Citations and evidence

Last modified: 2026-11-07 21:34:25 (PST)

Every claim in scientific writing should be supported by either citations to relevant sources or direct evidence from data or experiments. This principle is fundamental to maintaining credibility, enabling verification, and building on the accumulated knowledge of the scientific community.

Why claims need support

Scientific writing aims to convey truthful, verifiable information. Unsupported claims undermine this goal in several ways:

- **Credibility**: Readers cannot assess the reliability of assertions without knowing their basis
- **Verification**: Claims without support cannot be independently verified or challenged
- **Reproducibility**: Other researchers need sources to understand the foundation of your work
- **Intellectual honesty**: Proper attribution gives credit where it is due and distinguishes your original contributions from existing knowledge
- **Learning pathway**: Citations provide readers a roadmap to learn more about a topic

When you make a claim, you are asking readers to accept it as true. They deserve to know why they should trust that claim.

What constitutes adequate support

Different types of claims require different types of support:

Empirical claims

Claims about observable phenomena or data should be supported by:

- Direct presentation of the data (tables, figures, statistical analyses)
- Citations to published studies that collected the relevant data
- Description of experimental methods that generated the evidence
- Links to publicly accessible datasets

Theoretical claims

Claims about concepts, models, or interpretations should be supported by:

- Citations to papers that developed or validated the theory
- Logical arguments with clearly stated premises
- Mathematical derivations or proofs
- References to review articles that synthesize relevant theoretical work

Methodological claims

Claims about appropriate methods or best practices should be supported by:

- Citations to methodological papers or textbooks
- Empirical evidence of method performance
- Expert consensus statements
- Validation studies

Common knowledge

Not every statement requires citation. Well-established facts that are common knowledge within your field (for example, "DNA is a double helix" in molecular biology) can be stated without citation. However, when in doubt, provide a citation—over-citing is preferable to under-citing.

Do not substitute flippancy for support

Some phrases assert that a claim needs no support instead of supplying any:

- “it is obvious that”
- “clearly, it follows that”
- “it should be familiar to you that”
- “the familiar formula”
- “of course”
- “as everyone knows”

The problem is the appeal, not the individual words. “Clearly” is fine as an adverb of manner (“the methods section clearly describes the sampling frame”), and “of course” is fine when it flags an assumption you have already established. Each becomes a problem only when it stands in for an argument you have not made.

These phrases do no work for the reader. A reader who already knows the result gains nothing from being told that it is familiar, and a reader who does not is told that the gap is their own fault, without being given any way to close it. They also cost the writer something. Working a step out is how you find out whether it holds; calling it obvious settles the question by assertion, and leaves any error in the step where it is. So if you cannot supply a derivation or a citation for a step, treat that as a reason to work the step out, not as a reason to call it obvious.

Replace such phrases with one of the following:

- **A direct derivation.** Show the intermediate steps, or state which identity or algebraic manipulation takes you from one line to the next.
- **A specific citation.** Name the source that establishes the result, and give a chapter, section, page, or equation number whenever the source is long enough that a bare citation would leave the reader searching.
- **A cross-reference.** When you established the result earlier in the same document, point to it by number (see Defining terms clearly for the Quarto cross-reference syntax).

Deleting the phrase and keeping the claim is also usually an improvement: “the log-likelihood is a sum of three terms” makes the same claim about the log-likelihood as “the log-likelihood is obviously a sum of three terms”, and it invites the reader to check the claim rather than to defer to it.

Example 0.1 (Replacing an appeal to familiarity)**.**

> ✗ Substituting the normal density into the likelihood gives the familiar three-term formula.
>
> ✓ For an independent sample $x_1, \dots, x_n$ from a normal distribution with mean μ and variance σ^2, substituting the normal density into the likelihood and

> taking logarithms gives
>
> $$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2.$$
>
> The three terms are, in order:
>
> - a constant that does not involve the parameters;
> - a term that depends on σ^2 but not on μ;
> - the sum of squared deviations of the data from μ, divided by $-2\sigma^2$.

The second version gives the reader the formula and where it came from, so they can check the algebra themselves. The first version asks them to take the writer's word for it, and offers no help to anyone who has not seen the result before.

