**Using measurement overlap to find significant results**

*Definition*

Overestimating the magnitude of an effect that is largely due to measurement overlap of two seemingly distinct constructs.

*Alias(es) and related concepts*

Leveraging the jangle fallacy

*Umbrella term(s)*

None

*Examples*

1. Researcher is using similar items in two seemingly different constructs and concludes that the constructs have a high correlation.
2. Researcher finds a positive relationship between depression and suicidality, but the positive relation is in part due to the depression scale already containing items about suicide.
3. A researcher estimates a positive relation between construct A and construct B in a meta-analysis, but the meta-analysis includes studies that have used the same task to operationalize either construct.

*Potential damages*

* Inflated type I or type II error
* Inflated effect size estimates
* Reduced replicability
* Misleading conclusions

*Remedies*

* Be transparent about similar items
* Consider whether the measures used in a study distinctly operationalize different constructs
* Account for covariance due to similar items

*Detectability*

Yes

*Clues to detect*

* Items or tasks are similar for the correlated constructs

**Performing inappropriate power analysis**

*Definition*

Selection of inappropriate parameters or methods for power analysis and/ or misinterpreting/misusing power analysis results.

*Alias(es) and related concepts*

*Umbrella term(s)*

None

*Examples*

1. A researcher chooses a large sample size to get negative results in a study about the harmful effects of smoking.
2. A researcher uses default parameters for the power analysis to show that power analysis had been performed, however, the analysis is uninformative.

*Potential damages*

* Inflated type type II error
* Reduced replicability
* Misleading conclusions
* Inflated credibility of the study

*Remedies*

* Being transparent about how power analysis was conducted
* Use meaningful parameters that are based on field standards, literature, or common sense

*Detectability*

Yes

*Clues to detect*

* Relatively low sample size
* Details of power analysis not reported or justified
* Parameters of the power analysis are generic and do not fit the study
* The results of power analysis are misinterpreted
* Negative results are reported in a controversial field where others find positive results

**PARKing**

*Definition*

Pre-registering a hypothesis or analysis after knowing the outcome of the analysis.

*Alias(es) and related concepts*

“Pre-registering after the results are known”

*Umbrella term(s)*

Faking open science

*Examples*

1. Researcher realizes that their paper won’t be accepted without a pre-registration, so they create one post-hoc and link it to their study.

*Potential damages*

* Inflated credibility of the study

*Remedies*

* Pre-register before analyzing the data
* Being transparent about when the pre-registration was done
* Disclose the familiarity (if any) of the researcher with the data

*Detectability*

Yes

*Clues to detect*

* Data collection occurred previous to the pre-registration date
* Date of pre-registration is unrealistically close to the first submission without the paper being a registered report

**Selective sampling**

*Definition*

Collecting data in a way that biases the findings.

*Alias(es) and related concepts*

Purposive sampling

*Umbrella term(s)*

None

*Examples*

1. Researcher tests the likeability of chocolate on a group of children only in order to find that everyone loves it.
2. Using uncomparable groups: The researcher tests if men are more aggressive than women. For comparison, women from a university are compared with men from a prison.
3. Picking a subsample of a panel dataset to find the desired results.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Misleading conclusions
* Reduced generalizability
* Reduced replicability if sampling bias is not disclosed

*Remedies*

* Use a sampling method that doesn’t bias the results
* Make sure that the sample represents the population
* Use comparable (or matched) groups
* Report the sampling process transparently
* Consider using statistical control for confounding variables
* Pre-register the sampling process

*Detectability*

Yes

*Clues to detect*

* Unclear rationale for sample selection
* Compared groups are from different populations
* Convenience sample

**Grooming participants**

*Definition*

Affect participants to make them give responses that support the desired narrative.

*Alias(es) and related concepts*

*Umbrella term(s)*

Influencing participants

*Examples*

1. Researcher tells the participants that he believes the treatment will work.
2. Researcher uses a briefing document that is trying to influence participants’ attitudes about the topic that is assessed in the study.
3. During data collection the same organization that runs the study also runs a marketing campaign to influence public opinion on the same topic.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Misleading conclusions
* Reduced replicability

*Remedies*

* Blind the experimenter where possible
* As a researcher interacting with participants, remain neutral and follow scripts during testing
* Use automated research procedures (e.g. research presentation software) instead of human research assistants wherever possible
* Avoid exposing participants to any cues that might influence their responses

*Detectability*

No

*Clues to detect*

* Absence of blinding
* Absence of procedure scripts
* Suggestive questions in the survey

**Using biased measurements**

*Definition*

Assessing a construct with an instrument that is biased or invalid, in order to support the desired narrative.

