



Preparing better tables

Wei Liu

Chinese Journal of Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, China

Correspondence to: Wei Liu. Dongfeng Road East 651, Guangzhou 510060, China. Email: helen_lw550@hotmail.com.

Abstract: In biomedical research papers, tables are commonly used to present data, to help readers to understand the research, especially to interpret the results. A well-designed table includes an informative title, reasonable column and row headings, understandable data field, and sufficient footnotes.

Keywords: Table; title; column; row; data field; footnotes

Received: 28 December 2017; Accepted: 11 January 2018; Published: 24 January 2018.

doi: 10.21037/jphe.2018.01.01

View this article at: <http://dx.doi.org/10.21037/jphe.2018.01.01>

“Tables are for communication, not data storage.”

—Howard Wainer (1943–), American statistician, researcher, and author (1).

Introduction

Tables are a visual medium commonly used to present background information or research results in articles reporting public health research. Many readers look at the tables (and figures) before reading the text of an article, and all readers want to understand the table without having to refer back and forth to the text. Therefore, it is important that tables have informative titles, a clear structure, and reliable data to allow readers to easily recognize and interpret important information or patterns in the data. Well-designed tables should help readers to identify the most important findings of a study, whereas poorly designed tables may not present the data clearly and can even mislead readers. In this article, I tell when you need a table, how to design a good table, and several common mistakes you need to avoid when designing a table.

When you need a table

First, read the journal’s Author Guidelines to find out whether the journal limits the number, content, structure, and any other aspects of tables. Some journals count tables and graphs in the word count of an article, for example.

Tables usually take more spaces than words and so should be used only when they communicate more efficiently than

words. That is, tables should save readers’ time in finding the data of interest and in understanding the messages conveyed by the data. Generally, tables should not be used to present a great amount of raw data in a research paper although such tables may be included as supplemental materials deposited on-line.

Unlike figures, which may have panels (e.g., “Figure 1A and 1B”), tables do not have parts. Instead of “Table 1A and 1B”, either reorganize the table (e.g., by using cut-in headings) or create two separate tables.

Table components

A table has a table number, a title, row and column headings, a field of data (numbers or symbols or text), perhaps some expanded abbreviations, and footnotes (*Figure 1*). A well-designed table includes an informative title, accurate column and row headings, an organized data field, and additional information sufficient for readers to understand the table.

Each table should have a purpose, such as collecting background information, presenting an amount of raw data, showing comparisons, revealing a trend, and so on. The specific purpose determines how the table should be organized (2).

Writing the title

A table title should identify the data in the table; it should not summarize the results (2). To make the table understandable independent of the text, the table title

Table No. title				
Column heading	Spanner heading			Column heading
	Column heading	Column heading	Column heading	
Row heading	234.9			↓
Row heading	220.0			↑
Stub heading	The data field			
Row heading		●		
Row heading		○		
Row heading		◐		
A “cut-in heading”				
Row heading			Adverse events included headache and fever	
Row heading			No adverse events reported	
Expanded abbreviations go here				
Footnotes go here. Common footnote symbols: ^{a, b, c, d, e} ; and *, †, ‡, §, , ¶, **, ††, ‡‡, etc.				

Figure 1 The components of a table. In addition to numbers, tables can use symbols and text to communicate information.

should provide enough information and be specific. For example, a title such as “Cancer incidence” is too broad and contains no specific information, whereas “Cancer incidence in China, 2000 through 2010, by age group and sex” clearly identifies the data in the table.

The topic of variables and study subjects need to be defined in the title. Naming the variables in the title duplicates the row and column headings. Instead, use a general term to cover all variables will make the title concise. For example, in the title “Healthcare resource information of the participating hospitals in 10 provinces in China”, the term “healthcare resource information” covers the variables, such as the numbers of physicians, nurses, and hospital beds; daily outpatient visits; the number of surgeries; financial data; payment models, and so on.

When a table contains data of both independent and dependent variables, they both should be mentioned in the title (3). For example, the title “Relationship between effective predictors and mortality of cancer patients analyzed using the multiple logistic regression model” defines the independent variable “effective predictors” and the dependent variable “mortality.”