"Trivial" and "trivially"

"Trivial" and "trivially" usually work the same way: they announce that a step is too small to write down, which relieves the writer of writing it down. Treat them as a prompt to check the step, and to write it out if it takes only a line.

The word has one defensible use, which is describing how a general expression reduces to a simpler one in a special case. For example, a weighted average reduces to an ordinary average when all of the weights are equal. Even in that use, "minimal case" or "degenerate case" names the situation more precisely, and avoids the dismissive tone that "trivial" carries.

What makes a citation relevant

A relevant citation is one that actually supports the specific claim you are making. Common problems with citation relevance include:

- **Citing review papers indiscriminately**: While review papers are valuable for general background, cite the original research when making specific claims about particular findings
- **Citing papers that don't address the claim**: Ensure the cited work actually discusses the point you're making
- **Over-generalizing from narrow studies**: Don't cite a study with limited scope to support a broad general claim
- **Using outdated sources**: When current knowledge has superseded older findings, cite more recent work

To ensure relevance:

- Read the papers you cite (at least the relevant sections)
- Verify that the cited work actually supports your specific claim
- Cite the most direct source available
- When citing for general background versus specific claims, make the distinction clear

What makes a source trustworthy

Not all sources are equally reliable. Consider these factors when evaluating trustworthiness:

Peer review

Peer-reviewed publications in reputable journals have undergone expert scrutiny. This doesn't guarantee correctness, but it provides a baseline level of quality control.

Reputation of authors and institutions

Work from recognized experts and well-regarded institutions tends to be more reliable, though this should not be the sole criterion.

Replication and consensus

Findings that have been replicated by independent groups or that represent scientific consensus are more trustworthy than isolated claims.

Transparency and reproducibility

Studies that:

- Clearly describe their methods
- Share their data and code
- Disclose potential conflicts of interest
- Have been successfully reproduced by others

are more trustworthy than those lacking these features.

Preprints and non-peer-reviewed sources

Preprints can be valuable for accessing cutting-edge research, but they have not undergone peer review. When citing preprints:

- Note that they are preprints
- Check whether a peer-reviewed version has since been published
- Exercise extra scrutiny of the methods and conclusions

Sources to generally avoid

Some sources typically lack the rigor needed for scientific writing:

- Wikipedia and general encyclopedias as the only source for a claim (cite them beside the primary source, as Cite Wikipedia, beside the primary source explains)
- Popular press articles about scientific findings (cite the original research instead)
- Blogs and social media posts (unless discussing public discourse or documenting specific claims made in those venues)
- Predatory or pay-to-publish journals without genuine peer review
- Retracted papers

Copy-paste with caution

Copying exact wording — whether from another writer or from your own earlier work, and even with a citation — can cause problems: plagiarism, if the wording is not clearly marked as a quotation, and a shift in voice that makes the text harder to read.

A few rules of thumb:

- **Do not copy-paste prose from anyone who is not a coauthor** of the document you are currently writing. If you must reproduce someone else's exact words, mark them as a direct quotation and cite the source.
- **Prefer paraphrasing over quoting.** Restating an idea in your own words, with a citation, usually reads better than a block quotation, and it shows that you understand the idea.
- **Do not recycle your own published text without thought.** Reusing passages from your earlier papers (self-plagiarism) can violate journal policies; cite the earlier work instead of repeating it.

A related case is repetition *within* your own document. When a passage you are tempted to copy already appears elsewhere, cross-reference it instead of repeating it: this keeps a single source of truth and improves concision.

Link liberally

A link lets a reader check a claim or learn more without searching for the source themselves. When in doubt, add the link; leave one out only for a specific reason. Link these things whenever they appear:

- a term to the place where it is defined, in the same document or another one;
- a result to the theorem, equation or section that establishes it;
- a definition, theorem or named method to its Wikipedia article, when one exists (see Cite Wikipedia, beside the primary source);
- a cited source to its DOI or official URL;
- software, packages and datasets to their documentation or home page;
- a related section elsewhere in the document, instead of repeating its content;
- a back-reference, such as "as defined above" or "the model from the previous section", to the content that it points back to.

