A causerie on statistical/epidemiological analysis and reporting, with discussion

Henrik Stryhn [1,2], Charles Caraguel [2,3]

[1] Centre for Veterinary Epidemiological Research, University of PEI

[2] Centre for Veterinary Education, University of Sydney

[3] University of Adelaide, School of Animal and Veterinary Science

Nov 14, 2025

An unexpected(?) Resource for Statistics: Reporting Guidelines

Reporting guidelines give recommendations on /requirements for how to write up your study:

- specific study types: STROBE(-Vet), REFLECT, PRISMA(-DTA)...,
- (basic) statistics: SAMPL (Statistical Analyses and Methods in the Published Literature),

and scientific journals usually also have their own guidelines for authors, e.g. *animal* (see also zenodo.org/records/12200635)

Animal 18 (2024) 101291

Contents lists available at ScienceDirect

Animal
The international journal of animal biosciences

ELSEVIER

Editorial: The reporting of statistics in research articles is key to the understanding and reproducibility of good research in animal science

I. Ortigues-Marty [a,*], H. Stryhn [b], E. Paquet [c], B. Ampe [d], C.A. Montoya [e], J. Fenlon [f]

[a] UMR1213 Herbivores, INRAE, Université Clermont Auvergne, VetAgro Sup, 63122 Saint-Genès-Champanelle, France
[b] Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, PE, Canada
[c] Département des sciences animales, Université Laval, Québec G1K 1E8, Canada
[d] Animal Sciences Unit, Institute for Agriculture and Fisheries Research (ILVO), 9090 Melle, Belgium
[e] Smart Foods and Bioproducts Group, AgResearch Ltd, Palmerston North, New Zealand
[f] Dept. of Statistics, University of Warwick, Coventry CV4 7AL, UK

Animal-Statistical Guidelines for Authors-v3.pdf

1 of 31 Automatic Zoom

animal
The international journal of animal biosciences

STATISTICAL GUIDELINES FOR AUTHORS
Last updated June 2024

Contents

Plan for today's session:

- selected topics from these guidelines (Henrik),
- discussion and clarification (Charles),
- open discussion (for everyone).

Applied Example: Trial Comparing Dietary Treatments for Calves after Transport

Background: young calves are transported to an auction market or collection centre before being sent to a veal unit.

Concerns about the health and welfare of these animals during and after transport, especially potential dehydration.

J. Dairy Sci. 107:9735–9751
https://doi.org/10.3168/jds.2024-24769

Relative merits of offering a milk replacer, glucose-electrolyte, or whey-based diet on the blood composition and health of unweaned calves after transport

M. S. Cockram,[1]* H. Stryhn,[1] A. Abdallah,[2] and S. Buczinski[2]
[1]Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, PE C1A 4P3, Canada
[2]Département des Sciences Cliniques, Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, QC J2S 2M2, Canada

Study design: randomized controlled trial,

- 6 calves selected within each of 29 loads of calves (a total of 174 calves),
- calves were randomly assigned to one of 3 liquid diets (treatments),
- **sampling** occurred after arrival (prior to treatment), 2 hours and 4 hours after feeding,
- **measurements**: clinical assessment and blood samples, analyzed for blood chemistry parameters (e.g., glucose and protein concentrations).

Study objective: “To examine the health and welfare implications of offering different types of liquid feed to unweaned calves after transport to meet their energy and water deficits before a subsequent journey to a veal unit.”

Reporting Guidelines: Introduction

Recommendations:

#1 State your research hypothesis and/or describe your research objective [...]. The hypothesis or objective should follow directly from the rationale and critical analysis of available knowledge presented in the Introduction.

#2 The research hypothesis should match the study design (whether experimental or observational) and be linked to the data, the model and the statistical analysis, as appropriate.

Explanation:

It is often helpful to think of investigations as occurring in the following steps:

1. formulation of research questions, or sometimes hypotheses;
2. search for relevant data, often leading to
3. design and implementation of investigations to obtain appropriate data;
4. analysis of data; and
5. interpretation of the results, i.e., the translation of findings into a subject-matter context or into some appropriate decision.

(quote from an applied statistics textbook)

The editor's/reviewer's perspective:

What is reported in a paper, must be linked to the study objective.

