

The NCBI Style Guide

Last Updated: 2004



National Center for Biotechnology Information (US)
Bethesda (MD)

National Center for Biotechnology Information (US), Bethesda (MD)

NLM Citation: The NCBI Style Guide [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2004-.

Many questions about grammar and consistency arise throughout the course of a day's work, and having one place to go for help with these questions speeds the completion of the task at hand.

People approach their tasks from two viewpoints—one of producing an NCBI website or Bookshelf content, and one of producing content to be published outside of NCBI. “The NCBI Style Guide” is geared toward helping people produce NCBI websites and Bookshelf content by providing specific examples of uses typical to NCBI. There are portions of “The NCBI Style Guide” that provide advice on technical writing in general, as well as a list of other suggested style guides that may be helpful to people producing content to be published elsewhere. Bear in mind that for content published elsewhere, the suggestions for specific styles may not apply.

“The NCBI Style Guide” is basically divided into two main parts, General Writing Skills and NCBI-specific Styles, to provide some assistance to people in both groups.

Table of Contents

Preface	vii
Part I General Writing Skills	1
Chapter 1 How to Begin	3
Chapter 2 Standards and Guides to Style	5
Chapter 3 Grammar Reminders and a Little Technical Writing	7
Anthropomorphism	7
Common Errors	7
Gender Neutral	9
Hedging	9
Point of View: First Person, Second Person, Third Person	9
Scientific English: Past, Present, and Simple	10
Subject-Verb Agreement	11
Voice: Active <i>versus</i> Passive	11
Wordy Constructions to Avoid, Substitutions of Common Phrases, Plainier Synonyms	12
Part II NCBI-specific Styles	21
Chapter 4 Writing for the Web Environment	23
Going Public	23
Constructing Searchable Web Pages	23
Keep Your Web Pages Current	24
Glossaries on the Web	25
Itemized Lists	25
Outlines in Web Help Documents	27
Web Pages without Reference Lists: Making Citations	27
NCBI Standard Abbreviations and Frequently Used Word Forms	28
Chapter 5 Style Points and Conventions	35
Amino Acids	35
Capitalization	35
Centrifugation	35
Colons	35
Commas	36

Concentrations	36
Dashes	36
Decimal Points	37
Degrees	37
e.g.	37
et al.	37
etc	37
Emphasis	37
Equations	38
Foreign Words	38
Genes and Protein Designations	38
Hyphens	38
-ical	39
i.e.	39
in vitro, in vivo	39
Italics	39
Latin	40
Lists	40
Math	40
Numbers/Numerals	40
Probability	41
Proofreaders' Marks	41
Punctuation	41
Quotation Marks	41
Restriction Enzymes	42
Semicolons	42
Spelling	42
Statistics	43
Symbols	45
Time	46
Trademarks, Trade/Supplier Names	46
Units of Measure	46
versus	46

X-ray	46
Chapter 6 Special Content	47
Appendices	47
Boxes	47
Figures	47
References	48
Tables	49

Preface

The NCBI Style Guide will help you communicate your thoughts effectively by providing advice on guides to style, examples of common word forms at NCBI, and helpful grammar and language tips.

Part I. General Writing Skills

Chapter 1. How to Begin

Before beginning any writing project, review the [guidelines for plain language](#). These guidelines will provide the foundation for your writing project.

Begin your writing project with a short list of topics to cover; more ideas will come to you once you've taken this small step.

The brief pointers in the bulleted list will help you now and later.

- Prepare your materials for presentation to an international audience.
- Use common words, where possible, but be precise.
- Use a combination of active and passive voice.
- Present text, tables, graphics, and other materials clearly, concisely, and consistently.
- For the Web environment in particular, divide your thoughts into manageable sections.

Chapter 2. Standards and Guides to Style

It is not possible to remember all of the standards set by official bodies, but yet you must follow the rules and be consistent. You can turn to these guides for help.

We use the following resources as standards and guides to style.

For overall guidance, use:

- Council of Biology Editors. *Scientific style and format*. 6th ed. Cambridge (UK): Cambridge University Press; 1994.

This book is good because it covers many scientific disciplines in detail and references the decisions of important international committees. The 7th edition will be available June 2006. You can [preview some of the changes](#).)

For questions of reference style, use:

- Patrias K. [National Library of Medicine recommended formats for bibliographic citation](#). Bethesda (MD): U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 1991.
- Patrias K. [National Library of Medicine recommended formats for bibliographic citation. Supplement: internet formats](#). Bethesda (MD): U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 2001.

This pair of documents is probably the most complete recording of reference requirements you may ever see. PDFs for these documents are [available](#). By the end of 2006, an updated, combined electronic version will be available on the Bookshelf as “Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers”.

For a quick overview of reference styles in “The NCBI Style Guide”, see Special Content, References.

For questions of chemical style, use:

- International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology. IUPAC-IUBMB Joint Commission on Biochemical Nomenclature and Nomenclature Commission of IUBMB. Available from: [IUPAC](#).

This Union is the nomenclature authority.

Chapter 3. Grammar Reminders and a Little Technical Writing

This chapter contains basic information, such as “Subject-Verb Agreement”, as well as more unusual aspects of writing, such as “Anthropomorphism”.

Anthropomorphism

Anthropomorphism is “an interpretation of what is not human or personal in terms of human or personal characteristics” (Merriam-Webster's Collegiate Dictionary, Tenth Edition, Springfield (MA): Merriam-Webster, Inc.; 2002). You can read more about it in the [Merriam-Webster Online Dictionary](#).

Anthropomorphism seems to occur more frequently in texts about software.

Example:

A function to determine all features in a particular region of a Bioseq **need not care** what types of features **they** are.

Rewrite as:

A function to determine all features in a particular region of a Bioseq **is not affected by** what types of features **there** are.

Although text that has been anthropomorphized may be entertaining, an international audience is better served by presenting text in a straightforward manner.

Common Errors

Here are the top seven errors, with specific examples of correct usage underlined.

1. “as/because”

Do not use “as” when “because” is meant.

- a. Supplied as: The polyA tail of an mRNA (cDNA) sequence should never be trimmed as it provides a useful landmark.
- b. Should be: The polyA tail of an mRNA (cDNA) sequence should never be trimmed because it provides a useful landmark.

2. “comprise” and “compose”

Remember that “the whole comprises the parts”. The following examples use the words correctly.

- a. The *FMRI* gene comprises 18 exons interspersed over ~40,000 base pairs of sequence.
- b. The Alu repeat family comprises short, interspersed elements present in multiple copies in the genomes of humans and other primates.
- c. The second model comprises three closely spaced exons.
- d. OMIM comprises descriptive, full-text MIM entries, a tabular synopsis of the Human Gene Map, clinical synopses, and mini-MIMs.
- e. Segmented Bioseqs are *composed of* pointers to other raw sequences in GenBank.

Never use “comprised of”. Use “composed of” instead.

3. “due to”

Do not use “due to” as a substitute for “because of” or “attributable to”. These sentences are correct.

- a. ...or whether the similarity observed is attributable to chance alone.
- b. Much of the success of the project is attributable to the flood of new molecular data.
- c. The term refers to similarity attributable to descent from a common ancestor.
- d. Many of them exist in the coding region of a gene, but the precise location cannot be resolved because of an error in the alignment of the exon.
- e. If a gene has multiple transcripts because of alternative splicing, then a variation can have several different functional relationships to the gene.
- f. GenBank annotation staff would respond to a request for a list of Accession numbers that are due to appear in upcoming issues of a publisher’s journal(s).

4. “from... to ...” and “between... and ...”

The “from... to ...” construction **contains** the end points, but the “between... and ...” construction **excludes** the end points.

- a. No space is allowed between the value and the field name.
- b. FMR1 resides somewhere between markers DXS532 and DXS7389.
- c. The bars represent matches from residues 3 to 60.

5. “it’s” versus “its”

“It’s” is a contraction of “it is”. If you are having trouble using “it’s”, just substitute “it is” in your sentence and see if it makes sense. If so, then you are using “it’s” correctly. “Its”, on the other hand, shows possession or refers to itself. Most errors occur because of the incorrect use of the contraction—“it’s”. The confusion probably arises from the way in which possessives are formed in English, for example, Belinda’s cat, Jeff’s boat, but its mechanism, etc.