*Alias(es) and related concepts*

*Umbrella term(s)*

Influencing participants

*Examples*

1. Researcher uses loaded, leading, or suggestive questions, e.g., “Does the lack of respect schoolchildren have for their teachers, in your opinion, influence everyday teaching methods in schools?”
2. Researcher assesses a psychological construct using an ad hoc questionnaire with no proven validity.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Misleading conclusions
* Reduced replicability

*Remedies*

* Frame questions in a neutral way
* Use questionnaires that are unbiased and psychometrically sound
* Get external opinions on questionnaires/study materials OR Registered report

*Detectability*

Yes

*Clues to detect*

* Ad hoc questionnaires are used instead of validated instruments
* No discussion of the psychometric properties of the instruments
* Measurement items use suggestive or biased language
* Measurement is based on single-item questions

**Optional stopping**

*Definition*

Monitoring hypothesis tests during data collection, and stopping when statistical inference is favorable, without controlling for sequential testing.

*Alias(es) and related concepts*

Peeking, Data peeking

*Umbrella term(s)*

None

*Examples*

1. Researcher is collecting responses and tests the hypothesis after every participant - when significance is reached, the researcher stops collecting data.

*Potential damages*

* Inflated type I error
* Reduced replicability
* Misleading conclusions

*Remedies*

* Pre-register the estimated sample size
* Pre-register stopping rules and adjustments for type I error-inflation

*Detectability*

Maybe

*Clues to detect*

* Vague or absent reason for sample size
* Low sample size
* Absence of pre-registration
* P-values are just below the significance threshold (usually 0.05)
* Relatively large effect size compared to other studies in the field

**Retaining pilot data**

*Definition*

Including data from a pilot study if the results support the hypothesis.

*Alias(es) and related concepts*

Double dipping

*Umbrella term(s)*

None

*Examples*

1. Researcher conducts a pilot study to check the protocol and analyzes pilot data. The pilot data and main study data are aggregated in the data analysis if pilot data results are in line with expectations.

*Potential damages*

* Inflated type I or type II error
* Reduced replicability
* Inflated or deflated effect size estimates
* Misleading conclusions

*Remedies*

* Don’t include pilot data in the final data set
* If for resource reasons you need to include pilot data report pilot data separately

*Detectability*

No

*Clues to detect*

None

**Missing data hacking**

*Definition*

Choosing the strategy to handle missing data based on the impact on the results.

*Alias(es) and related concepts*

Favorable imputation

*Umbrella term(s)*

P-hacking

*Examples*

1. A researcher tries three ways of handling missing data, for example, listwise deletion, multiple imputation, and inverse probability weighting. The expected results only appear with inverse probability weighting. The researcher reports only this strategy in the paper and leaves out results with listwise deletion and multiple imputation.
2. Can also be within a single method, specifically multiple imputation, since it uses one or more variables to replace missing data, and the choice of these variables is up to the researcher, but can also be statistically based.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Misleading conclusions
* Reduced reproducibility
* Reduced replicability

*Remedies*

* Pre-registration of missing data approach
* Perform sensitivity analysis

*Detectability*

Maybe

*Clues to detect*

* No mention of the missing data approach or the missing data at all
* Lack of rationale or references supporting the missing data approach
* Unexplained discrepancy between the recruited and analyzed sample sizes
* Absence of open data

**Using ad hoc exclusion criteria for participants**

*Definition*

Exclusion of participants without proper justification and transparent reporting.

*Alias(es) and related concepts*

*Umbrella term(s)*

P-hacking

*Examples*

1. Researcher finds that a correlation between two variables is not significant. After removing two participants - who should be included - the association becomes significant. Then the researcher comes up with post hoc exclusion criteria for those participants.
2. A researcher doesn’t find an expected association between perceived stress and personality. When looking only at the top 25% of perceived stress scores, the association is there. They go on to report the top 25% scores as their population of interest and do not disclose that they looked at the rest of the sample population.

*Potential damages*

* Inflated type I or type II error
* Reduced generalizability
* Reduced replicability
* Reduced reproducibility
* Inflated or deflated effect size estimates
* Misleading conclusions

*Remedies*

* Pre-registration
* Use valid reasons to exclude participants
* Report post hoc changes in exclusion criteria
* Perform sensitivity analysis

*Detectability*

Maybe

*Clues to detect*

* Absence of pre-registration
* Absence of open data
* Unexplained discrepancy between the recruited and analyzed sample sizes
* Sample too narrow for recruitment methods

**Excluding data points**

*Definition*

Exclusion of data points or outliers without proper justification and transparent reporting.