Organizing the body of the table

The body of a table should be designed to help readers find

and interpret the data. Thinking about how readers would understand the table is important before you start to design the table. The body of a table is composed of column heads, row heads, and the data field.

Column heads and row heads should identify the variables being presented. The heading of the first column identifies the row heads below. For example, “Disease type” as a heading of the first column covers various types of disease that are listed in row heads. Spanner headings, which apply to two or more columns, allow data to be presented for subgroups (Table 1).

Similar to column headings, row headings should identify the data in the rows. When several row subheads are needed, they can be indented under a stub heading or introduced with a “cut-in” heading that crosses the entire data field (Figure 1).

Columns and rows need to be organized logically according to the purpose of data presentation. Although not always possible, ordering row or column headings alphabetically may help readers find specific data faster (Table 2). On the other hand, ordering the data from the highest to the lowest value, or vice versa, may help reveal trends (Table 3).

Tables are efficient in presenting a large amount of data and may be designed to show comparisons, trends, or patterns. Numbers, words, letters, and symbols can be

Table 1 Lung cancer incidence in the United Kingdom

Period	Incidence per 100,000 people in rural counties			Incidence per 100,000 people in urban counties		
	Total	Men	Women	Total	Men	Women
2000–2004						
2005–2009						
2010–2014						

Table 2 Mortality rates of the 10 most common cancers in China in 2012

Cancer site	Crude mortality (1/10 ⁵)	ASRMC (1/10 ⁵)	ASRMW (1/10 ⁵)	Cumulative rate (%)
Bladder				
Brain				
Colorectum				
Esophagus				
Liver				
Lung				
Lymph nodes				
Pancreas				
Prostate				
Stomach				

Table 3 Mortality rates of the 10 most common cancers in China in 2012

Rank	Cancer site	Crude mortality (1/10 ⁵)	ASRMC (1/10 ⁵)	ASRMW (1/10 ⁵)	Cumulative rate (%)
1	Lung				
2	Stomach				
3	Liver				
4	Esophagus				
5	Colorectum				
6	Bladder				
7	Prostate				
8	Pancreas				
9	Brain				
10	Lymph nodes				

Table 4 Baseline characteristics of breast cancer patients diagnosed with and without symptoms

Characteristic	Symptomatic group [cases (%)]	Asymptomatic group [cases (%)]	P value
Total			
Age (years; mean \pm SD)			
Sex			
Men			
Women			
T stage			
T1–2			
T3–4			
...			

used in the data field. Conventionally, comparisons are made between adjacent columns or between adjacent rows; interlaced comparisons are difficult and should be avoided (3).

The units of measurements should be included in either row or column headings as appropriate and need not follow individual values in the table. Units in the column or row headings must apply to every cell in the column or row. When the units of only one or two variables differ from those of the other variables in a table (*Table 4*), these variables may be removed from the body of table and presented either in a footnote or the text. For example, when presenting demographic and clinicopathologic data of patients, the unit of measurements for most variables is the number of patients with percentage in parentheses, whereas age is reported as mean and standard deviation or median with range in parentheses. In such a case, “age” could be removed from the body of table, with data being presented either in table footnotes or the text.

The precision of data (the number of decimal points given) of the same variable should be consistent, such as means, standard deviations, percentiles, and confidence intervals. Unnecessary precision increases the time needed to recognize the trend or make comparisons, and thus is usually discouraged. Generally, one or two decimal places after the decimal point is enough. For example, incidence rates of 37.8903 per 100,000 population in city A and 18.6247 per 100,000 population in city B should probably be rounded to 37.9 per 100,000 and 18.6 per 100,000, making it easier to see that the incidence rate in city A is twice of that in city B. The data precision as well as the units of measurement used in tables should be consistent with those used in the text.

Group sizes can be either presented in a single row or column or included after column or row headings to help readers understand the data.

