Current trends in statistical reporting: improving transparency, reproducibility, and impact

Clark T. Holdsworth, Ph.D.

Doctorate from Kansas State University

Associate Editor

Oxygen Partial Pressure (mmHg)
- Air 149
- Alveolar 100
- Arterial 95-98
- Mean capillary 39
- Muscle myoglobin associated 3.1

Atmosphere ⟶ Mitochondria
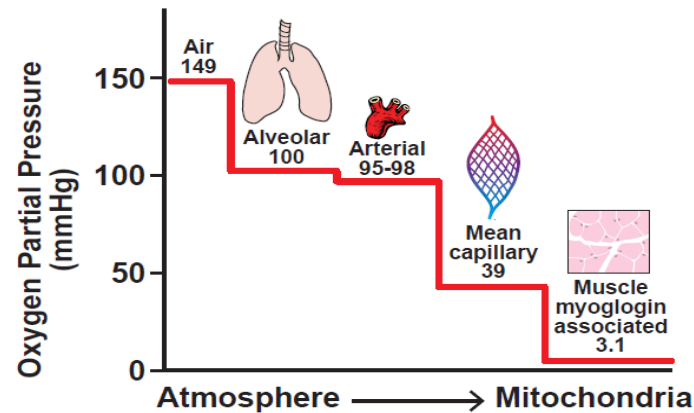
- Herein, we aim to provide you with a **manuscript-centered evaluation** of that statistical component of science.

- Namely, we want to delineate the relationship between conventional **scientific communication** and the statistical methods for processing empirical data.

- This interface between the **conducting and reporting** of analyses is ultimately the point of judgment for the reviewer. Therefore, language becomes an important tool.

"Language is a tool adequate to provide any degree of precision relevant to a particular situation."

-Kenneth L. Pike- Linguist and anthropologist

- This relationship has evolved over the history of scientific communication, and new trends have emerged in the peer-review process.

- We will provide details of these trends with the intention of pre-empting modern reviewer criticisms in data analysis and reporting.
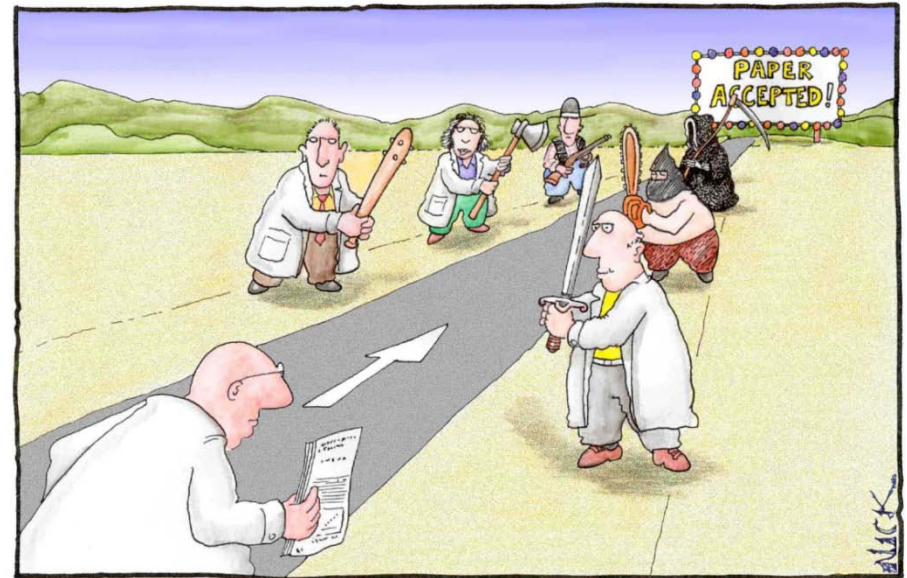
Most scientists regarded the new streamlined peer-review process as 'quite an improvement.'

Q: Why talk about statistics?

1) Despite appearances it represents are large portion of the manuscript

2) Statistical analysis is the crux of the entire experimental approach

*When testing hypotheses it provides a quantitative approach for reducing bias and providing an objective assessment of the data.

1. Hypothesis
2. Study design
3. Statistical analysis
4. Results
5. Tables and Figures
6. Conclusions

"Hypotheses are nets: only he who casts will catch."