*Alias(es) and related concepts*

*Umbrella term(s)*

P-hacking

*Examples*

1. Removing individual reaction time trials based on post hoc criteria.
2. Trying different outlier cut-off criteria until an effect is statistically significant.

*Potential damages*

* Inflated type I or type II error
* Reduced generalizability
* Reduced replicability
* Reduced reproducibility
* Inflated or deflated effect size estimates
* Misleading conclusions

*Remedies*

* Pre-registration
* Use valid reasons to exclude data points or outliers
* Report post hoc changes in exclusion criteria
* Perform sensitivity analysis

*Detectability*

Maybe

*Clues to detect*

* Absence of pre-registration
* Absence of open data
* Unexplained discrepancy between the recruited and analyzed sample sizes and degrees of freedom

**Redefining group membership rules**

*Definition*

Post hoc (re)definition of grouping criteria without proper justification and transparent reporting.

*Alias(es) and related concepts*

*Umbrella term(s)*

P-hacking

*Examples*

1. Collapsing the multicategorical variable of sexual orientation into heterosexual and non-heterosexual.
2. Trying different age ranges in cross-sectional age comparisons to maximize group differences.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Reduced replicability
* Reduced reproducibility
* Misleading conclusions
* Reduced generalizability

*Remedies*

* Pre-registration
* Use valid reasons to define group membership
* Report post hoc changes in grouping rules and report results using original grouping rules as well
* Perform sensitivity analysis

*Detectability*

Maybe

*Clues to detect*

* Absence of pre-registration
* Absence of open data
* Response options in materials/methods different than reported groups
* Oversimplified sample description

**Discretizing continuous variables**

*Definition*

Taking a continuous variable and making it categorical without proper justification and transparent reporting.

*Alias(es) and related concepts*

Dichotimizing variables, Median split

*Umbrella term(s)*

P-hacking

*Examples*

1. Researcher doesn’t find an association between depression and continuous age variables, and recodes age into young and old categories. After that, age groups show a significant association with depression. An independent samples t-test is reported instead of a correlation.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Reduced replicability
* Misleading conclusions
* Reduced generalizability

*Remedies*

* Pre-registration
* Use original measurement levels
* Discretize only with strong justification
* Perform sensitivity analysis

*Detectability*

Yes

*Clues to detect*

* Absence of open data
* Scale or response options in materials or methods do not match how they are reported in the results
* Test statistics do not match expected data analysis strategy

**Modifying measurements**

*Definition*

Changing the properties of a measure/measurement to produce favorable results without proper justification and transparent reporting.

*Alias(es) and related concepts*

*Umbrella term(s)*

P-hacking

*Examples*

1. Researcher uses only a portion of the items from a longer scale.
2. Researcher combines items from different scales into a single measure.
3. Researcher chooses which EEG electrodes to aggregate based on the results.

*Potential damages*

* Reduced replicability
* Reduced reproducibility
* Reduced content validity of the measure
* Reduced construct validity of the measure
* Inflated or deflated reliability of the measure
* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Misleading conclusions

*Remedies*

* Publish study materials
* Pre-registration
* Use conventional measurements/measures
* Describe and justify any modifications on measurements
* Perform sensitivity analysis

*Detectability*

Maybe

*Clues to detect*

* Absence of open data
* Absence of open study materials
* Discrepancy between the reported version of measurement/measure and original or conventional measure

**Variable transformation fishing**

*Definition*

Selecting variable transformations that produce favorable results without proper justification and transparent reporting.

*Alias(es) and related concepts*

*Umbrella term(s)*

P-hacking

*Examples*

1. Researcher runs a statistical test using several different transformations (e.g., changing levels of measurement, log-transformations, rescaling) of the outcome, and only reports the one that produces a significant result.

*Potential damages*

* Reduced replicability
* Reduced reproducibility
* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Misleading conclusions

*Remedies*

* Pre-register under which conditions transformations will be performed
* Describe and justify any variable transformations
* Blinded data analysis
* Perform sensitivity analysis

*Detectability*

Maybe

*Clues to detect*

* Abscence of open data
* Using transformations that are unconventional for the measure
* Transformation is applied without justification
* Reported values are outside of regular range

**Using ad hoc covariates**

*Definition*

Addition or removal of covariates to influence the estimates or significance for the effect of interest.

