StatReviewer Report

stat reviewer

Generated: 2022-12-01 09:56:22 -0600

Issues Found

1

The subject of the study is very likely to be a keyword in a search engine. To make your study more relevant to your readers, include the subject of the research (e.g., rats, individuals with chronic pain, etc.) in the actual study title.

More Information About This Issue

Most scientific topics cover a diverse set of subjects. Readers often wish to know if a particular study was done using animals or humans. Within human research participants, there is a vast array of subgroups such as age, racial, diagnostic, etc. To help your readers identify the relevance of your study to their interests it is helpful to report the research subject(s) directly in the title.

Examples

This title: "Tumor size and lymph node status on survival"
Can be revised to this: "Tumor size and lymph node status on 5-year survival in women with breast cancer"

Additional Resources

For more information, the University of Southern California Libraries has a list of excellent suggestions to assist authors in devising their titles:
http://libguides.usc.edu/writingguide/title

2

When reporting a proportion in the abstract (e.g., "12.0%"), it is important to also report the numerator and denominator. For example, 50/100 (50%) patients recovered fully. Providing this information will allow your readers to evaluate the sample size and any missing data.

More Information About This Issue

Sample sizes can vary from outcome to outcome, so reporting the actual sample sizes used to estimate the proportion can help eliminate uncertainty. Reporting the numerator/denominator (%) such as "25/100 (25%)" provides substantial information about the the frequency counts used to estimate the proportion.

Examples

This text: "40% of patients responded to the survey".
Can be revised to: "230 of 575 (40%) of patients responded to the survey".

This text: "The success rate was 22% of cells".
Can be revised to: "Out of 309 cells, 68 (22%) were successful".

This text: "33% of the placebo group and 55% of the active treatment..."
Can be revised to: "33/100 (33%) of the placebo group and 550/1000 (55%) of the active treatment..."

3

Although your readers can often infer the design, It is helpful to report a specific study design directly in the abstract (e.g., case-control, meta-analysis, lab experiment).

More Information About This Issue

The abstract should express a clear definition of the study design using a commonly accepted group of terms. These terms differ widely across different research communities, but clearly stating the research design using any such terms is a crucial element for communicating the methods. Basic science or bench research studies are often described as "laboratory studies" or "experimental designs". Studies using humans can often be described as either "Retrospective" or "Prospective" in addition to reporting the specific design (e.g., "Cohort Study").

Examples

Example 1: "This laboratory study involved N = 45 mice..."
Example 2: "A prospective, longitudinal cohort design"
Example 3: "A systematic review and meta-analysis"
Example 4: "Diagnostic accuracy study"

Additional Resources

For clear reporting on a range of research designs, the Equator-Network provides unparalleled resources:
http://www.equator-network.org/

For examples of common design types from different fields:
https://en.wikipedia.org/wiki/List*of*psychological*research*methods
https://en.wikipedia.org/wiki/Research_design

4

When human subjects are involved, the study setting or location is an important part of understanding the methods of the study. Please ensure that the study setting or location is reported to the reader in the abstract.

More Information About This Issue

For studies involving human participants, it is important to report the setting of the conducted research. Depending on the nature of the research, this could be the setting where the individuals were recruited, treated, or studied. The nature of the setting often plays an important role in understanding the results of the study so care should be taken to report any aspect of the setting that might impact the results.

Examples

Example 1: "This case-control study recruited patients receiving treatment at a specialty clinic..."
Example 2: "This cross-sectional survey included children enrolled at a public high school"
Example 3: "Records were obtained for any individual living in France between..."

5

Unless the statistical tests are a focus of the research (e.g., "regression"), it is recommended that they be omitted from the abstract.

More Information About This Issue

Summarizing the results of a study can be difficult and requires presenting your readers with only the most essential information. Often, introducing the statistical tests that were used to analyze the data is not essential in the abstract. Further, many journals enforce strict word limitations in the abstract that require authors to judiciously describe only the most core elements of the methods. Unless the statistical tests are a focus of the research, it is recommended that they be omitted from the abstract.

Examples

This: "The data were analyzed using regression".

Could be entirely omitted from the abstract.

6

It is not necessary to report the software that was used to conduct the statistical analyses in the abstract (e.g., "pass").