-Novalis- early mineralogist and philisopher

H_0: **Null hypothesis**
H_1: **Alternative (experimental) hypothesis**

		Population	
		H_0 is TRUE	H_0 is FALSE
Decision ↑ (Sample)	**Reject H_0**	**Type I Error** α FALSE POSITIVE	✔ 1-β
	Accept H_0	✔ 1-α	**Type II Error** β FALSE NEGATIVE

Typically, the null is represented as a hypothesis of no effect or no detectable differences.

▶ Testable

- Falsifiable
- Specific
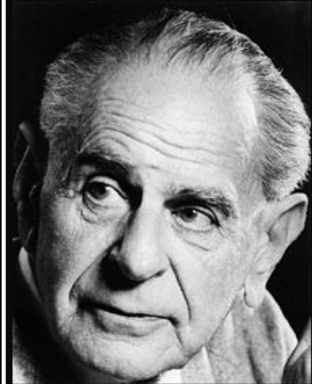
...no matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white.
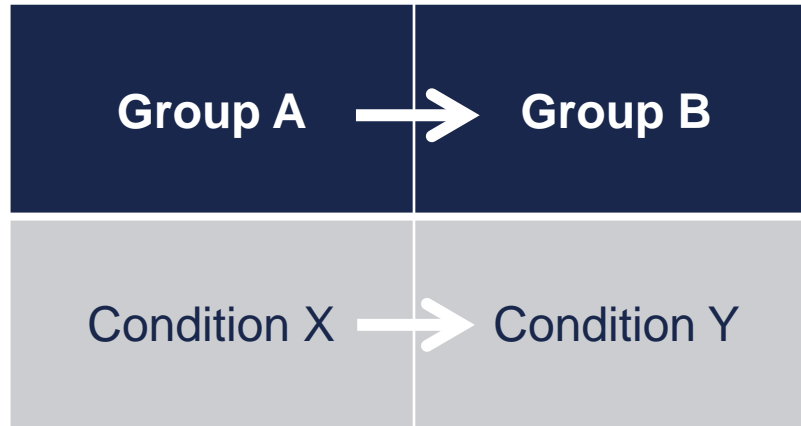
(Karl Popper)

izquotes.com

- Typically a difference between an indicator of central tendency (e.g. means)

Group A → Group B	
Condition X → Condition Y	

- The design should be evident in the hypothesis

*Sometimes the hypothesis is dependent on a subjective threshold. For example, the quantification of a model fit or the variance accounted for by a regression analysis.

These require the threshold to be established *a priori*. This is more robust if the criteria has been leveraged from a previous consensus in the field.

Cite, cite, cite!

ACCDON

- **Study design**
  - ► Information pertinent to the statistical analysis

    - • Sample size

      http://www.gpower.hhu.de/en.html

    - • Number of groups, conditions, time points, etc.

  - ► Schematics can be extremely helpful for the reader/reviewer

Again, high specificity of these descriptions. The specific circumstances dictate the conclusions inferred from the results. Specificity also improves the accuracy of falsification (i.e. statistical power).

- Rigorous reporting of the study design also assures that reproducibility is possible.

- Many journals have abolished length restrictions for the methods.

- Discrepancies in study design during replication studies can lead the community to believe:

  -at best, the authors were incompetent in their statistical analysis

  -at worst, the authors intentionally altered results

- Preregistration of study design and statistical analysis

- Linking between published manuscript and repository

- <u>Statistical analysis</u>

- Lack of transparency can lead the community to believe:

  -at best, the authors were not careful with the selection/rigor of their sample.

  -at worst, the authors intentionally omitted aspects of the experimental design that might affect generalizability.

Establish the level of significance and be consistent

- Student's T test
  - ► Unpaired
  - ► Paired
    - Greater power due to experimental design

  - ► Robust against violations of equal variance and normality for large sample sizes (central limit theorem)

ACCD N
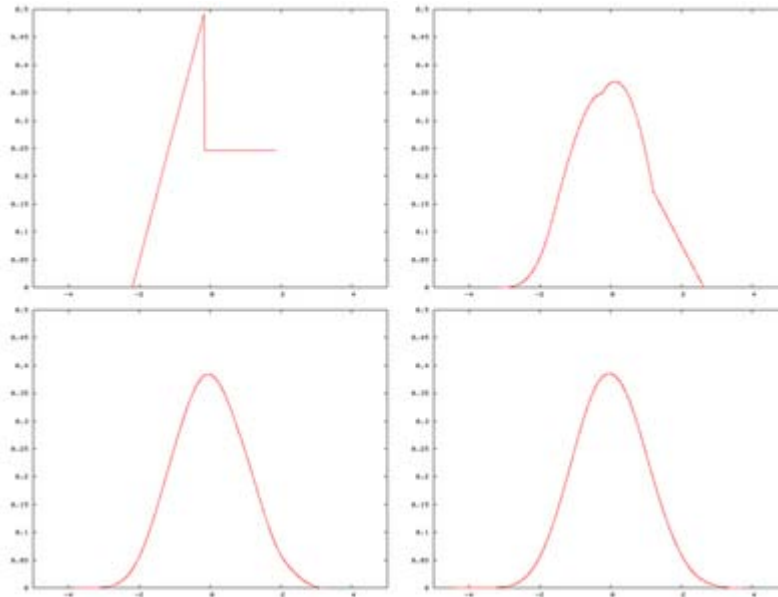
- Regression
  - ► Pearson Product Correlation