*Alias(es) and related concepts*

Selectively including control variables

*Umbrella term(s)*

P-hacking

*Examples*

1. A researcher opportunistically decides which background variables (e.g., age, gender) to control for, without a causal theory or a pre-registration.
2. A researcher decides whether to control for a baseline value in an experimental design depending on the results of statistical tests.
3. Researcher avoids the inclusion (or measurement) of theoretically justified moderators (e.g., severity of a condition, or socio-economic status) to be able to imply greater generalisability.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Reduced replicability
* Reduced reproducibility
* Misleading conclusions
* Reduced generalizability

*Remedies*

* Provide reasonable arguments for inclusion/ exclusion of covariates
* Pre-register complete models/analytical plan
* Report robustness checks
* Follow clear guidelines

*Detectability*

Maybe

*Clues to detect*

* Different covariates in different analysis steps are used
* There is a lack of justification for the selection of the covariates

**Choosing unjustified p-value adjustment**

*Definition*

Not adjusting or over-adjusting p-values when running multiple tests.

*Alias(es) and related concepts*

Not adjusting or over-adjusting p-values

*Umbrella term(s)*

P-hacking

*Examples*

A researcher decides (1) whether or not to adjust for multiple tests (e.g., in an ANOVA) and (2) which adjustment method to use, and (3) which (i.e. how many) comparisons to include depending on results obtained. (4) Researcher uses Bonferroni correction when correlating several variables, to prove that an association does not exist.

*Potential damages*

* Inflated type I or type II error
* Misleading conclusions

*Remedies*

* Pre-register p-value adjustment plans
* Pre-register planned contrasts
* Perform sensitivity analysis

*Detectability*

Yes

*Clues to detect*

* P-value adjustment not mentioned
* Multiple tests are made that would require p-value adjustment
* Using p-value correction that is too strict (e.g., Bonferroni) without proper justification

**Choosing a poor model specification**

*Definition*

Creating too complex models on too small datasets causing the model to learn the noise and random fluctuations instead of generalizable patterns. Alternatively, creating too simple models that do not adequately fit the data.

*Alias(es) and related concepts*

Overfitting or underfitting models, Bias-variance tradeoff

*Umbrella term(s)*

P-hacking

*Examples*

1. Overfitting: The researcher fits a regression model with 25 predictors on a sample of 100 participants.
2. Underfitting: The researcher uses linear regression to investigate a non-linear association.

*Potential damages*

* Inflated type I or type II error
* Inaccurate predictions
* Misleading conclusions

*Remedies*

* Use a theoretically justified model in confirmatory studies
* Underfitting: Visualize data and the model
* Use methods that prevent overfitting in exploratory research, e.g. use separate train and test datasets, use cross-validation resampling methods, use regularization or other feature selection methods
* Perform sensitivity analysis

*Detectability*

Yes

*Clues to detect*

* The number of observations is low
* The number of included predictors in the model is large
* No mention of holdout (or test) dataset or cross-validation
* Overfitting: Very high R2 value (close to 1)
* Improper prediction selection (no regularization)
* Underfitting: Data visualization shows high model bias

**Neglecting assumptions for statistical models**

*Definition*

Using statistical models although requirements are not met.

*Alias(es) and related concepts*

*Umbrella term(s)*

P-hacking

*Examples*

1. Analyzing data, using parametric tests such as t-tests but the data requires a non-parametric test.
2. Analyzing dependent data using a statistical model that does not account for dependency.

*Potential damages*

* Inflated type I or type II error
* Reduced replicability
* Reduced reproducibility
* Misleading conclusions

*Remedies*

* Perform and report necessary assumption checks

*Detectability*

Yes

*Clues to detect*

* Not reporting assumption checks
* Evidence of assumption breaches (e.g., non-normality, non-independent data, largely different SDs by group)

**Selecting a favorable random number generator seed**

*Definition*

Trying different random seeds until getting a favorable result, potentially in combination with small number of replications

*Alias(es) and related concepts*

Resampling lottery

*Umbrella term(s)*

P-hacking

*Examples*

1. A researcher keeps on bootstrapping a confidence interval (e.g., for a mediation indirect effect) with different seeds until the 95% confidence interval just excludes 0.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Reduced replicability
* Reduced reproducibility
* Misleading conclusions

*Remedies*

* Only report significance if the results are robust across random seeds
* Use a large number of replications (e.g. bootstrap samples)

*Detectability*

Maybe

*Clues to detect*

* P-values just below the significance threshold (usually 0.05)

**Selective test reporting**

*Definition*

Repeatedly testing a hypothesis in different ways until the desired result is found, and then selectively reporting the findings that support the desired conclusion.