Rows or columns that contain identical values in all corresponding cells (that is, there is no variation in values in the row or column) can probably be removed from the table. Instead, the relevant data may be briefly mentioned in a footnote or in the text. For example, when the sample size was the same (e.g., 50 patients) for all groups and all values are presented as means and standard deviations, the column “Sample size” can be removed from the table, and a statement “All values are presented as mean \pm standard deviation for 50 patients in relevant groups” can be added to footnotes.

When a table is too wide to fit in a journal page, columns and rows can be reversed. Some journals may allow using wide tables and rotating them by 90 degrees to fit in the journal’s page. However, such typesetting is inconvenient to readers and is best avoided when possible (3).

The data field should not contain empty cells unless no value could logically be expected. When some data are missing, the missing status should be clearly indicated with an ellipsis (...) in the cell and accompanied by a footnote to indicate that the data were “not available”, “not detectable”, “not applicable”, or else, which can sometimes be abbreviated in the table. Numerical data need to be aligned consistently in cells by the decimal point for the convenience of comparisons. When the data contain mathematical symbols such as “ \pm ” (usually followed by standard deviation), “–” (generally used to indicate ranges if there is not enough space to write “to,” which is preferable), and “ \times ” (usually followed by a unit multiplier such as 100 or

Table 5 Detailed characteristics of the 6 studies on lung cancer included in the meta-analysis

Author	Nationality	Total cases	Tumor size (cm)	PDCD4 expression (cases)	Examine methods	Dichotomic criteria
Chen <i>et al.</i> (ref No.)	USA	124	2 [2–6]	21	IHC	Positive/negative
Mudd <i>et al.</i> (ref No.)	Germany	71	3 [1–7]	13	qPCR	High/low
Wang <i>et al.</i> (ref No.)	China	43	3 [2–6]	18	IHC	Strong, moderate, weak, negative
Kin <i>et al.</i> (ref No.)	Korea	84	3 [1–9]	16	IHC	Positive/negative
Wei <i>et al.</i> (ref No.)	China	79	4 [2–7]	40	qPCR	High/low
Motoya <i>et al.</i> (ref No.)	Japan	105	3 [1–10]	28	Western blotting	High/low

The data of tumor size are presented as median, with range in square brackets. PDCD4, programmed cell death protein 4; IHC, immunohistochemistry; qPCR, quantitative polymerase chain reaction.

10,000), they could be aligned by either the decimal point or the mathematical symbol to better illustrate differences among data. Text longer than a few words should be left-justified (*Table 5*).

Composing footnotes

Footnotes are placed below the body of a table, after any expanded abbreviations, and provide further information about some entry in the table. In public health and medical journals, footnotes are generally indicated with two sets of marks: superscript, lower-case letters used in alphabetical order (^a, ^b, ^c, ^d, ^e) and symbols (*, †, ‡, §, ||, ¶, **, ††, ††, etc.). Avoid using either regular or superscript numerals, which can be mistaken for exponents.

Statistically significant differences are sometimes indicated with symbols in the data field and explained in footnotes. For example, the superscript symbol “**” can be added after the data of the test group, and the explanation footnote can be written as “*, $P < 0.05$ vs. control group (or other groups) by Chi-square test.” The comparison (e.g., between test group and control group in the above example) should be clear in the explanation statement. P values should be expressed as equalities (e.g., $P = 0.03$) rather than as inequalities (e.g., $P < 0.05$) when possible. Also, many journals prefer 95% confidence intervals to P values because they keep the interpretation focused on the implications of the size and precision of the estimated result and away from chance as an explanation for the result.

When preparing any components of a table (i.e., table title, row and column headings, and footnotes), the key terms should be consistent with those used in the text (1), allowing readers to easily link the messages in the table with

the text. In research papers, some terms are exchangeable with unnoticeable differences (e.g., recurrence and relapse). No matter which specific terms are chosen, use them consistently in all parts (including tables and figures) of the paper would improve the understandability.