More Information About This Issue

Summarizing the results of a study can be difficult and requires presenting your readers with only the most essential information. Often, introducing the software that was used to conduct the analyses is not essential in the abstract. Further, many journals enforce strict word limitations in the abstract that require authors to judiciously describe only the most core elements of the methods. Unless the statistical software is a focus of the research, it is recommended that software be omitted from the abstract.

Examples

This: "The analyses were conducted using pass".
Could be entirely omitted from the abstract.

7

A study can be designed to examine many different aspects of a scientific problem. Please report the objectives or aims of the study (e.g., "The objective of the study was to..."), and if applicable the actual hypothesis of the study (e.g., "We hypothesized that...").

More Information About This Issue

The introduction should clearly express the objectives or aims of the study, and if relevant, the actual hypothesis(es) being tested. A study could have been conducted for a large range of objectives so it is important to clearly and succinctly state what the aims of the study were. These aims/objectives should be carefully crafted to be in accordance with what the study design is actually capable of demonstrating. For example, if an animal model is being used, the objectives of the study may not translate to objectives involving humans.

Examples

Example 1: "The objectives of the study were..."
Example 2: "The aim of the study was to.."
Example 3: "We hypothesized that..."

8

To properly interpret the statistical inferences, a reader needs to be able to evaluate the assumptions used in the statistical power calculation. If an a priori statistical power calculation was conducted, please report it. However, if a calculation was not conducted, please do not conduct a post hoc power calculation. Instead, simply state this fact (e.g., "No statistical power calculation was conducted prior to the study..."), and provide a rationale for how the sample size was selected (e.g., "The sample size was based on the available data", "The sample size was based on our previous experience with this design").

More Information About This Issue

When using null hypothesis testing (i.e., inferences using p-values), it is essential to consider the statistical power of the tests. Statistical power is the chance of obtaining statistically significant results ($p < 0.05$) assuming that there is a certain difference or association among groups or variables. Failing to consider the statistical power for a study makes interpreting findings that are statistically non-significant very difficult. This is the case because it is difficult to discriminate between the lack of a meaningful effect versus the mere lack of power to detect it. All power calculations must report on the following assumptions:

α-LEVEL (TYPE-I ERROR RATE)
To calculate statistical power, the statistical significance level must be specified (alpha). A widely used convention for levels of type-I error is α= .05. There may be instances where a researcher deems that a 5% type-I error rate is either too high or too low and chooses to set their significance level to either a much more stringent α= .01 (1% type-I error rate) or lax α= .10 (10% error rate) level. This must be done judiciously, however, as higher type-I error rates may raise doubts about the veracity of the findings.

β-LEVEL (TYPE-II ERROR RATE)
Another component of a statistical power calculation is the theoretical rate for failing to reject a false null hypothesis (β: type-II error). A widely used convention for acceptable levels of power (1 - β) is .80. Conceptually, this is where the researcher has an 80% chance of finding statistically significant differences.

ONE-TAILED VERSUS TWO-TAILED SIGNIFICANCE TESTS
To calculate statistical power, the nature of the hypothesis has to be specified. The proposed hypothesis and corresponding statistical test dictates whether directional (one-tailed) or nondirectional (two-tailed) significance tests are conducted. In essence, the researcher is able to allocate statistical power to detect differences in one direction (eg, response to treatment A > placebo) or in two directions (eg, response to treatment A ≠ placebo; either > or < than placebo). In most research, two-tailed tests are employed and these are recommended for most power calculations. Nevertheless, there are doubtless instances where unidirectional hypotheses may be appropriate such as in non-inferiority testing.

SAMPLE SIZE (N)
The sample size is required for estimating statistical power. Greater sample sizes lead to more statistical power. This relationship has profound consequences because any observed difference, no matter how clinically irrelevant, can be found to be statistically significant with a large enough sample size. Although the formulae used to calculate statistical power can be arranged to solve for any of the factors used in the calculation (ie, N, α, β, effect size), power analyses are usually calculated to find the required N to achieve some value of acceptable power (eg, power = .80).