- a. The sequence of each exon is an individual raw Bioseq in its own right.
- b. It’s the lock that is lost, not its combination.

6. “since/because”

Use “since” when referring to time.

- a. During each build, we cluster the data submitted since the last build into the existing refSNPs and form new refSNPs when necessary.
- b. The set of new data entering each build typically includes all submissions received since the close of data in the previous build.
- c. Data are presented by quarter since operations began on July 25, 2000.
- d. Because the execution of the search algorithm is decoupled from the formatting, the results can be delivered in a variety of formats without rerunning the search.
- e. XML is also structured output but can be produced from ASN.1 because it has equivalent information.

7. “which/that”

The word “which” usually introduces information that is “extra” (or interesting but not the point being made at the time), meaning that the sentence is complete and imparts what is necessary without the information that is introduced by “which”. The word “that”, on the other hand, introduces information that is necessary to understanding the meaning/intent of the sentence.

- a. The resulting file, which PHRAP outputs in “.ace” format, consists of the sequence itself plus the associated quality scores.

- b. The alignment information is extracted into a Seq-align, which is packaged as annotation (Seq-annot) associated with the BioSeqSet.
- c. An Entrez “node” is a collection of data that is grouped together and indexed together.
- d. Entrez is designed to infer relationships between different data that may suggest future experiments or assist in interpretation of the available information.

Gender Neutral

Use gender-neutral language.

For example, change:

mankind → humanity

best man for the job → best person for the job

the common man → average person, ordinary people

mailman → mail carrier

congressman → congressional representative

policeman → police officer

Other ways to handle this problem are to recast a sentence in the plural, reword to eliminate gender problems, or replace the masculine pronoun with “one” or “you”.

Hedging

Hedging takes place when an author wants to “evade the risk of commitment especially by leaving open a way of retreat” (Merriam-Webster's Collegiate Dictionary. Tenth Edition. Springfield (MA): Merriam-Webster, Inc. 2002). Usually, the sentence will read something like, “This assay may provide a better way of identifying mutant genes”. It is reasonable to temper one's confidence about a new method.

Sometimes authors do not realize they are using a kind of hedging when they unintentionally express a lack of confidence by using certain phrases, such as “intends to”, “aims to”, and “has the goal of”.

Examine this sentence: “GenBank aims to provide a searchable database of sequences.”

Contrast that sentence with this one: “GenBank is a searchable sequence database”.

This is a more confident statement, and the reader is not distracted from the main thought.

Point of View: First Person, Second Person, Third Person

The point of view describes whether a person is:

- **describing events as a participant**, such as “I raced to catch the train before it could enter the tunnel” (told in the first person because it uses “I”)
- **giving instructions to someone directly**, such as “You must catch the train before it reaches the tunnel” (“you” is used for second person)
- **relating events about others**, such as “He failed to catch the train before it reached the tunnel” (third person because of “he”)

The most common error occurs when different points of view are mixed; usually, the second-person “you” is mixed with a third person point of view:

Example:

“The data are entered as field names. One must be careful to end each field name with a delimiter code. You don't have to worry about whether the delimiter code is in uppercase or lowercase. One must then exit this subroutine to run the next portion of the program.”

In the example above, the second-person “you” has been used in a paragraph of third-person sentences. To keep the same point of view, one could have said, “One does not have to worry about whether the delimiter code is in uppercase or lowercase.”

Sometimes “one” is considered a little formal. If a formal approach is not needed, consider this approach, which uses an implied “you”:

“Enter the data as field names, being careful to end each field name with a delimiter code. Either uppercase or lowercase may be used in delimiter codes because they are case insensitive. After entering all of the data, exit the subroutine and continue with the next programmed task.”

First person uses: **I, we**

Second person uses: **you**

Third person uses: **he, she, it, one, they**

Check your text for consistency in the use of point of view. Check the use of “you” first.

Scientific English: Past, Present, and Simple

One of the finer points of science writing in English is the expression of the degree of acceptance of information by the use of present tense and past tense. Established knowledge (statements of fact, natural laws, universal (established) truths, or a widely held tenet of science) is given in the **simple present tense**, and the new results and findings of an author's own research are given in the **simple past tense**. Once an author's findings are published **and** become an established truth (generally accepted by experts in the field), they are referred to in the present tense.

These ideas are complex but are illustrated below:

1. Simple present tense:
 - PubMed Central **is** the National Library of Medicine's digital archive of full-text journal literature.
 - Entrez **is** the text-based search and retrieval system used at NCBI for all of the major databases.
2. Simple past tense (isolates the event in the past):
 - The cells **were** produced by continuous proliferation of the hybridoma cell line.
 - The lytic ability of the T8 cell subset **was** examined.

Some authors, especially those whose native language is not English, are not as aware that the use of past and present tenses signals the status of the presented information. Because the use of the simple present tense signals that something is a general truth, the reader is given some feeling of certainty about the information. A misinterpretation of the significance of the findings could occur if the wrong tense is used. The use of the simple present tense when first presenting **one's own findings** could be perceived as arrogant as well.

Even the demonstration of significant results in one's own study is **not a reason** for using the present tense in the Discussion section of the paper first reporting those results. However, explanations of graphical presentations of data, such as tables and figures, should use the present tense. For example, when referring to Table 3, it is fine to say: “Table 3 **shows** that the mean yield is lower than expected.”

Note to Authors: To avoid misunderstanding, report **your findings** in the past tense.

A fascinating, detailed account of this convention is provided in:

- Burrough-Boenisch J. Examining present tense conventions in scientific writing in the light of reader reactions to three Dutch-authored discussions. *English for Specific Purposes* 2003;22(1):5-24.

Additional information about the use of tenses in science is provided in:

- Matthews JR, Bowen JM, Matthews RW. *Successful scientific writing: a step-by-step guide for biomedical scientists*. 2nd edition. Cambridge: Cambridge University Press; 2001.

Acknowledgment: My sincere thanks go to Joy Burrough-Boenisch for reviewing this section on “Scientific English: Past, Present, and Simple”. Any errors that remain are entirely my own.---Belinda Beck

Subject-Verb Agreement

Subject and verb agreement means that if the subject is singular, the verb must be singular. If the subject is plural, the verb must be plural. This sounds very simple, but people run into trouble because of intervening phrases (phrases that come between the subject and the verb). For example, this type of sentence gives many people trouble:

- The number of birds was impressive.

The subject of the sentence is “number” (singular), and the verb is “was” (singular). The plural word “birds” is part of a prepositional phrase (that intervenes between the subject and the verb) and therefore is not considered when checking for subject-verb agreement.

Voice: Active versus Passive

When the subject of the sentence performs the action of the verb, the sentence has been constructed in active voice.

- I hit the ball across the yard.

If the subject is being acted upon, then the sentence has been constructed in the passive voice.

- The ball was hit across the yard.

The choice of active or passive sentence structure affects the emphasis in the sentence (see also “Emphasis” in the section Style Points and Conventions). The **beginning** of the sentence receives more emphasis in the active voice. The **end** of the sentence receives more emphasis in the passive voice. Both can be used to advantage.

Passive Voice

Advantages of passive voice:

- minimizes or leaves out the role of the person performing an action (for cases where you do not know who did it, do not want to mention who did it, or who did it is irrelevant)
- softens the tone or makes the message less personal

Disadvantages of passive voice:

- makes the sentence wordy and awkward
- adds formality

Sentences constructed in passive voice often begin with “There is” and “There are”. Most sentences can be improved by changing to active voice:

- **Instead of:** “There are dozens of research resources at NCBI.”
- **Consider:** “NCBI offers dozens of research resources.”

Active Voice

Advantages of active voice:

- Shorter sentences are easier to read (“was” and “by” are usually eliminated).
- Sentences are more dynamic, forceful, clearer, stronger, and concise.

Disadvantage of active voice:

- Sentence overuse emphasizes the doer of the action.

Mixing Voices

Do not mix active and passive constructions in the same sentence. An easy way to check for passive voice is by asking, “Who did that?” If you cannot give a definite answer, the sentence is passive.

Scientific writing most often uses passive sentence construction to give the author less importance and to focus attention on the facts at hand. But scientific writing can be made more engaging by using some sentences in active voice as well.

