Example of a Transparent, Reproducible Research Manuscript

Matthew J Parkes [1] , [2] *

1 Centre for Epidemiology Versus Arthritis, Faculty of Biology, Medicine, and Health, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK.
2 NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

* Corresponding author: matthew.parkes@manchester.ac.uk

Abstract

This paper is an example of how a fully transparent and reproducible research manuscript can be produced using the R packages `RMarkdown`, `rticles`, and `knitr`, written using RStudio, and compiled through `pandoc`. There are multiple other ways in which this output can be achieved, but this is a simple, easy way to generate manuscripts using one interface. The PLOS LaTeX article template (from the `rticles` package, originally adapted from [https://journals.plos.org/plosone/s/latex]) has been used only as a guide, but the `rticles` package allows options from other journals.

.pdf format based on PLoS LaTeX sample manuscript, see 1
http://journals.plos.org/ploscompbiol/s/latex, using rticles package 2

Preface   3

Typical word processing software is designed to write documents comprised of text,   4
images, and tables all run together continuously. They use a WYSIWYG ('What you   5
see is what you get') model for the interface, which means that the content of the   6
document and the formatting are intertwined. This is to allow writers to change how   7
the document looks in real-time whilst writing, and gives great flexibility in how a   8
document can look.   9

However, the advantages of using WYSIWYG word processing software does not   10
benefit academic writing. Academic writing is a niche task: The formatting is usually   11
irrelevant to the document's content (as this is usually controlled by journals), and   12
manuscripts will almost always contain fixed structures such as the header and title   13
information, an abstract, and some form of structured layout (e.g. Introduction,   14
Methods, Results, and Conclusion, as well as disclosures and/or acknowledgements). In   15
addition, academic manuscripts contain other elements that word processing software   16
alone cannot typically do: reference management, and analyses.   17

This document demonstrates how a fully reproducible manuscript can be written in   18
RMarkdown, a special purpose language that actually amalgamates two languages - R   19
and Markdown - to allow analysts/writers to produce one continous document which   20
interweaves structured formatted writing with live code which is executed when the   21
document is compiled.   22

Table 1. Regression Model Output

group	baseline_mean	baseline_sd	at3Months_mean	at3Months_sd	at12Months_mean
0	26.71	16.78	NA	NaN	22.34
1	24.58	14.12	NA	NaN	16.25

The additional R package `knitr` compiles and runs the code, and in combination with the Pandoc syntax translation software, allows RMarkdown scripts to be compiled into a .pdf document, ready for publishing.

Introduction

This is an where you'd write the introduction. Citations can be easily entered using Markdown's syntax. For example, the original article from which the dataset used in this manuscript originates is cited here[1]. The original trial that generated the dataset used in this paper is cited here[2].

Doing Analyses with RMarkdown

It is possible to produce a 'live' CONSORT diagram, however given time constraints, one is not included in this demonstation. It is hoped that a revised version of this paper will include one, generated entirely from the dataset and code.

Results

This is an example of a reproducible results section. The results included below are not copied and pasted from another source, but generated live from analysis code when the .pdf is compiled. The code for this table is written directly in the RMarkdown file, in a code 'chunk'. Authors can control whether the code and output is displayed (by including `echo = TRUE` at the top of the code block), or just the output alone (by including `echo = FALSE` at the top instead, most likely the default for most publications).

```
## [1] "<table class=\"Rtable1\">\n<thead>\n<tr>\n<th class='rowlabel firstrow
```

You can also create lists in Markdown: - Item 1 - Item 2 - Item 3 - Sub-item 1 - Sub-sub-item 1 - Sub-item 2

References

1. Vickers AJ. Whose data set is it anyway? Sharing raw data from randomized trials. Trials. 2006;7: 15. doi:10.1186/1745-6215-7-15

2. Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, et al. Acupuncture for chronic headache in primary care: large, pragmatic, randomised trial. BMJ. 2004;328: 744. doi:10.1136/bmj.38029.421863.EB