    ***Exercise caution when interpreting this statistic.
    Avoid explicit statements of causation.**

  - ► Report the exact regression analysis
    - One-way ANOVA
    - Two-way ANOVA
    - Mixed ANOVA
    - ANCOVA
    - MANOVA

    **How variables were selected
    Stepwise – forward selection or backward elimination**

- **Correcting for multiple comparisons**

  - ▶ Bonferroni
    - Conservative, particularly for large numbers of hypotheses or correlated test statistics
  - ▶ Sidak
    - Higher power, but less corrective for negatively dependent tests
  - ▶ Tukey
    - Only for pairwise comparisons

■ **Results**

► Avoid redundancy
  • Results should be reported in the text or in a table/figure. Instances of overlap will be revised and consistent overlap could be used to triage the manuscript.

► Do not re-summarize the methods

- A result is called statistically significant if a p value is below a certain threshold, most commonly <0.05 or <0.01.

  - This designation is a construct of scientists, not statisticians.

- Do not violate the stated level of significance.
  - "X was greater than Y, but the result did not reach statistical significance."
  - "X was marginally significantly different from Y."

- **Transparent reporting**

  - ▶ There is a trend toward reporting **exact** p values.

The ASA's Statement on p-Values: Context, Process, and Purpose (2016)

"P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone."

ASA
AMERICAN STATISTICAL ASSOCIATION

- **Therefore, p values are useful only as decision making tools. Their value is maximized by the context of the study and the addition of estimates.**

ACCDON

- **Reporting confidence intervals**

  - ► Unlike a test for differences between means, the confidence interval provides a useful estimate of the sample means relative to the population mean.

  - ► This extends the utility of the p value by providing context (in the form of an interval estimate of the population parameter) for the differences between means.

- Reporting correlations

  ▶ R value: Strength is context dependent

  ▶ R^2 value: Fraction of variance explained
    - Include regression equation with graphs where possible

  ▶ Provide sum of least squares

***Exercise caution when interpreting this statistic. Avoid explicit statements of causation.**

  ▶ P value is of severely limited use
    - Most correlations will be significant given a reasonable sample size
    - Be wary of strong relationships where clustering occurs at the extremes

- **Figures and Tables**

  ► Figures should be understandable independent of the main text. Where possible, a descriptive title is encouraged.

  ► Where possible, reference each table and figure in its entirety before moving to the next one. Readers prefer to not constantly go to back to previous tables and figures as the narrative develops.

  ► Panels should be used for comparison only, otherwise split.

- Careful truncating axes.

