Writing papers for publication

Neil Ferguson
Contents

■ Why write papers?
■ Basic conventions
■ Style: Precision, clarity & economy
■ Planning a paper
■ Picking your target journal/audience
■ Journal types
■ The submission/reviewing process.
Engage the reader, by telling a story!
Why write papers?

- Scientific research is a shared process of discovery and communication.
- Journal articles are the standard medium for such communication.
- They should enable your work to be:
  - Understood
  - Critiqued
  - Reproduced
  - Extended
- While good papers can give personal recognition, this should not be the primary goal!
Basic paper structure

- Title
- Abstract
- Introduction
- Methods (*sometimes at end*)
- Results
- Discussion
- Conclusions
- Acknowledgments
- References
- *Figures & Tables*
A good paper...

...tells a story

Anything can be made interesting with sufficient work
Title

- Make your title specific enough to describe the contents of the paper, but not so technical that only specialists will understand. The title should be appropriate for the intended audience.

- The title usually describes the subject matter of the article: “Effect of Smoking on Academic Performance”

- Sometimes a title that summarizes the results is more effective: “Increased transmissibility explains the third wave of infection by the 2009 H1N1 pandemic virus in England”.

- Optimally, by reading the title, the work being reported should be clear to the reader without having to read the paper itself

- e.g.: "The Effects of Light and Temperature on the Growth of Populations of the Bacterium, Escherichia coli".

- This title reports exactly what the researcher has done by stating three things:
  - 1. The environmental factors that were manipulated (light, temperature).
  - 2. The parameter that was measured (growth).
  - 3. The specific organism/system that was studied (e.g. the bacterium, Escherichia coli).
Abstract

The abstract section in a scientific paper is a concise digest of the content of the paper.

An abstract is more than a summary:
- A summary is a brief restatement of preceding text that is intended to orient a reader who has studied the preceding text.
- An abstract is intended to be self-explanatory without reference to the paper, but is not a substitute for the paper.

Should present the purpose of the paper, the most important, general materials and methods, summarized results, and the major conclusions.

Do not include any information that is not contained in the body of the paper. Exclude detailed descriptions of materials and methods.

Tables or figures, references to tables of figures, or references to literature cited usually are not included.

Often written last - extract the most important points from each section of the paper and then use these to summarise.
Introduction

- The Introduction is the statement of the problem that you investigated.
- It should give readers enough information to appreciate your specific objectives within a larger theoretical framework.
- After placing your work in a broader context, you should state the specific question(s) to be answered.
- Can include background information/past research and how your work aims expands the knowledge in this general area.
- Need to explain why the research was done, relating this research to other relevant work, justifying the hypotheses to be tested.
- Go from the general, theoretical framework to your specific question, without being too broad. Present only the most relevant ideas.
- Except in special cases (generalist journals), assume that the reader is at least moderately familiar with the general subject of the paper.
- All background information gathered from other sources should be appropriately cited.
Methods

- Explains how and, where relevant, when the study was done, study design, methods of gathering data and statistical methods/mathematical models used to analyse the data.

- For epidemiological studies, need to describe the study population, and inclusion/exclusion criteria. Study end-points and measured variables also need to be carefully described. Mention relevant ethical issues if human or animals subjects were used.

- Should be detailed enough so that any reader knowledgeable in the subject area could duplicate the study.

- If you used a complicated protocol/model, it may helpful to include a diagram, table or flowchart to explain the methods you used.

- If new models or statistics were developed, describe them precisely, using equations as well as explanatory text. Cite relevant past work.

- Include any preliminary results that were used to design the main experiment that you are reporting on. ("In a preliminary study, I observed the owls for one week, and found that 73% of their locomotor activity occurred during the night, and so I conducted all subsequent experiments between 11 pm and 6 am.").
Results (1)

- Present *summarized* data/analyses using narrative text, tables and figures.

- Only the results are presented. Little interpretation and no conclusions are usually given.

- Data/analyses are summarised in tables and/or figures to supplement the text and to present the data in an easily understandable form. Raw data rarely presented.

- If tables and/or figures are used, they must be accompanied by narrative text. Don’t repeat data shown in tables and figures. But, don’t just make passing reference either (e.g. "Results are shown in Table 1." is not appropriate.).

