StatReviewer Report

Created For: em@ariessys.com
Manuscript: [redacted]
Created On: 03/01 07:07

Title

1 Avoid words that do not convey precise meaning in your title. For example, "A study of..." is not needed to communicate the research design.

More Information About This Issue

Nearly all studies can be stated to be a "study of..." or to be examining the "effects of...". Therefore these phrases can be omitted from the title to better isolate the topic of the research.

Examples

The title, "The clinical characteristics in the study of dementia" can be rewritten as "Clinical characteristics of early-stage dementia". Similarly, unless the term "effects" is being used as the subject (e.g., "estimating indirect effects in meta-analysis") there is often a more specific way to represent the topic of the research. For example, "The effects of nitroglycerine on cellular bodies in the heart" can be rewritten as "Nitroglycerine deterioriates cellular bodies in the heart.".

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 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 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

Abstract

> 3 The abstract provides essential information for interpretation and must represent the study independently of the rest of the manuscript. Although presenting the results using a narrative style can be informative, reporting actual descriptive statistics (e.g., mean, SD) can provide much more information. Please do not just report p-values, and define the descriptive statistics directly in the abstract (e.g., "mean +- SD").

More Information About This Issue

Many readers only read the abstract of a study and do not actually read the entire manuscript. Therefore, it is crucial to present the results of the study using the most amount of information as possible. To allow your readers to evaluate the findings of the study, it is important to report the numerical results of the study using descriptive statistics or other measures of effect size.

Examples

This sentence: "The groups significantly differed in response to the experimental condition"
Can be revised to: "The mean (SD) of Group A 4.5 (2.1) differed significantly from Group B 1.2 (0.7), p = 0.034."

This sentence: "There was a correlation between income and job satisfaction."
Can be revised to: "There was a correlation between income and job satisfaction, r (30) = 0.65, p = 0.002".

This sentence: "The risk of mortatility was increased for those with higher morbidity."
Can be revised to: "The risk of mortatility was higher for those with for those with higher morbidity (5%) than for those with lower morbidity (1.2%), RR: 4.2 (95%CI: 1.2 to 7.2)."

Additional Resources

The type of numerical results that can be reported in an abstract differs from study to study. However, most study designs naturally lead to a certain type of data being presented in the abstract. For example, for a randomized controlled trial (RCT) it is recommended that authors report the primary outcome measurements (and their differences) directly in the abstract. For meta-analyses, it is recommended that authors report the aggregated effect sizes in the abstract. See the Equator Network for reporting recommendations specific to each type of study design.

http://www.equator-network.org/

4

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

5

The study setting or location is an important part of understanding the methods of the study. Please ensure that the study setting or location are 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..."

> 6 Reporting that a finding is "significant" in the abstract without any numerical support (e.g., means, OR, etc.) can be misleading and is very difficult to interpret. Please ensure that descriptive statistics and p-values are reported for any findings that are purported to be "significant".

More Information About This Issue

Describing a finding using the term "significant" has become a very common practice in all of science. However, this term communicates very little information by itself (e.g., "the differences were significant..."). It is far superior to replace this term with the actual findings using descriptive statistics or hypothesis testing results.

Examples

This: "The differences between groups were significant".
Can be revised to: "The differences between the treatment and control group was 4.5% (95%CI: 2% to 7%), p = 0.003".

Introduction

> 7 A study can be designed to examine many different aspects of scientific problem. Please report the 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..."

Methods

> 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 $\alpha = .05$. 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 $\alpha = .01$ (1% type-I error rate) or lax $\alpha = .10$ (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 calcuation 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 $(1 - \beta)$ 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) beign 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 assist your readers in understanding the study's design, it is helpful to summarize the methods using a common study design term (e.g., laboratory research, clinical trial, etc.).

More Information About This Issue

The methods 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

Statistical Analyses

> **10** Please report the nature of the hypothesis testing (e.g., two-tailed testing is used by convention).

More Information About This Issue

The nature of 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). Two-sided tests are used for the vast majority of hypothesis testing, so unless a good reason exists to use a one-sided test, it is recommended that two-sided testing be used.

Examples

Example 1: "Two-sided hypothesis testing was used".
Example 2: "A two-sided test of significance was use with $p < 0.05$".

Additional Resources

A great tutorial on one and two-sided testing is avaible at the help pages of the UCLA Institute of Digitial Research and Education:

https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-the-differences-between-one-tailed-and-two-tailed-tests/

> 11 If parametric descriptive statistics (e.g., mean (SD)) are used, it is important to verify that the assumptions required to interpret the statistics have been verified. Please report the methods used to evaluate these assumptions (e.g., histograms, QQ-plots, etc.).