EFFECT SIZE (ES)
Effect size is perhaps the most important yet least understood concept of power analysis. When conducting null hypothesis testing, the researcher is actually examining the degree of difference between the experimental conditions and has, perhaps unknowingly, made assumptions about the effect size that is being studied. Significance tests do not simply test the presence or absence of an effect; they are conditional on the effect size. An effect size is best understood in the context of the outcome measure (ore dependent variable) being studied. For example, a researcher can examine differences in means, medians, proportions, odds, etc (See: https://en.wikipedia.org/wiki/Effect_size).

Examples

Example 1: "We assume that the control group will exhibit a mean (SD) of 45 (7) and the treatment group will exhibit a mean (SD) of 40 (7). Assuming a two-sided alpha = 0.05, enrolling 32 individuals per group will yield power = .80 to detect this clinically significant difference using a independent samples t-test".

Example 2: "Our preliminary data suggest that the A group will experience 5% mortality rate and that the B group will experience 10% mortality. Enrolling n = 575 patients/group (N = 1,150) will provide 90% power to detect this difference assuming a two-sided hypothesis test using a difference in proportions".

Additional Resources

For more information about effect sizes:
https://en.wikipedia.org/wiki/Effect_size

For more information about statistical power:
https://en.wikipedia.org/wiki/Statistical_power

There are many professionally developed software packages to estimate statistical power. For example:
https://www.ncss.com/software/pass/

Additionally, several sites have free online statistical power calculators:
http://powerandsamplesize.com/Calculators/

9

To allow your readers to interpret the context of this analysis in light of previous examinations of these data, it is important to report the nature of the current analysis (e.g., "This is the primary analysis of these data..."), or to explicitly report if this is a secondary or subgroup analysis of these data (e.g., "This analysis is a secondary analysis of previously collected data..."). The preplanned (i.e., a priori) versus post hoc (i.e., derived after initial examination of these data) nature of this analysis should also be reported. Finally, to allow evaluation of the previous reports of these data, please cite any published manuscripts that report any aspects of the data used in this study.

More Information About This Issue

When communicating an analysis in a manuscript, it is very important to provide the context of the analysis to your readers. Studies that have been designed to examine a certain question or hypothesis are interpreted very differently than studies that were conducted for another purpose and are only secondarily being used to examine a question. Furthermore, the nature of how the current analysis fits with previous publications from the same data set, or on the same individuals in the study, can be crucial for interpreting the study.

Examples

Example 1: "This is the primary analysis of these data and there are no previous publications of these data".

Example 2: "This is a planned subgroup analysis of these data. The primary analysis has been previously published in the parent manuscript (Smith, 2017)".

Example 3: "This is an unplanned post hoc secondary analysis of these data. The original study was designed to examine group differences in A, while this study analyzes predictors of the original primary outcome (see: Smith, 2017). There are 3 other previous publications from this study (Smith 2017, Smith 2016, Smith 2015)".

Additional Resources

Recommendations from the International Committee of Medical Journal Editors on analyses from the same database: http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/overlapping-publications.html

Authors are encouraged to consult with their target journal to review policies about overlapping publications or secondary analyses from the same data set.

10

Regression analyses can be done for many different reasons including prediction, forecasting, confounder control, or isolating an effect of interest. Therefore, reporting the intended purpose for conducting a regression analysis can be very helpful for your readers.

More Information About This Issue

There are many reasons that an analyst could employ a regression model. For example, a model could be used to examine the univariate (i.e., crude) association between one predictor and one outcome. Alternatively, a model could be used to isolate the influence of one predictor while adjusting (or controlling) for the influences of other variables. Regression models can be used for prediction, description, and many other purposes, so it is important to introduce why such models are being used.

Examples

Example 1: "A multivariable logistic regression model was used to develop a prediction equation to estimate risk of side-effects..."
Example 2: "A linear regression model was used to examine the same-day association between IL-6 levels and pain scores after adjusting for the previous day's IL-6 levels"

Additional Resources

For information on regression analysis: https://en.wikipedia.org/wiki/Regression_analysis

---

11

In multiple regression, a high degree of correlation between two predictors is called multicollinearity (or collinearity). When present, collinearity drastically alters the observed estimates of two correlated predictors in the same model. To assure your readers that collinearity was not a problem in the regression analysis, it is often a good idea to report the steps taken to consider this issue (e.g., condition index, VIF, correlation matrix, etc.).