- The text calls attention to the important results that the researcher will refer to in the Discussion & Conclusions.
Results (2): figures & tables

- Figures and tables should help the reader to understand more easily than a written description. They should not duplicate text, and text should only highlight the major points to be noted on graphs/tables.

- The same data should not be presented twice, so decide which form helps you tell your readers what you want them to know. All figures and tables must be referred to before they are explained in the text.

  – **Bad example**: "The results are shown in Graph 1." This is not a summary of the results, and the graphs should be referred to as "Figures."

  – **Another bad example**: "Growth rates under low fertilizer had an average of 3.2 g with a 95% confidence interval of ± 1.4 g, and growth rates under high fertilizer treatment had an average of 6.5 g with a 95% confidence interval of ± 1.0 g (see Figure 1). Since the intervals do not overlap, they are significantly different at the 0.05 level." The text here repeats specific information that is also shown in the figure, & it is not necessary to explain why non-overlapping confidence intervals indicate a significant difference.

  – **Good example**: "We found that higher levels of fertilizer resulted in significantly larger growth rates, as determined by the 95% confidence intervals (Figure 1)."
Results (3): graphs

- Use a graph to illustrate a relationship or pattern in your data.

- Be sure that the type of graph you choose is appropriate for the type of data you wish to display. What assumptions are you making by using a particular type of graph? e.g. a line graph with the points connected indicates that the variables are continuous over the range displayed.

- The axes of a graph must have clear, concise labels. If there is more than one line or bar on the graph, each must be clearly identified.

- All figures must have clear and specific legends. A legend is usually written as an incomplete sentence with only the first word capitalized - "Figure 1. The . . ." If required for clarity, you may include several more sentences. It should be placed below the figure.
Results (4): good and bad graphs

- Reader can’t tell from the figure whether the bars are mean values, and whether the difference between them is significant. There needs to be some indication of the variation within each treatment group, i.e. error bars. Note as well that the Y axis has no units and the figure legend is not explanatory.

- Better version of this graph.

Figure 1: The impact of treatment on the frequency of clinical episodes over 12 months of follow-up. Mean and 95% confidence intervals are shown.
More on figures

- Often need to use multi-part figures, making optimal use of limited space.

- Important to group related material together – e.g. data in one figure, parameters estimates in another, predictions in another.
Results (5): tables

- Tables are used to present matrices of data or results. If it is important to show a pattern or trend, use a graph instead of a table.

- Do not present raw data and expect your reader to do the arithmetic before s/he can understand the contents of a table.

- Try to avoid large tables - no one will read through them. Perhaps the information can be presented better in several smaller tables.

- All tables must have legends that explain their contents sufficiently that they can stand alone (much like figures). It is sometimes appropriate to have footnotes for a table.

<table>
<thead>
<tr>
<th>Table 1 Projections for the eventual total number of vCJD cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean i.p. (years)</td>
</tr>
<tr>
<td>&lt; 20</td>
</tr>
<tr>
<td>2,900</td>
</tr>
<tr>
<td>5,500</td>
</tr>
<tr>
<td>30–60</td>
</tr>
<tr>
<td>≥ 60</td>
</tr>
</tbody>
</table>

The range of total number of cases is given that is consistent with the number of cases recorded in 2000 and with the average annual number of cases in the next three years, calculated by the conditional mean incubation period (i.p.). Results include concentrated sampling of parameter combinations that result in large epidemics.
Discussion

- The discussion should interpret and explain the meaning of your results and usually proceeds from the specific to the general.

- Begin with a summary of your results in a sentence or two. Remind the reader of important trends, etc.

- Don’t forget that "negative" results can be important too, since they may suggest that your hypothesis was incorrect.

- Some relevant questions to address (in order):
  - How do these results relate to the original question? Do the data support your hypothesis?
  - Are your results consistent with what other investigators have reported? If your results were unexpected, try to explain why.
  - Is there another way to interpret your results?
  - What further research would be necessary to answer the questions raised by your results?
  - How do your results fit into the big picture?
Acknowledgements

Acknowledgments are used to thank any persons who contributed to the study, but at a level not warranting authorship.

