

LaTeX And SAS For Creating And Assembling Data Monitoring Committee Reports

Adam Omidpanah, Axio Research

ABSTRACT

Data Monitoring Committees (DMCs) periodically review unblinded clinical data to assess the safety and conduct of clinical trials. The Reporting Statistician is a non-voting, unblinded, independent party who receives study data and randomization information and generates a DMC report. In addition to ensuring the DMC report is correct, the Reporting Statistician must assure that the report is easily interpretable. This paper focuses on integrating SAS With LaTeX for DMC report generation.

INTRODUCTION

The DMC report is a collection of tables, listings, and figures (TLFs) that summarize incidence and prevalence of safety and protocol events in a clinical study. The DMC relies on a well laid out report to make accurate and fast judgement as to any possible study concerns¹. A Portable Document Format (PDF) file is an ideal presentation format. The content is static and is organized by bookmarks, shortcuts, and sections. PDFs need fewer resources than other digital media, and they can be shared via secure file transfer. Most importantly, PDFs can be rendered from markdown language.

LaTeX is a computer programming language for typesetting technical documents. Custom SAS macros can be easily written to export numerical summaries as LaTeX markdown (.tex files), which reliably produce TLFs. Furthermore, .tex files or their PDF outputs can be assembled into a unified report, and automatically annotated with coversheets and tables of contents.

SAS currently supports 4 LaTeX tagsets for reproducing SAS output as LaTeX markdown². However, in the author's opinion, these tagsets are too highly specialized to replicating SAS ODS output to be modified and used successfully. The constructive approach outlined here will provide a minimum working example; a basis where tools, tips, and tricks can be easily developed.

We present a few simple examples of using SAS with LaTeX to generate DMC reports.

POPULATING A TABLE

LaTeX uses environments to isolate formatting breaks within a report. `tabular` is a powerful environment for reporting results in a readable way. Columns are delimited by the `&` escape and newlines by the `\\` escape. Converting numeric values to character is highly recommended to prevent truncation of trailing decimal zeros. In `tabular` the user has precise control of the column labels and formatting. This code block creates LaTeX markdown for a table. The table displayed in Figure 1.

```
proc report data=adam.adsl out=textable(drop=_BREAK_);
  columns region enrlnfn randfn saffn trt01a, saffn ;
  define region / group 'Geographic region';
  define enrlnfn / analysis sum 'Enrolled' ;
  define randfn / analysis sum 'Randomized';
  define trt01a / across 'Treatment received';
  define saffn / analysis sum 'Exposed to study treatment';
quit;

data textable2;
  set textable;
  enrlnfn2 = put(enrlnfn, 8.0);
  randfn2 = put(randfn, 8.0);
  saffn2 = put(saffn, 8.0);
  _c5_2 = put(_c5_, 8.0);
  _c6_2 = put(_c6_, 8.0) || "\\ " ;
run;
```

```

data _null_ *;
  set texttable2 nobobs = n;
  file "C:\MyTex.tex" dlm="&";

  if _n_ = 1 then do;
    put "\documentclass{standalone}";
    put "\begin{document}";
    put "\begin{tabular}{lrrrrr}";
    put "& & & \multicolumn{2}{c}{Treatment received} \\";
    put "Geographic Region & Enrolled & Randomized & Exposed to treatment &
    Red & Blue \\ \hline";
  end;

  put region enrlnfn2 randfn2 saffn2 _c5_2 _c6_2;

  if _n_ = n then do;
    put "\hline \end{tabular}";
    put "\end{document}";
  end;
run;

```

The output can be pasted into a .tex document, or the LOG option to the FILE statement in the DATA step can be replaced an external file name. The LaTeX wikibook³ covers advanced formatting topics, such as captions, footnotes and margin notes, breaking tables across pages, indenting, math type, etc.

