

ARTICLE

Methods, Tools, and Technologies

Writing statistical methods for ecologists

Amy J. Davis¹  | Shannon Kay² 

¹United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center, Fort Collins, Colorado, USA

²United States Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA

Correspondence

Amy J. Davis
Email: amy.j.davis@usda.gov

Funding information

U.S. Department of Agriculture, Forest Service

Handling Editor: John Humphreys

Abstract

The *Methods* section is a key component of any ecology research publication containing detailed information on how the data were collected and analyzed. However, descriptions of which statistical methods were used and how they were applied can substantially vary and may not provide enough information for the analyses to be reproducible. Computational and statistical programming advances have allowed ecological researchers without a strong statistical or mathematical background to access and use increasingly complex statistical methods. Thus, statistical methods are written by and need to be accessible to researchers across a range of quantitative expertise. Poorly written *Methods* sections can incorrectly inflate the strength of or call into question the results of an analysis. Although there are resources available, we have not found one that is specific to writing statistical methods, includes all the elements we discuss, and is targeted for ecologists. Here we provide guidelines for ecological researchers when writing statistical methods and review frequent errors made in *Statistical Methods* sections. We highlight some common dos and don'ts when writing *Statistical Methods* sections and present a simple checklist to help guide authors with their writing to ensure reproducibility. We illustrate the use of this guidance with two examples.

KEYWORDS

ecological communication, quantitative ecology, scientific writing, statistical methods, statistics education, writing clarity

INTRODUCTION

Statistics are a cornerstone of ecological research. Unfortunately there have been high levels of anxiety associated with taking statistics courses (Onwuegbuzie & Wilson, 2003) and many tend to dislike or avoid statistical or mathematical courses (Hogg, 1991; Prayoga & Abraham, 2017; Sloomaeckers et al., 2014). Ecological researchers have a diversity of statistical backgrounds

from having formal training to learning on the job. A survey of mostly early-career ecologists found that 75% “were not satisfied with their understanding of mathematical models” (Barraquand et al., 2014). For many ecologists, *Statistical Methods* sections are challenging and time-consuming to fully understand and may be skimmed, or skipped altogether, when reading articles. If readers cannot comprehend the *Statistical Methods* sections as written, they will likely be unable to repeat the analyses, and perhaps more importantly, will be unable to judge the strength of the results. Many aspects

Amy J. Davis and Shannon Kay contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Published 2023. This article is a U.S. Government work and is in the public domain in the USA. *Ecosphere* published by Wiley Periodicals LLC on behalf of The Ecological Society of America.

of a *Statistical Methods* section may be too “in the weeds” for some readers, and it has been suggested (Fawcett & Higginson, 2012) and refuted (Fernandes, 2012) that more equations in biological papers equate to fewer citations. However, explaining the rationale for each statistical approach in a way that provides context for all readers and refraining from relegating the “scary math” to appendices is critically important and aids in the advancement of new theories (Chitnis & Smith, 2012).

Recent statistical and computational tools have been developed to help broaden the accessibility of different methods to a larger pool of ecological researchers. The use of “Plug-and-play” statistical software (e.g., OpenEpi—Dean et al., 2013) allows for analyses to be conducted by answering simple questions. Graphical user interface (GUI) options for common ecological analyses (e.g., Buckland et al., 2001; Rexstad & Burnham, 1991; White & Burnham, 1999) allow users to conduct more complex analyses in a user-friendly manner. The rise of freeware, coding-based statistical software (e.g., R; R Core Team, 2021) has allowed for greater use of complex statistical approaches by researchers without advanced statistical training. The use of R has become widespread in ecology, with nearly 60% of articles published in ecology journals in 2017 reported using R as their primary

tool in data analysis (Lai et al., 2019). The availability of freeware Bayesian analysis options (JAGS—Plummer, 2003; STAN—Gelman et al., 2015; NIMBLE—de Valpine et al., 2017) has led to increases in the use of Bayesian approaches in ecology (Anderson et al., 2021; Hooten & Hobbs, 2015). There has been an increase in the use of sophisticated and computationally intensive statistical approaches in ecology over time (Touchon & McCoy, 2016). Despite this rise in the use of advanced statistical methods, many graduate programs have not kept pace by requiring statistics coursework (Touchon & McCoy, 2016). To ensure the statistical approaches in ecological publications are comprehensible, repeatable, and properly cited, clear guidance is needed on best practices for writing statistical methods.

Proper use of statistics is critically important and there are extensive resources available for guidance on which methods are appropriate (e.g., Fox et al., 2015; Gotelli & Ellison, 2004; McGarigal et al., 2013; Steel et al., 2013; Zuur et al., 2010, 2007), but our focus is specifically for *writing* and *reporting Statistical Methods* sections. There are resources that focus on how to effectively communicate mathematical concepts to scientists in other fields (Table 1). Recently, Shoemaker et al. (2021) provide general guidance for writers and readers on

TABLE 1 References of interest for deeper dives into related concepts to writing statistical methods (not an exhaustive list).

