

# Publication Quality Tables in MS Word

## Creating RTF files

Kieran McCaul

WA Centre for Health & Ageing  
University of Western Australia

Australian and New Zealand  
Stata User's Group Meeting, 2009

# Introduction

- All of my work involves using MS Word to produce manuscripts for submission to journals
- These will be peer-reviewed, so I want the tables to look “nice”
- You can, however, generate tables as RTF files that Word can read directly
- Other users have produced `ado` files to do this (Roger Newson, Ben Jann)
- I decided to have a go myself

# Some Background

- In the past I would just use tab-delimited files
  - Use the Stata commands `file open`, `file write`, and `file close` to open a text file and write out the contents of the table to a tab-delimited file
  - Then open this file in my text editor and copy it to the clipboard and open Word and paste it into a document
  - Then highlight these tab-delimited lines of text and use the Word commands, `Insert table -> Convert text to table...`, to format the text as a Word table
  - Then write in headings for each column of the table, adjust the width of the columns, put borders on the top and bottom of the table, etc, etc
- Then redo it all again if you have to change the results that you are tabulating

# Example

- Suppose I have some data on diabetics and non-diabetics
- I want to look at the differences between diabetics and non-diabetics with respect to a number of different covariates (sex and body mass index in this example)
- How many ways could I tabulate this?

# Table of results in Word

Table 1: small table ¶

Covariate		Diabetes	No-Diabetes
Sex	Women	1,119	11,157
	Men	1,416	10,311
BMI	Underweight	20	489
	Normal	815	9,713
	Overweight	1,013	7,692
	Obese	506	2,314

¶

# Table of results in Word

Table 1: small table

Covariate		Diabetes	
		Yes	No
Sex	Women	1,119	11,157
	Men	1,416	10,311
BMI	Underweight	20	489
	Normal	815	9,713
	Overweight	1,013	7,692
	Obese	506	2,314

# Table of results in Word

Table 1: small-table

Covariate		Diabetes	
		Yes	No
Sex	Women	1,119 (9.1)	11,157 (90.9)
	Men	1,416 (12.1)	10,311 (87.9)
BMI	Underweight	20 (3.9)	489 (96.1)
	Normal	815 (7.7)	9,713 (92.3)
	Overweight	1,013 (11.6)	7,692 (88.4)
	Obese	506 (17.9)	2,314 (82.1)

# Table of results in Word

Table 1: small table

Covariate		Diabetes	
		Yes	No
Sex	Women	1,119 (9.1)	11,157 (90.9)
	Men	1,416 (12.1)	10,311 (87.9)
BMI			
	Underweight	20 (3.9)	489 (96.1)
	Normal	815 (7.7)	9,713 (92.3)
	Overweight	1,013 (11.6)	7,692 (88.4)
	Obese	506 (17.9)	2,314 (82.1)



# Table of results in Word

Table 1: small table

Covariate		Diabetes			
		Yes		No	
		N	(%)	N	(%)
Sex	Women	1,119	(9.1)	11,157	(90.9)
	Men	1,416	(12.1)	10,311	(87.9)
BMI	Underweight	20	(3.9)	489	(96.1)
	Normal	815	(7.7)	9,713	(92.3)
	Overweight	1,013	(11.6)	7,692	(88.4)
	Obese	506	(17.9)	2,314	(82.1)

# Automating table production is problematic

- Tables are not like graphs
  - How the results are arranged in the cells of a table is a matter of personal style
  - Journals also have differing style requirements for tables
  - You are constrained by the size of the page
- To take these all into account, a program will need many options

# Simple Example

- Consider the table below
- The RTF code for creating this table is not complex

Table 1: small-table

Covariate		Diabetes	
		Yes	No
Sex	Women	1,119 (9.1)	11,157 (90.9)
	Men	1,416 (12.1)	10,311 (87.9)

