis from individual studies (e.g., an effect size to be included in a meta-analysis)	-
FAIR criteria	Acronym for data sharing standards: Findable, Accessible, Interoperable, Reusable.	http://www.nature.com/articles/d41586-019-01720-7
False positive	Statistically significant result obtained by chance when the effect being investigated does not exist.	https://doi.org/10.1038/nmeth.2738
False negative	Statistically non-significant result obtained when the effect being investigated genuinely exists.	https://doi.org/10.1038/nmeth.2738

Term	Definition	Reference
Feasibility	Likelihood that an experiment can be conducted under the given constraints (number of experimental units available, laboratory equipment, time, personnel)	
Independent variable of interest	Factor that a researcher manipulates within a controlled environment in order to test its impact on the outcome measured. Also known as: predictor variable, factor of interest.	https://learningstatisticswithr.com/book/studydesign.html
Internal validity	Internal validity refers to how far measurements in an experiment reflect causal conclusions or mechanisms. That is, do we really measure what we want to measure? A proper experimental design will aim for high internal validity reducing potential risks of bias by introducing for example randomization and blinding.	
Knowledge claim	"A formal knowledge claim [refers to] a statement that a research project/study has established new knowledge, or consolidated existing knowledge, with sufficient certainty that that knowledge can now be acted upon. The required level of certainty might depend on the nature (risk and potential benefits) of the possible action." Knowledge claiming research refers to studies for which any outcome would be considered diagnostic evidence about a claim from prior research.	https://doi.org/10.7554/eLife.63294 https://doi.org/10.1371/journal.pbio.3000691
Long-term safeguarding of research data	Preservation of data over long time scales (at least 10 years). See also data security, data management	https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf

Term	Definition	Reference
		https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/guidelines_research_data.pdf
Metadata	Data about experimental datasets outlining content, timestamp, authors, explanation of variables.	https://doi.org/10.1038/sdata.2016.18 https://www.nature.com/articles/sdata201618 https://www.jmir.org/2022/1/e25440/
Methods against bias	Procedures to reduce bias in a study like randomisation and blinding but also improved statistical methods	-
Multicenter study	Procedure where multiple labs collect data for the same study design to investigate variability between labs to estimate generalisability of results	https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2003693
Nuisance variable	Sources of variability or conditions that could potentially bias results. Also known as: confounding factor, confounding variable	https://learningstatisticswithr.com/book/studydesign.html
Null Hypothesis Significance Testing	NHST is a method of statistical inference to assess the probability that the observed data (or more extreme) will occur given a null hypothesis of no effect of our variable of interest.	https://www.nature.com/articles/nmeth.2698

Term	Definition	Reference
Observational (or measurement) unit	The observational unit (OU) is the entity on which observations (or measurements) are made. If observational units are not the experimental unit and thus mistaken as sample size (N), pseudo-replication is introduced.	https://doi.org/10.1017/9781139696647
Open access	Freely available research articles for anyone, no subscription needed.	https://doi.org/10.12688/f1000research.8460.3
Open data	Freely available and reusable data, without barriers see FAIR criteria	https://book.fosteropenscience.eu/en/02OpenScienceBasics/02OpenResearchDataAndMaterials.html
Outcome measure	Any variable recorded during a study to assess the effects of a treatment or experimental intervention. Also known as: dependent variable, response variable	https://learningstatisticswithr.com/book/studydesign.html
Outlier	Outliers are data-points which are thought to present failures in measurement since they are too extreme to present a valid response to an intervention. Conventionally, univariate outliers are defined in relation to the central tendency and measure of dispersion of an outcome variable. The threshold values employed (e.g., 1.5 times the interquartile range) should be calculated from the distribution of a variable across all experimental groups and not separately for each experimental group to avoid type I error inflation.	https://doi.org/10.31234/osf.io/fqxs6
Statistical Power	Probability that a significance test detects an effect (i.e. a deviation from the null hypothesis), if an effect exists (i.e. true positive result).	https://doi.org/10.1177/0956797616647519

Term	Definition	Reference
Preclinical research trajectory	A preclinical research trajectory comprises a cumulative series of experiments (including exploratory and confirmatory) that generate evidence to enable a decision to carry a newly developed intervention forward to clinical testing.	https://prod--journal.elifesciences.org/articles/62101/figures#content
Pre-registration of the study, pre-registration of study protocol	Formal registration of methods, experimental design, and analysis procedures prior to data-collection	https://doi.org/10.1177/1745691618771357
Preprint	Sharing of a manuscript prior to peer-review mostly on dedicated repositories like BioArxiv.	https://www.plos.org/why-preprint
Publication bias	Studies remain often unpublished when an outcome is contrary to the initial hypothesis. This leads to an overrepresentation of positive findings impeding a meta analytic assessment of the investigated effect.	https://catalogofbias.org/biases/publication-bias/
Quality management	Procedures in a laboratory to increase study quality through standard operating procedures that ensure rigorous experimental design and reduction of measurement errors.	https://doi.org/10.15252/embr.201847143
Quality-assessment of primary studies	Retrospective procedure to evaluate whether studies follow reporting standards and whether they are at a risk of bias.	https://training.cochrane.org/handbook/current