Strike a balance between the use of active voice and passive voice. Both have their uses. Consider the emphasis in the sentence and whether the sentence needs to be more dynamic. The use of more constructions in the active voice makes text more engaging.

Wordy Constructions to Avoid, Substitutions of Common Phrases, Plainer Synonyms

To write concisely, you must eliminate superfluous detail. Sometimes, this is just a matter of eliminating some phrases altogether. Take the phrase “is designed to”. Unless the design is the point of the sentence, drop this phrase. Type this search in Google: “is designed to” site:ncbi.nlm.nih.gov. Look at the results and try mentally dropping the phrase “is designed to”. Was the sentence adversely affected or improved?

The following list contains helpful suggestions for writing more succinctly.

a number of → few, many, several, some

a majority of → most

accordingly → therefore

accounted for the fact that → because

actually → (consider eliminating this word)

additional → more, added, other

additionally → and, also

adequate number of → enough

adjacent to → close to, near, next to

advantageous → useful, helpful

along the lines of → similar to

an innumerable number of → innumerable, countless, many

an order of magnitude → 10 times

antecedent to → before

anterior to → before

anticipate → expect

any particular type of → any

are of the same opinion → agree

as a consequence of → because of

as a means to → to

as a result of → because of, from

as far as our own observations are concerned, they show → we observed

as regards → about, concerning

as to → on, for, about, of

ascertain the location of → find

at an early date → soon

at present → now, currently

at such time as → when

at the place where → where

at the present moment → now

at the present time → now

at the time that → when

at this point in time → now

basically → (consider eliminating this word)

because of the fact that → given that

by means of → with, in, by

by reason of → because of

by virtue of → by, under, because of

caused injuries to → injured

completely filled → filled

concerning the matter of → on, for

concur → agree

connect together → by

consequence → result

consequently → therefore

consider all factors carefully → consider

consolidate → combine, join

constitute → make up

deem → consider, think, treat as

definitely → (consider eliminating this word)

definitely proved → proved

demonstrate → show, prove

designate → appoint, name, choose, set

despite the fact that → although

diminish → lessen, reduce

discontinue → stop

due to the fact that → because

during such time as → while

during the course of → during, while

during the period from → from

during the time that → while, when

elucidate → explain, clarify

enumerate → list, name

equally as → equally

equivalent → equal, the same

excessive number of → too many

fabricate → build, make

facilitate → make easier, help

fewer in number → fewer

following → after

for a period of → for

for all intents and purposes → (consider eliminating this phrase)

for the purpose of → to, for

for the purpose of examining → to examine

for the reason that → because

frequently → often

from the point of view of → from, for

furthermore it would seem → and

generally → (consider eliminating this word)

give proper consideration to → consider

give rise to → cause

goes under the name of → is called

future plans → plans

has a tendency to → tends

has the ability to → can

has the capability of → can, is able

has discretion to → may

have a negative impact → hurt, harm

hence → therefore

identical → same

if conditions are such that → if, when

illustrate → show

impact → affect, influence

implement → carry out, begin, start, create, set up

in a precise manner → precisely

in a satisfactory manner → satisfactorily

in a situation where → if, when

in addition to → besides, also

in all cases → always, invariably

in an adequate manner → adequately

in back of → behind

in case → if

in close proximity to → near

in connection with → with, about, concerning, for

in excess of → more than, over

in favor of → for

in lieu of → instead of, rather than

in order to → to, for

in proximity to → close to, near

in reference to → on, for

in regard to → on, for

in relation to → on, for

in spite of the fact that → although

in terms of → in, for, about

in the absence of → without

in the amount of → for

in the case in which → when

in the case of → in, with

in the course of → during, while

in the event that → if, when

in the nature of → like, similar to

in the near future → soon

in the neighborhood of → about, roughly

in the place in which → where

in this instance, however → but

in view of the fact that → because

in the vicinity of → close to, near

inasmuch as → given that

inception → start, beginning

indicate → show, suggest

indication → sign

individual → person

initial → first

initiate → begin, start, set up

institute → begin, start, set up

is able to → can

is authorized to → may

is capable of → can

is in a position to → can, may

is permitted to → may

it has been reported by Smith → Smith reported

it is (omit)

it is believed that (omit)

it is felt that (omit)

it is important that → must, should

it is necessary that → must, should

it is often the case that → often

it is possible that the cause is → the cause may be

it is probable that → probably

it is this that → this

it is worth pointing out that → note that

it would appear that → apparently

it would thus appear that → apparently

kind of → rather (or consider eliminating this phrase)

lacked the ability to → could not

large amounts of → much

large in size → large

large numbers of → many

located in → in

located near → near

magnitude → size

manufacture → make

methodologies → methods

modification → change

modify → change

necessitate → require

necessitates the inclusion of → needs, requires

necessity → need, requirement

no later than September 30 → before October 1

notification → notice

notwithstanding → despite

notwithstanding the fact that → although

numerous → many

objective → goal, aim

on a daily basis → daily, every day

on account of → because

on behalf of → for

on or before September 30 → before October 1

on the basis of → by, from, because of

on the grounds that → because

on the part of → by

on the occasion of → on

on the part of → by

ongoing → continuing, active

opt for → choose

optimum → best

option → choice

oval in shape, oval-shaped → oval

owing to the fact that → because

parameter → limit, boundary, guideline, condition

particular → (consider eliminating this word)

past history → history

preclude → prevent

preferable → best, better, preferred

posterior to → after

previous → earlier, last, past

previous to → before

previously → before, earlier

principal → main, chief

prior → earlier

prior to → before

provided that → if, but

really → (consider eliminating this phrase)

referred to as → called

regarding → about, on, for

represents → is, makes up, stands for

responsible for → causes, has charge of

results so far achieved → results thus far, results to date

retain → keep

round in shape → round

serves the function of being → is

situated → placed, sits, lies

smaller in size → smaller

sort of → (consider eliminating this phrase)

subsequent to → later, after

subsequently → later, afterwards, then

sufficient → enough

sufficient number of → enough

take into consideration → consider

terminate → end, stop

termination → end

the organism in question → this organism

the majority of → most

the question as to whether → whether

the reason why is that → because

the tests have not as yet → the tests have not

the treatment having been performed → after treatment

there are (omit)

there can be little doubt that this is → this probably is

there is (omit)

there is a need for → must, should

this is a subject that → this subject

this is a topic that → this topic

there is no doubt but that → doubtless, no doubt

through the use of → by, with (not “via”)

throughout the entire area → throughout the area

throughout the whole experiment → throughout the experiment

transpire → happen

type of → (consider eliminating this phrase)

two equal halves → halves

under circumstances in which →

under the provisions of → under

until such time as → until

used to fulfil → needed

utilize → use

visualize → think of, imagine

warrant the use of → must or should have, require

was of the opinion that → believed

with a view to → to, for

with a view to getting → to get

with reference to → about

with regard to → about, concerning

with respect to → on, for

with the exception of → except for

with the object of → to, for

with the result that → so that

without some kind of → without

Part II. NCBI-specific Styles

Chapter 4. Writing for the Web Environment

Our Web pages must communicate our ideas clearly and effectively, and the visual aspects are as important as the text. Navigation must be easy, and some thought must be given to the interaction experience of the user.

Going Public

NCBI has an international audience. Because humor does not translate well into other languages, keep instructions clear and concise.

Avoid jargon, slang, and nonstandard usages.

- Do not use “**flavors**” when you mean styles, formats, or varieties. Instead of writing “GenBank submission tools come in three flavors”, write “ NCBI provides three GenBank submission tools”.
- Avoid using “**allow**” to mean that computer resources or controls provide a service. Instead of writing “My NCBI allows users to store searches”, write “My NCBI stores searches for users”.
- Do not write that a computer “**crunches**” data---“processes”, or a similar term, is more appropriate.

Read about anthropomorphism and think about how computer jargon may relate to it.

Before making your Web pages public, have your colleagues, a person naive to the project, and an editor review your pages. Use the most accurate representations of symbols, with consideration of the capabilities of various Web browsers. Consider layout and color, as well as content, as means of making your ideas understandable and engaging.

