Using Tables and Graphs for Reporting Data

Article *in* Journal of the Association of Physicians of India · October 2015

CITATIONS
8

READS
17,103

1 author:

Sandeep B Bavdekar

199 PUBLICATIONS **2,443** CITATIONS

SEE PROFILE

ART OF WRITING

Using Tables and Graphs for Reporting Data

Sandeep B Bavdekar

Abstract

For presenting the study results, authors need to choose from one of the three formats: text, tables and graphs. The underlying consideration of any author while making the choice is to ensure that readers find it easy to understand, to assist the author present data in a way that would catch the reader's eye, hold his interest and enhance his understanding. The choice should, therefore, be based on the advantages and limitations of each of these formats. Most of the observations are reported in textual format. Tables are useful when comparisons are to be shown. Graphs attract readers' attention better and the data they depict remains in the reader's memory. The type of graph used is dependent upon the nature of data that is to be shown.

Every research paper has a lot to say through the data that has been collected and analyzed. However, this data needs to be presented in a logical, coherent and easy-to-understand manner. The way data is presented can influence the reviewers' and editors' decision. It can also make the difference between a paper being appreciated by the readers or being neglected or even trashed by them. Tables and Graphs are the two important tools that authors use to make the data more presentable and easy to read and understand.

Tables

Tables are an arrangement of words, numbers, or signs, in parallel columns, used to depict data or relationships. Tables provide a summary of a set of data or parameters.[1] They are primarily included in the Results section, but can also be seen occasionally in the Methods and Discussion sections. In the Results section they are used for providing comparison of baseline characteristics of the study and control groups and to depict observations of the research study. They are good for presenting large amounts of data that would otherwise be very unwieldy and confusing to be described in the textual format.[2] In the Methods section, Table is often used to provide definitions of the terms used in the protocol (e.g. diagnostic criteria, etc.), while in the Discussion Section they are used to depict results of similar studies allowing the reader to compare the results of several studies at a glance.

Anatomy of a Table

Tables generally have five components. The title, heading or caption of the Table, the row headings (also called as stubs), the column headings, the data fields and the footnote. Sometimes, the Table contains a spanner, which is a common heading for some or all the column headings.[2] Authors should be aware of these terms as reviewers, copy editors and printers frequently request authors to amend the tables in the manuscript using some of these terminologies (Figure 1).

The title of a table should be informative. It should tell the

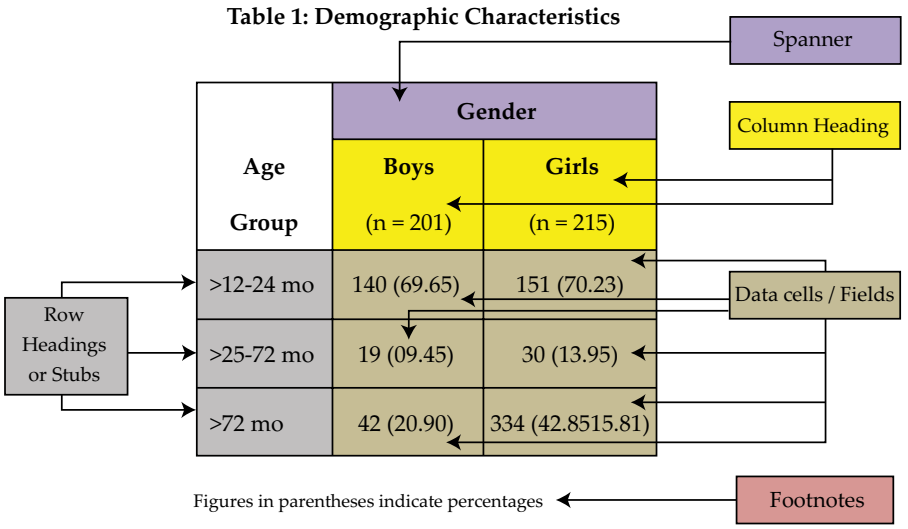

Fig. 1: Anatomy of the Table

Professor and Head, Department of Pediatrics, Topiwala National Medical College and BYL Nair Ch. Hospital, Mumbai, Maharashtra
Received: 15.07.2015; Accepted: 25.07.2015

reader about what data to expect in the table. However, it should not describe the data presented in the table and provide details of the data or declare what has been found. For example, a title mentioning adrenaline and rats is too cryptic and does not provide any worthwhile information to a reader. Such a title will force the reader to scan the methodology section to find out about the context of the data presented in a table. On the other hand, a title stating "The effect of drug A on weight loss (statistically significant, P< 0.001) as compared to Drug B with no significant effect on cholesterol levels" is saying too much. The titles may include names of the drugs used, time period, endpoint measured; but should not include experimental details, actual data, names of statistical tests used or results. Annesley[2] puts this succinctly when he says "Data belong in the data fields of the table" and not in the title.