# RTF code

```
1 (\rtf1\ansi\deff0
2 (\fonttbl
3 (\f0 Times New Roman;)
4 (\f1 Symbol;)
5 )
6 \paperw11909\paperh16834\margl1440\margr1440\margt1440\margb1440
7 (\pard\fs22 Table 1: small table \par)
8 (\pard\fs22
9 \trwd\trgaph144\cbrdr\brdr\cellx2432\cbrdr\brdr\cellx4700\cbrdr\brdr\cellx9236
10 \pard\intbl\ql Covariate\cell
11 \pard\intbl\ql \cell
12 \pard\intbl\qc Diabetes\cell
13 \row
14 \trwd\trgaph144\cbrdr\brdr\cellx2432\cbrdr\brdr\cellx4700\cbrdr\brdr\cbrdr\brdr\cellx6968\cbrdr\brdr\cbrdr\brdr\cellx9236
15 \pard\intbl\ql \cell
16 \pard\intbl\ql \cell
17 \pard\intbl\qc Yes\cell
18 \pard\intbl\qc No\cell
19 \row
20 \trwd\trgaph144\cellx2432\cellx4700\cellx6968\cellx9236
21 \pard\intbl\ql Sex\cell
22 \pard\intbl\ql Women\cell
23 \pard\intbl\qc 1,119 (9.1)\cell
24 \pard\intbl\qc 11,157 (90.9)\cell
25 \row
26 \trwd\trgaph144\cellx2432\cellx4700\cellx6968\cellx9236
27 \pard\intbl\ql \cell
28 \pard\intbl\ql Men\cell
29 \pard\intbl\qc 1,416 (12.1)\cell
30 \pard\intbl\qc 10,311 (87.9)\cell
31 \row
32 \trwd\trgaph144\cbrdr\brdr\cellx2432\cbrdr\brdr\cellx4700\cbrdr\brdr\cbrdr\brdr\cellx6968\cbrdr\brdr\cbrdr\brdr\cellx9236
33 \pard\intbl\ql \cell
34 \pard\intbl\ql \cell
35 \pard\intbl\ql \cell
36 \pard\intbl\ql \cell
37 \row
38 )
39 )
40 )
```

# RTF code

```
1 (\rtf1\ansi\deff0
2 (\fonttbl
3 (\f0 Times New Roman;)
4 (\fi Symbol;)
5 )
6 \paperw11909\paperh16834\margl1440\margr1440\margt1440\margb1440
7 (\pard\fs22 Table 1: small table \par)
8 (\pard\fs22
9 \trwd\trgaph144\cbrdr\brdr\cellx2432\cbrdr\brdr\cellx4700\cbrdr\brdr\cellx9236
10 \pard\intbl\ql Covariate\cell
11 \pard\intbl\ql cell
12 \pard\intbl\ql Diabetes\cell
13 \row
14 \trwd\trgaph144\cbrdr\brdr\cellx2432\cbrdr\brdr\cellx4700\cbrdr\brdr\cbrdr\brdr\cellx6968\cbrdr\brdr\cbrdr\brdr\cellx9236
15 \pard\intbl\ql cell
16 \pard\intbl\ql cell
17 \pard\intbl\ql Yes\cell
18 \pard\intbl\ql No\cell
19 \row
20 \trwd\trgaph144\cellx2432\cellx4700\cellx6968\cellx9236
21 \pard\intbl\ql Sex\cell
22 \pard\intbl\ql Women\cell
23 \pard\intbl\ql 1,119 (9.1)\cell
24 \pard\intbl\ql 11,157 (87.9)\cell
25 \row
26 \trwd\trgaph144\cellx2432\cellx4700\cellx6968\cellx9236
27 \pard\intbl\ql cell
28 \pard\intbl\ql Men\cell
29 \pard\intbl\ql 1,416 (12.1)\cell
30 \pard\intbl\ql 10,311 (87.9)\cell
31 \row
32 \trwd\trgaph144\cbrdr\brdr\cellx2432\cbrdr\brdr\cellx4700\cbrdr\brdr\cbrdr\brdr\cellx6968\cbrdr\brdr\cbrdr\brdr\cellx9236
33 \pard\intbl\ql cell
34 \pard\intbl\ql cell
35 \pard\intbl\ql cell
36 \pard\intbl\ql cell
37 \row
38 )
39 )
40 )
```

# Simple Example

Table 1: small table

Covariate		Diabetes	
		Yes	No
Sex	Women	1,119 (9.1)	11,157 (90.9)
	Men	1,416 (12.1)	10,311 (87.9)