Constructing Searchable Web Pages

All static and dynamically generated Web pages on public servers must be designed to maximize usability and searchability. HTML documents must have title and meta tags in the <head> section, and links must be unbroken:

- **TITLE** The document title (of 50 characters or fewer, including spaces) must be unique within NCBI. In addition to the document title, consider adding a title attribute to each meaningful image on the page. This is important if you use an image with graphical text as the heading of your page. The title should briefly describe what is in the image. Most visual browsers preferentially show image title text as a tooltip over the image. If the image is absent, the browser uses the ALT tag. If the ALT tag is absent, there is no pop-up. Voice browsers also speak the title tag, if any, or the ALT tag, if any, or nothing. Document titles are displayed on the search results page of a search engine, so the title needs to be human readable and should clearly convey what the page contains.
- **AUTHOR** Use your group's project name for the author value.
- **KEYWORDS** Although some search engines do not use keywords, keywords are still helpful in intranet searches, such as the NCBI Site Search. Use no more than 25 single keywords or some combination of 2-3-word phrases that are specific to the content of the page in a comma-delimited list. Think of different ways to describe your content; people often use different words and phrases to search for the same content (e.g., medical school or school of medicine, vitamin C or ascorbic acid). Start with a base of "national center for biotechnology information, ncbi, national library of medicine, nlm, database, archive" (searches of keywords are case insensitive) and then add words more specific to the content of your NCBI pages. Use a thesaurus or the [Medical SubHeading \(MeSH\)](#) browser to explore additional keywords, as well as a dictionary to ensure correct spelling and to find common alternate spellings. This approach will be useful for both internet and intranet searches by search engines.

- **DESCRIPTION** The description, like the title, must be unique. The best descriptions are short and specific sentences with fewer than 25 words, with the most relevant words near the beginning. Descriptions are displayed on the search results page of a search engine.
- **LINKS** Links may take you out of the present document or take you to another part of the same document. Links out of the document must reference existing content. These links include hyperlink “href” attributes, such as `label`, image src tags, such as ``, and linked stylesheets and Javascript files, such as `<link rel=“stylesheet” type=“text/css” href=“styles.css”/>` and `<link rel=“javascript” type=“text/javascript” href=“myscript.js”/>`. All links within the document must refer to existing anchors.

Example

```
<head>
```

```
<title>Single Nucleotide Polymorphism (SNP)</title>
```

```
<meta name=“author” content=“snpdev”>
```

```
<meta name=“keywords” content=“national center for biotechnology information, ncbi, national library of medicine, nlm, national institutes of health, nih, database, archive, single nucleotide polymorphism, SNP, sequence variation, genome, gene mapping, population structure, dbSNP, evolution, microsatellite repeat, phenotype”>
```

```
<meta name=“description” content=“The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms.”>
```

```
</head>
```

Make sure that the content on your page is visible on browsers that have JavaScript disabled. Not all search engines can index content contained in scripts.

Keep Your Web Pages Current

Dates are an important part of your Web document because they help readers understand where your information fits in the timeline of relevant events. In some Web documents, dates serve merely to let readers know when the pages were checked. In other Web documents that report on fast-changing information, dates assume more importance.

When information is time sensitive, place the date at the top of the Web document. Otherwise, place the date at the bottom of the page.

It is also helpful to distinguish these two types of dates. “Last Updated” conveys some sense of information that is time sensitive, whereas “Revised” seems to suggest more of a “maintenance” approach. For the sake of consistency in the presentation of these dates, use the following forms:

- Last Updated: November 21, 2003.
- Revised: November 21, 2003.

When using “Last Updated”, place the date at the **top** of the Web document. When using “Revised”, place the date at the **bottom** of the page.

Note the periods at the end of the entries.

Glossaries on the Web

There are two styles for glossaries on the NCBI Web pages:

word to be defined - fragment and/or sentence definition.

word to be defined

fragment and/or sentence definition.

The first one saves a little space, if the definition is short enough. Note that both definitions begin with a lowercase letter and end with a period.

Itemized Lists

Many NCBI Web pages contain itemized lists. It is important to consider the elements of an itemized list and then decide on the type of list that you will need. For more information, also see Lists in Style Points and Conventions.

There are three considerations for making lists:

1. **Type of list:** Is the **order** of the list items important? (If Yes, use a **Numbered List**; if No, use a **Bulleted List**.)
2. **Whether list items are fragments or complete sentences:** Will the list items be fragments or complete sentences? (Sentence-style punctuation is used for complete sentences.)
3. **Capitalization:** When should list items be capitalized? (Action verbs that begin lists and lists that follow headings without additional explanatory text begin with capital letters.)

Below are six examples of the two types of lists. There are two style choices in the **Numbered List** and four style choices in the **Bulleted List**. Please apply your style choice to your Web document consistently.

Numbered List

In a Numbered List, the order of execution of the items is important and/or emphasis on the number of items is being made.

When providing instructions or directions that must be followed in a particular order, capitalize the first word of the entry and end it with a period. Each item of this list begins with an “action” verb (sometimes called “commands”), meaning that the person reading the list is being told what to do and must take action to get the wanted results. See Example 1.

Example 1:

1. Press the Validate and Continue button.
2. Enter the number of CDS features.
3. Scroll down to the CDS Feature box.
4. Complete the CDS Feature subsections.

When the items in the list are not complete sentences, no punctuation is used at the ends of the items. See Example 2.

Example 2:

For each complete submission you have made to us, you will receive by email:

1. an automatic preliminary GenBank flatfile

2. a GenBank Accession number
3. a completed GenBank flatfile

Bulleted List

In a Bulleted List, the order of execution of the items is not important (it is not a step-by-step procedure), and “counting” the items is not necessary.

In both Examples 3 and 4 that follow, the list items merely continue the idea of the introductory sentence fragment. Contrast the use of the verbs in Example 4 with the action verbs in the Numbered List, Example 1. Note that these verbs in Example 4 are not capitalized because they are not giving directions for actions as commands. Note also that there are no periods, semicolons, or commas at the ends of the entries in Examples 3 and 4, because each item entry presents a part of the thought (each item entry is only a part of the whole idea, with the “idea” being the introduction to the list by the introductory sentence fragment and the list entries themselves). Because of the layout of the list itself, ending punctuation is not needed; the list presents the idea clearly.

Example 3:

Use BankIt if:

- you have one or a few sequence submissions
- you prefer to use a WWW-based submission tool
- your sequence annotation is not complicated
- you do not require sequence analysis tools to submit your sequence(s)

Example 4:

Failure to recognize foreign segments in a sequence can:

- lead to erroneous conclusions about the biological significance of the sequence
- waste time and effort in analysis of contaminated sequence
- delay the release of the sequence in a public database
- pollute public databases with contaminated sequence

When a list of complete sentences immediately follows a heading, use paragraph indents with filled bullets at the first level. (Your Browser may control the types of symbols used with nesting.) Begin with a capital letter and end the sentence with a period. See Example 5.

Example 5:

BankIt: GenBank Submission by WWW

- GenBank provides annotation examples and descriptions for several types of sequence submissions.
- Each sequence must be submitted individually and will receive its own BankIt number.

When a list of fragments immediately follows a heading, use paragraph indents with filled bullets. Begin with a capital letter and do not use periods, semicolons, or commas as closing punctuation for the fragments. A combination of fragments and complete sentences is fine. If a complete sentence is used, use a closing period. See Example 6 for use of this combined form.

Example 6:

Requirements for Every Submission

- Contact information

- Release data information
- Input DNA sequence
 - A contiguous nucleotide sequence of at least 50 base pairs
 - Type of molecule sequenced
 - Description of the sequence:
 - Click on the specific examples in the left column.
 - If the sequence submitted does not fall into one of the categories on the left, please use the available features or provide a complete description.

Outlines in Web Help Documents

There are many ways to present information clearly and effectively. Consider using an outline when writing your Web Help documents. If you do plan on using an outline, an outline at the beginning of your Web Help document is a good way to orient your readers so that they can quickly understand the layout of your information. An outline is another visual aid to supplement the information supplied by the sidebars and toolbars. One good outline method, shown below, uses an indent system shown with specific symbols at each level of indent, a system used by Word and other wordprocessors.