More Information About This Issue

When creating a multiple regression model it is important to consider the relationship between predictors in the model. Although the predictions from the model are not impacted, the interpretations of individual predictors are radically altered when individual predictors are highly correlated with each other. There are several ways to examine the correlation among predictors, and care should be taken to ensure that individual predictors are not highly correlated when interpreting their individual influence is of interest. Methods that could be used include correlations, variance inflation factors (VIF), condition index, and several others.

Examples

Example: "A multiple regression model was estimate to predict blood pressure from several risk factors. The variance inflation factors of the predictors were evaluated to ensure that the predictors in the model were not causing interpretation difficulties".

Additional Resources

Multicollinearity: https://en.wikipedia.org/wiki/Multicollinearity

Penn State has a nice introduction to the consideration of multicollinearity: https://onlinecourses.science.psu.edu/stat501/node/347

12

Regression analyses involves specifying which variables are in the model and then estimating regression weights for each. Clearly reporting how predictors were selected for inclusion in the model is extremely important for interpreting the model (e.g., forced entry, stepwise variable selection, backwards elimination, etc.).

More Information About This Issue

Predictors in a multiple regression can be selected for inclusion in many different ways. Predictors can be forced into the model (i.e., forced entry) based on theoretical considerations. Alternatively, predictors can be selected for inclusion based on some empirical criteria (e.g., p-value, BIC, AIC). When a predictor is selected based on improving model fit, the selection is called a data-driven approach and can be conducted using several different forms of a stepwise variable selection method (e.g., forward selection, backward removal). The methods used to select variables for inclusion in the model should be carefully reported to the reader.

Examples

Example 1: "Predictors in the model were selected based on their theoretical importance and were forced into the model".
Example 2: "The candidate predictors were selected for inclusion based on a stepwise variable selection method..."

Additional Resources

Stepwise variable selection methods: https://en.wikipedia.org/wiki/Stepwise_regression

A nice tutorial on variable selection methods: http://www.biostat.jhsph.edu/~iruczins/teaching/jf/ch10.pdf

---

13

For complex regression models, it is often a good idea to examine how well model predictions match the observed data. This can involve the examination of model residuals (i.e., the goodness of fit) or through formal testing (i.e., Hosmer-Lemeshow test for logistic regression).

More Information About This Issue

The agreement between the predicted values and observed values is a crucial element for any statistical model. Especially for prognostic models (i.e., models predicting future states) and models evaluating statistical interaction (i.e., predictor x predictor), the calibration of the models results with reality is an important consideration. Because of this, it is important for authors to report how they considered the calibration of their models during the analysis.

Examples

Example 1: "The calibration of the logistic regression model was examined using the Hosmer-Lemeshow test".

Example 2: "A Q-Q plot was used to examine relationship between the predicted scores and the observed scores".

Additional Resources

The Hosmer-Lemshow test is a commonly used calibration test for logistic regression:
https://en.wikipedia.org/wiki/Hosmer%E2%80%93Lemeshow_test

For more information on calibration: https://en.wikipedia.org/wiki/Calibration_(statistics)

For calibration of probabilistic predictions using the Brier score: https://en.wikipedia.org/wiki/Brier_score

---

14

Although there are many ways to communicate the level of measurement for variables under study, using a term like "continuous" is not as precise as it could be. Authors are encouraged to utilize descriptions of their variables as being either nominal (categorical), ordinal, interval, or ratio, as these terms can provide your readers insights as to how they will be analyzed and reported. For example, "Histological measurements were rated on a 1 to 4 ordinal scale and compared between groups using a Mann-Whitney U test and summarized using median [IQR]. Interval level data were compared using a t-test and described using mean (SD)."

More Information About This Issue

The level of measurement of a variable describes the type of information contained in it. Although there are many ways to classify types of measurement, a commonly used system that utilizes four levels of increasing information: nominal, ordinal, interval, and ratio. Many fields utilize measurement levels such as "quantitative" or "qualitative", and these are perfectly acceptable, but they are not as precise as they could be (i.e., there are differing levels of quantitative information).

Nominal: As the names suggest, nominal (or categorical) level of measurement describes variables where the distinct types of information are grouped in the variable. Examples of nominal variables are sex, disease classification, mortality (alive or dead), etc.