A reader who skipped the earlier section, or who opened the page from a link to its middle, can jump to the content that a back-reference points to only if the back-reference is a link. Put the link on the words that name the target, once per paragraph. Where you can, use the target's name instead of "above" or "earlier", since a word that names a position becomes wrong when a later edit moves the content (see Avoid forward references).

Example 0.1 (Linking a reference back to earlier content)**.**

> ✗ The likelihood of the logistic regression model from the previous chapter is a product of Bernoulli probabilities.
>
> ✓ The likelihood of the logistic regression model is a product of Bernoulli probabilities.

In the ✓ version, the link takes the reader to the model's definition.

In Quarto, use a cross-reference such as `@def-confidence-interval` for a definition or result in the same document, so the link updates when numbering changes (see Defining terms clearly). For a target on another page, use an ordinary link to that page's anchor.

Leave a link out only for a reason such as one of the following:

- the target is private, such as a repository of graded homework;
- the same target is already linked earlier in the same paragraph;
- the target is unstable, and a stable alternative such as a DOI exists.

Example 0.2 (Adding links a reader would use)**.**

✗ We fit the model with the lme4 package and report a 95% confidence interval for each fixed effect.

✓ We fit the model with the lme4 package and report a 95% confidence interval for each fixed effect.

A reader who does not know lme4, or wants the exact meaning of "confidence interval" in this guide, can now follow the link instead of searching.

Cite Wikipedia, beside the primary source

Wikipedia is free to every reader, and its articles on definitions, theorems and methods are usually reliable enough to orient a newcomer. So cite the Wikipedia article for each definition, theorem or named method whenever one exists: the citation gives readers a second, free presentation without leaving them to search for it.

The Wikipedia citation adds to the traditional reference and never replaces it. Cite it through the bibliography like any other source, with a `@misc` entry in `references.bib` (author `{Wikipedia contributors}`, the article's URL, the access date) keyed `wikipedia_<article_title>`, and put the citation right after the content it covers, as `[@wikipedia_poisson_distribution]`. When the content is the name of a concept you define, link those words to the article instead, as described in Link the words, not a trailing citation. Keep the primary source's citation beside the claim it supports. Never list a Wikipedia article in a Further reading section. Wikipedia appears only as a citation beside the content it covers. A Further reading section lists primary sources, textbooks, courses and other resources instead.

- Cite the article for the specific concept, not a broader article that mentions it.
- Read the article's definition before citing it, because notation and conventions vary, and a reader who follows the link should find the same concept you defined. If the article's convention differs from yours, say how.
- Use a stable URL for the article, not a search result.
- Cite an article once where its term is defined, not at every use; later uses link to your own definition.

Example 0.1 (Citing a definition's Wikipedia article)**.**

✗ A *Poisson random variable* counts events in a fixed interval.

✓ A *Poisson random variable* counts events in a fixed interval (Wikipedia contributors 2026).

A reader who wants another presentation of the Poisson distribution now has one, and the definition itself still comes from your own text and its traditional reference.

Link the words, not a trailing citation

When a source backs the name of a defined concept or a named result, and the source has a stable web page, hyperlink the words themselves to that page. Do not write the words and then add a numbered citation or a parenthetical link after them. A trailing number such as [3] makes the reader leave the sentence to find out what it points to, and it separates the link from the words it supports. A parenthetical such as `(see the Wikipedia article)` has the same fault. A dataset reference follows the same rule (Ezra, 2026-11-04): link the dataset's name to the dataset's description, as in `Use the [`Smarket` data](#dat-smarket).`, not `Use the `Smarket` data (@dat-smarket).`, whose cross-reference renders as a parenthetical "(Dataset 1)" after the name.

Link the longest phrase that names the thing, in the sentence that uses it. Inside a bolded name, put the link inside the bold, as `**[set](url)**`. Keep the numbered `[@key]` citation for a claim that no phrase names, for a source without a stable page, such as a book, and for a locator such as `p. 5`. Where a link replaces the citation, the link is the credit, so the page needs no bibliography entry.

Example 0.1 (Linking the defined term to its Wikipedia article)**.**

✗ `A graph [@wikipedia_graph_theory] is a set of vertices joined by edges.`

✓ A graph is a set of vertices joined by edges.

The second version gives the reader the article from the word they were already reading.