This is an example of the indent system:

- Overview
 - Pubmed Searches
 - Author Names

An alternative method uses a combination of numbers and symbols:

1. Overview
 - PubMed Searches
 - Author Names

Both methods present Help documentation clearly and effectively.

Many people will print out documents to read off-screen. Unless the outline is on a Web page by itself, the assumption is that all of the material contained in the Outline will print out as one document. If sections of the Outline are links to other pages, add a parenthetical statement to the Outline section, such as “(open and print this link separately)”. Note that the words are lowercase. In general in Outlines, use lowercase for parenthetical information. A good example of this is shown on the [OMIM Help page](#).

Web Pages without Reference Lists: Making Citations

Link every citation in the text to PubMed, PubMed Central, and Bookshelf, where possible. Where this is not possible, all information must be provided for the reader in [National Library of Medicine \(NLM\)](#) format:

- ... was described previously ([Lipman et al. 2002](#); [Benson et al. 2000](#)) ([Abramovitz et al. Chemosphere 1990;21:1221-1229](#)) ([Salton G. Automatic text processing. Reading \(MA\): Addison-Wesley; 1989](#)).

In the example above, both the Lipman journal citation and the Benson journal citation are linked to PubMed. The journal article by Abramovitz is not in PubMed, but enough information is given in this format for the reader to find it from other sources. Likewise, the book by Salton contains enough information for the reader to contact the publisher or bookseller. When a reference list is not being provided on the Web page, complete citations must be provided in the text itself. The four citations given in the sample above are separated because of punctuation issues and concern for readers to be able to more easily understand the amount of information presented with the citations.

NCBI Standard Abbreviations and Frequently Used Word Forms

NCBI has an international audience of all educational levels. As an author, you must consider your audience, meaning that if you are presenting a biology paper of great complexity that is not likely to be read by a general audience, it is probably not necessary to write out DNA. If there is any doubt, a good method would be to provide the written-out form of the word, followed by the abbreviated form in parentheses:

- ... is called deoxyribonucleic acid (DNA).

Thereafter, use the abbreviated form only. Begin with your written-out forms again for every page that loads on the Web site, unless the pages that follow are a clear continuation of the initial page. Obviously, you must use your judgment on this.

Frequently Used Word Forms

a, an - article used according to the sound of the abbreviation (e.g., an STS; a sequence tagged site); this style point is an exception to the CBE rule, which states that the article is used according to the sound of the written-out form.

Jump to: [A-E] [D-F] [G-L] [M-O] [P-R] [S-Z]

[A-E]

AIDSLINE - AIDSTrials.

ASN.1

AVLINE

BAC-end sequence - (note hyphen).

BankIt

base pair - (two words).

BLAST - PHI-BLAST, PSI-BLAST, QBLAST, RPS-BLAST, IgBLAST, MegaBLAST.

BLink - BLAST link.

BLOSUM62

Bookshelf

Boolean - (capitalized).

by-product(s)

C terminus, COOH terminus, C-terminal

catalog - (no “-ue”).

CATLINE

CCAP - Cancer Chromosome Aberration Project.

CD-Search

CDD - Conserved Domain Database.

CGAP - Cancer Genome Anatomy Project.

CGH - comparative genomic hybridization.

check box

ChemID plus

cis - (italic).

ClinicalTrials.gov

ClustalW

Cn3D

co - co-dominant, co-inheritance, co-segregation, cofactor, cooperative.

COGnitor

COGs - Clusters of Orthologous Groups.

cross - crossing-over (recombination); but crossover (events), crossovers, crosslink, crossbreeding.

cutoff - (noun).

cut off - (verb).

[D-F]

3D - (for three-dimensional; not 3-D).

DART - Domain Architectural Retrieval Tool.

data - this is a plural word (datum is singular).

database-mining

dataset

dbEST

dbGSS

dbSNP

dbSTS

DDBJ

DDBJ/EMBL/GenBank - (alphabetical order as agreed by collaborators).

DDD - Digital Differential Display.

DGED - Differential Gene Expression Displayer.

drop-down menu

email - consider putting the email address in boldface text.

EMBL

employs - change to “uses” unless specifically referring to the employment of a person.

Entrez

Entrez Genomes

Entrez Nucleotides

EST - expressed sequence tag.

eukaryote - (no “c”).

ExPASy

fa2htgs - not to be capitalized, even beginning a sentence.

false-positive results

FASTA

FISH - fluorescence *in situ* hybridization.

flatfile - (one word).

fruit fly

FTP

[G-L]

GDS - GEO DataSets.

GenBank

GeneMap’99

GenPept

GEO - Gene Expression Omnibus.

GI - “Geninfo Identifier”.

GLS - Gene Library Summarizer

Help desk - NCBI Help desk

homepage

homolog - (no “-ue”) .

HomoloGene

Human–Mouse Homology - (note en dash).

ID - sequence ID, PubMed ID, MEDLINE UID.

in vitro - (italic).

in vivo - (italic).

kb - kilobase.

LOCATORplus

LocusLink

log in - (verb).

login - (noun).

look-up table

lowercase - (one word).

[M-O]

MAGEML - MicroArray Gene Expression Markup Language.

Map Viewer

matrices

Mb - megabase.

MD - (no spaces, no periods; “Medical Doctor”).

MEDLINE

MEDLINEplus

MGC - Mammalian Gene Collection.

MGD - Mouse Genome Database.

MGED - Microarray Gene Expression Database.

MinSeq - Minimal Sequence.

MMDB - Molecular Modeling Database.

Model Maker

multi - multi-step.

N terminus, NH₂ terminus, N-terminal

NEXUS

NLM Gateway

non - non-coding, non-cumulative, non-identical, non-redundant, non-synonymous, nonhuman.

OLDMEDLINE

OMIM

online

ONO - order and orientation information

[P-R]

pairwise

paralog - (no “-ue”).

PDB - Protein Data Bank (Brookhaven).

PDBeast

Pfam

PhD - (no spaces, no periods).

PHYLP

PopSet

post - post-translational.

pre - precalculate, precomputed, predetermine, pre-existing.

PREMEDLINE

ProbeSet

prokaryote - (no “c”).

PROW - Protein Reviews On the Web.

psort2

PSSM - position-specific scoring matrix.

PubMed

PubMed Central

pull-down menu

RCSB - Research Collaboratory for Structural Bioinformatics.

RefSeq - Reference Sequence.

RepeatMasker

RFLP - restriction fragment length polymorphism.

[S-Z]

SAGE - serial analysis of gene expression.

SAGEmap

Sequin

SERLINE

SKY/CGH - Spectral Karyotyping and Comparative Genomic Hybridization database.

SMART

SNP - single nucleotide polymorphism.

SOFT - Simple Omnibus Format in Text.

species - spp. (plural; note period).

Spidey

SSLP - simple sequence length polymorphism.

stand-alone program - (note hyphen).

STS - sequence tagged site.

SWISS-PROT

TaxBlast

Taxonomy Browser

TaxPlot

tbl2asn - not to capitalized, even beginning a sentence.

TinySeq - Tiny Sequence.

ToolKit

ToolBox

trans - (italic).

UniGene

UniSTS

UniVec

uppercase - (one word).

UserID

VAST - Vector Alignment Search Tool.

VecScreen

versus - (italic); OK to abbreviate as “vs.” in tables.

VNTR - variable number of tandem repeats.

website

Whole Genome Shotgun sequencing

wild card

XML

xProfiler

X-ray

ZFIN - Zebrafish Information.

Be consistent with these frequently used word forms.

Chapter 5. Style Points and Conventions

Please let me know if any of the style points seem awkward to you. What works in one field may not be as useful in another. A style point is something that is decided upon by the person(s) in charge or by a committee (that decides to do something because of the historical use of doing something more one way than another or after analysis of the needs of all members of the group), but by whatever means agreement is reached for style points, the main use of style points is to improve consistency.

Please note: Styling of Appendices, Boxes, Figures, References, and Tables is in Special Content.

Amino Acids

There are only four nucleotides (A, C, G, and T), but a sequence of three of the four nucleotides specifies an amino acid. The three nucleotides, sometimes called triplets, make up the codon.

Use the three-letter abbreviations for amino acids in chains, as well as with codon numbers and mutations.