Term	Definition	Reference
Randomisation	Assignment procedure to control for random variations between experimental units included in a study.	https://eda.nc3rs.org.uk/experimental-design-allocation
Registered Report	Type of article that undergoes a two-stage review. Before data collection reviewers assess the scientific premise and in depth methodology including an analysis plan. Upon acceptance in stage 1, articles are in principle accepted independent of the results. After data collection, reviewers only assess whether experiments were conducted according to the plan from stage 1.	https://www.nature.com/articles/s41562-021-01193-7
Reliability	Reliability refers to the consistency in a measurement. In a broader scientific context, this means that a result is reliable if it is consistently replicated. A reliable measurement is not necessarily valid; for example, if results are reproducible, but do not reflect the studied disease pathology. The sample size directly influences the reliability of a result as uncertainty about the measured effect is (in most cases) decreased with increased sample size.	
Replicability (sensu results reproducibility)	Replicability is the ability to obtain similar results by repeating an experiment with the goal of confirming previous empirical evidence. The replication experiment can use either the same or closely resembling methods (also referred to as direct replication), or it could include additional controls and/or conditions, a larger number of samples or animals than in the original study, with the aim of increasing and reaffirming the reliability of previously observed results. If the replication is performed in the same lab, it is referred to as within-lab replication; if at a different lab, it is referred to as between-lab replication. In the framework of preclinical research trajectories, replications are considered part of a confirmation process. Given that a series of experiments is needed to	https://doi.org/10.1371/journal.pbio.3000691 https://doi.org/10.1126/scitranslmed.aaf5027

Term	Definition	Reference
	confirm a hypothesis about a directional relationship, replications incrementally solidify evidence supporting (or refuting) an initial claim.	
Repository	Online database where information like experimental data or analysis code is deposited and archived	-
Reproducibility (sensu results reproducibility)	Reproducibility refers to the understanding of experimental procedures and analyses in such detail that researchers can engage in a replication. Through this, researchers can distinguish between variability in results that arise from methodological differences and variability due to sampling variability.	
Rigor	The strict application of the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation and reporting of results. This includes full transparency in reporting experimental details so that others may reproduce and extend the findings.	https://grants.nih.gov/reproducibility/faqs.htm#4828
Sample size	Number of experimental units per group, also referred to as N number. See also Statistical Power and Sample Size Determination	-
Sample size determination	Procedures to determine how many experimental units to include in a design. Depends on power, effect size, and significance level	https://doi.org/10.1038/nmeth.2738

Term	Definition	Reference
Scientific premise	Rigor of the prior research: strengths and weakness of the data and previously performed work upon which a study is built upon	https://nexus.od.nih.gov/all/2016/01/28/scientific-premise-in-nih-grant-applications/
Search strategy	Strategy to search databases for relevant studies, typically done systematically with a pre-specified set of searches and search terms.	-
Sharing of data	The extent to which data is made available to others, see also Open Data	-
Study quality	The assessment of a study's quality with respect to design, measurement, question	-
Systematic heterogenization	The process of introducing biological variation in the experimental design with the goal of increasing external validity and, consequently, amplifying the inference space. Although there is no consensus on which factors should be systematically heterogenized, suggestions of feasible and relevant variables include sex and age of the animals used and the timing of experiment. Systematic heterogenization can be introduced in exploratory and/or in confirmatory stages of the preclinical research trajectory but strategies for designing and analysing these experiments should be determined based on its objectives.	https://doi.org/10.1038/s41583-020-0313-3

Term	Definition	Reference
Translational validity	Translational validity refers to the extent to which a scientific finding can be translated from preclinical to clinical contexts. For this, knowledge of limitations and characteristics of the studied (animal) model is crucial. Moreover, clinical biomarkers, pharmacodynamics/pharmacokinetics, drug dosing can be used as proxies to estimate how well the experimental evidence reflects the human condition and identify limits.	
Triangulation	Triangulation refers to the combination of theories, methods, or observers in a research study to increase the trustworthiness and validity of research findings. Herein, we refer to methodological triangulation in which different data collection methods (flanking experiments) are used to support findings from a (primary) outcome measure. Flanking experiments can confirm or refute a scientific hypothesis where one set of findings confirms, increasing the credibility of research findings, or refutes another set. Triangulation can also be a powerful tool to explain differing aspects of a mechanism of interest.	https://www.nature.com/articles/d41586-018-01023-3 https://ebn.bmj.com/content/22/3/67
Within-lab-replication	The process of performing a replication using the same team and infrastructure from the original experiment. It implies that the experimental conditions will be very similar between both experiments, so it should be used to increase reliability of the results but not to increase external validity, unless changes in the experimental design are implemented.	