Figure 1: Table of Study Disposition

Geographic Region	Enrolled	Randomized	Exposed to treatment	Treatment received	
				Red	Blue
Eastern Europe	176	176	174	92	82
North America	224	224	215	109	106
Western Europe	233	233	227	111	116

CONCATENATING A DMC REPORT

Using the above approach, or through other means, one may arrive at a collection of files which must be assembled into a final report⁴. Within a LaTeX master file, markdown from separate files can be sourced using the `\input{<filename.tex>}` command. Alternately, PDF outputs can be assembled in the master file using the command `\includepdf[pages=-]{filename.pdf}` and its accompanying `pdfpages` package. The master file allows the user to organize content by order and under sections. Using a revised version of the above output, the below example shows how to the master file assembles a report.

```

\documentclass[hidelinks]{article}
\usepackage[landscape]{geometry}
\usepackage{hyperref}
\usepackage{pdfpages}

\newcommand\invisiblesection[1]{%
  \refstepcounter{section}%
  \addcontentsline{toc}{section}{\protect\numberline{\thesection}#1}%
  \sectionmark{#1}}

\newcommand\invisiblesubsection[1]{%
  \refstepcounter{subsection}%

```

```

\addcontentsline{toc}{subsection}{\protect\numberline{\thesubsection}#1}%
\sectionmark{#1}}

\begin{document}

\tableofcontents
\newpage

\invisiblesection{Study Conduct}
\invisiblesubsection{Disposition}
\includepdf[pages=-]{mytex.pdf}
\end{document}

```

Renders the following table of contents in Figure 2, starting on page 2 of course because the ToC is page 1:

Figure 2: Automatically generated table of contents

Contents	
1	Study Conduct 2
1.1	Disposition 2

CONVERTING MICROSOFT DOCUMENTS TO PDF FORMAT

As was noted earlier, there are other situations in which the Reporting Statistician may find himself or herself juggling disparate DMC files. For example, many sponsors prefer to perform critical programming, with the Reporting Statistician simply merging a real-randomization assignment. Even if SAS code is available, it may not be possible to perform the necessary modifications to even convert the output engine to PDF. Output from external programming may be in rich text format (RTF) file, as was commonly the case in the days of ODS preceding the PDF engine. Word documents are generally unreliable tools for reporting data and results, they are easily modified even by accident, they are unwieldy in terms of the memory and disk demand when a document broaches a remotely cumbersome length, and the automation tools, like mail-merge, are somewhat unpopular within the user community. One interesting solution is the use of shell-scripting to batch convert a directory of files to PDF. Windows users may delight to know that Powershell is a reasonably capable tool for doing so. Below is a script that converts each RTF file in a directory to a PDF file using Microsoft Word as a COM object

```

$path = "<path to files>"
$word = New-Object -ComObject word.application
Get-ChildItem -Path $path -Filter *.rtf | ForEach-Object {
    $document = $word.Documents.Open($_.FullName)
    $pdf_filename = "$($_.DirectoryName)\${_.BaseName}.pdf"
    $document.SaveAs([ref] $pdf_filename, [ref] 17)
    $document.Close()
}
$word.Quit()