Examples:

Arg-506

Arg506Gln

Met-Phe-Val-Asn-Gln-His

Note: Single-letter abbreviations should be used sparingly.

Example:

MFVNQH (instead of Met-Phe-Val-Asn-Gln-His)

Amino acid abbreviations are determined by IUPAC.

Capitalization

See CBE, pp. 149-165.

Note: When a sentence begins with an abbreviation or designation that has an initial lowercase letter, the abbreviation or designation retains its lowercase.

Example:

Our study determined the protein to be p53. p53 has been studied extensively.

Centrifugation

Use *g* force (3000*g*), not rpm. Note that there is no “times” sign.

Colons

According to CBE, there are four general uses of colons and nine specialized uses (pp. 44-46). Here, I will provide you with a few examples of the usual things.

Example:

Note: The images were scanned from photographs.

In the example above, note the capital “T” after the colon.

Example (from a Table title):

Table 2. LocusLink query terms: controlled terms.

In the example above, note the lowercase “c” after the colon.

Example (from a Reference):

Al-Shahrour F, Diaz-Uriarte R, Dopazo J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*. 2004 Mar 1;20(4):578-80. Epub 2004 Jan 22.

In the example above, note the lowercase “a” after the colon.

Example (from a Title):

GenBank: The Nucleotide Sequence Database

In the example above, note the capital “T” after the colon.

Commas

Refer to CBE, pp. 48-52, for a complete discussion. The purpose of the comma is to make the sentence clear. It is not a symbol of conversational pause.

In general, use a “serial comma” to avoid ambiguity. A serial comma sets off a “simple series of more than two elements” (see the commas set in bold in the next example).

Example:

MIM is organized into autosomal, X-linked, Y-linked, and mitochondrial catalogs, and MIM numbers are assigned sequentially within each catalog.

The comma not set in bold in the example above is a comma used to separate independent clauses (complete sentences). This comma makes the sentence easier to read and understand.

Another comma error is the use of a comma between the subject and the verb. The separation of the subject from the verb by an intervening comma is called “comma separation”.

Example:

WRONG: Suggested corrections and additions are welcome for future updates of these pages, and should be sent to the author.

RIGHT: Suggested corrections and additions are welcome for future updates of these pages and should be sent to the author.

Concentrations

Show liter as “l” with combinations of units but written out as “liter” otherwise. Show gram as “g” (and kilogram as “kg”). Abbreviate mol/l as “M”.

Dashes

There are two types of dashes, the en dash (–) and the em dash (—). (A hyphen is shorter than an en dash.)

en dash

used to separate ranges and words of equal weight

Examples:

63–65, Mann–Whitney, dose–response curve

em dash

used to break a thought in a sentence (which is often a long pause, an aside, or a title fragment)

Example:

Once he made the decision—did I say that already?—nothing stopped him from reaching his goal.

Contrast these uses with hyphens (see Hyphens).

Decimal Points

Insert 0 in numbers less than 1 (0.05).

Degrees

Close up symbol to number (55°C).

For educational degrees, such as MD and PhD, do not use periods.

e.g.

Use to mean “for example”. Do not italicize; use periods and follow with a comma.

See also Italics.

et al

Do not italicize. Provide all names in the “References” section. In other formats (where the display of information is an issue because of space), reference citations may use et al. with only one author.

See also Italics.

etc

Substitution of “and others” or removal of the term is preferred, if possible.

Emphasis

Underlining is not preferred for emphasis within the text to avoid confusion with underlining that is used to show Web links. (If hyperlinks are shown in a different color of underlining or in a different manner so that no confusion would exist, underlining for emphasis would be OK.)

Use quote marks to show that a word is used in an unusual way. If emphasis is needed to show how an acronym was derived, use boldface letters for the beginning of each word, e.g., **D**eleted in **C**olorectal **C**arcinoma (DCC). When the derivation of the acronym is obvious, do not use the boldface letters. See also Voice: Active *versus* Passive.

Equations

If equations are numbered, cite them as Equation 1, etc. Capitalize the word Equation and write it out in the text, figures, tables, and boxes. A sentence (or sentence fragment) introducing an equation ends with a colon. The equation itself does not receive any type of sentence punctuation, such as a period, semicolon, or comma, at its end.

Italicize variables, but use roman type for SD, SE, CV, etc.

See also Math and Statistics.

Foreign Words

See Italics.

Genes and Protein Designations

Genes are italicized, and proteins are set in roman type (not italic). Use where reasonable. Web pages with long lists of genes in italics would be distracting. In this case, use roman type.

Hyphens

Hyphens are used for many purposes, such as with:

- prefixes and suffixes
- compound terms
- compound modifiers of several types
- spelled-out fractions
- written-out forms of numbers from 21 through 99
- representation of single bonds in chemical or molecular formulas, peptide bonds in residues, and linkage of nucleotides

See CBE pp. 61-64 for many examples of the specific uses of hyphens. Below are a few common examples to make your search quicker.

Unit Modifiers Are Hyphenated

e.g., best-case scenario, false-positive results, high-risk behavior, model-fitted values

Note

Two-word phrases that are easily understood as a unit are not hyphenated.

No Hyphen: amino acid residues

Use a Hyphen with Some Prefixes

e.g., anti-oncogenes, mis-segregation, non-redundant, pre-existing, BUT nonlinear, pretreatment

Most “co” prefixes are closed up: coeluted, comigrate, copolymerize

Exceptions

co-injected, co-worker

Compounds with Phosphorus

No hyphen if the first word in the name is a noun. A group or adjectival form takes a hyphen, e.g., glucose 1-phosphate, cytosine 3[prime]-phosphate, glycerol-1-phosphate (see [IUPAC-IUBMB](#)).

Contrast hyphens with en dash and em dash (see [Dashes](#)). See also [Amino Acids](#).

-ical

English uses a couple of adjective endings that are sometimes confused: **-ic** and **-ical**. There are differences in meaning, and examples will best illustrate these differences.

Take the pair of words **classic/classical**. You may own a classic car, but you listen to classical music. “Classic” refers to something being traditional and enduring, or serving as a standard of excellence—perhaps an example of an outstanding model. “Classical”, on the other hand, refers to something that is a historically important form or genre of architecture, music, art, etc. Of the two forms, “classic” seems to apply more to something specific, and “classical” seems to apply to something broader, more general but distinguishable.

Consider **graphic/graphical**. If you watch TV, you may see a lot of graphic violence, where “graphic” more refers to the effects on the emotions, but if you choose, instead, to read [The NCBI Handbook](#), you will be learning from its many graphical images, where “graphical” more refers to the format of the material. Merriam-Webster's Dictionary does not make this particular distinction, so this is a style point. Here we are using “graphic” to mean providing details of perhaps a shocking nature, whereas “graphical” is used to mean a drawing, illustration, or art product.

i.e.

Note punctuation. Do not italicize.

See also [Italics](#).

in vitro, in vivo

Use italics.

See also [Italics](#).

Italics

There are general and scientific uses of italics. [On the use of italic and roman fonts for symbols in scientific text](#) is available, as well CBE coverage on pages 169-171.

For our most common purposes, use italics for genus/species names, genes, loci, and alleles; parts of chemical names as appropriate (including *cis*, *trans*, *ortho*, *meta*, and *para*); all variables (e.g., probability (*P* or *p*)); and written-out Latin forms (such as *a priori*, *ad libitum*, *de novo*, *in situ*, *in utero*, *in vitro*, and *in vivo*).

CBE considers whether a Latin word is common and suggests that “in vitro” and “in vivo” not be italicized. This, then, becomes a style point. I think it would be more confusing to selectively choose some written-out Latin words to be italic than to go along with the old standard of italicizing all of the Latin words.

What of “e.g.,” “et al.,” “i.e.,” and “etc.”? Well, these are not written-out forms, AND they are very common; therefore, there is no need to italicize “e.g.,” “et al.,” “i.e.,” and “etc.”. More basically, the handling of Latin and other foreign words is influenced by the “acceptance” of a term as “common” as to whether an italic form is used. As

specific foreign words become more common, the trend is for the italic style to be dropped, as a simplification. For example, uncommon French words used in English would still be shown in italics, but French words that have been used in English for a long time, such as “à la carte”, are readily understood by most people and would not be set in italics (and some go so far as to drop the accent mark). Because differences of opinion can arise as to whether a foreign word has become common enough in English to go without italics, style points are then described in style guides, for the sake of consistency.