Reference	Description
Other sources on methods writing	
Adams-Huet and Ahn (2009)	A guide for writing statistical methods targeted at clinical investigators.
Azevedo et al. (2011)	A guide for writing methods in general with a section on data analysis, focusing on medical methods.
Ghasemi et al. (2019)	Guidance on writing methods for biomedical fields, provides many useful tables, boxes, and references.
Lang and Altman (2015)	Detailed guidance on statistical reporting specific to analytical method used, targeted for biomedical researchers.
Michel et al. (2020)	Specific guidance with examples on how to report data analysis and statistical methods for experimental biology.
Worthy (2015)	Guidance on statistical analysis and reporting for medical fields. Provides recommendations on methods, results, tables and figures, and discussion.
Sources on how to effectively communicate mathematical concepts	
Crowe and Cash (2023)	Short guide to writing <i>Methods</i> sections for broad range of researchers with focus on common pitfalls to avoid and pearls to include.
Fallon (2016)	Multidisciplinary writing guide for researchers using a series of prompting questions with a final section of examples.
Scheinerman (2011)	Descriptions of mathematical notation and usage, how to produce mathematical symbols in LaTeX, and links to further reading on topics.
Shoemaker et al. (2021)	Recommendations for both readers and writers of mathematical ecology focusing on bridging the gap between advanced modeling techniques and understandability.
Sources for reviewers	
Triaridis and Kyrgidis (2010)	Provides guidance on how to review articles and particularly statistical methods. Helpful when writing to see how methods may be critiqued.
Parker et al. (2018)	A checklist to help reviewers ensure transparency.

communicating mathematics. Other resources provide guidance on writing *Methods* sections in general or how to write and report specific statistical details (Table 1). These publications provide some useful overlap for our objective in terms of ensuring that mathematical elements are clearly linked to the ecological concepts and to use illustrations and diagrams to help convey ideas. However, there is a need for clear guidance on how to comprehensively and clearly communicate statistical methods used by ecologists.

Recognizing that there are errors or inconsistencies in reporting statistical methods (Michel et al., 2020; Worthy, 2015), there has been a push, particularly in the biomedical fields, to develop uniform guidance on how to present statistical methods and results (Crowe & Cash, 2023). This has led to more detailed recommendations as to how to report statistical methods in these fields (e.g., Indrayan, 2020; Lang & Altman, 2015; Worthy, 2015). Although *Methods* sections in general have long been used to enable study replication, recent understandings of the complexities of statistical methods have engendered new standards to help ensure reproducibility in statistical work by putting a greater emphasis on publishing data and corresponding code for analyses with articles (Fidler et al., 2017; Peng, 2011). We support these efforts and additionally recommend more uniformity in how statistical methods are presented to improve clarity and reproducibility in ecological research.

Our objective is to provide recommendations for ecological researchers of all levels of statistical knowledge on how to write clear, repeatable *Statistical Methods* sections. We start by identifying common pitfalls in statistical writing, then give guidance for authors describing what information should be included in any account of statistical analyses performed. Like other resources (e.g., Makin & Orban de Xivry, 2019; Parker et al., 2018; Zuur & Ieno, 2016), we include a simple checklist for authors to help guide writing statistical methods and include some simple dos and don'ts to aid in avoiding any oversights (Schroter et al., 2008). Finally, we demonstrate how to use this guidance by working through two ecological examples of *Statistical Methods* sections.

PITFALLS IN STATISTICAL WRITING AND HOW TO AVOID THEM

There are several common pitfalls ecologists make when writing *Statistical Methods* sections in publications such as not clearly defining each test or analysis, not providing enough information to make it reproducible, or not properly citing methods or the software used to implement the

analyses (Makin & Orban de Xivry, 2019; Miller, 2006). Likely the most important consideration is ensuring that each analysis appears in the *Methods* section and enough information is provided so that others fully understand the results and could repeat the same analysis. Results of analyses that were conducted post hoc are sometimes presented in the *Results* section or even the *Discussion* section but were never described in the *Methods* section, which does not allow readers to replicate these analyses. Even if an analysis or hypothesis test was conducted after the planned analyses, perhaps to test a specific follow-up question, the approaches for these need to be described in the *Methods* section. This does not equate to requiring one to narrate every bit of exploratory data analysis conducted or to describe common statistical methods in detail. Novel statistical methods need more detailed descriptions than more general or well-known methods, and citations should be provided when available (Worthy, 2015).

Another pitfall is to use terminology incorrectly to describe statistical methods or metrics. Precise terminology importantly avoids confusion in statistical writing. *Parameters* can be either *estimated* directly from a model or *derived* from model estimates. For example, we could say that we *estimated* monthly survival rates from a survival analysis, and we *derived* annual survival rates from the monthly survival estimates. If descriptions in the *Methods* state “we estimated parameter X” but the parameter was derived from the model, the reader may misunderstand the nature of the model. It is common for people to refer to “running models”; however, it is more appropriate to say models are *fit* to the data. It is appropriate to say “we *run* simulations” based on a model or set of assumptions to aid in prediction. It is important to be mindful of which terms are and are not interchangeable to avoid confusion.

Another pitfall is confusion between the statistical method and the tool or program used to carry out the analysis. We often see authors reporting the function itself (e.g., `lm` in R) rather than the actual analysis (e.g., linear regression). Similarly, if equations are included, they should reflect the mathematical form (e.g., $y = \beta_0 + \beta_1 x + \varepsilon$) and not the software program's function formula (e.g., $y \sim x$). However, both the method and the tool used to implement the method should be reported and properly cited (Worthy, 2015).