```

CONCEPTUAL GRAPHICS

Conceptual graphics are a visual display of information that uses objects and relations. Study specific information may be annotated to represent progress toward key milestones. Consort diagrams are conceptual graphics. Although ODS graphics have geometry tools to create conceptual graphics, identifying the correct dimensions of objects by guesswork makes for tedious and error prone coding. LaTeX supports a powerful graphical library called TikZ. TikZ uses complex programmatic drawing that is capable of automatically setting node and edge parameters given the user inputs. An example below shows how macro variables can be used to annotate such a graphic on an ongoing basis producing the result in Figure 3:

```

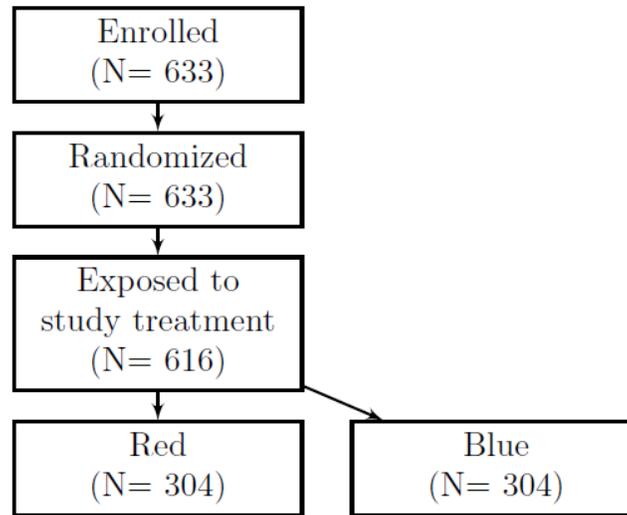
proc sql;
  select count(distinct usubjid) into :enr from adam.adsl;
  select count(distinct usubjid) into :ran from adam.adsl where randfl = "Y";
  select count(distinct usubjid) into :exp from adam.adsl where saffl = "Y";
  select count(distinct usubjid) into :red from adam.adsl where saffl = "Y"
and trt01a = "Red";
  select count(distinct usubjid) into :blue from adam.adsl where saffl = "Y"
and trt01a = "Blue";
quit;

%let tikz = \documentclass{article}
\usepackage{tikz}
\usetikzlibrary{shapes,arrows}
\begin{document}
\begin{center}
  \begin{tikzpicture}[auto,
    block/.style={rectangle, draw=black, thick, fill=white, text centered,
text width=8em},
    line/.style={draw, thick, -latex'}]
    \matrix [column sep=5mm,row sep=3mm, every node/.style=block] {
      \node (enr) {Enrolled \\\(N=&enr)\}%quote(;) & \\\
      \node (ran) {Randomized \\\(N=&ran)\}%quote(;) & \\\
      \node (exp) {Exposed to study treatment \\\(N=&exp)\}%quote(;) & \\\
      \node (red) {Red \\\(N=&red)\}%quote(;) & \node (blue) {Blue
\\(N=&red)\}%quote(;) \\\
      }%quote(;)
    \begin{scope}[every path/.style=line]
      \path (enr) -- (ran) %quote(;)
      \path (ran) -- (exp) %quote(;)
      \path (exp) -- (red) %quote(;)
      \path (exp) -- (blue) %quote(;)
    \end{scope}
  \end{tikzpicture}
\end{center}
\end{document};
%put &tikz;

```

Generates the output in Figure 3. The use of SAS macros to annotate the nodes of a conceptual graphic significantly expands the possibility of output that can be created without any change to initial programming. For minimal reproducibility, the output focuses on numbers, but can in fact include lists, and formatting specifications, such as the incidence and reasons for study discontinuation. Importantly, the margins of the node are calculated based on the text supplied to the legend.

Figure 3: Consort diagram using TikZ



CONCLUSION

LaTeX is an ideal companion to SAS for generating highly accessible and reliable DMC report outputs. The reporting statistician has a responsibility to the DMC to be versant with typesetting at a level to anticipate and meet their demands. LaTeX provides straightforward solutions to common issues in creating and assembling DMC reports.

REFERENCES

1. Fleming, T. R., Ellenberg, S. S., & DeMets, D. L. (2018). Data Monitoring Committees: Current issues. *Clinical Trials*, 15(4), 321–328. <https://doi.org/10.1177/1740774518764855>
2. <https://support.sas.com/rnd/base/ods/odsmarkup/latex.html>
3. <https://en.wikibooks.org/wiki/LaTeX>
4. DeMets, D. (2018). The independent statistician model: How well is it working? *Clinical Trials*, 15(4), 329–334. <https://doi.org/10.1177/1740774518772841>

ACKNOWLEDGMENTS

This work was supported by Axio Research.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Adam Omidpanah
Axio Research
206 577 0276
adamo@axioresearch.com
<https://www.axioresearch.com/>