When working with long lists of genus/species names, genes, loci, and alleles, particularly in tables, the italics become more of a distraction than a help for readers. In these cases, use your best judgment about the use of italics.

Per changes by IUPAC in 2003, do not italicize the first three letters of restriction enzymes (EcoRI, BamHI).

Latin

See Italics.

Lists

Paragraph format: Within paragraphs, use *(a)*, *(b)*, *(c)*, etc. and then *(i)*, *(ii)*, *(iii)*, etc.

Outside of the paragraph format:

- “Bulleted” lists are fine. Be sure that all items with a similar weight use the same symbol.

In bulleted lists, entries of phrases, words, or sentence fragments do not use a period at the end of the entry. Do not connect list items with conjunctions such as “and” or “or”.

If the bulleted entries are complete sentences, use sentence capitalization and a period at the end of each entry.

- Use a “numbered” list only when the items must follow a sequence. When steps of a procedure are being referred to, use Arabic numbers (OK as a subsection within the bulleted list).

For more information, see Itemized Lists in Writing for the Web Environment.

Math

Follow CBE.

Numbers/Numerals

Write out at the beginning of a sentence. Write out one to nine except with units of measure. Use numerals for 10 and above.

Note: If a study is of 10 or more patients, items, things, etc., use numerals in sentences referring to parts of the total group, even when less than 10.

Example:

We studied 36 patients. In the first group of 6 patients, only 2 were anemic.

Quick List

1. Written-out numbers *versus* as numerals:
 - a. For numbers of 10 or more that do not refer to units of measure, use numerals.

- b. Units of measure require numerals (3 mm, 4 months).
 - c. Avoid beginning a sentence with a numeral.
2. Fractions: one-half, two-thirds, one-tenth, 1/32
3. Ordinal numbers: 1st, 2nd, 3rd
4. Series: if the series contains numbers 10 or higher, use all numerals (3 mice, 4 rats, and 11 hamsters)
5. Either side of the decimal point: use 0.1 not .1 (all cases); use 6.0 not 6 (only if significant)
6. Commas: 3,000; 14,000; 3,333,331
7. Time 0 (contrast zero time)
8. Decimals instead of fractions with measurements: 0.5 volume (not one-half volume)
9. Bases use numerals (even those fewer than 10): 5 bases

Probability

Use an italic *P* or *p*. Use spaces around math symbols.

See also Statistics.

Proofreaders' Marks

Some of you wonder about the shorthand-looking marks made on hard copies of your Web pages. This is a way of showing corrections and changes. There are just a few marks to learn, and they come in handy!

Here is a nice link to view the marks: [Proofreaders' Marks](#).

Punctuation

See CBE, pp. 36-71.

See also Colons, Commas, Dashes, Hyphens, Quotation Marks, and Semicolons.

Quotation Marks

Place quotation marks inside of punctuation.

Use double quote marks. For quote marks inside of quote marks, use single quotes on the inside of the phrase and double quote marks on the outside of the phrase.

Example:

The NCBI Handbook addresses PubMed's handling of Stopwords, specifically that “PubMed ignores Stopwords, such as ‘about’, ‘of’, or ‘what’...”, when processing simple searches.

Use double quotation marks around phrases that some of the words begin with lowercase.

Example:

You may also want to display your citations using “Send to Text” to eliminate the sidebar menu and toolbars before printing your results.

In the example above, the “Send to Text” phrase would not need quote marks if it were a hotlink because of the way links show up on the Web pages. The blue font would be enough to make it distinctive.

Also use double quotation marks to avoid confusion, as well as to show that you are implying a meaning somewhat different than that usually assumed (or that you doubt the validity of the usual meaning).

Restriction Enzymes

Do not use italics for the first three letters and close up the entire name, e.g., AccI, HaeII. Removal of italics is a change made by IUPAC in 2003.

Helpful Tip

Visit [The Restriction Enzyme Database \(REBASE\)](#). Select “Enzyme Navigation Tables” and “All enzyme names...” for a complete list of restriction enzymes.

Semicolons

See CBE, pp. 46-48, for the four general uses and five specialized uses of semicolons. Here are examples of common uses:

Example:

Certain treatments were more effective than others in eliminating the contaminating DNA; however, to achieve this there was a decrease in sensitivity.

Example:

Links are provided to related Web sites including: chromosome databases (e.g., the Mitelman database); other NCI (e.g., CGAP and CCAP) and NCBI (e.g., the Map Viewer (Chapter 20) and LocusLink (Chapter 19) resources and PubMed (Chapter 2)) sites; The Jackson Laboratory; and several other CGH sites.

Spelling

American spellings are preferred over British spellings.

Differences in spellings, from British to American, usually involve:

- swap of the letters “r” and “e”, such as “centre” (British) and “center” (American)
- addition of the letter “l”, such as “cancelled” (British) and “canceled” (American)
- use of letter “s” instead of letter “z”, such as “characterise”/“analyse” (British) and “characterize”/“analyze” (American)
- addition of letter “a” or “o” with letter “e”, such as “haematology”/“foetus” (British) and “hematology”/“fetus” (American)
- addition of letter “u”, such as “tumour”/“mould” (British) and “tumor”/“mold” (American)
- use of “ph” for “f”, such as “sulphate” (British) and “sulfate” (American)

Table 1 gives a few examples of British and American spellings.

Table 1. British and American spellings.

British	American
analyse	analyze
artefacts	artifacts
cancelled	canceled
centre	center
characterise	characterize
colour	color

Table 1 continued from previous page.

British	American
criticise	criticize
defence	defense
emphasise	emphasize
encyclopaedia	encyclopedia
favour	favor
foetus	fetus
fulfill	fulfil
grey	gray
haematology	hematology
immortalise	immortalize
judgement	judgment
licence	license
mould	mold
neighbour	neighbor
organisation	organization
oxidise	oxidize
prioritise	prioritize
realise	realize
recognise	recognize
specialise	specialize
stylised	stylized
sulphate	sulfate
summarise	summarize
symbolise	symbolize
tumour	tumor

Interesting Fact

PubMed “translates” users' queries with British spellings into American spellings behind the scenes. For more information, read the [NLM Technical Bulletin](#).

Statistics

In general, variables are italicized, such as P for probability test, t test, U test; R for regression coefficient, r for correlation coefficient; n for total number.

Distributions could be:

- exponential (duration)
- geometric (time to first success)
- Poisson (counting, rare events)
- normal (measurement error, asymptotic, approximations)

- uniform
- Bernoulli (coin toss)
- binomial (sum of coin tosses)

Some common variables used in bioinformatics are: *CR* (critical region); *D* (divergence); *E* (*E*-value), *err* (average training error); *Err* (true test error); *F* (distribution function); *g*; *H* (entropy); *I* (mutual information); *K*, *k*; *P*, *p* (probability); *S* (sample space); *T*, *t*; *V* (variance); *X*, *x*; *Y*, *y*; *z*.

Common terms are:

- aggregating, model averaging
- bagging, bootstrap aggregating
- Bayes theorem
- best-case scenario
- BIC, Bayesian information criterion
- bootstrap, replicate learning sets
- Central limit theorem
- Chebyshev's inequality
- CI, confidence interval
- COV, covariance
- CV, coefficient of variance
- cross-validation
- eigen value
- FDA, flexible discriminant analysis
- Gaussian
- Gini index
- GLM, generalized linear model
- Hessian
- hyperplane
- K-fold cross-validation
- kernel functions
- kNN, *k* nearest neighbor
- IRLS, iterated reweighted least squares
- Kolmogorov–Smirnov
- Lagrangian multipliers
- Laplace approximation

- LDA, linear discriminant analysis
- least squares
- LLR, likelihood ratio
- logodds
- Mahalanobis distance
- MAR, missing at random
- MCAR, missing completely at random
- Markov chains
- MDL, minimum description length
- MLE, maximum likelihood error
- model-fitted values
- MoM, method of moments
- MSE, mean square error
- Newton–Rhapson
- Neyman–Pearson
- nonlinear
- QQ-plots
- PDA, penalized discriminant analysis
- SD, standard deviation
- SE, standard error of the mean
- Softmax, dummy regression
- SRS, simple random sample
- WLLN, weak law of large numbers
- z-score

Symbols

An [abbreviated list of quantities, units, and symbols used in physical chemistry and related disciplines](#) is available from IUPAC.