- The table has five rows
- In the first row, the third and fourth cells are merged
- The remaining rows have four cells
- Borders:
  - Along the top of the first row
  - Along the top of the last two cells of the second row
  - Along the bottom of the second row and the last row
- Some cells are left-justified, some are centred

## Document characteristics

```
{\rtf1\ansi\deff0
{\fonttbl
{\f0 Times New Roman;}
{\f1 Symbol;}
}
\paperw11909\paperh16834\margl1440\margr1440\margt1440\margb1440
{\pard\fs22 Table 1: small table \par}
```

- Declares that the file is RTF version 1, ANSI
- Builds a font table
  - Times New Roman and Symbol fonts in this example
- Sets the page dimensions and margins
- Writes the Table title

## First row

```
\trowd\trgaph144\clbrdrt\brdrs\cellx2432\clbrdrt\brdrs\cellx4700  
\clbrdrt\brdrs\cellx9236  
\pard\intbl\ql Covariate\cell  
\pard\intbl\ql \cell  
\pard\intbl\qc Diabetes\cell  
\row
```

- Defines the characteristics of the first row
  - Three cells with a border along the top
- Fills in the contents of each of these cells
  - First cell contains "Covariate" left-justified (\ql)
  - Second cell is empty
  - Third cell contains "Diabetes", centred (\qc)



## Second row

```
\trowd\trgaph144\clbrdrb\brdrs\cellx2432\clbrdrb\brdrs\cellx4700  
\clbrdrt\brdrs\clbrdrb\brdrs\cellx6968\clbrdrt\brdrs\clbrdrb\brdrs\cellx9236  
\pard\intbl\ql \cell  
\pard\intbl\ql \cell  
\pard\intbl\qc Yes\cell  
\pard\intbl\qc No\cell  
\row
```

- Defines the characteristics of the second row
  - Four cells with a border along the top of the last two cells
  - Border along the bottom of all cells
- Fills in the contents of each of these cells
  - First two cells are empty
  - Third cell contains "Yes", centred (\qc)
  - Fourth cell contains "No", centred (\qc)

## Third row

```
\trowd\trgaph144\cellx2432\cellx4700\cellx6968\cellx9236  
\pard\intbl\ql Sex\cell  
\pard\intbl\ql Women\cell  
\pard\intbl\qc 1,119 (9.1)\cell  
\pard\intbl\qc 11,157 (90.9)\cell  
\row
```

- Defines the characteristics of the third row
  - Four cells with no borders
- Fills in the contents of each of these cells
  - First cell contains "Sex", left-justified
  - Second cell contains "Women", left-justified
  - Third cell contains "1,119 (9.1)", centred
  - Fourth cell contains "11,157 (90.9)", centred

The RTF code has a basic structure:

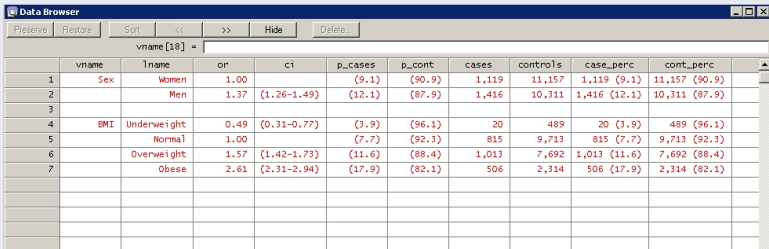
- There are a set of initial commands that only appear once - initializing commands, page size, etc
- Each different row of a table needs to be defined
- Writing to a cell

This suggests that each of these could be handled by ado files

- I've written four (so far): `rtfwrite`, `rtfrowdef`, `rtfrow`, and `rtfcell`
- They assume that a summary dataset has already been constructed
- Each observation in this dataset is a row of data in the table