Common problems with tables

- ❖ The amount of data does not justify presenting them in a table. When you have only a few data points, present them in the text rather than in a table;
- ❖ The title does not describe the data;
- ❖ The title unnecessarily repeats the row or column headings. Such titles usually look too wordy. Use terms that can summarize the topic of involved variables will make the title concise;
- ❖ The heading of the first column is missing;
- ❖ Row or column heads are not organized sensibly;
- ❖ Column headings do not clearly define the nature of data in the column. For example, in *Table 6*, the column headings “2004–2010” and “2011–2017” are two periods, which failed to define the nature of the data (scores) in those two columns. As a solution, a spanner heading “Google Trends score of market shares” is used in *Table 7*;
- ❖ Headings are missing units of measurements. Without units of measurements, the data cannot be interpreted correctly;
- ❖ Column or row totals are missing or incorrect;
- ❖ Column or row integrity has been violated by one or more cells that are inconsistent with the row or column headings;
- ❖ Entries that should be included in the same cell (e.g., means and standard deviations) are presented in

Table 6 Google Trends score of market shares between 2004 and 2010 and between 2011 and 2017 for four drugs

Drug	2004–2010	2011–2017	P
Drug A			
Drug B			
Drug C			
Drug D			

Table 7 Google Trends score of market shares between 2004 and 2010 and between 2011 and 2017 for four drugs

Drug	Google Trends score of market shares		P
	2004–2010	2011–2017	
Drug A			
Drug B			
Drug C			
Drug D			

Google Trends scores are presented as median with interquartile range in parentheses.

different cells;

- ❖ The “Enter” key is used to create returns to make table cells, or spaces are used to simulate table cells. Although such tables may look normal on computer, their structures are not stable and may be changed during typesetting, leading to errors in data presentation and misunderstanding of the results;
- ❖ Blank cells exist in a table, which could be interpreted by different readers as not detected, detected but not detectable, data not available, method not applicable, other possibilities, or even errors caused during typing or typesetting. When no data could be presented in certain cells, a symbol may be used in those cells and be defined in footnotes;
- ❖ Inconsistencies between tables and the text widely exist (4). Such inconsistencies will lead to questions on the reliability of the research. Except for the reasons of unethical issues and typing errors that shouldn’t have happened, data inconsistency may be resulted from incorrect data updates. Especially in large-scale ongoing studies, data updates happen frequently during a study. Paper revisions may need data updates. However, some authors may have updated the data in the text but not in the tables, or vice versa. Cross checking all data

throughout the whole paper (i.e., checking data in the Abstract against the text and checking data in the text against visual media such as tables and figures) is always necessary;

- ❖ Data are repeatedly presented in the text and tables as well as figures. Such duplicate presentation should be avoided (2). For example, when all mortality data are presented in *Table 3*, it reads better to present the result “The crude mortality was the highest for lung cancer, followed by gastric cancer and liver cancer” than to repeat the data “The mortality was XX for lung cancer, YY for gastric cancer, and ...” in the text;
- ❖ Abbreviations are not spelled out below the data field. Abbreviations may be used in column and row headings as well as the data field. They need to be defined according to the sequence (from left to right and from the top down) they appear in the table.

Acknowledgments

I highly appreciate Mr. Tom Lang for his critical comments and helpful edits on this article.

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the the Guest Editor (Thomas A. Lang) for the series “Publication and Public Health” published in *Journal of Public Health and Emergency*. The article has undergone external peer review.

Conflicts of Interest: The author has completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/jphe.2018.01.01>). The series “Publication and Public Health” was commissioned by the editorial office without any funding or sponsorship. The author has no other conflicts of interest to declare.

Ethical Statement: The author is accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International

License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Wainer H. Understanding Graphs and Tables. Ed

doi: 10.21037/jphe.2018.01.01

Cite this article as: Liu W. Preparing better tables. J Public Health Emerg 2018;2:3.

- Researcher 1992;21:14-23.
2. Lang TA, Secic M. How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers. 2nd Edition. American College of Physicians, 2006.
 3. Zeiger M. Essentials of Writing Biomedical Research Papers. 2nd edition. McGraw-Hill Companies, 2000.
 4. Thimbleby H, Cairns P. Reducing number entry errors: solving a widespread, serious problem. J R Soc Interface 2010;7:1429-39.