Ordinal: this level of measurement describes a variable where rank orders between levels is recorded, but the difference between these levels is unknown. Examples of ordinal data are class ranking, responses to a 5-choice Likert item, or any time units are ranked on another variable.

Interval: this level of measurement describes a variable where the degree of difference between items is evaluable but not the ratio between them. For example, IQ scores can differ by 10 points, but someone with an IQ of 150 cannot be thought to have twice the IQ of someone with 75.

Ratio: this level of measurement contains the most amount of information given that items can be placed in ratio to one another (i.e., they have a true zero and equal distance between levels). For example, weight (kg) is measured on a ratio scale given that someone who is 100 kg is twice as heavy as someone who is 50 kg.

Examples

This: "The data are reported using mean (SD) or frequency (%)"
Could be revised to this: "The interval data are reported using mean (SD) or frequency (%)"

This: "Quantitative variables are described with mean (SD)."
Could be revised to this: "Ordinal variables are described with median [IQR] and variables with at least interval level scaling are described using mean (SD)."

Additional Resources

Wikipedia has a very nice introduction to several theories of levels of measurement:
https://en.wikipedia.org/wiki/Level*of*measurement

15

To properly interpret the study, a reader must be able to evaluate potential bias due to lost, missing, or excluded data. In that regard, please report if any data were missing or lost for any reason. It is important to report the extent of missing data using frequency counts (e.g., "12 patients did not have data for BMI and were excluded from the analysis"). If there were no missing data, this fact should be stated (e.g., "There were no missing data").

More Information About This Issue

Missing data, or data that is incompletely observed, is very common in all types of research designs. In bench research there can be failed experiments, corrupted or flawed measurements, or human error in collecting measurements. In research involving direct interaction with human participants, missing data can occur for the same reasons or for omitted responses by the participants. In systems research, missing data may occur due to a host of reasons beyond the investigators control. No matter the reason for the missing data, it is crucial to report the existence of missing data, and the extent to which observations were missing in the analysis.

All statistical analyses involve the selection of observations from a larger population. Failure to consider the biasing effect of missing observations can often lead to erroneous conclusions.

Examples

Example 1: "There were 7 cases that were excluded from the analysis due to missing values".

Example 2: "There were no missing data and all animals were included in the analysis".

Additional Resources

For an introduction to missing data and methods to address bias:
https://en.wikipedia.org/wiki/Missing_data

16

There is at least one error in an odds ratio and its 95% confidence interval. The upper and lower bounds of at least one confidence interval cannot be accurate given the odds ratio. Please double check all of the values for accuracy. For example: aOR=.52, 95% CI: .37–.74

More Information About This Issue

The software checks confidence intervals for accuracy. When reporting a 95%CI for an odds ratio, the natural log of the point estimate (OR) must reside half-way between the natural log of the upper and lower bounds of the interval. Double check the confidence intervals in your manuscript for accuracy and note that rounding might impact the software.

17

Because it is theoretically impossible to observe a p-value equal to 0, and because it is confusing to report a p-value equal to 1.0, it is recommended that such extreme p-values (e.g., "p=1.000") be reported as "p < 0.0001" and "p > 0.999".

18

When reporting risk ratios (e.g., HR, RR, or OR) it is useful to consider the number of digits of accuracy. Reporting too few digits for very small ratios (e.g., OR: 0.1) does not provide enough precision (OR: 0.125 is better), while reporting too many digits for high ratios (e.g., OR: 45.54) can be distracting and not required for interpretation (OR: 45.5 is better). For ratios between 0.40 and 4.00 two digits are recommended. Please reconsider the number of digits used in these ratios: aOR=10.29

More Information About This Issue

A useful rule to remember when reporting risk ratios (e.g., hazard ratios, odds ratios, and risk ratios) is the "Rule of four". Following this rule ensures that the rounding error can be no greater than 1.3%, no matter what value is being reported. This rule states that:

“Round the risk ratio to two significant digits if the leading non-zero digit is four or more, otherwise round to three;” (Cole, 2015)

This means that:
- ratios between 0.040 and 0.399 are rounded to three digits (e.g., HR: 0.345)
- ratios between 0.40 and 4.0 are rounded to two digits (e.g., OR: 2.34)
- ratios between 4.0 and 40.0 are rounded to one digit (e.g., RR: 35.0)
- ratios greater than 40 are not presented with any digits (e.g., OR: 75)