Number bracketed citations

On the lab's notes sites, a bracketed citation, `[@key]`, prints as a number, such as [1] or [1, p. 5]. An in-text citation, `@key`, prints the author name and the number, such as "Hastie et al. [2]". Write `[@key]` when the source supports a claim, and `@key` when the author is the subject of the sentence. Never start a sentence or a list item with a bare `@key` unless the author is its subject. The author list becomes the subject, so a list of such items prints the same author name, such as "Wikipedia contributors [3]", again and again. Write the description first, and end the sentence with the bracketed `[@key]`. The sites get this from the shared citation style and filter in Morrison-Lab/macros.

Place a citation right after the claim it supports

Place citations immediately after the claim or claims they support. A citation at the end of a sentence reads as support for the whole sentence, so when it supports only one claim in it, put it beside that claim. When one source supports every claim in a sentence, the end of the sentence is right after the claims.

Example 0.1 (Placing a citation after a definition)**.**

> ✗ A *Poisson random variable* (Wikipedia contributors 2026), or *Poisson count*, counts events in a fixed interval.
>
> ✓ A *Poisson random variable*, or *Poisson count*, counts events in a fixed interval (Wikipedia contributors 2026).

The article presents the definition, and the definition is the claim, so the citation follows the definition rather than interrupting it after the term.

Cite a source once per passage

Cite a source once in each div or passage, at the claim it supports. A second citation of the same source in the same div adds a number that the reader must look up again and tells them nothing new. Cite again only when the second claim rests on a different part of the source, and then give a locator, such as `[@key, p. 5]`. The rule to cite a Wikipedia article once, where its term is defined, is a case of this one.

Example 0.1 (Citing one source once in a definition)**.**

> ✗ A *Poisson random variable* counts events in a fixed interval (Wikipedia contributors 2026). Source: Wikipedia contributors (2026).
>
> ✓ A *Poisson random variable* counts events in a fixed interval. Source: Wikipedia contributors (2026).

Both versions credit the same article. The second one does it once, in the source line that closes the div.

Best practices

To effectively support your claims:

1. **Cite as you write**: Add citations immediately when making claims, rather than planning to "add references later"
2. **Use citation management tools**: Software like Zotero, Mendeley, or BibTeX helps organize and format references correctly
3. **Check your citations**: Before submitting, verify that every citation supports its associated claim
4. **Provide context**: Help readers understand why a source is relevant (for example, "Smith et al. (2020) demonstrated that…" rather than just "(Smith et al., 2020)")
5. **Balance primary and review sources**: Use primary sources for specific findings, reviews for general background
6. **Stay current**: Supplement foundational older references with recent work showing the current state of knowledge
7. **Cite diverse sources**: When possible, include work from different research groups and perspectives

Quote the original

Quote and cite the source that said it, not a later source that quotes it. Check the wording and the page against the original.

> ✗ Mitchell (1997), as quoted in Goodfellow et al. (2016), defines …
>
> ✓ "A computer program is said to learn …" (Mitchell 1997)

Credit a quotation to whoever speaks it in the primary source, even when the speaker is quoting someone else. When attribution is in doubt, read the primary source (a transcript, the original page) first. Uncertainty about who said a line is not evidence about who said it.

Copy organization names from their source

When a bibliography entry names an organization, copy the name from the organization's own page (its about page, README or footer) rather than expanding an acronym from memory.

Verify each bibliography entry

Check each detail of a bibliography entry against an official source. Look the work up through Crossref search or its REST API, check the title, volume, issue and pages against the record, and add the DOI. Check a constant in a cited theorem, such as the factors in an inequality, against a second free, official source, and say in the pull request what you checked against what.

Cite an alias directly

When a source uses another name for a defined term, write the alias in bold and cite the source after it, as in `also called the **alias** [@key, section]`. The citation already says that the source uses the name, so do not add a sentence such as "The source uses that name".

✗ `also called the **alias**. The cited textbook uses that name.`

✓ `also called the **alias** [@key, section]`

Cite the original for named results

When you present a result that carries a person's name, or whose first appearance is well documented, cite the publication where it first appeared alongside a modern textbook that teaches it. The original shows readers where the result comes from and credits its authors. The textbook is where a student will actually learn it.