The row headings (or the stubs) are listed in the far left column of the table. Here, one or more variables are listed and generally these are used to name independent variables. The row headings should be left-aligned. The other columns on right of the row, link the respective row heading with individual column heading. The row-heading could also include the units of measurement if it is common to the whole row. The columns generally represent dependent variables. It is also acceptable to state the units in the column heading if the data in the columns have common units of measurements. Column headings are generally 'centered'. No data cell or field should be kept vacant. It is difficult for the reader to understand what a missing entry represents. If the data is no applicable, it must be clearly stated.[2] The alignment of entries in the data fields could vary and depend upon the type of the data. Words in the data cells should be left aligned, whole numbers could be right-aligned while fields containing decimal points could be aligned on these points.[2] When the text in the row heading goes into a second line, the data in the corresponding columns should align with the top line of the text of the row heading.

Footnote is an important part of the table. As the name suggests, it is positioned at the bottom of the table. Footnote includes the explanatory matter. It can be used to explain the non-standard abbreviations. Each footnote is placed in a separate line. One can use symbols, super-scripted Arabic numbers or super-scripted alphabets to indicate the matter that is being explained. Usually, authors do not have a choice regarding this. The "instructions to authors" indicate what is to be used. If symbols (*, †, ‡, §, ||, ¶, **, ††, ‡‡) are preferred by the Journal, most of them even prescribe the order in which they are to be used. Usually, journals prefer to use super-scripted letters or alphabets to numbers; since the latter could be confused with an exponent.[2] Appropriate placement of symbol indicating a particular footnote is also crucial. A symbol placed in the column identifier applies to the entire column and that placed in the row heading applies to the whole row. If a footnote applies to a particular cell, then the symbol should be placed only in the concerned data field after the entry for that field.[2] If a footnote applies to the entire table, the corresponding symbol should be placed after the title.

Functions of a Table

As stated earlier, a table presents the data in an organized fashion, making it easier for the reader to read and understand it. Thus, tables come in handy when there is an enormous amount of data to be presented or when data is highly complex. Tables are useful, especially when various parameters in two or more groups are to be compared. Thus, a table comparing baseline characteristics is generally the first table in the Results section of an article reporting a randomized control trial. Tables are useful for a casual reader who intends to just browse through the article. For such a reader, just a glance at the Tables is sufficient for grasping the data presented in the research article. And it is for this reason that every Table has to be a self-sufficient and stand-alone one: It should be designed in such a manner that a reader perusing it should be able to understand the data presented in the table, without having to read the sections of Methods and Results sections. A table serves another important function for the authors. They are an efficient way of conveying data. Several paragraphs will have to be written to explain what can be presented in a single table. In addition, many authors find it very difficult to be brief and restrict the manuscript word count to the prescribed limits. As words and numbers in tables are not counted in the calculation of manuscript word count, it is useful to have tables for this purpose. In addition, a footnote being a part of the table, words in the footnote are not counted in the manuscript's "word count limit". Many authors exploit this concession to the hilt, by placing a large amount of explanatory material in the footnotes. One should be aware of the limitations of tables as well. They are not useful for conveying ideas. Also, they are not good at depicting changes occurring over time (trends). Graphs do that job better. Also, graphs have an immediate and greater visual impact than tables.

While incorporating tables in the article, the authors should decide whether the data requires to be presented as table or textual matter would suffice. It is worth remembering that factors such as the type and complexity of data and the precision with which it needs to

Table 1: How to keep table small?

