

## **Some style and grammar tips for biostatistics and statistics students.**

**Rod Little**

**Revised January 30, 2013**

I have read many drafts of theses and papers written by Biostatistics and Statistics students and others, and certain recurring issues of English style arise. Below are some suggestions, offered to assist future writers and make life easier on future readers. Style is to some extent subjective, so the suggestions below are certainly not “theorems” (though some, like using a spell-checker, are close). Also, a frustrating characteristic of the English language is that there are always exceptions to any rule. Some of the problems discussed below are particularly prevalent amongst non-native English speakers, for whom the problem of grappling with the statistics is compounded by the need to find ways of expressing things in a strange language. You have my genuine sympathy. But whether native speaker or not, developing a good style should be a high priority, since it is an important ingredient for success in academia or industry. The best way to improve is for non-native speakers to speak English, and for all to read widely and practice writing at every opportunity.

I plan to update this memorandum from time to time, and welcome comments and suggestions for additions or corrections. This memorandum is nowhere near comprehensive. Good sources on English style include “Style, the Basics of Clarity and Grace”, 4th edition, by Williams and Colomb (2006), and the “Chicago Manual of Style”.

**1. Draft early and often.** If you have done some work, or have an idea, write it down in a memorandum – to someone else who is interested (like your adviser), or otherwise to yourself! Don’t keep work in your head until the end and then write it all out in one go. There are many reasons for this: the drafting of memoranda is a way of practicing writing, and it improves ideas by forcing clarity. An early draft can be continuously refined in a word processor, allowing wrinkles to be ironed out. Memoranda can form the bases for longer articles and eventually a thesis, so you have something to start with when writing them. Writing a thesis is an arduous task and should be spread out as much as possible, rather than waiting until the end.

**2. Less is more!** Scientific writing should be clear and concise. When you write anything, go over it carefully and attempt to reduce the length by 20% without reducing content, simply by more concise wording, removing adjectives that do not add anything, and avoiding repetition. Strive to say things just once clearly, rather than several times unclearly. (Avoid “Hmm, that didn’t come out quite right, let’s write it again another way...!”) Some repetition is useful in lectures and more leisurely expositions, but for journals like JASA or Biometrics or JAMA space is at a premium and repetition should be avoided. The result of this will be both shorter and better style.

When you have edited the piece once, go over it again and try to shorten it by another 10%!

Here are two examples:

Original: “The repeated-measures design has been recognized to permit direct study of change over time within individuals, and thus considered to be a better approach than the cross-sectional study design in estimating effects of covariates.” (36 words)

Edited: “Unlike cross-sectional designs, repeated-measures designs permit direct study of individual-level changes over time, and effects of covariates on these changes.” (23 words)

Original: “In this section we apply our proposed models to the analysis of the data with compliance as the outcome of interest. The main objective of this analysis is to investigate how treatment preference and assignment preference affect the compliance status.” (40 words)

Edited: “In this section we apply our proposed models to the data with compliance as the outcome. The objective is to investigate how treatment and assignment preferences affect compliance.” (28 words)

**3. Know your audience.** Before starting to write, make sure that you have a very clear idea of who is your intended audience, what you believe they already know, and what you think they might not know. Students often overestimate what the actual audience knows about the topic, since they are too close to what they are writing, and their initial level of interest. As a result, more information on setting and motivation is often needed.

**4. Avoid a pompous wordy style.** Much scientific writing is pompous, complex and wordy, as if this makes the author seem clever and the ideas important. Wrong! People love clear and direct writing.

**5. Avoid a chatty conversational style.** On the other hand, scientific writing is not like a dinner-time conversation. Non-natives who took classes in conversational English sometimes bring too much of that style to their scientific writing. This should be relatively formal, avoiding phrases like “Now let’s prove this theorem ...”, “the estimate worked pretty well”, “the anemia profile differed a lot for infants...”, etc.

**6. Use active rather than passive verbs.** A major feature of pompous, bad style is to distort the sentence order and use passive rather than active verbs, apparently thinking that this somehow sounds more objective and scientific. Try to use an active verb, and avoid starting a sentence with long adjectival clauses, as is common in legal documents. Thus

“The author calls this a prior likelihood approach” (8 words), NOT “This approach is referred to by the author as a prior likelihood approach” (13 words)

“We choose a prior for  $f(t)$  similar to that used by Wahba (1978)” (13 words) NOT

“based on a similar idea in Wahba (1978), we choose the same prior for  $f(t)$ ” (15 words). (It was pointed out to me that “used by Wahba” is better than “in Wahba”, since the reference should be to the person, not the paper).

This raises the “I” vs. “we” question. Personally I am happy with “I” for single-authored works, but some editors seem to prefer the royal “we”. I am not sure there is a clear consensus on this. For grants, avoid I and we – third person is more accepted.

