HOW TO PRESENT RESEARCH DATA?

Tong Seng Fah *MMed (FamMed UKM)***, Aznida Firzah Abdul Aziz** *MMed (FamMed UKM)*

Department of Family Medicine, Universiti Kebangsaan Malaysia

Address for correspondence: Dr. Tong Seng Fah, Department of Family Medicine, Medical Faculty, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, Cheras, 56000 Kuala Lumpur, Malaysia. Tel: 603-91733333 Ext 2831, Fax: 603-91738153, Email: tsf@mail.hukm.ukm.my

INTRODUCTION

The result section of an original research paper provides answer to this question "What was found?" The amount of findings generated in a typical research project is often much more than what medical journal can accommodate in one article. So, the first thing the author needs to do is to make a selection of what is worth presenting. Having decided that, he/she will need to convey the message effectively using a mixture of text, tables and graphics. The level of details required depends a great deal on the target audience of the paper. Hence it is important to check the requirement of journal we intend to send the paper to (e.g. the Uniform Requirements for Manuscripts Submitted to Medical Journals[1]). This article condenses some common general rules on the presentation of research data that we find useful.

SOME GENERAL RULES

o   Keep it simple. This golden rule seems obvious but authors who have immersed in their data sometime fail to realise that readers are lost in the mass of data they are a little too keen to present. Present too much information tends to cloud the most pertinent facts that we wish to convey.

o   First general, then specific. Start with response rate and description of research participants (these information give the readers an idea of the representativeness of the research data), then the key findings and relevant statistical analyses.

o   Data should answer the research questions identified earlier.

o   Leave the process of data collection to the methods section. Do not include any discussion. These errors are surprising quite common.

o   Always use past tense in describing results.

o   Text, tables or graphics? These complement each other in providing clear reporting of research findings. Do not repeat the same information in more than one format. Select the best method to convey the message.

TEXT

Consider these two lines:
1. Mean baseline HbA_{1c} of 73 diabetic patients before intervention was 8.9% and mean HbA_{1c} after intervention was 7.8%.
2. Mean HbA_{1c} of 73 of diabetic patients *decreased* from 8.9% to 7.8% after an intervention.

In line 1, the author presents only the data (i.e. what exactly was found in a study) but the reader is forced to analyse and draw their own conclusion ("mean HbA_{1c} decreased") thus making the result more difficult to read. In line 2, the preferred way of writing, the data was presented together with its interpretation.

o   Data, which often are numbers and figures, are better presented in tables and graphics, while the interpretation are better stated in text. By doing so, we do not need to repeat the values of HbA_{1c} in the text (which will be illustrated in tables or graphics), and we can interpret the data for the readers. However, if there are too few variables, the data can be easily described in a simple sentence including its interpretation. For example, the majority of diabetic patients enrolled in the study were male (80%) compare to female (20%).

o   Using qualitative words to attract the readers' attention is not helpful. Such words like "remarkably" decreased, "extremely" different and "obviously" higher are redundant. The exact values in the data will show just how remarkable, how extreme and how obvious the findings are.

Example:
"It is clearly evident from Figure 1B that there was significant different (p=0.001) in HbA_{1c} level at 6, 12 and 18 months after diabetic self-management program between 96 patients in intervention group and 101 patients in control group, but no difference seen from 24 months onwards."
[Too wordy]

The above can be rewritten as:
"Statistical significant difference was only observed at 6, 12 and 18 months after diabetic self-management program between intervention and control group (Fig 1B)". [The p values and numbers of patients are already presented in Figure 1B and need not be repeated.]

o   Avoid redundant words and information. Do not repeat the result within the text, tables and figures. Well-constructed tables and graphics should be self-explanatory, thus detailed explanation in the text is not required. Only important points and results need to be highlighted in the text.