*Alias(es) and related concepts*

*Umbrella term(s)*

P-hacking Cherry-picking

*Examples*

1. Researcher analyzes the data using (1) multiple statistical methods (multiple t-tests, ANOVAs, different random structures in LMEMs) and/ or (2) multiple data eligibility specifications. Based on the results, they choose to present only one analysis that gives a significant result.

*Potential damages*

* Inflated type I or type II error
* Misleading conclusions
* Reduced reproducibility
* Reduced replicability

*Remedies*

* Pre-register data processing (e.g. missing data approach) and statistical analysis strategy
* Blinded data analysis
* Use simulated data to plan analysis (write code)
* Perform specificity curve analysis

*Detectability*

Maybe

*Clues to detect*

* Absence of pre-registration
* Arbitrary data processing steps and/or statistical methods
* P-values just below the significance threshold (usually 0.05)

**Using unjustified references**

*Definition*

Selectively citing works by specific researchers or journals to inflate citation metrics or boost a journal’s impact factor.

*Alias(es) and related concepts*

*Umbrella term(s)*

Citation engineering

*Examples*

1. Researcher selectively cites their own publications for boosted citation metrics.
2. Citation networks: Researcher cites a colleague’s unrelated work in order to get cited in a similar way.

*Potential damages*

* Inflated credibility of publications
* Inflated credibility of journals
* Misleading conclusions

*Remedies*

* Cite only relevant studies
* Provide comprehensive coverage of related scholarly literature

*Detectability*

Yes

*Clues to detect*

* Publications are cited without relevance to the claims
* Specific authors or journals are cited disproportionately frequently

**Citing unreliable research**

*Definition*

Citing an unreliable publication to support the study’s narrative.

*Alias(es) and related concepts*

*Umbrella term(s)*

Citation engineering

*Examples*

1. Researcher cites a publication that presents low-level evidence to support a claim with no reference to the study’s limitations or no reference to other studies.
2. The researcher cites a retracted paper.

*Potential damages*

* Increasing the credibility of low-evidence research
* Inflated credibility of statements

*Remedies*

* Only cite publications that properly support their claims
* Never cite a retracted study

*Detectability*

Yes

*Clues to detect*

* Cited publication provides no or low-quality evidence to its claims
* The cited publication is retracted or otherwise discredited or its claims are refuted

**Using irrelevant references**

*Definition*

Using citations that are not connected to the claims to increase the credibility of a statement.

*Alias(es) and related concepts*

*Umbrella term(s)*

Citation engineering

*Examples*

1. Researcher supports a statement with three citations and two of them are unrelated to the statement.

*Potential damages*

* Inflated credibility of statements
* Misleading conclusions

*Remedies*

* Cite only relevant studies

*Detectability*

Yes

*Clues to detect*

* Publications are cited without relevance to the claims

**Making unsupported conclusions**

*Definition*

Interpreting research findings or their implications in a way that is not backed by evidence.

*Alias(es) and related concepts*

*Umbrella term(s)*

None

*Examples*

1. Researcher concludes that a treatment is effective for groups and contexts that were not considered in the study.
2. Researcher concludes that a treatment worked, however, the treatment effect did not differ from the effect of the control condition (or no control condition was used).
3. Researcher implies causality based on a research design or that does not allow causal inference.

*Potential damages*

* Misleading conclusions

*Remedies*

* Make sure that every interpretation is properly supported by evidence
* Make clear that the evidence is limited to certain contexts
* Use conditional statements where evidence is weak or the researcher uses extrapolation

*Detectability*

Yes

*Clues to detect*

* The chosen methodology and statistical analysis do not allow to answer the hypothesis
* The statistical results do not match the conclusions
* Causal claims are made without the methodology or analysis allowing causal inference
* Results are generalized to contexts outside of the study’s scope

**Not disclosing deviations from pre-registration**

*Definition*

Deviating from the pre-registration without transparency and proper justification in the publication.

*Alias(es) and related concepts*

*Umbrella term(s)*

Faking open science

*Examples*

1. Researcher pre-registers that they would collect data from a non-student sample, but ends up including students, and does not disclose this deviation.
2. Researcher pre-registers a data analysis using linear regression but used robust regression instead without reporting the discrepancy.
3. See also Example 1 in Selective reporting of hypotheses.