GUIDANCE FOR AUTHORS

In Box 1 we provide some dos and don'ts for writing a *Methods* section for ecologists. Box 2 has a checklist to guide statistical methods writing for each analysis. We also provide examples of “sufficient” and “insufficient”

BOX 1 Dos and don'ts in writing statistical methods for ecologists

Dos

1. Do imagine a nonscientist reading your article—would they understand what the objectives/rationale were for each analysis?
2. Do include enough information for reproducibility: could another researcher repeat the statistical analyses you performed from your description?
3. Do describe EVERY analysis used in the paper, including post hoc analyses or tests
4. Do clearly describe which data or subset of data are used in each analysis.
5. Do review guidance for authors from journals for standards specific to their audiences

Don'ts

1. Don't include results or discussion of statistical approaches that were not first described in the *Methods* section.
2. Don't equate a software function with a statistical method.
3. Don't forget to give credit where it is due! Be sure to cite statistical methods, software, and packages.

descriptions of methods and an example of how to use the checklist to guide statistical methods writing. These guidelines should help authors ensure that their analyses are clearly defined, understandable, and repeatable. Following these guidelines will help ecologists be more effective in their communication, which may lend more credence to their results and help readers make their own conclusions regarding inferences made from the analyses.

The first step is to consider the target audience and the typical analytic background of the intended readers. It can be helpful to consult a journal's *Guidance for Authors* section to determine if there are specific recommendations for statistical methods. If there is no specific advice from the journal, then looking at previous publications from the journal will give the author an idea of what may be expected. However, published articles are not without faults (Michel et al., 2020; Nuijten et al., 2016) and lack of detail in published articles should not be treated as justification not to include detail in one's own writing. Even when targeting a highly

BOX 2 Statistical methods writing checklist for ecologists

Items in boldface are necessary; items in italics are helpful to include when appropriate.

For each analytical approach be sure to include:

1. **Objective:** Describe the rationale for the analysis and how it relates to the study objective.
2. **Variables:** Define the experimental unit and the response and predictor variables clearly.
 - **Response:** Specify the response variable distribution, why it is appropriate, and any link functions.
 - **Predictors:** Specify which predictor variables were examined and which were fixed or random effects.
 - *Meaningful difference:* State what is an ecologically meaningful difference in the response variable.
3. **Statistical method:** Describe the statistical method (e.g., *t*-test, generalized linear model).
 - **Model fit:** Specify how model fit was assessed.
 - *Model comparison:* Specify the model comparison approach and how many models were fit.
4. **Implementation:** Describe the function, package, and software used and include any appropriate citations (e.g., *glmer* in package *lme4*, in R).
 - **Cite:** Cite software and packages.

quantitative audience, statistical methods should be written with sufficient justification and context to allow a general reader to follow the logic of the approach even if the technical details are beyond their training. Ultimately, the *Results* section should flow clearly and directly from the *Methods* section. If many analyses were conducted, consider adding subheadings to aid the reader in matching each analysis with its result.

Checklist

Our checklist for statistical methods writing is designed to guide ecologists on how to write comprehensive statistical methods. For each analysis included in a study, the author should include these four primary components: the objective or rationale of the analysis, the variables

used in the analysis, the statistical method, and how the analysis was implemented. Details for each element are described below. Once authors have a clear understanding of each element, they can use the checklist (Box 2) to help guide them in their writing. Some statistical methods have elements of the checklist implied (e.g., logistic regression implies a binomial response variable) or elements that are not applicable (e.g., model selection is not needed for a *t*-test) and therefore they do not need to be explicitly stated. In the checklist, we distinguish between necessary elements (our four primary components) and helpful elements that should be included when appropriate. The order of elements does not always have to follow the order in which we present them; it is more important that all elements are included. We do recommend starting with the objectives to orient the readers and ending with how the analyses were implemented to not confuse the tool (program/package) with the statistical method, but this is not essential.

Objectives

Every statistical approach used in a study should be conducted to help achieve the study objectives. Writers should take care to connect the statistical method that is described to the biological/ecological metric of interest and its relevance in the study (Lang & Altman, 2015; Worthy, 2015). Making these connections clear helps both scientists and nonscientists understand the *Methods* sections. Additionally, the better these sections are explained and connected to the implications of the study, the easier it is for the reader to assess the inferences made. By making sure you can clearly articulate the rationale for your statistical approaches, and link them to the study objectives, the statistical approach will be more comprehensible, and your results will be more impactful.

Variables

All relevant variables should be clearly described in the *Methods* section including the units of measurement (e.g., km², %, mm) and the scale at which each variable is measured (e.g., an individual, a site, a population; termed the experimental unit). The experimental unit of each variable should clearly link back to the experimental design. The response variable (also called the dependent variable), the characteristic you are interested in predicting or investigating relationships with, should be identified in addition to all the predictor variables (also called independent variables, covariates, or explanatory variables), the characteristics that may relate to

the response variable. If using a parametric model (i.e., a model that makes distributional assumptions), the distribution that is appropriate for the response variable and any link functions should be stated and which variables were used as random or fixed effects should be included (Gelman & Hill, 2006). For simple linear regression, the response variable is modeled as a normal or Gaussian distribution, which is implied in simple linear regression and does not need to be reiterated. However, for generalized linear models or other modeling approaches (e.g., Bayesian), the appropriate distribution may be one that only allows for positive data or discrete data. Describing the distribution of the response will ensure the approach is repeatable and explaining the rationale for the distribution will help lay people follow the logic of the analyses.