Such contributions include, but are by no means limited to:

- help in study design, provision or analysis of data
- help in preparing aspects of the manuscript
- critiquing a draft of the manuscript.
- financial support of the work.
References

- Poor referencing can alienate reviewers and readers.
- There is no excuse for poor literature reviewing. Before starting a research project (never mind a paper), you need to thoroughly review the existing literature to avoid reinventing the wheel.
- Citation is a learned skill – rarely possible to cite everything in the literature related to a topic, so have to select the key works (primary data publications, first papers developing a relevant method/model).
- Some self- or group-citation inevitable, but try to be fair & include citations to other groups even if differences of opinion/emphasis exist.
- Do not cite purely on the basis of an abstract (i.e. read papers, if only briefly).
- Follow the journal style for references.
- Use bibliography database software (e.g. Endnote, Reference Manager, BibTeX).
Plagiarism

- Never copy & paste text from other papers.
- Don’t use other people’s text to construct a draft with the intention of changing it later.
- If you want to quote, type out quote (in quotes) and reference immediately.
- Journals check for plagiarism, and Imperial treats it as serious scientific misconduct.
- Ignorance of the rules or “I forgot”, “I didn’t mean to” not an excuse.
- Consequences can be severe for a scientific career.
Style: precision

- Scientific writing must be accurate. Make sure you say what you mean.
  Instead of: The rats were injected with the drug. (was a syringe was filled with drug and ground-up rats and both were injected together?)
  Write: Drug was injected into each rat.

- Be careful with commonly confused words:
  Temperature has an effect on the reaction.
  Temperature affects the reaction.

  Less food (can't count numbers of food)
  Fewer animals (can count numbers of animals)

  A large amount of food (can't count them)
  A large number of animals (can count them)
Style: economy

- Use verbs instead of abstract nouns
  Instead of "take into consideration", write "consider"

- Use strong verbs instead of "to be"
  Instead of: "The enzyme was found to be the active agent in catalyzing...", write: "The enzyme catalyzed..."

- Consider using shorter words (however variety is also good).
  
<table>
<thead>
<tr>
<th>Instead of:</th>
<th>Write:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possess</td>
<td>have</td>
</tr>
<tr>
<td>sufficient</td>
<td>enough</td>
</tr>
<tr>
<td>utilize</td>
<td>use</td>
</tr>
<tr>
<td>demonstrate</td>
<td>show</td>
</tr>
<tr>
<td>assistance</td>
<td>help</td>
</tr>
<tr>
<td>terminate</td>
<td>end</td>
</tr>
</tbody>
</table>

- Use concise terms.
  
<table>
<thead>
<tr>
<th>Instead of:</th>
<th>Write:</th>
</tr>
</thead>
<tbody>
<tr>
<td>due to the fact that</td>
<td>because</td>
</tr>
<tr>
<td>in a considerable number of cases</td>
<td>often</td>
</tr>
<tr>
<td>the vast majority of</td>
<td>most</td>
</tr>
<tr>
<td>during the time that</td>
<td>when</td>
</tr>
<tr>
<td>in close proximity to</td>
<td>near</td>
</tr>
<tr>
<td>it has long been known that</td>
<td>I'm too lazy to look up the reference</td>
</tr>
<tr>
<td>In other words</td>
<td>just in case you were too dim to understand</td>
</tr>
<tr>
<td></td>
<td>first time</td>
</tr>
</tbody>
</table>

- Use short sentences. A sentence made of more than 40 words should probably be rewritten as two sentences.
Style: clarity

- Tense: a paper is a report about something that has been done in the past:
  - Most of the paper should be written in the past tense.
  - The present tense is used when stating generalizations or conclusions.
  - Present tense is mostly used in the Introduction, Discussion and Conclusion.

- Use the active voice. It's clearer and more concise than the passive voice.
  Instead of: An increased appetite was manifested by the rats and an increase in body weight was measured.
  Write: The rats ate more and gained weight.

- Use the first or second person in place of passive voice.
  Instead of: The samples were analyzed (by God?).
  Write: We analyzed the samples

- Avoid dangling participles.
  Instead of: "After incubating at 30 degrees C, we examined the petri plates." (You must've been pretty warm in there!)  
  Write “We incubated the petri plates at 30 degrees C, then examined them.”