*Potential damages*

* Reduced replicability
* Potential HARKing
* Inflated confidence in the research
* Misleading conclusions

*Remedies*

* Disclose and justify every divergence from the pre-registration
* Avoid vagueness in pre-registration
* Use methods that will provide robust results even when pre-registration is not specific at points (e.g. blind data analysis, cross-validation)

*Detectability*

Yes

*Clues to detect*

* Pre-registration and published study differ on important aspects
* Link to the pre-registration in the manuscript does not work/ leads to a private OSF project

**HARKing**

*Definition*

Presenting a hypothesis that is based on observed results (post-hoc or a posteriori) as if it was presumed before obtaining results (a priori).

*Alias(es) and related concepts*

“Hypothesizing after the results are known”, Texas sharpshooter fallacy, Post hoc ergo propter hoc

*Umbrella term(s)*

None

*Examples*

1. Researcher claims to have predicted an unexpected result.
2. Researcher has no hypotheses originally and forms hypotheses after exploring the data and presenting the hypotheses as they had those from the beginning.
3. Researcher has a hypothesis (e.g., a mediation hypothesis) and tests it, and if the results do not confirm the hypothesis but rather indicate an alternative pattern (e.g., a moderation), the researcher claims that this is what they hypothesized all along.
4. Post-hoc directional hypotheses: The researcher presents a hypothesis as if it was uni-directional (i.e. group A’s mean is larger than group B’s, or a correlation will be positive), although the original hypothesis was bi-directional. This change will make the hypothesis test significant.

*Potential damages*

* Reduced replicability
* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Misleading conclusions

*Remedies*

* Formulate hypotheses before analyzing the data
* Pre-registration of confirmatory studies
* Clearly separate exploratory and confirmatory findings
* Use robust exploratory research practices (e.g. holdout dataset, cross-validation, multiverse analysis, blinded data analysis, etc.)

*Detectability*

Maybe

*Clues to detect*

* Absence of pre-registration
* Unexplained and unconventional choices in the methods and results section

**Selective reporting of hypotheses**

*Definition*

Reporting hypothesis test only if it fits the researcher’s expectation.

*Alias(es) and related concepts*

Cherry-picking hypotheses, Chrysalis effect, Fishing expedition

*Umbrella term(s)*

Cherry-picking

*Examples*

Researcher formulates five hypotheses of which only three are supported by the data - only these 3 get reported in the final research report (Chrysalis effect). (2) Fishing expedition - The researcher surveys college students about the outfit they are wearing and their scores on several tests which allows for many possible analyses (examining different colors, types of clothing, tests, score cutoffs, etc.). They end up reporting only a subset of findings to claim college students perform significantly better on tests when they are wearing green. See also Modifying measurement, Selective reporting of indicator variables, and Selective reporting of outcomes.

*Potential damages*

* Potential HARKing
* Inflated confidence in the research
* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Misleading conclusions

*Remedies*

* Report all hypotheses in the write-up regardless of whether they were confirmed or not
* If some hypotheses get left out due to the scope of the write-up be transparent about it
* Pre-registration

*Detectability*

Maybe

*Clues to detect*

* Number of hypotheses in pre-registration (or dissertation) exceeds the number in the publication

**Selective citing**

*Definition*

Avoiding to mention studies that do not support the hypothesis of the research or even those that do support the hypotheses to make the study appear more novel.

*Alias(es) and related concepts*

Cherry-picking citations

*Umbrella term(s)*

Citation engineering Cherry-picking

*Examples*

1. Researcher overly cites empirical work that supports their hypotheses and withholds citing work that did not find the effect at all or even the opposite.
2. Researcher omits other null findings to maximize the perceived value of a null finding.

*Potential damages*

* Inflated confidence in the research
* Misleading conclusions

*Remedies*

* Provide comprehensive coverage of related scholarly literature

*Detectability*

Yes

*Clues to detect*

* Important studies and experts are missing from references
* Cited studies only point into one direction
* Systematic reviews and meta-analyses are not cited

**Omitting important details of the scientific process**

*Definition*

Not reporting important details of the methodology and statistical analysis.

*Alias(es) and related concepts*

Incomplete methods or results section

*Umbrella term(s)*

Cherry-picking

*Examples*

1. Researcher omits sample characteristics.
2. Researcher reports correlations without specifying the type, e.g. Spearman.
3. Researcher reports having conducted an online study and does not reveal it was an MTurk-study.
4. Researcher does not fully disclose compensation etc. for participants during data collection.
5. Researcher does not share study materials on request.
6. Researcher does not report exact questionnaire items.