All predictor variables should be explicitly listed with their justification for inclusion in the analyses; and, if relevant, their data sources should be provided (e.g., remote sensing data, habitat data, or climate data). Understanding the motivation for including certain predictor variables will help the reader follow the story the authors are trying to tell and allow the readers to consider any additional predictor variables that may be lacking in an analysis. The assumed functional form of the predictor variables (linearity, interactions, etc.) should also be stated. For Bayesian models, in addition to the distributions for the variables in the model, the distributions for the priors should be specified. Explaining which predictor variables or combinations of predictors are considered, the approach can be replicated, and readers can better assess the strength of the approach.

An often-overlooked feature that should be included in the *Methods* section is to identify what magnitude of a difference in the response variable is ecologically meaningful (Lang & Altman, 2015). There is an important difference between statistical significance and biological/ecological significance (Nakagawa & Cuthill, 2007). Reporting effect sizes (e.g., Cohen's *d*; Cohen, 1988) is a good way to evaluate whether a statistically significant result is also relevant in a practical sense. Relatedly, the range of values of the predictor variables used is helpful to provide. It is possible to not detect a relationship with a predictor that is frequently considered important by other research, but the range examined in a study might not be wide enough to make a meaningful difference. A power analysis could be conducted to determine the likelihood of detecting a biologically meaningful difference, if so, the rationale and approach should be described. By priming the reader with what type of difference is ecologically meaningful, the reader will be better prepared to interpret the results appropriately.

Statistical method

An important element to include for each analysis is the statistical method being used. As mentioned previously, there can be confusion between the statistical method (e.g., linear regression) and the tool that is used to implement the statistical method (e.g., `lm` in R). It is important to distinguish the statistical method from the tool used to conduct the analysis. By conflating the method and the tool, it may come across as a lack of foundational understanding of the approaches used and may lead to lack of confidence in the work.

For most *Statistical Methods* sections, equations do not need to be provided. Mathematical equations are the basis for statistical analyses, but, depending on the audience, they may not be the most effective ways of describing an approach. Do include equations when modifications are being made to standard approaches or for custom Bayesian hierarchical models. If a novel statistical approach is being used, then equations are usually necessary to demonstrate the fundamentals of the approach. Similar to others (e.g., Cronin & Schoolmaster, 2018; Shoemaker et al., 2021), we encourage the use of diagrams or directed acyclic graphs (i.e., a visual representation of model specification and parameter relationship) to help convey novel mathematical/statistical approaches to general audiences. However, for established statistical approaches, references to those approaches or even equations in other publications are usually sufficient.

In addition to model specifics, it is important to describe which models were fit, how models were compared, and how model fit was assessed. There are many good guides on how to conduct model selection in ecology (Doherty et al., 2012; Hooten & Hobbs, 2015; Johnson & Omland, 2004; Tredennick et al., 2021). It is important to be mindful of how parameters from models are interpreted (Arif & MacNeil, 2022; McElreath, 2020; Tredennick et al., 2021). If any model selection was performed, make sure to identify the approach used, explain why it is appropriate for the situation, describe what a meaningful difference in model metrics would be (e.g., $\Delta 2 \text{AIC}_c$ suggests a meaningful difference in model parsimony), and provide a citation for the approach used (e.g., Burnham & Anderson, 2002). It should also be clear how model selection uncertainty would be addressed (e.g., model averaging, cumulative covariate weights; Burnham & Anderson, 2002).

Model selection methods often provide relative measures of goodness of fit (e.g., AIC, BIC, WAIC, cross validation; Burnham & Anderson, 2002; Hooten & Hobbs, 2015). That is, they *compare* the fit of the set of models considered but do not provide an absolute measure of how well the “best” model fits the data. Essentially, a goodness-of-fit

metric is one that compares the expected values from a model to the observed data. Goodness-of-fit metrics are important to include to demonstrate to the reader how good the model is at fitting the data and, therefore, how reliable the results are. Goodness-of-fit metrics are specific to the model being used (e.g., R^2 for simple linear regression, area under the curve for logistic regression), and in many cases there are no predefined goodness-of-fit metrics available. There are options to assess the goodness of fit even when a predefined metric does not exist for an analysis, including bootstrapping, simulations, or a type of cross validation to assess model fit (Bonett & Bentler, 1983; Boyle et al., 1997; Stute et al., 1993). For Bayesian models, researchers can use posterior predictive checks (Gelman et al., 2013).

In addition to model fit specifications, for Bayesian models, authors should describe how convergence was evaluated. The reliability of Bayesian models depends on good mixing and convergence in the posterior distributions. Understanding how convergence was assessed allows readers to have confidence in the results. For nonparametric models, the values of tuning parameters should be reported—especially if not using defaults (e.g., “`mtry`” in a random forest, the number of variables selected at each split, defaults to $\sqrt{p}$ for classification and $p/3$ for regression where p is the number of predictor variables). Similarly, in a simulation study, all the specific parameter values used should be reported. Note that including some of the more detailed method information in a Supplementary section is generally an option.

Implementation

Not only is it important to describe and cite the statistical methods used, but also the tools used to implement the analysis. This is critical for ensuring reproducibility as there can be differences in optimization routines and the presentation of results across programs. If there are different options for estimation approaches and an author uses a nonstandard approach, for example, using restricted maximum likelihood (REML) instead of maximum likelihood (ML) in fitting a generalized linear model, authors should report the method used. As mentioned previously, if unique code was developed as part of a study, authors should provide the code along with the manuscript (Culina et al., 2020; Filazzola & Lortie, 2022). Many readers find it useful to walk through analyses with the code provided, leading to a better understanding of what was done. Furthermore, credit should not just be given for the statistical method being used, but also for the tools used to apply the method including any software, packages, and functions. Even within the same software, there

can be many ways to perform the same analysis using different functions that may have options for means of estimation. Additionally, it is important to include the version number for programs and packages used as estimation approaches may vary by version. By properly reporting and citing these tools, the analyses will be reproducible. The citations associated with programs/packages can also be great references to provide background on the statistical methods being used.