TABLES

Tables are useful to highlight precise numerical values; proportions or trends are better illustrated with charts or graphics. Tables summarise large amounts of related data clearly and allow comparison to be made among groups of variables. Generally, well-constructed tables should be self explanatory with four main parts: title, columns, rows and footnotes.

o   *Title.* Keep it brief and relate clearly the content of the table. Words in the title should represent and summarise variables used in the columns and rows rather than repeating the columns and rows' titles. For example, "Comparing full blood count results among different races" is clearer and simpler than "Comparing haemoglobin, platelet count, and total white cell count among Malays, Chinese and Indians".

o   *Columns and rows.* Columns are vertically listed data, and rows are horizontally listed data. Similar data ought to be presented in columns. Often these are dependant variables and allow clearer comparison among groups. Compare Table 1A and 1B, the dependant variables in Table 1B are waist circumference, HbA_{1c}, SBP and etc. Table 1B shows a better comparison of dependant variables among ethnicity than Table 1A. The first column to the left is usually a list of its independent variables i.e. Malay, Chinese, Indian and others, as the example of Table 1B. A table with too many dependent variables would become too wide for a page. There are two alternatives to this problem. We can list the dependant variables in the first left column and independent variables across the top. However, doing so should not compromise clarity of the message we want to get across. The second alternative is to cut down unnecessary columns, which, can be replaced by footnotes explaining their definition. For example, we can eliminate the columns on *p,* student-t test and chi-square values (see Table 2). Significant test result can be marked using * or # with a footnote. Presenting

exact values of statistical data with no significant difference is rarely useful.

Table 1A: Baseline waist circumference, HbA_{1c}, Blood pressure and LDL-cholesterol level among Malay, Chinese, Indian and others races

	Malay	Chinese	Indian	Others
Waist circumference	98	102	105	95
HbA1c	8.89	8.66	9.0	8.7
SBP	165.1	164.0	170.34	168
DBP	98.5	101	99.3	97.6
LDL-C	3.8	3.9	3.4	3.1

Table 1B: Mean (SD) baseline diabetic metabolic control among different races

	WC* (SD)	HbA$_{1c}$ (SD), %	SBP† (SD)	DBP‡ (SD)	LDL-C§ (SD)
Malay	98 (15)	8.9 (1.5)	165 (21)	98 (13)	3.8 (0.9)
Chinese	102 (18)	8.7 (2.1)	164 (28)	101 (15)	3.9 (0.7)
Indian	105 (22)	9.0 (1.8)	170 (36)	99 (22)	3.4 (1.2)
Others	95 (28)	8.7 (2.5)	168 (40)	97 (28)	3.1 (1.0)

*WC, waist circumference (in cm); † SBP, systolic blood pressure (in mmHg); ‡DBP, diastolic blood pressure (in mmHg); §LDL-cholesterol (in mmol/L)

Table 2: Comparison of the presenting symptoms among patients with and without thrombocytopaenia

Symptom	Platelet count (%)		OR* (95% CI)
	Normal	Thrombocytopaenia	
Presented at or after day 3 of fever	26 (65)	31 (93.9)	5.88(1.20-28.8)†
Myalgia	32 (80)	23 (82.1)	1.09 (0.51-2.31)
Headache	25 (64.1)	22 (78.6)	1.56 (0.75-3.26)
Nausea/vomiting	18 (46.2)	23 (76.7)	2.24 (1.12-4.50)‡
Arthralgia	26 (40)	13 (54.2)	1.43 (0.76-2.69)
Retro-orbital pain	9 (23.1)	6 (26.1)	1.11 (0.54-2.29)
Rash	5 (12.5)	8 (24.2)	1.47 (0.88-2.49)

*Odds ratio (95% confidence interval), †p=0.04, ‡p=0.01,

o   Footnotes. These add clarity to the data presented. They are listed at the bottom of tables. Their use is to define unconventional abbreviation, symbols, statistical analysis and acknowledgement (if the table is adapted from a published table). Generally the font size is smaller in the footnotes and follows a sequence of foot note signs (*, †, ‡, §, ||, ¶, **, ††, ‡‡).[1] These symbols and abbreviation should be standardised in all tables to avoid confusion and unnecessary long list of footnotes. Proper use of footnotes will reduce the need for multiple columns (e.g. replacing a list of p values) and the width of columns (abbreviating waist circumference to WC as in table 1B)

o Body of the table. We can improve the clarity of data presented in body of the tables by the following:

- Consistent use of units and its decimal places. The data on systolic blood pressure in Table 1B is neater than the similar data in Table 1A.
- Arrange date and timing from left to the right.
- Round off the numbers to fewest decimal places possible to convey meaningful precision. Mean systolic blood pressure of 165.1mmHg (as in Table 1B) does not add much precision compared to 165mmHg. Furthermore, 0.1mmHg does not add any clinical importance. Hence blood pressure is best to round off to nearest 1mmHg.
- Avoid listing numerous zeros, which made comparison incomprehensible. For example total white cell count is best represented with 11.3 $X10^6$/L rather than 11,300,000/L. This way, we only need to write 11.3 in the cell of the table.
- Avoid too many lines in a table. Often it is sufficient to just have three horizontal lines in a table; one below the title; one dividing the column titles and data; one dividing the data and footnotes. Vertical lines are not necessary. It will only make a table more difficult to read (compare Tables 1A and 1B).
- Standard deviation can be added to show precision of the data in our table. Placement of standard deviation can be difficult to decide. If we place the standard deviation at the side of our data, it allows clear comparison when we read down (Table 1B). On the other hand, if we place the standard deviation below our data, it makes comparison across columns easier. Hence, we should decide what we want the readers to compare.
- It is neater and space-saving if we highlight statistically significant finding with an asterisk (*) or other symbols instead of listing down all the *p* values (Table 2). It is not necessary to add an extra column to report the detail of student-t test or chi-square values.

GRAPHICS

Graphics are particularly good for demonstrating a trend in the data that would not be apparent in tables. It provides visual emphasis and avoids lengthy text description. However, presenting numerical data in the form of graphs will lose details of its precise values which tables are able to provide. The authors have to decide the best format of getting the intended message across. Is it for data precision or emphasis on a particular trend and pattern? Likewise, if the data is easily described in text, than text will be the preferred method, as it is more costly to print graphics than text. For example, having a nicely drawn

age histogram is take up lots of space but carries little extra information. It is better to summarise it as mean ±SD or median depends on whether the age is normally distributed or skewed. Since graphics should be self-explanatory, all information provided has to be clear. Briefly, a well-constructed graphic should have a title, figure legend and footnotes along with the figure. As with the tables, titles should contain words that describe the data succinctly. Define symbols and lines used in legends clearly. Some general guides to graphic presentation are:

o Bar charts, either horizontal or column bars, are used to display categorical data. Strictly speaking, bar charts with continuous data should be drawn as histograms or line graphs. Usually, data presented in bar charts are better illustrated in tables unless there are important pattern or trends need to be emphasised.

o Avoid 3-D graphs and charts. Three dimensional graphics are impressive in slide show and easily capture the attention of the audience. In medical writing, they are not effective because it is difficult to read the exact value on the Y axis (the height of the bars) accurately (Figure 1A).

o Line graphs are most appropriate in tracking changing values between variables over a period of time or when the changing values are continuous data. Independent variables (e.g. time) are usually on the X-axis and dependant variables (for example, HbA_{1c}) are usually on the Y-axis. The trend of HbA_{1c} changes is much more apparent with Figure 1B than Figure 1A, and HbA_{1c} level at any time after intervention can be accurately read in Figure 1B.

o Pie charts should not be used often as any data in a pie chart is better represented in bar charts (if there are specific data trend to be emphasised) or simple text description (if there are only a few variables). A common error is presenting sex distribution of study subjects in a pie chart. It is simpler by just stating % of male or female in text form.

o Patients' identity in all illustrations, for example pictures of the patients, x-ray films, and investigation results should remain confidential. Use patient's initials instead of their real names. Cover or blackout the eyes whenever possible. Obtain consent if pictures are used. Highlight and label areas in the illustration, which need emphasis. Do not let the readers search for details in the illustration, which may result in misinterpretation. Remember, we write to avoid misunderstanding whilst maintaining clarity of data.