- Do not use jargon where regular language will do just to make your writing seem more technical, scientific, or academic.
Many graduate students initially struggle with the formal, grammatically precise style of journal article writing.

Key issues:

- Incorrect punctuation (where to place commas, the use of colons & semi-colons).
- Incorrect use of word types (adjectivising nouns, misuse of pronouns, prepositions & conjunctions).
- Poor sub-clause construction.
- Unambiguous subjects & objects.
- Plurals (e.g. ‘data are’ NOT ‘data is’).
- Use of slang or contractions (DON’T).
- Use of undefined abbreviations.
Style: use of style manuals

- Frequently used in US and by non-native English speakers.
- The best can be very useful in giving insight into the process of writing and how to make a paper flow naturally for the reader.
- Many on-line resources (some of which have been shamelessly plagiarised here!): e.g:
  - http://www.me.vt.edu/writing/
  - http://faculty.uca.edu/~jmurray/BIOL4425/lab/reports/guide/
  - http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWgeneral.html

- Books:
Planning a paper

- When you first start a writing project, make an outline of the major headings.
- List the key ideas to be covered under each heading.
- Organize your thinking and the logic of your arguments at this level, not when you are trying to write complete, grammatical, and elegant sentences.
- Separate out the three tasks of:
  - figuring out what you want to say – **tell a story**
  - planning the order and logic of your arguments
  - creating the exact language in which you will express your ideas.

- It can be useful when making an outline to attach page lengths and time lines to each subsection.
  - For instance, section 2.4 may be “Prevalence estimates for pre-school children”
    To this you might append, '3 more days analysis, 4 days writing; 10 pages.'

- Such time estimates are usually inaccurate, but the process of establishing them is quite useful.
What to include

Key to a good paper is focus:

– a large research project (e.g. Ph.D.) will typically produce multiple (2-3) papers.

– trying to include too much content makes papers diffuse, as the coherence of a single message can be lost.

– chopping up a project too much (‘salami-slicing’) causes papers to lose substance - may temporarily boost a publication list, but doesn’t enhance long-term reputation.

– publishing the same work in multiple journals is bad practice/unethical.

In epidemiology, a paper usually consists of

– results of a single epidemiological study.

– development of a new statistical/mathematical method.

– application of a mathematical/statistical model applied to a certain disease system and population.

– a meta-analysis.
Authorship

- Disputes over authorship occur frequently and can cause bitter and long-lived divisions.
- Best to be clear about expectations from the start - both in terms of who is likely to have authorship, and author order.
- Typically a PhD student will be first author on papers arising out of work in their thesis.
- Typically supervisors will go last. If you feel your supervisor has not contributed sufficiently to warrant publication, discuss the issue before removing them!
- In larger or longer projects, additional people may become involved. As they do, discuss their expectations as to authorship at an early stage.
- ‘Honorary’ authorship is increasingly discouraged (i.e. just because someone is head of dept./group or a colleague of your supervisor does not mean they should be an author). All authors should have made a substantive contribution to any paper. Some journals ask for contributions to be listed.
Journal types: generalist

- Nature, Science, PNAS, Nature Medicine, The Lancet, NEJM, PLoS Medicine, ELife...

  - **Advantages:** mostly high impact, very widely read, can have significant benefit to your career.

  - **Disadvantages:** very hard to get published in the top journals. Extreme space limitations (to the extent that Nature/Science/PNAS articles are rarely subdivided into Introduction/Methods/Results/Conclusions).

  - Worth trying, *provided* your work has come up with a highly original or interesting result. Topicality helps.

  - The art of writing good Science/Nature papers is very different from writing good papers for specialist journals. Readers potentially from a wide range of disciplines. Avoid assuming specialist knowledge.

  - Easiest if there is a single key result you’re wanting to convey, not a complex set of interrelated concepts/results (but there are exceptions).

  - Don’t go over the top – e.g. ‘innovative and novel approach’ – novelty and interest is assumed in these journals.

  - Not easy to get methodological papers published.
Journal types: specialist

- Normally a large number in any discipline, but usually 1 or 2 that stand out for quality/impact factor (e.g. PLoS Pathogens, Journal of Infectious Diseases, Epidemiology, Am. J. Epidem.).

