Introduction to Scienific Writing

Lennon Ó Náraigh [1]

[1] School of Mathematics and Statistics, University College Dublin, Belfield, Dublin 4, Ireland

27th February 2023

Aim of Talk

A brief overview of how to write a scientific report. It applies to writing your final reports but also, to writing a scientific paper more generally.

What this talk is not:

- Why we write;
- Why we use the Scientific Method;
- Writing ethically.

But see the website for some background materials (YouTube links).

Scientific writing framework

There is a simple framework for reporting on the results of an experiment or a piece of modelling[1]:

- Research Question under investigation
- Methodology
- Results
- Discussion and Conclusions

You should have this flow in the back of your mind. The structure of a scientific report will be based on this framework, but the sections of the report might hot have exactly these headings.

Not all-inclusive, e.g. review articles.

[1] Gastel, B. and Day, R.A., 2022. How to write and publish a scientific paper. ABC-CLIO.

Plan of Talk

- Go through the framework for Scientific Writing
- Talk about style
- Look at implementation of the framework (e.g. splitting ‘theory’ and ‘methodology’, where to put a literature review...)
- Look at good and bad writing practices

Research Question

The research question should be clear and amenable to a quantitative or a binary answer.

- A fair supervisor in a 4th-year module will provide a good RQ to the students;
- In academic research it is not always so clear cut, sometimes a good RQ comes after messing around for a long time.

Example:

> *Does an SEIR model provide a good fit to the data in case of the outbreak of Covid 19 in Ireland (first wave, March 2022-July 2022)?*

- Amenable to a yes/no answer.
- An answer can be generated using quantitative methods – e.g. compare SEIR to other models and make a ranking. Good data are available.
- Precise – the RQ is time-bound.

The RQ does not care about your feelings.

If you get a different answer to the RQ than you expect, you should:

- Check your methodology, coding, etc.
- Double-check the literature to see if what you are seeing is consistent with other studies.

If the answer stays the same after these checks, and you feel disappointed that things have not worked out the way you expected, that is still fine, as the RQ does not care about your feelings. Also,

- A null answer is fine in a report;
- In the context of academic research, a null answer can sometimes be salvaged as a 'methodology paper'.
- 'Positivity Bias'

Methodology

A description of the methods you use to answer the RQ. Low-level things:

- EDA,
- Computational tricks (what tricks you needed to use to generate convergent solutions to your ODE model)

High-level things:

- Choice of model, justification for same.

Innovative things:

- A novel model, requiring a theoretical formulation.

Methodology – Pitfalls

- Time spent on research: about 80% methodology and 20% results.
- Writing about research inverts this ratio.

Reason:

- It's a 'given' that you can use a Statistical Model or do R or Python programming, so in-depth discussions about these thing should be left out.
- If you have discovered some new computational trick, it can be included in an Appendix (or a GitHub repository).

Results

Assuming all goes well, results will help to answer the RQ. The answer should be written up in a persuasive way. Back to our example:

> *Does an SEIR model provide a good fit to the data in case of the outbreak of Covid 19 in Ireland (first wave, March 2022-July 2022)?*

If the answer looks like a 'yes', then the results section should include, e.g.

- Time series of data compared with model;
- Table showing RMSE of SEIR model compared with other models;
- Tables showing CIs of SEIR model compared with other models

It is not enough to shove tables and figures at the reader: the accompanying text should be a narrative, with the tables and graphs as supporting evidence.

Results – Pitfalls

Some of the following phrases[2] are fine if they can be justified in the Methodology section, or if further explanation is provided. Some of the other phrases are downright suspicious. Bottom line: Tell the truth!