Examples

Example 1: This difference resulted in an increased risk, OR: 3.45 (95%CI: 3.40 to 3.50), p < 0.001
Example 2: This difference resulted in a decreased risk, OR: 0.457 (95%CI: 0.407 to 0.507), p < 0.001
Example 3: This difference resulted in a massive increase in risk, OR: 120 (95%CI: 107 to 133), p < 0.0001

Additional Resources

For a clear explanation of the "Rule of four" see: http://www.bmj.com/content/350/bmj.h1845

19

There appears to be at least one proportion (%) reported in the manuscript that is incorrect (i.e., the numerator and denominator do not combine to form the reported percentage). Please check the results for accuracy. For example this value(s) deserve further attention:

Fifty patients (65.8%) achieved 90-day functional independence (mRS≤2), and 27 (35.5%)

More Information About This Issue

When reportting proportions (%), it is important to report the numerator and denominator that create the proportion. For example, 50 of 100 (5%) patients were women. The software checks to ensure that the reportted proportion (%) is actually the numerator divided by the denominator.

Examples

Example: 45/100 (45%) of the animals...
In this example, the numerator is '45', the denominator is '100' and the proportion is 45 divided by 100 = 45%

20

Describing an effect as "significant" is commonplace but not very informative. It is important to distinguish between effects that are "statistically significant" versus those that are also "clinically significant" or "practically significant". Simply finding "significant" effects is not equivalent to finding effects that also have practical importance.

More Information About This Issue

The term "significant" really has two meanings that can often be in conflict in a scientific manuscript. A "significant" event in common language use is one that is important. However, in hypothesis testing, a "significant" finding only needs to have a p-value less than 0.05. A "statistically significant" finding can be of great importance, but unfortunately for readers of science, it can also refer to a finding of no practical or clinical significance of any kind. For example, the 50% reduction of a head lice colony after a new treatment is found to be statistically significant (i.e., this value is greater than zero), but because the head lice colony was not eliminated and the surviving lice will repopulate, the finding is not of any practical or clinical significance. Authors are encouraged to assist their readers in distinguishing between "significant" findings in the hypothesis testing sense and "significant" findings in the importance sense.

Examples

This: "The group difference was found to be significant".
Could be revised to this: "The group difference was found to be clinically significant as it altered the course of the disease in meaningful ways".

This: "The differences in sediment type was statistically significant..."
Could be revised to this: "The differences in sediment type was statistically significant but the small observed differences were of no practical significance."

Additional Resources

The university of Ottawa has an introduction to this issue:
https://www.med.uottawa.ca/sim/data/Statistical*significance*importance_e.htm

---

21

The term "multivariate" is reserved for models with multiple dependent variables or multiple outcomes in the same model. When multiple predictor variables are used to predict a single outcome, the term "multiple" is a more apt descriptor. For example, "A multiple regression model was employed using several predictors..."

More Information About This Issue

Multivariate statistics is a special branch of statistics that focuses on the multidimensional space need for more than one outcome variable. Because these analyses are very specialized, it is important to only use the term "multivariate" when is truly applies; otherwise, readers might be confused about the applied methods.

Examples

This: "The predictors were entered into a multivariate statistical model".
Could be revised to this: "The predictors were entered into a multiple logistic regression model...".

Additional Resources

For an introduction to multivariate analysis, and several commonly used approaches:
https://en.wikipedia.org/wiki/Multivariate_statistics

22

It is very often assumed that if outliers are not mentioned in the manuscript that none were detected or that no action was required. However, this makes it difficult for a reader to discriminate between poor reporting (i.e., outliers were excluded from the analysis but never reported), poor methods (i.e., aberrant values were not evaluated for validity), or unremarkable findings (outliers were evaluated but no action was necessary). To assist your readers in interpreting the analysis, it is recommended that a statement about the consideration of outliers be provided in the statistical methods section.