EXAMPLES

Here we present ecological research examples and demonstrate how to effectively write a *Statistical Methods* section using the guidance we have provided. We start with general design setups but recognize that a true *Methods* section would be more in-depth; our objective is simply to provide context for the statistical methods writing guidance.

Example 1: Comparison of sufficient and insufficient statistical methods writing

Suppose we are interested in identifying influential environmental factors and habitat characteristics associated with the number of eggs laid by a particular raptor species. We conducted annual surveys of 100 randomly selected historic nest locations across the species' breeding range. We recorded the number of eggs laid at each nest and then added climate data (i.e., average annual temperature and average precipitation) and habitat data (i.e., dominant habitat type) around each nest to use in fitting a model to determine which of these weather or habitat factors may be significantly affecting clutch size. Assuming we have already described in the general methods how the nest data were collected (e.g., site selection, sampling protocols, details of data recorded, experimental unit) and how the study design relates to the objectives (e.g., we designed a study to examine how a range in environmental and habitat variables relate to clutch size to help understand potential impacts of climate change on species X), we start off our *Statistical Methods* section by defining our objectives and type of study design along with the variables we are going to include:

We analyzed the effects of climate variables and habitat variables on nest data of bird X,

but this does not clearly articulate our objective or provide specific information about the variables we are using. We revised the text as follows:

We were interested in identifying whether climate variables, specifically, annual precipitation (mm) and the average annual temperature ($^{\circ}\text{C}$), as well as the dominant land cover type (including shrub, deciduous forest, evergreen forest, grassland, and wetland), were significantly related to the number of eggs laid by bird species X. Based on population viability analyses for this species ([citation]), an increase in 0.5 eggs per nest would be a meaningful difference.

Now that we added the specifics on which climate variables and habitat variables, we included and what differences would be ecologically meaningful, we need to define what statistical method was used and how each variable was treated.

We used glm in R to run a model on precipitation, temperature, and land cover type.

This sentence is not appropriate because it is confusing the statistical method with the tool, we incorrectly refer to "running" a model instead of "fitting" a model, and we do not define what distribution was used in the GLM or the functional form of variables. We also need to specify whether any variables were included as random effects. Updating the text, we have:

We used a generalized linear mixed model (GLMM) to examine the relationship of our response variable, the number of eggs laid in each nest, and the fixed effect predictor variables of precipitation, temperature, and land cover class. We used a Poisson distribution for our response variable as the number of eggs must be a discrete number and a log link because the number of eggs cannot be negative. We included a polynomial term for temperature because we hypothesized that the number of eggs laid by bird X may increase with temperature to a certain point before decreasing at higher temperatures. Additionally, we hypothesized that an interaction may exist between precipitation and land cover class where precipitation may have a greater effect on the number of eggs laid in some habitats compared to other land cover classes. We included a random effect for a mated pair to account for multiple nesting attempts by the same pair over time.

Notice how we included the rationale for using the polynomial term on temperature and the interaction

between precipitation and land cover class. Now we need to describe how we assessed model fit, whether we did model selection, and how we implemented our analyses.

We used simulated residuals (Dunn and Smyth 1996) to visually assess model fit by plotting residuals against each predictor variable. We used 10-fold cross validation to evaluate the goodness of fit of the model. Analyses were implemented using the statistical software package R (R Core Team 2021)

using the lme4 (Bates et al. 2015) and DHARMA (Hartig 2022) packages.

Note that not only did we cite what software was used, but we also included a citation for each package we used within the software.

In summation, we have touched on (1) our objective, (2) what variables were included and how they were treated in the analysis, (3) what type of statistical method was used and how we assessed fit, and (4) how we implemented the analysis with proper citations (Box 2).

Statistical Methods Writing Checklist for Ecologists: Example—Occupancy

☒ Objective	We wanted to understand the spatial patterns of presence and absence of amphibian species Y in ponds across our study area.
☒ Variables ☐ Response ☒ Predictors ☒ <i>Meaningful difference</i>	[The response variables of an occupancy analysis are implied as the presence/absence of the species and the detection/non-detection of the species. Therefore, they do not need to be stated explicitly in the methods section.] We used a categorical variable specifying if the pond was large or small to examine if pond size was an important predictor of occupancy. We included a continuous variable for pH level as we expected a relationship within the range of pH levels observed. We included an indicator variable for precipitation on detection to account for an effect of potential declines in detectability when it rained. We considered a 10% difference in occupancy meaningful in our study.
☒ Statistical method ☒ Model fit ☒ <i>Model comparison</i>	We used an occupancy analysis which estimates occupancy and accounts for imperfect detection typical of wildlife species (MacKenzie et al., 2017). We assessed goodness of fit using a bootstrapped chi-squared statistic outlined by MacKenzie and Bailey (2004). We compared models using Akaike Information Criterion corrected for small sample size (AICc; Burnham & Anderson, 2002). Models within 2 delta AICc from the top model were considered competitive and if there was model uncertainty, we used model averaging (Burnham & Anderson, 2002).
☒ Implementation ☒ Cite	Occupancy analyses were conducted in Program MARK (White & Burnham, 1999).