- Consider if all the data that has been included in the table, needs to be presented. Remove non-essential data
- Check if all the columns and rows need to be included. Remove columns and rows that have limited utility or information
- Combine information into a single column or row. For example, information regarding age and gender ratio can be put in a single column
- Whenever possible, use abbreviations
- Reorient the Table by reversing the variables
- Place common units in the row-headings or column-headings
- Condense contents: Consider if the text in the stub can shortened
- Check if data or information in a particular row/ column can be placed the Footnote
- Consider placing the less important data in the online version only as supplementary material
- Split the table.

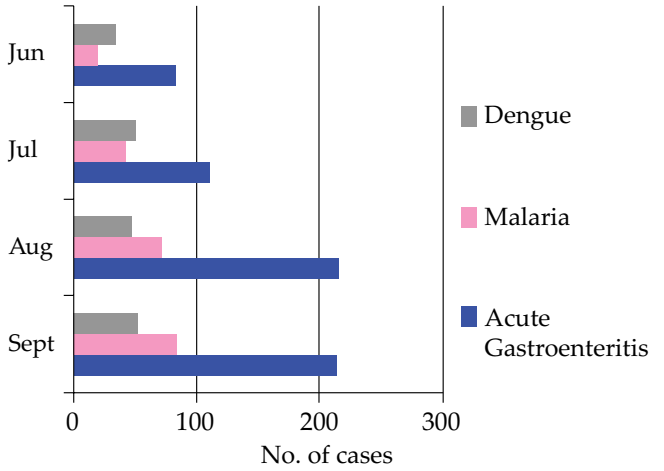

Fig. 2: Bar graph

be reported, will determine whether to depict the data in the form of textual matter, tables or graphs. In any case, data depicted in the table should never be repeated in the textual format. The authors should ensure that the manuscript has only as many tables as required to tell the story. Tables should also be kept small in size. Tables that run into pages are unlikely to be read and in the process the message that authors wish to convey could get lost. Certain checks and maneuvers will help authors ensure that tables are kept small (Table 1).

Graphs

Graphs help readers see, understand and remember the data, better. They are used for depicting outcomes, relationships and trends. The authors need to know when and what graph to use. This decision is based mainly on the type of data. Graphs consist of axes, labels, scales, an origin, symbols (or data points) and leg-end. A good graph should have easily readable data points and connecting lines, visually balanced and clearly decipherable axes and legible legends. These qualities ensure that readers understand the data shown in the graph without having to refer to the textual matter. But the most important attribute of a good graph is that it draws attention to the data and not to itself![3]

The graphs that are commonly used in medical literature include: line graphs, bar diagrams, histogram, pie charts and scatter diagrams. ***Line graphs*** (also called as time series graphs) use line segments that connect various data points to show changes over time. For example, a line graph could effectively depict changes in blood sugar levels over a period of time after a meal. Generally, X-axis would be used for time while ordinate would provide the value of the variable. A number of lines (usually up to five) could be used to depict changes in the concentration in different populations (say, normal adult males; normal adult females, pregnant women, adult diabetic males on insulin, adult diabetic females) over time.

Histograms constitute a graphical representation of the distribution of numerical data. The frequencies are shown as adjacent rectangles, erected over discrete intervals (also called as bins). The area of each rectangle is proportional to the frequency of the observations in the interval. The categories are usually specified as consecutive, non-overlapping intervals of a variable. The categories (intervals) must be adjacent, indicating that the original variable is continuous. When intervals are chosen to be of the same size, the heights of the rectangles can also denote the proportional frequencies of the intervals. A glance at a histogram helps the reader determine if the pattern of distribution of the data is symmetrical, skewed (left or right), unimodal, bimodal or multimodal. A ***Frequency polygon*** is formed by joining the midpoints of each rectangle in a histogram

A ***Bar graph*** is used to depict Qualitative Data (Figure 2). The categories or traits are listed on one axis and the frequencies on the other. A bar graph emphasizes the relative sizes of each of the categories using horizontal or vertical bars. They are most informative when the authors want to draw the readers' attention of actual value of the variable than its confidence interval (CI).[4] While using bar graphs, certain style issues need to be addressed. The space between the bars should be narrower than the bars so that gaps do not dominate the figure. If space within the bars is white and that in the gaps is also white, readers would find it difficult to grasp the difference. Hence, shading the bars is useful, but the shading within the bars should be pleasing to the eyes. One must desist from the use of suppressed zero scale (scale that does not include zero), as it exaggerates the differences among groups.[4] Three-dimensional bars look impressive, but they usually do not add value and many a times, reduce the clarity.[4]