When the credit is only a citation, end the definition or theorem div with a line `Source: @key [locator]`. A reference adds no commentary to the statement, and a separate remark that holds only a citation is clutter. When the credit needs substantial prose about the source or history of the content, put that prose in a short remark after the statement and its proof, not inside the definition or theorem itself, so that each theorem div still holds one result.

Apply the rule to:

- named theorems, inequalities, and algorithms;
- definitions that a specific paper introduced;
- results whose usual name is not the whole story, such as Chebyshev's inequality, which Bienaymé derived earlier.

Do not apply it to:

- routine lemmas;
- textbook exercises;
- results with no identifiable first source.

When the credit is disputed or the first source is unclear, say so in one sentence and cite a history of the result instead of naming a source you have not checked.

Check every detail of the original against an official source, such as:

- the publisher;
- the journal's archive;
- a library such as Numdam.

Never cite an unauthorized copy, and never fill in a title, volume, or page number from memory.

Example 0.1 (Crediting Chebyshev's inequality)**.**

✗ Chebyshev's inequality bounds the probability that …

✓ Chebyshev's inequality bounds the probability that … Bienaymé derived the bound (Bienaymé 1853), and Tchébychef proved a more general form (Tchébychef 1867). A modern treatment is in Blitzstein and Hwang (2019).

Put the content first

Headings and sentences should name the topic, not its author: "A formal definition of machine learning", not "Mitchell's definition". Follow a quotation with its citation alone, without commentary on who said it or where it was repeated.

Adapting another course's material

When teaching material draws on another course, write your own version of the content rather than summarizing how that course presents it. Credit the source in a note at the start or end of the adapted section.

Put that note in a div, never in running prose. An attribution is not part of the argument, and a reader skimming the section should be able to tell it apart from the content it credits. In Quarto, use a collapsed `.callout-note` titled "Source", which frames the credit and keeps it out of the way on the page. A `::: notes` div keeps the credit out of the slides, but on an HTML page it renders as a plain unstyled `<div>`, which looks like running prose; use it only on a site whose stylesheet sets `.notes` apart.

The same applies to license notices and to statements of how the adaptation differs from the source.

✗ A plain paragraph between the heading and the first definition:

```
This section is adapted from @zhou2024, used under the MIT License.
```

✓ The same credit in a collapsed callout:

```
::: {.callout-note collapse="true"}
#### Source

This section is adapted from @zhou2024, used under the MIT License.
:::
```

Material from your own work

Do not formally acknowledge material you borrow from yourself or your own group. "Your group" means you and the other members of the labs that maintain the work, such as the Morrison Lab and UCD-SERG. Link to it, but do not credit it, except for the credits and citations listed below. This covers:

- Material your group wrote that moves between repositories you maintain. When lecture notes move from one course repository to another, leave the sister-repository credit out of the new work. Material in those repositories that someone outside the group wrote keeps its credit (see below).
- Text, figures or code reused from work that you or your group wrote elsewhere, such as a lab manual your group maintains (the UCD-SERG lab manual, for example), your own papers, or your own software.
- A template or tool of your group's that you built the site from.

A formal acknowledgment tells the reader the material came from someone else, so it does not belong on material that is your own, wherever that material is stored. A co-author outside your group is someone else: when you reuse a co-written paper or manual, the work itself gets no credit (a publisher-required citation aside, see below), but credit each outside co-author by name, in a sentence or an author line. The rule concerns acknowledging reuse of your own material. Citing a published paper or released software of yours as the original publication of a result, or as a tool you used, is ordinary citation and stays. A passage or figure reused from a published paper of yours follows the publisher's reuse terms, which may require a citation; that is the only citation the reuse itself can require.

Do not cite your own course notes or sites as the source of a definition or result either. Instead, cite the original publication where you can identify one, and a modern textbook (see Cite the original for named results), and link to your own page as further reading.