FIGURE 1 Checklist applied to writing statistical methods for occupancy analysis in [Example 2](#).

Example 2: Using the checklist to guide statistical methods writing

The setup

Suppose we are interested in investigating the presence/absence of an amphibian species across a wetland landscape. We want to know if this species occurs more often in large or small ponds to help inform management. We also suspect that ponds with higher pH levels will be less hospitable to this species. We think that precipitation may impact detection. We conducted an occupancy analysis. Post hoc, we conducted a likelihood ratio test to corroborate if pond size was an important factor. We will use the checklist to ensure we have clearly described both the occupancy analysis and the likelihood ratio test. Starting with the occupancy analysis, follow along in Figure 1 as we provide the elements needed (right side) that correspond with the checklist items (left side). Elements in brackets are implied by or unnecessary for the approach and do not need to be written out in a *Methods* section. We have included them in the checklist for completeness.

The checklist should be used for each analysis. The only other analysis in our study is the likelihood ratio test. Follow along in Figure 2 to see how we develop the write-up for the likelihood ratio test.

Putting it all together, we decided that describing the statistical method before the variables flowed better. We also edited to eliminate any redundancies that resulted from writing out each component of the checklist and omitted any elements implied by the method.

We wanted to understand the spatial patterns of presence and absence of amphibian species Y in ponds across our study area. We used an occupancy analysis which simultaneously estimates the probability of occurrence and accounts for imperfect detection typical of wildlife species (MacKenzie et al. 2017). We used a categorical variable specifying if the pond was large ($x = 1$) or small ($x = 0$) to examine if pond size was an important predictor of species Y occupancy. We considered a difference of 0.1 in occupancy meaningful in our study. We included a continuous variable for pH level as we expected a linear relationship within the narrow range of pH levels observed. We included an indicator variable for an effect of precipitation on detection to account for potential declines in detectability when it was raining.

Statistical Methods Writing Checklist for Ecologists: Example—Likelihood ratio test

☒ Objective	We wanted to test if pond size was an important factor in species Y occupancy.
☒ Variables ☐ Response ☒ Predictors ☒ Meaningful difference	[Likelihood ratio tests are commonly used, and it is understood the response ratio is χ^2 distributed. It does not need to be stated as it is implied by the approach.] We conducted the likelihood ratio test on the full model and the full model without the pond size factor. We used a significance level of 0.05.
☒ Statistical method ☐ Model selection ☐ Model fit	We used a likelihood ratio test. [Model selection is not needed for hypothesis tests. Fit is implied.]
☒ Implementation ☒ Cite	The likelihood ratio test was conducted in program MARK (White & Burnham, 1999).

FIGURE 2 Checklist applied to writing statistical methods for the likelihood ratio test in [Example 2](#).

We compared models using Akaike Information Criterion corrected for small sample size (AIC_c ; Burnham and Anderson 2002). Models within 2 delta AIC_c values from the top model were considered competitive (Burnham and Anderson 2002). If there was model uncertainty (i.e., if more than one model was within 2 AIC_c of the top model), we used model averaging (Burnham and Anderson 2002). We assessed goodness of fit using a bootstrapped χ^2 statistic outlined by MacKenzie and Bailey (2004). Occupancy analyses were conducted in program MARK (White and Burnham 1999).

We used a likelihood ratio test to specifically test if pond size was an important factor in species Y occupancy. We conducted the likelihood ratio test on the fully parameterized model and the full model without the pond size factor. We used a significance level of 0.05 as our cutoff. The likelihood ratio test was conducted in the program MARK (White and Burnham 1999).

CONCLUSION

Writing clear, comprehensive, and repeatable *Statistical Methods* sections is an important skill for ecological researchers. We want our research to be accessible; therefore, it is important to write our articles to be digestible regardless of the educational background of the reader. Statistical components of a manuscript can often deter some from wanting to read an article or prevent some from understanding what was done. For authors, writing *Statistical Methods* sections can be daunting. Herein, we provide clear guidance for ecologists to ensure their methods are interpretable by a broad audience—by clearly explaining the objectives of each analytical approach—and their methods are repeatable—by describing statistical methods, defining the variables, and stating the program/packages used to implement the analyses.

ACKNOWLEDGMENTS

We want to thank the multiple researchers who provided comments on previous drafts of this manuscript, including S. Baggett, E. Beval, F. Buderman, A. Burns, M. Combs A. Feuka, R. Giglio, E. Helmer, J. Morissette, C. Pierce, S. Shriner, and S. McKee. The findings and conclusions in this report are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy. This research was supported [in part] by the U.S. Department

of Agriculture, Forest Service. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

No data were collected for this study.