Confirmatory and exploratory objectives/analyses should be distinguished, possibly also primary and secondary pre-specified hypotheses (STROBE-Vet).

Diets for young calves study:

- main interest is in comparing the three diets —
  - but on what outcome(s)?,
  - how are those outcomes measured?,
- also of interest to describe the health status of the calves upon arrival and relate it to "risk factors" → second(ary) objective.

Reporting Guidelines: Material and Methods — Study Design

Recommendations:

#1 Define the nature of your data (experimental, observational or from a collated database), and describe its structure as a two-dimensional array, e.g., how factors of variation, variables and experimental/observational units are set up in columns and rows of your data spreadsheet.

#2 For designed studies, name the experimental design you have used (e.g., completely randomised or randomised block design).

#6 Explain how the replication / sample size was chosen.

#8 Indicate the population of inference that your experimental unit is representative of.

#9 Report the *a priori* power of your experiment if you can.

Explanation:

The experimental unit is the smallest unit to which an individual treatment is imposed. By contrast the observational unit is the smallest unit on which a response will be measured.

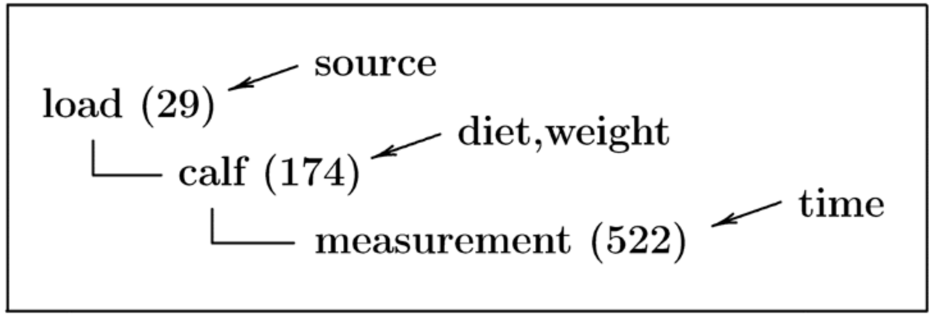
When there are multiple measurements on the same experimental unit (e.g., animals within groups, repeat samples on the same animal or samples over time) then there is no true replication, and these measurements are observational units.

The editor's/reviewer's perspective:

> For epidemiological studies, it is also important to identify the study type and the relationship between predictors (STROBE).

Diets for young calves study:

Complex designs are perhaps best understood if visualized in a diagram, e.g.,



- described in the paper as a "3-factor (diet, source, weight) randomized controlled trial".

More Details on ... Sample Size

A common source of confusion/frustration (despite the resource: Stevenson (2021)[1]),

- not some "statistical magic" that delivers a definitive number for sample size (42?),
- all formal procedures require pre-decided statistical model, involving choices of:
  - targeted outcome and scale for its analysis,
  - targeted parameters and/or hypotheses of interest,
  - assumptions involved in model/analysis,

  as well as detailed prior knowledge (estimates or guesses) about the outcomes,
- recommended to focus on a single research question formulated in the simplest possible way,
- power calculations are commonly used but not always advisable,[2]
- formal sample size calculations may not be feasible (in a meaningful way); some alternatives:
  - refer to similar previous studies,
  - consider your study as a pilot study.

Diets for young calves study:

Sample size calculations reported for comparison between two diets of changes in blood glucose, at 3 sizes of true differences.

[1] Sample size estimation in veterinary epidemiologic research, *Frontiers in Veterinary Science*, doi:10.3389/fvets.2020.539573.

[2] Alternatives may be better for study planning (Bland (2009), *BMJ* 339, 1133 – 1135); power calculations after the study has been conducted (post-hoc), are less (little) meaningful.

Reporting Guidelines: Material and Methods — Modelling / Analysis

Recommendations:

#1 A thorough description of the model is imperative to ensure reproducibility.

#2 Describe the model for statistical analysis mathematically or in words, as well as the type of analysis used. In case of non-experimental study, justify the statistical model you have used.

#4 Specify treatment structure and the methods used to compare treatment means.

#6 For correlation and regression, clearly state if a causal relationship is assumed or only an association between variables.

#1 [...] Justify and describe any transformation of the data.

Explanation / additional comments:

Sufficient details must be given of the statistical methods used to provide an understanding of the statistical analysis and results, and to allow replication.