✗ A citation of the group's own notes as the source of a definition:

```
The Cauchy-Schwarz inequality bounds the inner product by the product of the norms [@morriso
```

✓ The original source and a textbook, with a cross-reference link to the group's page:

```
The Cauchy-Schwarz inequality bounds the inner product by the product of the norms [@original
See also the
[Cauchy-Schwarz inequality](https://morrison-lab.github.io/mds/linear-algebra-inner-products
in the sister notes.
```

Carry over what the source of your material itself credits:

- A book, paper or other work the material follows. For example, the Benjamin-Chung Lab Manual, which the UCD-SERG manual credits.
- A person outside your group whose lecture or notes that source adapted, credited by name. For example, the lds course credits Kameron Decker Harris, Brian Hutchinson and Logan Sizemore, so any repository that reuses those lectures credits them too.

A link to a sister site that sends the reader to further reading is a cross-reference, not an attribution, so it stays.

When you find an existing credit to your own work, remove it and add or keep a link to it.

Common citation errors to avoid

- **Citation needed**: Making claims without any supporting citation or evidence
- **Vague attribution**: Using phrases like "studies have shown" without citing specific studies
- **Appeal to obviousness**: Using "obviously", "clearly", or "as everyone knows" in place of a derivation or a citation
- **Circular citation**: Citing a paper that doesn't contain the claimed information but cites another paper that does (cite the original source)
- **Redundant citation**: Citing the same source more than once in one div or passage (see Cite a source once per passage)
- **Citation padding**: Adding citations that don't actually support your claims just to appear well-referenced
- **Selective citation**: Only citing work that supports your position while ignoring contradictory evidence
- **Ghost authorship**: Failing to cite work that directly influenced your ideas

Examples

Poor (unsupported claim)

> Machine learning models often perform poorly on small datasets.

Problem: This claim is stated as fact without any support.

Better (citation provided)

> Machine learning models often perform poorly on small datasets (Vapnik 1998; Hawkins 2004).

Improvement: Citations provide evidence for the claim.

Best (citation with context)

> Machine learning models often perform poorly on small datasets. Vapnik (1998) showed that the generalization error of learning algorithms typically decreases as training set size increases, and Hawkins (2004) demonstrated that complex models are particularly prone to overfitting when trained on limited data (Vapnik 1998; Hawkins 2004).

Improvement: The specific support each citation provides is explained.

Poor (irrelevant citation)

> Python is the most popular programming language for data science (Knuth 1984).

Problem: Knuth's 1984 paper on literate programming doesn't address Python or data science.

Better (relevant citation)

> Python is the most popular programming language for data science (Stack Overflow 2024).

Improvement: The citation is to a current survey of programming language usage.

Conclusion

Supporting claims with appropriate citations and evidence is not optional—it is essential to scientific communication. It allows readers to verify your claims, understand the foundation of your arguments, and locate resources for further learning. Always ask yourself: "How does my reader know this is true?" If your text does not answer that question, add a citation or present direct evidence.

Bienaymé, Irénée-Jules. 1853. "Considérations à l'appui de La Découverte de Laplace Sur La Loi de Probabilité Dans La Méthode Des Moindres Carrés." *Comptes Rendus de l'Académie Des Sciences* 37: 309–24. https://www.numdam.org/item/JMPA_1867_2_12___158_0/.

Blitzstein, Joseph K., and Jessica Hwang. 2019. *Introduction to Probability*. 2nd ed. Chapman; Hall/CRC. https://doi.org/10.1201/9780429428357.

Hawkins, Douglas M. 2004. "The Problem of Overfitting." *Journal of Chemical Information and Computer Sciences* 44 (1): 1–12. https://doi.org/10.1021/ci0342472.

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* (USA) 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.

Mitchell, Tom M. 1997. *Machine Learning*. McGraw-Hill.

Stack Overflow. 2024. *Stack Overflow Developer Survey 2024*. https://survey.stackoverflow.co/2024/.

Tchébychef, P. L. 1867. "Des Valeurs Moyennes." *Journal de Mathématiques Pures Et Appliquées*, 2nd series, vol. 12: 177–84. https://www.numdam.org/item/JMPA_1867_2_12___177_0/.

Vapnik, Vladimir N. 1998. *Statistical Learning Theory*. Wiley. https://www.wiley.com/en-us/Statistical+Learning+Theory-p-9780471030034.

Wikipedia contributors. 2026. *Poisson Distribution — Wikipedia, the Free Encyclopedia*. https://en.wikipedia.org/wiki/Poisson_distribution.