*Potential damages*

* Reduced replicability
* Reduced reproducibility
* Reduced generalizability
* Misleading conclusions

*Remedies*

* Report every important detail of the scientific process
* Pre-registration or written research plan before conducting the study
* Use a lab log during data collection to keep track of changes in the scientific process

*Detectability*

Maybe

*Clues to detect*

* Details that are usually shared are missing
* Replication is not possible from published methods
* Absence of open study materials

**Selective reporting of indicator variables**

*Definition*

Reporting only the indicator variables (or predictors, features, independent variables) that are used in analyses that produce expected results.

*Alias(es) and related concepts*

Cherry-picking indicator variables, Cherry-picking conditions/ groups

*Umbrella term(s)*

Cherry-picking

*Examples*

1. Researcher reports IVs that are associated with the outcome rather than including all measured IVs in the results section.
2. Researcher drops one or more conditions or groups/merges two or more groups into one / splits a group into more groups than were initially planned depending on statistical results.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Reduced replicability
* Reduced generalizability
* Misleading conclusions

*Remedies*

* Report all indicators
* Pre-registration

*Detectability*

Maybe

*Clues to detect*

* Number of pre-registered indicators exceeds the number of indicators in publication
* Reported mean time of participation does not match the number of reported measures
* Measures get reported in the methods section but not in the results section
* Indicators reported in Supplemental Material but not mentioned in main text

**Selective reporting of outcomes**

*Definition*

Reporting only the outcomes (or dependent variables) that are used in analyses that produce expected results.

*Alias(es) and related concepts*

Cherry-picking outcomes

*Umbrella term(s)*

Cherry-picking

*Examples*

1. Researcher uses several scales to measure the same construct but only reports the one that produces expected results.

*Potential damages*

* Inflated type I or type II error
* Inflated or deflated effect size estimates
* Reduced replicability
* Reduced generalizability
* Misleading conclusions

*Remedies*

* Report all outcomes
* Pre-registration

*Detectability*

Maybe

*Clues to detect*

* Number of pre-registered outcomes exceeds number outcomes in publication
* Reported mean time of participation does not match number of reported measures
* Measures get reported in the methods section but not in the results section
* Outcomes reported in Supplemental Material but not mentioned in main text

**Incorrect reporting of test statistics**

*Definition*

Not using statistical test reporting conventions to obscure exact results and assume that they are above or below threshold values.

*Alias(es) and related concepts*

*Umbrella term(s)*

None

*Examples*

1. Researcher is ‘rounding off’ a p-value in a paper (e.g., reporting that a p-value of .054 is less or equal to .05).
2. Researcher reports p-values only and conceals test statistics.
3. Researcher reports a correlation without disclosing the degrees of freedom, number of observations, or confidence interval, so it seems like the effect is large (for example r=.70, n=15, CI=[.01; .90]).
4. Researcher does a model comparison and only reports fit statistics that are in favor of the preferred model.

*Potential damages*

* Inflated type I or type II error
* Reduced reproducibility
* Misleading conclusions

*Remedies*

* Adhere to reporting conventions (e.g., APA)
* Use literary programming (e.g., RMarkdown, quarto, jupyter)
* Work with software that supports you in producing and checking your write-up (e.g., papaja, stat-check)
* Publish data
* Publish processing and analysis code

*Detectability*

Yes

*Clues to detect*

* Statistics are not reported according to conventions (e.g., three digits for p-values, reporting of df)
* Fit statistics are missing without proper explanation
* Anomalies in reported statistics, e.g., test statistics are incompatible with p-values
* Absence of open data
* Absence of open code

**Visualizing data in a misleading way**

*Definition*

Choosing suboptimal visualizations or altering figure properties in order to exaggerate or diminish effects.

*Alias(es) and related concepts*

*Umbrella term(s)*

None

*Examples*

1. Researcher truncates the y-axis so it is not starting at zero and/or does not add error bars. This makes differences seem larger and more significant than they are in reality.
2. Researcher uses arbitrary categories to present interval data on a map.
3. Researcher displays a pie chart with percentage numbers falling below or exceeding 100.