ORCID

Amy J. Davis  <https://orcid.org/0000-0002-4962-9753>

Shannon Kay  <https://orcid.org/0000-0001-8742-6570>

REFERENCES

- Adams-Huet, B., and C. Ahn. 2009. "Bridging Clinical Investigators and Statisticians: Writing the Statistical Methodology for a Research Proposal." *Journal of Investigative Medicine* 57: 818–24.
- Anderson, S. C., P. R. Elsen, B. B. Hughes, R. K. Tonietto, M. C. Bletz, D. A. Gill, M. A. Holgerson, S. E. Kuebbing, C. McDonough MacKenzie, and M. H. Meek. 2021. "Trends in Ecology and Conservation over Eight Decades." *Frontiers in Ecology and the Environment* 19: 274–82.
- Arif, S., and M. A. MacNeil. 2022. "Predictive Models Aren't for Causal Inference." *Ecology Letters* 25: 1741–5.
- Azevedo, L., F. Canário-Almeida, J. A. Fonseca, A. Costa-Pereira, J. Winck, and V. Hespanhol. 2011. "How to Write a Scientific Paper-Writing the Methods Section." *Revista Portuguesa de Pneumologia* 17: 232–8.
- Barraquand, F., T. H. Ezard, P. S. Jørgensen, N. Zimmerman, S. Chamberlain, R. Salguero-Gomez, T. J. Curran, and T. Poisot. 2014. "Lack of Quantitative Training among Early-Career Ecologists: A Survey of the Problem and Potential Solutions." *PeerJ* 2: e285.
- Bonett, D. G., and P. M. Bentler. 1983. "Goodness-of-fit Procedures for the Evaluation and Selection of Log-Linear Models." *Psychological Bulletin* 93: 149.
- Boyle, P., R. Flowerdew, and A. Williams. 1997. "Evaluating the Goodness of Fit in Models of Sparse Medical Data: A Simulation Approach." *International Journal of Epidemiology* 26: 651–6.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford: Oxford University Press.
- Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodel Inference—A Practical Information-Theoretic Approach*, 2nd ed. New York: Springer-Verlag.
- Chitnis, N., and T. A. Smith. 2012. "Mathematical Illiteracy Impedes Progress in Biology." *Proceedings of the National Academy of Sciences of the United States of America* 109: E3055–5.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.
- Cronin, J. P., and D. R. Schoolmaster, Jr. 2018. "A Causal Partition of Trait Correlations: Using Graphical Models to Derive Statistical Models from Theoretical Language." *Ecosphere* 9: e02422.
- Crowe, R. P., and R. E. Cash. 2023. "A Letter from the Editors: Pearls and Pitfalls for Writing a Methods Section." *Prehospital Emergency Care* 27: 117–20.

- Culina, A., I. van den Berg, S. Evans, and A. Sánchez-Tójar. 2020. "Low Availability of Code in Ecology: A Call for Urgent Action." *PLoS Biology* 18: e3000763.
- de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R. Bodik. 2017. "Programming with Models: Writing Statistical Algorithms for General Model Structures with NIMBLE." *Journal of Computational and Graphical Statistics* 26: 403–13.
- Dean, A. G., K. M. Sullivan, and M. M. Soe. 2013. "OpenEpi: Open Source Epidemiologic Statistics for Public Health." www.OpenEpi.com.
- Doherty, P. F., G. C. White, and K. P. Burnham. 2012. "Comparison of Model Building and Selection Strategies." *Journal of Ornithology* 152: 317–23.
- Fallon, M. 2016. *Writing up Quantitative Research in the Social and Behavioral Sciences*. Boston, MA: Sense Publishers.
- Fawcett, T. W., and A. D. Higginson. 2012. "Heavy Use of Equations Impedes Communication among Biologists." *Proceedings of the National Academy of Sciences of the United States of America* 109: 11735–9.
- Fernandes, A. D. 2012. "No Evidence that Equations Cause Impeded Communication among Biologists." *Proceedings of the National Academy of Sciences of the United States of America* 109: E3057–7.
- Fidler, F., Y. E. Chee, B. C. Wintle, M. A. Burgman, M. A. McCarthy, and A. Gordon. 2017. "Meta-research for Evaluating Reproducibility in Ecology and Evolution." *BioScience* 67: 282–9.
- Filazzola, A., and C. Lortie. 2022. "A Call for Clean Code to Effectively Communicate Science." *Methods in Ecology and Evolution* 13: 2119–28.
- Fox, G. A., S. Negrete-Yankelevich, and V. J. Sosa. 2015. *Ecological Statistics: Contemporary Theory and Application*. Oxford: Oxford University Press.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL: Taylor & Francis.
- Gelman, A., and J. Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Gelman, A., D. Lee, and J. Guo. 2015. "Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization." *Journal of Educational and Behavioral Statistics* 40: 530–43.
- Ghasemi, A., Z. Bahadoran, A. Zadeh-Vakili, S. A. Montazeri, and F. Hosseini-panah. 2019. "The Principles of Biomedical Scientific Writing: Materials and Methods." *International Journal of Endocrinology and Metabolism* 17: e88155. <https://doi.org/10.5812/ijem.88155>.
- Gotelli, N. J., and A. M. Ellison. 2004. *A Primer of Ecological Statistics*. Sunderland, MA: Sinauer Associates Sunderland.
- Hogg, R. V. 1991. "Statistical Education: Improvements Are Badly Needed." *The American Statistician* 45: 342–3.
- Hooten, M. B., and N. T. Hobbs. 2015. "A Guide to Bayesian Model Selection for Ecologists." *Ecological Monographs* 85: 3–28.
- Indrayan, A. 2020. "Reporting of Basic Statistical Methods in Biomedical Journals: Improved SAMPL Guidelines." *Indian Pediatrics* 57: 43–8.
- Johnson, J. B., and K. S. Omland. 2004. "Model Selection in Ecology and Evolution." *Trends in Ecology & Evolution* 19: 101–8.
- Lai, J., C. J. Lortie, R. A. Muenchen, J. Yang, and K. Ma. 2019. "Evaluating the Popularity of R in Ecology." *Ecosphere* 10: e02567.
- Lang, T. A., and D. G. Altman. 2015. "Basic Statistical Reporting for Articles Published in Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines." *International Journal of Nursing Studies* 52: 5–9.
- Makin, T. R., and J.-J. Orban de Xivry. 2019. "Ten Common Statistical Mistakes to Watch out for when Writing or Reviewing a Manuscript." *eLife* 8: e48175.
- McElreath, R. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL: CRC Press.
- McGarigal, K., S. A. Cushman, and S. Stafford. 2013. *Multivariate Statistics for Wildlife and Ecology Research*. New York: Springer Science & Business Media.
- Michel, M. C., T. J. Murphy, and H. J. Motulsky. 2020. "New Author Guidelines for Displaying Data and Reporting Data Analysis and Statistical Methods in Experimental Biology." *Drug Metabolism and Disposition* 48: 64–74.
- Miller, J. E. 2006. "How to Communicate Statistical Findings: An Expository Writing Approach." *Chance* 19: 43–9.
- Nakagawa, S., and I. C. Cuthill. 2007. "Effect Size, Confidence Interval and Statistical Significance: A Practical Guide for Biologists." *Biological Reviews* 82: 591–605.
- Nuijten, M. B., C. H. Hartgerink, M. A. Van Assen, S. Epskamp, and J. M. Wicherts. 2016. "The Prevalence of Statistical Reporting Errors in Psychology (1985–2013)." *Behavior Research Methods* 48: 1205–26.
- Onwuegbuzie, A. J., and V. A. Wilson. 2003. "Statistics Anxiety: Nature, Etiology, Antecedents, Effects, and Treatments—A Comprehensive Review of the Literature." *Teaching in Higher Education* 8: 195–209.
- Parker, T. H., S. C. Griffith, J. L. Bronstein, F. Fidler, S. Foster, H. Fraser, W. Forstmeier, J. Gurevitch, J. Koricheva, and R. Seppelt. 2018. "Empowering Peer Reviewers with a Checklist to Improve Transparency." *Nature Ecology & Evolution* 2: 929–35.
- Peng, R. D. 2011. "Reproducible Research in Computational Science." *Science* 334: 1226–7.
- Plummer, M. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, vol. 124, pp. 1–10. Vienna.
- Prayoga, T., and J. Abraham. 2017. "A Psychological Model Explaining Why We Love or Hate Statistics." *Kasetsart Journal of Social Sciences* 38: 1–8.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rexstad, E., and K. P. Burnham. 1991. *User's Guide for Interactive Program CAPTURE*. Fort Collins, CO: Colorado State University.
- Scheinerman, E. 2011. *Mathematical Notation: A Guide for Engineers and Scientists*. Scotts Valley, CA: Create Space.
- Schroter, S., N. Black, S. Evans, F. Godlee, L. Osorio, and R. Smith. 2008. "What Errors Do Peer Reviewers Detect, and Does Training Improve their Ability to Detect them?" *Journal of the Royal Society of Medicine* 101: 507–14.
- Shoemaker, L. G., J. A. Walter, L. A. Gherardi, M. H. DeSiervo, and N. I. Wisnoski. 2021. "Writing Mathematical Ecology: A Guide for Authors and Readers." *Ecosphere* 12: e03701.