More Information About This Issue

Good statistical analyses examine the nature of the data under evaluation. Often, this involves considering the nature of aberrant values or outliers. Such values can be caused by severe errors in the measurements, artifacts in mechanical processes, typographical errors in data entry, or even valid extreme values that are rare in their occurrence It is not common to report how such values were considered in an analysis, but it should be. Authors are encouraged to report how the data were evaluated for accuracy, how extreme or outlying values were evaluated, and if any action was taken to correct the record or alter the analysis. Please note that depending on the setting, the exclusion of outliers can be controversial so great care must be made addressing this issue.

Examples

Example 1: "Prior to conducting the analysis, the recorded data were evaluated for accuracy by several investigators blinded to the group treatments".

Example 2: "120 extreme values were detected the heart rate recordings. These values were deemed as outliers but were retained for the primary analysis. A sensitivity analysis was conducted that estimated the association while excluding these values".

Example 3: "Prior to conducting the regression analysis, an influence analysis was conducted using studentized residuals. The analysis did not identify any points with undue influence so all of the observed data were included in the analysis".

Additional Resources

For an introduction to outlier analysis:
https://en.wikipedia.org/wiki/Outlier

Bonus: Checklist Items from the STROBE guideline

23

The study's design should be clearly labeled in the title or abstract using a common study design term (e.g., case-control study, cohort study, cross-sectional study).

24

Please report the actual hypothesis of the trial (e.g., "We hypothesized that..."), or the objectives or aims of the trial (e.g., "The objectives were...").

25

Please ensure that the following items are reported:

- The study's design should be clearly labeled in the methods using a common study design term (e.g., case-control study, cohort study, cross-sectional study).
- The use of the term "prospective" or "retrospective" without a formal study design is not a sufficient precise description of the methods.
- It is important to report the context of the current analysis as either being the primary analysis of these data, a secondary analysis, or one of several analyses that have been published from the same effort.

26

To allow the study to be properly evaluated, it is important to clearly define all of the variables used in the analysis. In an observational study, there are several variables types that are often used to examine a hypothesis. Please ensure that the following variables are clearly defined in the manuscript:

- The confounding variable(s) or variables that might serve to confuse the effects. There are often many "predictors" or "variables" used in a statistical model, but it is important to clearly define what theoretical role each of these variables are thought to play (e.g., exposure versus confounder).
- The effect modifier, or variables that might alter the effect of the exposure on the outcome. These types of variables are not used in every study, so it is important to state which variables were examined as effect modifiers, or to state that no variables were analyzed as effect modifiers.

27

It is valuable to provide information about the actual reliability (e.g., inter-rater agreement, assay precision) or validity (e.g., correlations with reference standards) of the utilized measurements in the study.

28

In observational research, there are many sources of potential bias that can cause estimated values to deviate from their true values (e.g., information bias, selection bias). Thus, it is important to formally mention the sources of potential bias and to describe any methods or procedures used to address potential sources of bias.

29

To properly interpret the statistical inferences, a reader needs to be able to evaluate the assumptions used in the statistical power calculation. If an a priori statistical power calculation was conducted, please report it. However, if a calculation was not conducted, please do not conduct a post hoc power calculation. Instead, simply state this fact (e.g., "No statistical power calculation was conducted prior to the study..."), and provide a rationale for how the sample size was selected (e.g., "The sample size was based on the available data", "The sample size was based on our previous experience with this design").

30

Where appropriate, describe the way continuously scaled variables were handled in the analysis. This includes the choice and rationale of any variable groupings or categorizations, as well as the consideration of nonlinear associations or trends between the exposure variable(s) and outcome(s).

31

It is very common to have missing data in observational research. The assumptions underlying the reasons for missing data (e.g., missing at random) and the methods used to address missing data (e.g., multiple imputation, complete cases analysis) should be reported to the reader.

32

Sensitivity analyses are often used to evaluate the consistency of findings from an observational study in light of alternate assumptions or different analytical strategies. For example, how might changes to the definitions/methods used to define the exposures or outcomes have changed the estimates? Alternately, how might missing data have impacted the findings, or how might a different analytical strategy have impacted the results? It is important to report any sensitivity analyses that were employed .

33

It is important to discuss the extent to which the current results generalize, or are applicable, to patients or settings beyond this analysis. Sometimes referred to as external validity, this judgment pertains to the characteristics of the participants in the study, the study setting, the interventions examined, etc.