What is said	Do you really mean this?
It has not been possible to provide definite answers to the questions	An unsuccessful experiment, but I still hope to get it published
Three of the samples were chosen for detailed study	The other results didn't make sense
In my experience	Once
In case after case	Twice
Correct within an order of magnitude	Wrong
Much additional work is required to get a complete understanding	I don't understand
We hope that this study will stimulate further investigation in the field	I quit

[2] https://www.inf.unibz.it/∼calvanese/teaching/2020-03-PhD-RM/RM-2020-U1-calvanese.pdf

Discussion and Conclusions

Discussion:

- Anticipate questions or queries that an intellectual opponent (grader / peer reviewer) will have and answer them pre-emptively.
- Test results for robustness.
- Explore other relevant information.

E.g. (SEIR model):

- Re-do analysis with various data fitting methods (NL-LS/ML), check if results are the same;
- Re-do analysis with various case-count models (Gaussian, Poisson), ...
- Check if results are robust to frequency of case reporting (daily / weekly), ...

The RQ often lives to fight another day:

- Is there a better model out there which will provide a better fit to the data?

After 'stress-testing' your results, it's best to wrap up with some definitive conclusions.

Questions of Style

- The Scientific Method is humanity's best effort to overcome its cognitive biases (group affinity, availability heuristic, confirmation bias, ...), to understand the material world around us.
- As such, it is meant to function as a *bulls..t detection machine*.
- Therefore, there is no place in scientific writing for bulls..t.
- Clear writing is the order of the day.
- Don't try to gloss over points of uncertainty with flowery language; write what you mean, mean what you write (writing ethically).
- No harm to read a style guide once in your life.[3]

[3] Strunk Jr, W. and White, E.B., 2007. The Elements of Style Illustrated. Penguin.

Implementation of Framework

It is not always the case that a scientific report will have four sections (RQ/Methodology/Results/DC). But this logical structure will be in evdience in the background. A typical scientific paper (in Applied Maths) will involve the following elements:

1. Title
2. Abstract
3. Introduction
4. Theoretical Formulation
5. Methodology
6. Results 1
7. Results 2, ...
8. Discussion and Conclusions
9. Appendices involving detailed validation of methods.
10. Bibliography

Title and Abstract

The title should be succinct and address the RQ; the title

- should not be too short or too generic;
- should not be too long or specific;
- should not contain waste words (study on ..., results on ..., observations on ...).
- should not contain abbreviations or jargon.

The abstract is a concise, self-contained summary of the findings in the report. It is usually the last thing to be written. The abstract

- should be a single paragraph;
- should use the past tense, since it refers to work already done;
- should not contain: minor details, paper structure, acronyms, abbreviations, mathematics, citations.

What goes in the introduction

The title, abstract, and Introduction are part of *framing the RQ*. In the Introduction, very brutally, you are:

- Writing down the problem statement (which overlaps with the RQ);
- Reviewing previous solutions;
- Saying why the previous solutions are bad and why our new solution is better.

As part of the Introduction, you can include a literature review and a plan of the report / paper. Often, it is best to use sub-headings for these.

Literature Review

In the literature review, you are *placing the work in the context of the existing literature*.

- A summary of key papers in the area, the methodology they use, and the findings obtained.
- Explanation of how this 'sets the stage' for your own efforts.
- The emphasis here should be on highly-cited key papers. For a report, key textbooks may also be used as sources.

Results 1, Results 2, Discussion and Conclusions

Self-explanatory, but:

- Sometimes it can be good to break up the 'results' into different sections. Specific discussion can be included in these sections.
- This leaves some high-level discussion and conclusions to the final section, including:
  - A concise statement of the report's important results;
  - Stating or reiterating any limitations to the report and its methodology;
  - Looking to the future, to sketch out other problems that were not addressed, questions that were not answered, or variations that could also be explored.

Appendices and Bibliography

- Appendices can be used to describe important but minor details:
  - Validation of methods,
  - Computer code,
  - Computational 'tricks' that might be useful to future users
- Bibliography:
  - High-quality papers and textbooks are preferred.
  - Links to code repositories may be necessary as part of good citation practice.
  - Don't cite to support common knowledge.

Warning

Most of this information is common scientific knowledge (Robert Day, CS website of Diego Calvanese, my own experiences, ...). But some of it is discipline-specific.

If in doubt, ask your project supervisor.