A ***pie chart*** is a circular chart divided into sectors, illustrating

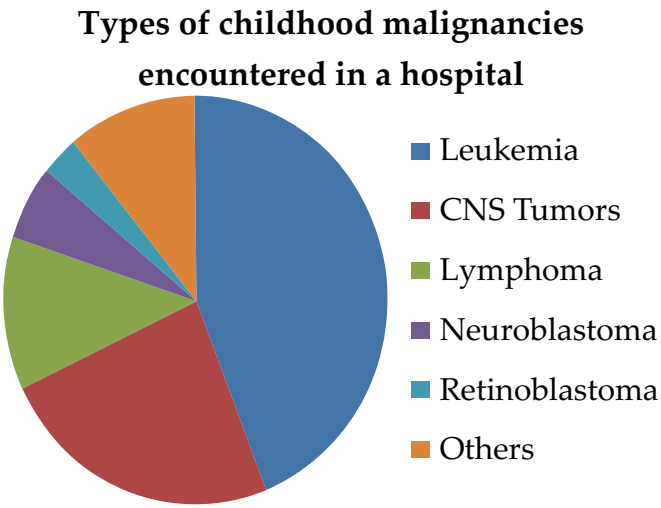

Fig. 3: A pie chart

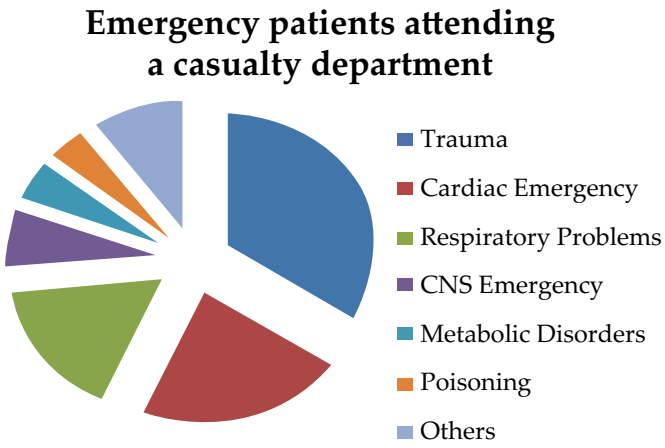

Fig. 4: An exploded pie chart

numerical proportion (Figure 3). The arc length of each sector (and consequently its central angle and area), is proportional to the quantity it represents. Pie charts, though popular, are not useful in most scientific papers.[4] A pie chart takes up relatively larger amount of page space as compared to bar graphs. It can be used when there are only a few sectors to depict. If there are too many sectors to be shown, the relative sizes cannot be discerned clearly. An exploded pie chart (Figure 4) and a donut (or doughnut) chart (Figure 5) are the varieties of pie chart. In an exploded pie chart, the sectors are separated from each other, while as the name suggests, the central portion is kept "hollow" in a doughnut chart. Other variants of pie chart such as polar area diagram, multilevel pie chart and spie chart are rarely used in medical literature.

A ***Scatter plot*** (also called as scatter plot) is useful when an author wants to depict relationship or association between two continuous variables. For example, it would be a good idea to use a scatter plot to depict the relationship between calf circumference and birth weight in the newborns (Figure 6). The scatter plot provides the readers with a visual impression about the type of association (positive-, negative-, complex- or no- correlation) between the variables studied.[5]

Some other graphs are also used in medical literature, albeit less frequently. These include a stem-and-leaf display (or a stemplot), box plot, a Forest plot and a Kaplan-Meier survival curve. A ***Stem plot*** is used for representing quantitative data in a graphical format. It is similar to a histogram, but has the additional advantage of preserving the original data. Stem plot was quite popular in the 1980s, as it could be made using typewriters. With computers making complex graphics possible, it has somewhat lost its attractiveness. A ***box plot*** is used for illustrating groups of numerical data through their quartiles. It shows the distribution of the data and the reader can get an idea about the dispersion of the data and whether it is symmetrical or skewed. It allows the readers to know about outliers, and estimate