- Can look up Journal impact ratings on Web of Science – http://wos.mimas.ac.uk/

- Also worth looking up how long papers take to get published – some stats journals particularly bad.

- Pick the right journal for your work – don’t always go for the highest impact journal on principle.

- If the work would benefit from a more extended, discursive style, or is theoretical/methodological, then lower profile but high quality journals may be the best option.

- A good rule of thumb is to pick the journals you most like reading papers from.

- Always read the ‘Information for Authors’ before starting on the paper.
Open Access

- Increasing numbers of journals are Open Access (or ‘Author pays’) – PLoS, BMC but also many others.

- Popular in many biomedical fields (e.g. global health) – leading to some journals rapidly gaining impact (e.g. PLoS).

- Many funders (Wellcome, NIH) mandate that papers need to be freely accessible - subscription journals manage this through ‘author copies’, but not an issue for Open Access.


- Some Open Access journal completely generalist – PLoS One – publishing anything which is scientifically robust, in any field.

- Many commercial journals now give an open access option – at a cost.

- Note open data rules – PLoS now demands all data is released.
The publication process

- Identify what you want to report.
- Pick target journal & read Information for Authors for that journal.
- Prepare draft manuscript in line with Information for Authors.
- Show to collaborators/colleagues/supervisor for comments/edits.
- Edit, edit & edit again!
- Where appropriate, identify 3 or 4 people you’d be comfortable to nominate as reviewers.
- Submit paper (nowadays, often electronically), with covering letter.
- Paper will then be considered by journal editors:
  - some journals reject >50% papers immediately (e.g. Nature).
  - most journals send (nearly) all papers out for review.
  - depending on reviewers’ comments, paper will be accepted outright (rare), revisions will be invited (common), or be rejected outright (common).
The submission letter

Important – need to convey in lay language (many editors are professional publishers, not research scientists):

- main results & why these are interesting.
- why the paper is suitable for the journal.
- any suggestions for reviewers.

Re: Implications of BSE infection screening data for the scale of the British BSE epidemic and current European infection levels.

This paper details the first epidemiological analysis of infection screening data for BSE in UK cattle. To allow simultaneous fitting to both passive surveillance data and clinical cases, the analysis required considerable development of the back-calculation methods used in earlier studies of the BSE epidemic. Sensitivity analyses were conducted to examine possible mechanisms that could give rise to the apparent under-ascertainment of clinical cases indicated by the screening data. The results indicate that 2 to 4 fold more animals were infected than previously estimated. We also estimate the incidence of infection in cattle born between 1993 and 1997 in other European Union countries using data on testing of cattle slaughtered for consumption.

We believe these results are of importance to scientists and policy-makers working in the field, and be of broad interest to the wider stakeholder community and general public, both in the UK and overseas. The results are robust, with rigorous statistical methods having been used to calculate confidence bounds on all estimates.

Potential referees include xxxxxxxx at the London School of Hygiene and Tropical Medicine, yyyyyyyyy at Warwick University, and zzzzzzzz at UCL.

Yours,
Responding to reviewers

- Do not treat the reviewing process as a battle.
- Thank the reviewers for their comments.
- Identify:
  - which criticisms are valid – can they easily be dealt with, or do you need to substantially modify the paper/study? Can you deal with the comments by stating assumptions/limitations of your work more clearly in the text?
  - which criticisms reflect a misunderstanding of your work – modify the text to be clearer in those areas.
  - which criticisms reflect differences in scientific point of view – acknowledge the differences (perhaps in the text), but if you think you’re right, give arguments for sticking to your guns!
  - which criticisms are completely unjustified – be robust if absolutely necessary, and if a review seems absurdly biased discuss with co-authors the best strategy (e.g. appealing to the journal editor for an additional review to be sought).
Conclusions

- Planning is key: know the story you want to tell before you start writing.
- Learn precision, clarity & economy in your writing style.
- *Think* about the optimal & clearest way to present results (text, figure or table).
- Pick the journal and know your audience.
- Clarify authorship at the earliest possible stage.
- Follow the Information for Authors.
- Know the literature to place your work into the bigger picture.
- Papers always need redrafting – get feedback from colleagues.
- Take reviewers/editors comments seriously but not personally.