Is truncating the Y-axis misleading?
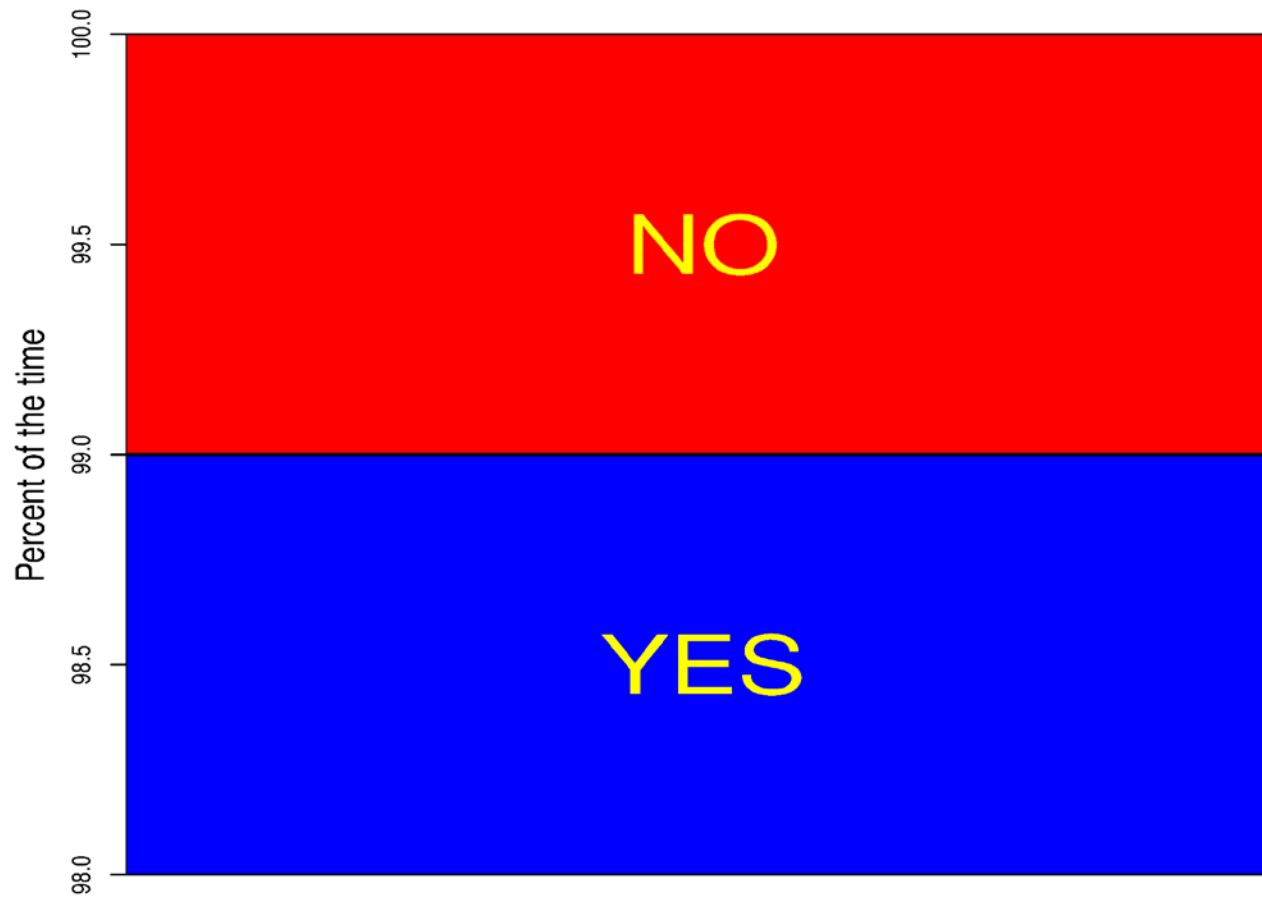
- Maximize visibility.

- Do:
  - ▶ Increase font size
  - ▶ Use **boldface**
  - ▶ Highlight

- Don't:
  - ▶ Use light text colors
  - ▶ Use thin font types

  *For line or scatterplots with nearly overlapping groups/conditions, try to balance individual line visibility with the ability to discern between lines where differences exist. Insets are useful.

- **Limitations**

  - ► Introducing sample size as a limitation can raise concern.
    - • Too large a sample size is a commonly ignored limitation.

      http://www.gpower.hhu.de/en.html

  - ► Confidence in the statistical analysis should always be established before the interpretation is conducted.

  - ► Do not undermine reader confidence at the end of the manuscript, without warning.

■ Conclusions

► Match the study design and statistical analysis

► Use of "suggest", "indicate", etc. is for inferences drawn from the results, not the results themselves

► Be cautious about statements of correlation
  • Independent variable must be manipulable
  • Bradford Hill criteria are sometimes useful

► Specify the exact conditions of the hypothesis when concluding

■ Seeking a statistical consult

▶ In a time of heightened accountability, editors are more prone to seek consults on manuscript issues.

▶ In the past, such consults were not uncommon for ethical concerns where a specialist was required.

▶ Now, statistical consults are employed for the same reason.

- Reviewer concerns regarding the statistical treatment of the data can be pre-empted by employing a statistical consult *a priori* and reporting it in the statistical analysis section.

- Procure some form of documentation and indicate in your cover letter that it is available on request.

- Very few reviewers have the statistical expertise to feel confident challenging a trained statistician.

- Available for statistical review and editing
  - ▶ Knowledgeable statisticians

Many thank to our gracious hosts at Qingdao Center for Resource Chemistry and New Materials!

Questions?