See pp. 815-818 in the CBE Index for a list of symbols.

CBE provides the correct symbol representations for many diverse fields of science.

In Chapter 15 (**Subatomic Particles, Chemical Elements, and Related Notations**, pp. 251-259), you learn that IUPAC recommends roman letters for particle symbols, but the Particle Data Group uses italics for the elementary particles. By IUPAC standards, “e” stands for “electron”, and “*e*” stands for “elementary charge”. The Particle Data Group uses *e* to mean “electron”.

Chapter 16 (**Chemical Names and Formulas**, pp. 260-304), p. 292, describes “Protein Designations Based on Gene Names”.

Example:

“Names for proteins representing a mutant that is characterized by replacement of a single amino acid are sometimes derived from the corresponding gene symbol. For example, an amino acid replacement with valine at position 12 of the Ras protein may be indicated by Ras^{Val12}.” (from CBE, p. 292)

More specific information on genes in particular can be found in Chapter 20 (**Cells, Chromosomes, and Genes**) on pp. 334-380. In general, genes are shown in italics and proteins are shown as roman, but when listing many, many genes, it is not helpful to use this style. You must use your judgment when deciding on when to show genes in an italic style.

Viruses are covered in Chapter 21 (pp. 381-394), **Bacteria** in Chapter 22 (pp. 395-410), **Plants, Fungi, Lichens, and Algae** in Chapter 23 (pp. 411-445), and **Human and Animal Life** in Chapter 24 (pp. 446-485). Within Chapter 24 is a detailed description of taxonomy.

There are other interesting chapters in the “Special Scientific Conventions” section of CBE as well.

Time

Convert military time (0600–1800) to AM/PM time (6 a.m.–6 p.m.; 6:30 a.m.–6:30 p.m.). Include time zone and daylight/standard time designations, if necessary.

Units of time: year, month, week, day, hour, min, sec. Only “min” and “sec” are abbreviated.

Trademarks, Trade/Supplier Names

Use superscript TM or ® when known, at first occurrence only.

At first mention of a company name, use its full form (e.g., Sigma Chemical Co.).

It is OK to abbreviate Co., Inc., Ltd., etc. Thereafter, refer to the company with a shortened version (e.g., Sigma). No need to supply city/state information. If items have been provided as gifts, provide city/state information (no postal code).

Units of Measure

See the CBE Index of “symbols(s)”, beginning on page 815.

Use “ml” for milliliter, “mol” for mole, and “M” for mol/l. “Liter” is written out when used alone.

versus

Use italics.

Write out “versus” in text, but it is OK to use “vs.” in tables.

X-ray

Use a capital X wherever the term appears.

Chapter 6. Special Content

Appendices, Boxes, Figures, References, and Tables are listed as “Special Content” because they have more styling issues associated with them, and it is harder to make them consistent (but not impossible!).

Appendices

Use Arabic numbers when there is more than one Appendix. Give a brief title, capitalizing the first word of the title and using lowercase thereafter. Capitalize the first word after a colon, unless it is a designation or abbreviation that is lowercase. Place a period after the title. Appendices appear after Acknowledgments and Addendums and before References.

Example:

Appendix 1. FASTA identifiers: Identification of sequence source databases.

Boxes

This format sets information in a separate box from the text, accessible from the text by a link and from the jump list on the left of the screen (sidebar).

Number the boxes consecutively with Arabic numbers followed by a period. Capitalize the first word of the title and use lowercase thereafter. If the title contains a colon, the word after the colon is capitalized, unless that word is a designation or abbreviation that begins with a lowercase letter. This is the same title style as Tables.

Example:

Box 1. Approaches to associate sequence data with LocusLink loci.

Figures

Write out the word “Figure” in the text and legends.

In legends, the title is written in bold on a separate line from the additional information. Italicize the locants and the parts of the figure. Locants may be in lowercase or uppercase, but be consistent.

The word “panel” can usually be avoided. Be sure that all symbols on the figure are described. Refer to symbols as “filled” (not “solid”) and to lines as “unbroken” (not “solid”).

Example:

Figure 7. Karyotype parser.

The short-form written karyotype (*a*) has been converted into a modified long-form karyotype (*b*).

Note the position of the locants in the figure legend. The locant follows the information it describes. Locants given in a series format are styled: *a*, fresh samples; *b*, desiccated samples.

When referring to the figure in the text, use roman-style locants (“Figure 7a”). When you want to refer to two parts of the figure (such as “Figure 7, a and b”) and the figure contains only two parts, change the text entry to read “Figure 7”. If the figure contains more than two parts and you are referring to two parts, show “Figure 7, a and b”.

References

Cite references in the text in the proper order (numerically) as full-size, on-line numbers within parentheses. Number the references in the References section consecutively. References cited only in tables and figures are listed after those that have been cited in the text.

References “in press” are cited in the References section of the paper. Unpublished work, manuscripts in preparation or submitted for publication, verbal reports given at meetings, and personal communications are cited in the text in parentheses (or as footnotes).

Use of et al.: References in a paper that are linked with a PubMed ID or to the NCBI Bookshelf should still list all authors. In projects where space is an issue (such as in a Table but not in the References), it is acceptable to use one author with et al.

Unlike titles in boxes, tables, and figures that use a capital in the first word after a colon, the part of the title in a reference after a colon is not capitalized (unless it is a proper noun, designation, or other form that takes a capital letter anyway).

Use the NLM style for format. See the styles below for examples of this format. See also [Sample PubMed Central Citations - XML Tagged](#) and select the “untagged version” from the page. For PDFs of the recommended formats, see [Recommended Formats for Bibliographic Citation](#) and [Supplement: Internet Formats](#).

Books

Single author--

1. Salton G. Automatic text processing. Reading (MA): Addison-Wesley; 1989.

More than one author—

2. Sankoff TF, Kruskal JB. Time warps, string edits and macromolecules: the theory and practice of sequence comparison. Reading (MA): Addison-Wesley; 1983.

With editors—

3. Altschul SF. Evaluating the statistical significance of multiple distinct local alignments. In: Suhai S, editor. Theoretical and computational methods in genome research. New York: Plenum Press; 1997. p. 1-14.

Chapter from a book—

4. Mizrachi I. GenBank: the nucleotide sequence database. In: McEntyre J, editor. The NCBI handbook [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2002 Oct. Available from: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books>

Journals

Single author—

1. Lipman DJ. Making (anti)sense of non-coding sequence conservation. *Nucleic Acids Res* 1997 Sep 15;25(18):3580-3583. Review.

More than one author—

2. Benson DA, Boguski MS, Lipman DJ, Ostell J. GenBank. *Nucleic Acids Res* 1997 Jan 1;25(1):1-6.

Internet

1. Patrias K. National Library of Medicine recommended formats for bibliographic citation. Supplement: internet formats [Internet]. Bethesda (MD): National Library of Medicine; 2001 Jul [cited 2001 Aug 8]. 106p. Available from: <http://www.nlm.nih.gov/pubs/formats/internet.pdf>

Tables

The word “Table” should be capitalized and spelled out, followed by an arabic number and a period. Give the table a brief title, followed by a period. If the title contains a colon, capitalize the word after the colon (unless it is a designation or abbreviation that is shown in lowercase). Additional descriptive material may be given in a paragraph below the title. Use superscript, lowercase, italic letters for footnote citations. Change abbreviations such as “n.d.” (“not done”), “n.s.” (“not significant”), and “n.a.” (“not applicable”) to ND, NS, and NA, respectively.

Example:

Table 1. ID system: Products.

Type	Source	ASN.1	GBFF ^a	Qscore	GenPept	Protein FASTA
Cumulative	GenBank	X		X	X	X
Incremental	GenBank	X		X	X	X
Incremental	GenBank ^b		X	X		
Cumulative	RefSeq	X	X		X	X
Incremental	RefSeq	X	X		X	X

^a GBFF, GenBank flatfile.

^b NCBI records only.