# Data formatted for table

- I use Roger Newson's `parvest` program to save regression estimates (if any)
- I format confidence intervals and percentages in the dataset as I want them to appear in the table



	vname	lname	or	ci	p_cases	p_cont	cases	controls	case_perc	cont_perc
1	Sex	Women	1.00		(9.1)	(90.9)	1,119	11,157	1,119 (9.1)	11,157 (90.9)
2		Men	1.37	(1.26-1.49)	(12.1)	(87.9)	1,416	10,311	1,416 (12.1)	10,311 (87.9)
3										
4	EMI	Underweight	0.49	(0.31-0.77)	(3.9)	(96.1)	20	489	20 (3.9)	489 (96.1)
5		Normal	1.00		(7.7)	(92.3)	815	9,713	815 (7.7)	9,713 (92.3)
6		Overweight	1.57	(1.42-1.73)	(11.6)	(88.4)	1,013	7,692	1,013 (11.6)	7,692 (88.4)
7		Obese	2.61	(2.31-2.94)	(17.9)	(82.1)	506	2,314	506 (17.9)	2,314 (82.1)

```
rtfwrite , rtfname(filename) [fonts(\f0 fontname [\f1 fontname [...]])  
    fsize(#) pagesize(A4 or letter) margins(####) title(text) inches  
replace ]
```

- Opens a file (foo.rtf)
- Allows a number of options:
  - page size – A4 is default
  - page margins – default is 1" for all margins
  - fonts to be used – default is Times New Roman and Symbol
  - font size – 11pt is default
  - inches – otherwise cm, the default, is expected

```
rtfrowdef , columns(#[#[...]]) rowdef(definition name) [brdt(numlist)  
brdb(numlist) brdr(numlist) brdl(numlist) ]
```

- Constructs a table row definition
- Options:
  - Width (in inches or cm) of each of the columns
  - Borders on top or bottom of cells (or the left and right of cells, if you must)
  - Provide a name for the row definition

```
rtfrow , rowdef(row definition)
```

- Once all the different type of rows in the table are defined, `rtfrow` simply writes the row definition to the RTF file
- Once that is written, each cell can be written to



```
rtfcell , [align(l, r, or c) text(text) row(#) var(varname) format(fmt)  
p(string) last ]
```

- Writes the contents of a table cell
- Options:
  - Cell alignment
  - What is written to the cell can be text or the contents of a variable
  - If is a variable, you can specify the format
  - If the variable contains p-values, you can specify the smallest p-value to be formatted (eg: p(0.001) would put "<0.001" in the cell if the p-value is smaller than 0.001

# Stata program

```
7 use "`$data\table\1\results.dta'", `clear'
8
9 RTFwrite, `title("Table 1: small table") `rtfname("`$programs\sex_diab.rtf") `replace'
10
11 RTFrowdef, `columns(4.29 4 8) `brdt(1(1)4) `rowdef(head1)'
12 RTFrowdef, `columns(4.29 4 4 4) `brdt(3 4) `brdb(1(1)4) `rowdef(head2)'
13 RTFrowdef, `columns(4.29 4 4 4) `rowdef(main)'
14 RTFrowdef, `columns(4.29 4 4 4) `brdb(1(1)4) `rowdef(lastrow)'
15
16 RTFrow, `rowdef(head1)'
17 RTFcell, `align(l) `text("Covariate")
18 RTFcell, `align(l)'
19 RTFcell, `align(c) `text("Diabetes") `last'
20
21 RTFrow, `rowdef(head2)'
22 RTFcell, `align(l)'
23 RTFcell, `align(l)'
24 RTFcell, `align(c) `text("Yes")
25 RTFcell, `align(c) `text("No") `last'
26
27 forvalues i = 1/2 {
28     RTFrow, `rowdef(main)'
29     RTFcell, `row(i) `align(l) `var(vname)'
30     RTFcell, `row(i) `align(l) `var(lname)'
31     RTFcell, `row(i) `align(c) `var(case_perc)'
32     RTFcell, `row(i) `align(c) `var(cont_perc) `last'
33 }
34
35 RTFrow, `rowdef(lastrow)'
36 RTFcell, `align(l)'
37 RTFcell, `align(l)'
38 RTFcell, `align(l)'
39 RTFcell, `align(l) `last'
40
```

## Problems

- This approach requires a user to format the data for the table
  - so it requires some Stata programming skills
  - but it reduces the number of options required to produce a table

## Still to do

- Finding the right column widths to use is fiddly at the moment
  - Working on an algorithm that would automatically calculate column widths required