More Information About This Issue

The appropriate use of parametric tests (e.g., t-test, ANOVA, correlation, regression) and parametric descriptive statistics (e.g., mean, SD) require that certain parametric assumptions be satisfied. These assumptions vary by the type or test or statistic, and it is important to consider these assumptions before using these approaches. When these assumptions have been considered, it is valuable to communicate this fact to your readers. This can be done using:

Histograms: A plots of the distribution of variables. This can be used to check if variables are normally distributed.

Q-Q plots: A plot examining how closely a distribution resembles a specified distribution.

Skewness & Kurtosis: A distribution can examined for how skewed (i.e., lopsided or extreme on one side) or how kurtotitic (i.e., how wide or tall is a distribution).

Formal tests: There are many tests such as D'Agostino-Pearson test, Jarque-Barre, Levene's test, Kolmogorov-Smirnov, and several others that can be used to examine how well a distribution(s) fit a certain disribution(s). These tests should not be used when sample sizes are small, but they can be helpful in appropriate sample sizes.

Examples

Example 1: "The distribution of variables were examined using histograms and desctiptive statistics".
Example 2: "Levene's test was used to examine the homogeneity of variances assumption".
Example 3: "Several variables were highly skewed and were thus analyzed using nonparametric tests".

Additional Resources

For a nice description of parametric assumptions, see: https://en.wikipedia.org/wiki/Parametric_statistics

For an introduction to D'Agostino test: https://en.wikipedia.org/wiki/D%27Agostino%27s*K-squared*test

The American Society for Quality has a nice introduction to the use of histograms in assessing distributions: http://asq.org/learn-about-quality/data-collection-analysis-tools/overview/histogram2.html

12 Please report the software that was used to conduct the statistical analyses.

More Information About This Issue

It is important to report the software that was used to conduct the analyses. The goal for the description of the plan of analysis is to allow replication of the results. Because different software packages possess different algorithms that may impact the results, reporting the software is an important aspect of the study's methods.

Examples

Example 1: "The analyses were conducted using SPSS 23.0 (IBM, Inc)".
Example 2: The statistical analyses were conducted using R 3.1.1 (R Core Team, Vienna, Austria)".

Additional Resources

For a list of common statistical software packages:
https://en.wikipedia.org/wiki/List*of*statistical_packages

---

13 To assist your readers in interpreting the data in the results and/or tables, it is helpful to define the measures of variability (SD, IQR) that will be used to describe the data.

More Information About This Issue

Descriptive statistics proivde a summary of the data in a sample. Generally, there are two types of descriptive statistics that should be reported, central tendency (e.g., mean, median, mode) and variability (standard deviation [SD], variance). In the statistical methods section, authors should clearly define how they will report descriptive statistics throughout the manuscript. There are many different ways that descrptive statistics can be reported. However, it is recommended that authors report their data using either mean (SD), median [25th, 75th], and frequency count (%). By clearly defining how these values will be displayed, the values can be presented in the results section (and any tables) with little chance for misinterpretation.

Examples

Example 1: "The data are presented using mean (SD) and frequency counts (%)".
Example 2: "Because the data were highly skewed, the data are presented using median [25th, 75th percentile]".

Additional Resources

For more information about descriptive statistics see:
https://en.wikipedia.org/wiki/Descriptive_statistics

Results

14 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., "6 animals did not survive 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 analyis 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

---

15 When reporting exact p-values, it is recommended that three decimal places of precision are used (e.g., p = 0.047). This is the case even for statistically non-significant p-values (p = 0.435). Please ensure that all p-values are reported with this level of precision.

---

16 A post-hoc power calculation is introduced to the reader. Sometimes, a power calculation is conducted after a study is completed to inform future investigators of the sample size required to have sufficient power in future studies. While this can be informative, authors should be careful to avoid the calculation of "observed power" of their current study. Observed power is based on dubious assumptions, only reflect the observed p value of the comparison, and should be avoided (for a complete discussion of this issue see: Hoening & Heisey, 2001: The

> abuse of power: The pervasive fallacy of power calculations for data analysis. American Statistician, 55[1]).

More Information About This Issue

The attention to the available statistical power for the study is certainly appreciated. However, the reported calculations are best described as post-hoc power calculations. Although widely used, post hoc power calculations are based on dubious assumptions, only reflect the observed p value of the comparison, and should be avoided (for a complete discussion of this issue see: Hoening & Heisey, 2001: The abuse of power: The pervasive fallacy of power calculations for data analysis. American Statistician, 55[1]).

General

> 17 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., abberant values were not evaluated for validity), or unremarkable finings (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 abberant 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 occurence. 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: "Priori to conducting the regression analysis, an influence analysis was conducted using studenized 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