*Potential damages*

* Misleading conclusions

*Remedies*

* Follow best practices on how to visualize data

*Detectability*

Yes

*Clues to detect*

* Visualization does not match the reported results in the text
* In a plot y-axis is starting at an arbitrary point
* Statistical uncertainty (e.g., error bars) is not shown on plots
* Only summary statistics are shown without individual data points
* Scale or response options in text do not match how they are presented in a plot
* Chartjunk (e.g., 3d elements, ornaments) is present on the plot

**Not linking the Pre-registration to the published study**

*Definition*

Creating a pre-registration but not associating it with the published study.

*Alias(es) and related concepts*

*Umbrella term(s)*

None

*Examples*

1. Researcher pre-registers a study and after conducting the research the pre-registration is not mentioned in the manuscript because of too many diversions.

*Potential damages*

* Potential HARKing

*Remedies*

* Always link the pre-registration to the manuscript and report discrepancies

*Detectability*

Maybe

*Clues to detect*

* A pre-registration that fits the study is findable

**Not making data accessible**

*Definition*

The datasets and/or codebooks are not made accessible to the public and/or peer-reviewers without justifiable cause.

*Alias(es) and related concepts*

*Umbrella term(s)*

None

*Examples*

1. Researcher doesn’t provide a publicly accessible repository link to the dataset.
2. Data repository link is accessible, but the data is not comprehensible (e.g. lacks cleaning and organization, codebook and instructions, etc), hence it is difficult or impossible to use to reproduce findings.

*Potential damages*

* Reduced reproducibility
* Prevents data reuse

*Remedies*

* Data should be shared based on the FAIR (findable, accessible, interoperable, and reusable) principles and legislative context of the researcher
* Use synthetic datasets when original data can’t be shared
* If confidential and personal information makes participants identifiable, apply masking and anonymization, and then share data

*Detectability*

Yes

*Clues to detect*

* No information in the publication on the availability of the data
* Data are not shared according to FAIR principles

**Declaring false authorship**

*Definition*

Attribution and arrangement of authorship that does not correspond to the authors’ contributions, in order to influence the publishing process, and increase the credibility of the study.

*Alias(es) and related concepts*

*Umbrella term(s)*

Citation engineering

*Examples*

Honorary authorship: Researcher adds a co-author who did not contribute to the manuscript. (2) Ghost authorship: The researcher excludes a co-author who significantly contributed to the project. (3) Controversial researcher writes a paper and publishes it under a pseudonym, so it seems that more than one person shares the same view.

*Potential damages*

* Inflated credibility based on the reputation of authors who were included in (or excluded from) the author list
* Contributing authors may not get the credit they ought to get

*Remedies*

* Include everyone who made a significant contribution to the project
* Only include authors who contributed to the project
* Explicitly declare contributions to the project (e.g., CRediT taxonomy)

*Detectability*

No

*Clues to detect*

None

**Publishing studies selectively**

*Definition*

Choosing which study to publish or share based on whether the findings fit expectations.

*Alias(es) and related concepts*

File drawer problem

*Umbrella term(s)*

Cherry-picking

*Examples*

1. Researcher runs a study and finds out the results do not support their hypothesis (e.g., no significant findings). Thus the researcher does not try to publish or share the study publicly.
2. Researcher runs several studies, and publishes only those that support the hypothesis in a multi-study paper.

*Potential damages*

* Creates publication bias
* Creates a false sense of confidence in a multi-study paper
* Misleading conclusions

*Remedies*

* Publish all studies, even when the findings do not support hypotheses
* Only pre-register on platforms, that will eventually publish all pre-registrations

*Detectability*

No

*Clues to detect*

* Publication bias can be estimated in meta-analysis

**Creating multiple publications from the same study**

*Definition*

Breaking up of research findings from the same dataset into several publications without proper justification and the disclosing of related papers.

*Alias(es) and related concepts*

Salami slicing

*Umbrella term(s)*

Citation engineering

*Examples*

1. Researcher conducts a study measuring several outcomes (or predictors) and publishes results in several papers with each paper focusing on just one outcome (or predictor), while not disclosing the other papers.
2. A study on cross-cultural differences with 20 participating labs from 20 countries results in 10 publications where in each one two countries are compared.

*Potential damages*

* Inflated type I or type II error
* Biased effect size estimates in meta-analyses (due to non-independence of results)
* May increase the number of unreliable studies

*Remedies*

* Publish study results in one single publication or disclose all related papers
* Pre-registration of publication strategy

*Detectability*

Maybe

*Clues to detect*

* The methods suggest a large study but the scope of the paper is narrow
* Several papers exist with similar outcomes or predictors based on the same dataset by the same researcher or lab
* Description of the sample is the same over several studies by the same researcher or lab
* Absence of open data