**7. Choose specific active verbs over generic verbs like “to do”. E.g.**

“The random-effects distribution has not yet been *modeled*” nonparametrically” NOT

“Modeling the random-effects distribution nonparametrically has not yet been *done*”.

**8. Don’t pack too many ideas in one sentence, and put the most interesting part for the reader at the end (the “stress position”).** It is generally wise not to have more than one important idea in a sentence. We tend to expect the “pay-off” or more interesting idea to be at the end of the sentence; if it is at the beginning, the rest of the sentence tends to be a let-down and may bore the reader. See <http://www.research.att.com/~andreas/sci.html> for more discussion of these points.

**9. Avoid repeating the same word in a short space.** This can usually be avoided by rephrasing.

**10. Seek a smooth flow of ideas, rather than jumping between topics.** Review what you have written and check that there is no jumping around between topics without good reason. If A1 and A2 are on the same or similar topics, and B1 and B2 are on the same or similar topics, then order as A1A2B1B2 rather than A1B1A2B2.

**11. Don’t switch tenses in the same paragraph.** For example, not:

“After a trend in time was determined, both time-varying and time-stationary covariates will be added to the model.” Rather

“After a trend in time was determined, both time-varying and time-stationary covariates were added to the model.”

In general I use the present tense to describe work rather than the past tense, though past tense is not wrong.

**12. Use a spell/grammar checker!** Typos are sloppy and many can be avoided using computer technology. If you are using Latex and don’t have a good spell-checker, read the text into a mainline word processor like Microsoft Word, and use the spell-checker in that program to pick up spelling and grammar problems in the pure text parts. Then save as a text file and read it back into Latex. You still need to check for typos – a spell-checker is not perfect.

**13. Think hard about the best choice of notation, and make sure *all* notation has been carefully and consistently defined.** Use a single consistent notation throughout – if covariates are  $x$  in section 1, they should still be  $x$  in section 5, not  $z$ . It takes work to get this right!

**14. Use section and equation numbers liberally and refer to them in the text.** Numbering is a precise way of guiding readers through the paper.

**15. Match singular/plural for subjects and verbs:**

“random effects are assumed normal”, not “random effects is assumed normal”.

“in this work” or “in these works” NOT “in these work”.

“data are”, not “data is” (note that the word “data” is plural).

**16. Articles for singular/plural nouns.** Singular nouns generally require a definite or indefinite article, “the” or “a”; plural nouns generally do not. (I know, this drives our Asian friends crazy!)

“lack of an explicit form for” NOT “lack of explicit form for”.

“lack of explicit forms for” NOT “lack of the explicit forms for”.

“it is difficult to evaluate the posterior loglikelihood of ...” NOT “it is difficult to evaluate posterior loglikelihood of...”

**17. Prepositions.** (I’ll think of some more for later drafts).

“distribution for  $X$ ” not “distribution on  $X$ .”

“denote by  $X$  the random variable” or “let  $X$  denote the random variable” NOT “denote  $X$  the random variable.”

**18. Don’t begin a sentence with a number.**

For example, start a sentence “Eighty percent of mothers...” NOT “80% of mothers ...”

**19. Miscellaneous pet peeves**

“The interesting parameters” or better “the parameters of interest” NOT “the interested parameters” (parameters are inanimate objects and do not have interests like reading or playing soccer!)

The word “data” is plural.

“criteria” is a plural noun. The singular is “criterion”.

“which” starts a nonrestrictive (descriptive) clause, “that” starts a restrictive clause. Often “which” is used when “that” would be more precise. Hint for native English speakers: if “that” sounds ok it is probably right. E.g.

“hence  $d$  is the maximum likelihood estimate, ~~that~~ which has the useful property of being asymptotically efficient.”

“Of these two estimators, the one ~~which~~ that satisfies the order restriction is better.”

“Alternative” (e.g. “find an alternative phrasing”) is often contracted to “alternate” (“find an alternate phrasing”), but “alternate” has a different meaning, namely going to and fro, as in “alternating current.”

“well” vs. “good” In informal speech, Americans often use the adjective “good” when correct usage is the adverb “well”. Hence “this was a good procedure” is OK, but “the procedure performed well”, NOT “the procedure performed good.”

The founder of modern statistics is R.A. Fisher, not R.A. Fischer.

"Future work" not "Future works". "Work" is here a collective noun, and the future research is collected into a singular collective term "work" not the plural "works". In the same vein, we say "in future research we plan to study ..." not "in future researches we plan to study"

“its” (no apostrophe) for possessives, “it’s” (with apostrophe) for contractions of “it is”. For example, “Although maximum likelihood is optimal asymptotically, ~~its~~ it’s not so good in small samples.” “Maximizing the likelihood with respect to ~~it’s~~ its parameters yields ...”.

Punctuation is a whole other story; for an amusing and informative read see “Eats, Shoots and Leaves: the Zero Tolerance Approach to Punctuation” by Lynne Truss.