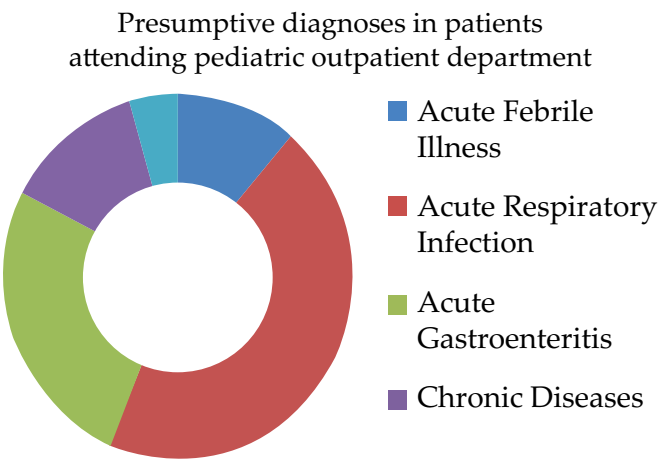

Fig. 5: A doughnut chart

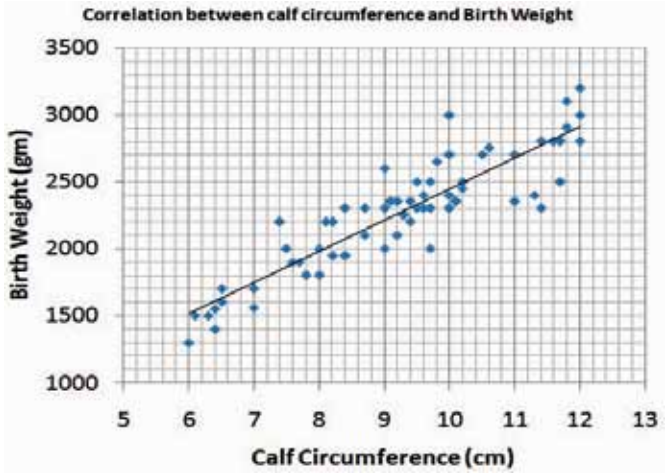

Fig. 6: Scatter plot

range and inter-quartile range. A ***Forest Plot*** is used mainly to depict data of meta-analysis of randomized controlled trials (RCTs) or epidemiological studies. It shows the information from the individual studies included in the meta-analysis, the amount of variation among the study results and an estimate of the overall result.[6] ***Kaplan Meier survival curve:*** Survival analysis represents a group of methods used for analyzing data where the outcome variable is "time to the occurrence of an event (of interest)". This data is graphically shown using the Kaplan-Meier Survival curve.

While deciding to present data in a research paper, the key question that authors face is what is the best way to show a particular set of data? They can do so using one or more of the three formats: in text or tables or through graphs (charts or diagrams). Every format has its own advantages and disadvantages and choosing the modality for describing the data is a matter of judgment. Textual format is most useful when the data is not voluminous or complex. Tabular format is most attractive when one needs to show comparison between groups (especially when details about the data are important). Graphs, figures and diagrams have the following merits: they catch the reader's eye, put the data in proper perspective and increase the reader's ability to understand and recall data. However, graphs are limited by the fact that minute differences in values cannot be appreciated easily, and readers can be misled if they do not pay attention to issues such as the scale used.[7] Once it is concluded that a graph is the best way to portray the data, the type of graph to be used is dictated by the kind of data.

References

1. Boyd JC, Rifai N, Annesley TM. Preparation of Manuscripts for Publication: Improving Your Chances for Success. *Clin Chem* 2009; 55:1259-64.
2. Annesley TM. Bring Your Best to the Table. *Clin Chem* 2010; 56:1528-34.
3. Annesley TM. Put your Best Figure Forward: Line Graphs and Scattergrams. *Clin Chem* 2010; 56:1229-33.
4. Annesley TM. Bars and Pies Make Better Desserts than Figures. *Clin Chem* 2010; 56:1394-400.
5. Scatter Diagram. http://web2.concordia.ca/Quality/tools/25scatter.pdf [Last accessed on June 19, 2015]
6. Lewis F, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001; 322:1479-80.
7. National Agriculture Innovation Project. Uses and limitations – simple, Multiple, Component and percentage bar diagrams–pie chart. http://agridr.in/tnauEAgri/eagri50/ STAM101/lec02.html [Last accessed on June 19, 2015].