Data transformation or other techniques are helpful when the residuals obtained after fitting regression or ANOVA are not normally distributed or when variance varies with effect size.

Data quality is critical to any analysis. Check assumptions. All data should be subjected to appropriate scrutiny [...].

The editor's/reviewer's perspective:

> Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. (SAMPL)

Diets for young calves study:

Lots of challenges in choosing and describing the statistical methods,

- different outcome types: quantitative, ordered categorical, dichotomous,
- different focuses of analyses, either change over time or at each time point (in particular, time 0),
- assumptions of complex models not easy to meet: how to deal with that uniformly across many analyses?

More Details on . . . Choosing and Validating Models

"Is (X) the right test for my data?" may be mistaken:

- "the right" — there is no unique way to analyse data,
- "test" — the focus should not only be on a test: estimates are usually more important,
- "data" — any choice depends also on the objective/focus.

When validating model assumptions, we should not base decisions on:

- the distribution of the data — assumptions are usually not about the (raw) data, but the errors (residuals),
- significance of normality tests — many procedures are robust to deviations from normality,[3]
- distributions of predictor variables (in regression) — assumptions are rarely made about predictors,

and non-parametric methods is often a poor (but simple) solution to trouble with model assumptions.

[3] And it's not a black-white decision (a wrong use of statistical tests; page 9).

Diets for young calves study:

One reviewer suggested that the data should be analysed for all three time points simultaneously; we explained why that was not done, referring to the study objective and the focus on the effects of the diets.

Models were validated by graphical evaluation of the residuals. Choice of transformation of quantitative outcomes was based on Box-Cox analysis and renewed assessment of residuals. One reviewer asked whether assumptions were tested; they were not, and we explained why.

Impact of abnormal observations (e.g., calves with little feed intake) were explored by sensitivity analysis.

Reporting Guidelines: Results

Recommendations:

#1 Report and justify deletion of data or outliers.

#2 Carefully consider the most appropriate illustration of your results.

#3 Report the descriptive statistics (e.g., in Supplementary materials). [...]

#4 Give estimates of the relevant statistics (mean values, regression coefficients, etc.) together with the appropriate SEs or Confidence Intervals of those estimates.

#5 Report [...] exact P-values except if P < 0.001.

#8 Be clear on which results are critical to your research hypothesis or question [...].

Explanation / additional comments:

Tables are recommended when absolute values may be of interest to readers. Figures are recommended to express trends (not necessarily linear), such as response with time or dose etc.

Descriptive statistics are recommended to make the readers understand the nature of your data.

The editor's/reviewer's perspective:

> The reporting of results should be clear and concise, in addition to having sufficient detail.

Diets for young calves study:

- 4 figures (all descriptive),
- 2 tables with (model-based) estimates, SE or CI, and *P*-values,
- additional results mentioned in text only (e.g., all time 0 results),
- no supplemental materials.

Interest in presenting results for changes from baseline affected the choice of analysis: multiple analyses (time 0 only, change time 0 to 2h, change time 0 to 4h) instead of one combined analysis (per outcome).

More Details on ... Statistical Hypothesis Testing and *P*-Values

The *P*-value (and statistical hypothesis testing) debate is a vast topic, and a wide range of opinions exist.

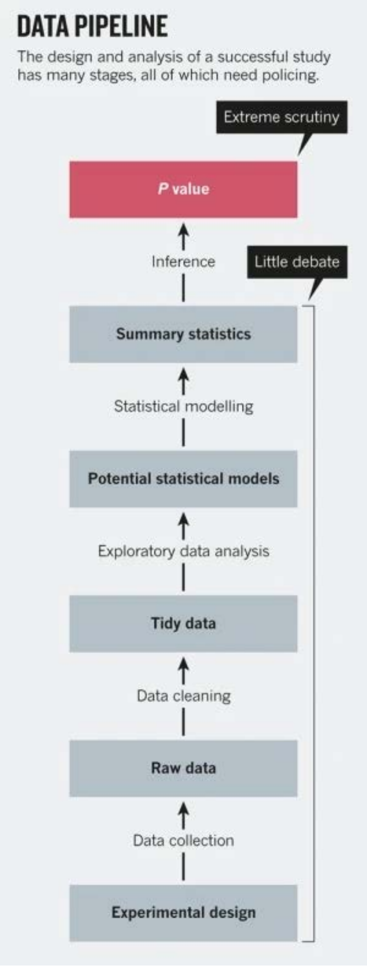
Some basic observations:

- we cannot avoid the issue: every manuscript must (implicitly) pick its position for analysis and presentation and align it with the intended journal,[4]
- misuses and misunderstandings[5] of hypothesis testing have contributed to the problems,
- statistical testing tempts users into (false) black-white interpretations, which are then critisized and blamed on the methodology,[6]
- arguably, the intense focus on this issue has been out of proportion and has overshadowed other important features of scientific work.

[4] The *animal* reporting guidelines are relatively soft and non-prescriptive on the issue; some journals are very prescriptive.

[5] Recommended comprehensive and in part technical discussion of misinterpretations: Greenland et al. (2016), *European Journal of Epidemiology* 31, 337 – 350.

[6] My personal view, and somewhat agreeing with a statement by Andrew Gelman (one of the big names in statistics): "People want something that they can't really get. They want certainty."; *Nature* 531, 151.



Diets for young calves study:

Main author's preference:

- detailed table of *P*-values,
- many *P*-values listed in text,

— justifiable (maybe) if interpreted as indicating the strength of the effects/data, not as determinants of conclusions.

See also later discussion on multiplicity.

(*Nature* 520, 612 (2015))

Reporting Guidelines: Discussion

Recommendations:

#1 Any shortcomings in design or analysis should be discussed with an indication of the possible effect on the results.

#2 Be clear on what your critical results are. Include a discussion of the biological relevance of the determined magnitude of effects for those results.

#3 Explicitly state the scope of inference, or generalisability and limits, of your results and conclusions.

Explanation / additional comments:

To consider: How general are the conclusions you are drawing from your results? To what extent do they depend on the set-up of your study?

In many studies, [...] there will also be other ancillary variables which are related to the main variables and may be highly correlated with them. As a result, multiple statistically significant effects are not independent often replicating the same response. Do not make the assumption that such multiple effects strengthen or reinforce your inference: they may simply reflect the same overall effect.

The editor's/reviewer's perspective:

> Conclusions and interpretations must be consistent with the results presented.

Diets for young calves study:

Discussion of results separated into results on arrival and effects of diets (i.e., changes from baseline) on:

- blood metabolites,
- dyhydration,
- risk of diarrhoa,
- health and vigour.

Actually, no discussion of the generalisability of results...

More Details on ... Multiplicity

The issue:

If we explore very many characteristics of a data set, something "interesting" is certain to come up (by chance)[7], due to some peculiarity of the data that is nowhere near reproducible.

— related to (but not limited to) limitations of statistical testing.

How can it be "bad" to extract all the information from the data?

- problem is not with extracting the information, but how we interpret it: a strong finding in the data that we were not necessarily looking for, may not be so strong after all,
- most statistical testing (or other decision making) is set up for pre-specified hypotheses, not patterns suggested by the data.

Some attempts to deal with the problem:

- requirements (in some journals/fields) of a protocol published prior to data collection,
- separation of study objectives into confirmatory (pre-defined hypothesis for which we seek evidence) and exploratory (may suggest relationships, but should be confirmed in other studies),
- adjustments for multiple comparisons in ANOVA: valid and recommended, but addresses only a small part of the problem (and possibly too strongly),
- other adjustments for multiplicity (e.g., multiple outcomes, multiple time points) — rare in practice.

[7] For a silly (funny) illustration: m.xkcd.com/882.

Thoughts on ... AI-assisted Data Analysis

Using AI for coding — can (perhaps)

- replace searches in help files, software documentation,
- replace Q&A searches on internet, across blog posts etc.,
- produce code that runs . . . (but must check it does the right thing).

Using AI for understanding — can (perhaps)

- generate nicely written explanations,
- replace searches in books and internet (Wikipedia)

— but need to ensure proper documentation.[a]

Using AI for designing studies or analyses is (perhaps) more challenging:

- how to ensure we understand and agree with its decisions?
- are we confident to trust its interpretations (based on correlation within its database)?
- how to avoid we implicitly create its answers (prompt bias)?

[a] Anecdotally, AI interfaces have produced references that do not exist.