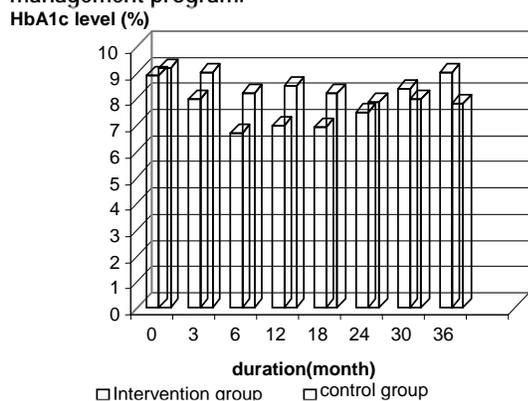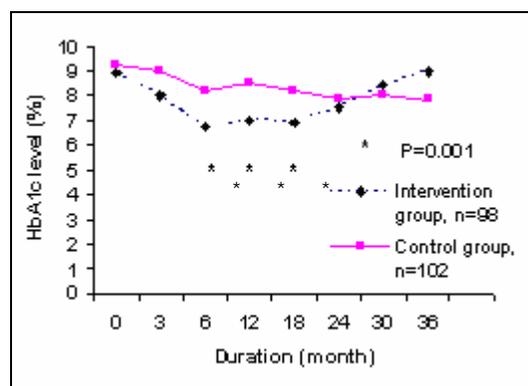**Figure 1A. Changes of HbA$_{1c}$ level after diabetic self-management program.**

Figure 1B. Changes of HbA$_{1c}$ level after diabetic self-management program.

STATISTICS

Papers are often rejected because wrong statistical tests are used or interpreted incorrectly. A simple approach is to consult the statistician early. Bearing in mind that most readers are not statisticians, the reporting of any statistical tests should aim to be understandable by the average audience but sufficiently rigorous to withstand the critique of experts.

o   Simple statistic such as mean and standard deviation, median, normality testing is better reported in text. For example, age of group A subjects was normally distributed with mean of 45.4 years old kg (SD=5.6). More complicated statistical tests involving many variables are better illustrated in tables or graphs with their interpretation by text. (See section on Tables).

o   We should quote and interpret p value correctly. It is preferable to quote the exact p value, since it is now easily obtained from standard statistical software. This is more so if the p value is statistically not significant, rather just quoting p>0.05 or p=ns. It is not necessary to report the exact p value that is smaller than 0.001 (quoting p<0.001 is sufficient); it is incorrect to report p=0.0000 (as some software apt to report for very small p value).

o   We should refrain from reporting such statement: "mean systolic blood pressure for group A (135mmHg, SD=12.5) was higher than group B (130mmHg, SD= 9.8) but did not reach statistical significance (t=4.5, p=0.56)." When p did not show statistical significance (it might be >0.01 or >0.05, depending on which level you would take), it simply means no difference among groups.

o   Confidence intervals. It is now preferable to report the 95% confidence intervals (95%CI) together with p value, especially if a hypothesis testing has been performed.

CONCLUSION

The main core of the result section consists of text, tables and graphics. As a general rule, text provides narration and interpretation of the data presented. Simple data with few categories is better presented in text form. Tables are useful in summarising large amounts of data systemically and graphics should be used to highlight evidence and trends in the data presented. The content of the data presented must match the research questions and objectives of the study in order to give meaning to the data presented. Keep the data and its statistical analyses as simple as possible to give the readers maximal clarity.

REFERENCES
1.   International Committee for Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: Writing and Editing for Biomedical Publication. [HTML] [PDF]

FURTHER READINGS
1.   Hall GM (ed). How to write a paper. 3rd Edition. London: BMJ Publishing Group, 2003.
2.   Grange RI. Saving time, effort and tears: a guide to presenting results. *Br J Urol.* 1998;81(2):335-9 [PubMed]
3.   Huth EJ. Writing and publishing in Medicine. 3rd ed. Maryland: Williams & Wilkins, 1999.
4.   Lang T, Secic M. How to report statistics in medicine: Annotated guidelines for authors, editors and reviewers. Philadelphia: American College of Physicians, 1997.