- Slootmaeckers, K., B. Kerremans, and J. Adriaensen. 2014. "Too Afraid to Learn: Attitudes towards Statistics as a Barrier to Learning Statistics and to Acquiring Quantitative Skills." *Politics* 34: 191–200.
- Steel, E. A., M. C. Kennedy, P. G. Cunningham, and J. S. Stanovick. 2013. "Applied Statistics in Ecology: Common Pitfalls and Simple Solutions." *Ecosphere* 4: art115.
- Stute, W., W. G. Manteiga, and M. P. Quindimil. 1993. "Bootstrap Based Goodness-of-fit-tests." *Metrika* 40: 243–56.
- Touchon, J. C., and M. W. McCoy. 2016. "The Mismatch between Current Statistical Practice and Doctoral Training in Ecology." *Ecosphere* 7: e01394.
- Tredennick, A. T., G. Hooker, S. P. Ellner, and P. B. Adler. 2021. "A Practical Guide to Selecting Models for Exploration, Inference, and Prediction in Ecology." *Ecology* 102: e03336.
- Triaridis, S., and A. Kyrgidis. 2010. "Peer Review and Journal Impact Factor: The Two Pillars of Contemporary Medical Publishing." *Hippokratia* 14: 5.
- White, G. C., and K. P. Burnham. 1999. "Program MARK: Survival Estimation from Populations of Marked Animals." *Bird Study* 46: S120–39.
- Worthy, G. 2015. "Statistical Analysis and Reporting: Common Errors Found during Peer Review and how to Avoid them." *Swiss Medical Weekly* 145: w14076.
- Zuur, A. F., and E. N. Ieno. 2016. "A Protocol for Conducting and Presenting Results of Regression-Type Analyses." *Methods in Ecology and Evolution* 7: 636–45.
- Zuur, A. F., E. N. Ieno, and C. S. Elphick. 2010. "A Protocol for Data Exploration to Avoid Common Statistical Problems." *Methods in Ecology and Evolution* 1: 3–14.
- Zuur, A. F., E. N. Ieno, and G. M. Smith. 2007. *Analysing Ecological Data*. New York: Springer.

How to cite this article: Davis, Amy J., and Shannon Kay. 2023. "Writing Statistical Methods for Ecologists." *Ecosphere* 14(5): e4539. <https://doi.org/10.1002/ecs2.4539>