Editorial

The Effective Use of Graphs

David J. Slutsky, MD, Editor-in-Chief[1]

[1] The Hand and Wrist Institute, Torrance, California; Assistant
Professor, Department of Orthopedics, Harbor-UCLA Medical Center,
Los Angeles, California

Graphs are a common method to visually illustrate relationships in the data. The purpose of a graph is to present data that are too numerous or complicated to be described adequately in the text and in less space. Do not, however, use graphs for small amounts of data that could be conveyed succinctly in a sentence. Likewise, do not reiterate the data in the text since it defeats the purpose of using a graph. If the data shows pronounced trends or reveals relations between variables, a graph should be used. If the data doesn't show any significant trend in the evidence, a graph is not the figure of choice.[1]

Although there are myriad computer programs that can generate a graph, the author must still heed some basic principles. A basic requirement for a graph is that it is clear and readable. This is determined not only by the font size and symbols but by the type of graph itself. It is important to provide a clear and descriptive legend for each graph. Graphs may have several parts, depending on their format: (1) a figure number, (2) a caption (not a title), (3) a headnote, (4) a data field, (5) axes and scales, (6) symbols, (7) legends, and (8) a credit or source line. For most purposes, design a graph so that the vertical axis (ordinate, Y axis) represents the dependent variable and the horizontal axis (abscissa, X axis) represents the independent variable. Hence, time is always on the X axis.[2] Graphs should always have at minimum a caption, axes and scales, symbols, and a data field. Plotting symbols need to be distinct, legible, and provide good contrast between the figure in the foreground and the background. Open and closed circles provide the best contrast and are more effective than the combination of open circles and open squares.[3] Like the title of the paper itself, each legend should concisely convey as much information as possible about what the graph tells the reader, but it should not provide a summary or interpretation of the results or experimental details. Avoid simply restating the axis labels, such as "temperature vs. time." It is crucial to choose the correct graph type based on the kind of data to be presented. If the independent and dependent variables are numeric, use line diagrams or scattergrams; if only the dependent variable is numeric, use bar graphs; for proportions, use bar graphs or pie charts. These are briefly described below.

A *scattergram* is used to show the relationship between two variables and whether their values change in a consistent way, such as analyzing the relationship between the concentration levels of two different proteins.

A *line graph* is similar to the scattergram except that the X values represent a continuous variable, such as time, temperature, or pressure. It plots a series of related values that depict a change in Y as a function of X. Line graphs usually are designed with the dependent variable on the Y-axis and the independent variable on the horizontal X-axis, such as a Kaplan-Meier analyses survival plots of time-to-event outcomes. The proportion of individuals is represented on the Y-axis as a proportion or percentage, remaining free of or experiencing a specific outcome over time.

A *bar graph* may consist of either horizontal or vertical columns. The greater the length of the bars, the greater the value. They are used to compare a single variable value between several groups, such as the mean protein concentration levels of a cohort of patients and a control group.

The *histogram*, also called a *frequency distributions graph,* is a specialized type of bar graph that resembles a column graph, but without any gaps between the columns. It is used to represent data from the measurement of a continuous variable. Individual data points are grouped together in classes to show the frequency of data in each class. The frequency is measured by the area of the column. These can be used to show how a measured category is distributed along a measured variable. These graphs are typically used, for example, to check if a variable follows a normal distribution, such as the distribution of protein levels between different individuals of a population.

A *pie chart* shows classes or groups of data in proportion to the whole data set. The entire pie represents all the data, while each slice or segment represents a different class or group within the whole. Each slice should show significant variations. The number of categories should be generally limited to between 3 and 10.

A *box plot* may be either horizontal or vertical. It is used to display a statistical summary of one or more box-and-variables, such as the minimum, lower quartile, median,

Address for correspondence
David J. Slutsky, MD, The Hand and
Wrist Institute, 2808 Columbia
Street, Torrance, CA 90503
(e-mail: d-slutsky@msn.com).

and maximum. It may also identify the outlier data. The spacing between the different parts of the box indicates the degree of dispersion and whether the data distribution is symmetrical or skewed.

Some common errors include the following: information in the text is duplicated in graphs, or information in graphs is duplicated in tables. The graph does not have proper legends. The wrong type of graph is chosen to represent the data. The graph is not plotted to scale. Data is not labeled, is inconsistent, interrupted, or exaggerated to produce the desired effect. Another common error is to include a line that suggests an unsubstantiated extrapolation between or beyond the data points. Connecting discrete data points with a continuous line, such as a series of average measurements taken from a group of patients, suggests that there are values between the age groups that fall on the lines, when, in fact, the author cannot know this. A better way to display separate values would be a bar chart, in which each column reflects the average value obtained from each age group.[4] If an extremely large range must be covered and cannot be practically shown with a continuous scale, indicate a discontinuity in the scale and the data field with paired diagonal lines (−//−) indicating a missing extent of the range.[2]

References

1 Ng KH, Peh WC. Preparing effective illustrations. Part 1: graphs. Singapore Med J 2009;50(3):245
2 Council of Science Editors' Scientific Style and Format for Authors, Editors, and Publishers. 8th ed. Chicago: University of Chicago Press; 2014
3 Manual of Style: A Guide for Authors and Editors. 10th ed. Oxford: Oxford University Press; 2007
4 Franzblau LE, Chung KC. Graphs, tables, and figures in scientific publications: the good, the bad, and how not to be the latter. J Hand Surg Am 2012;37(3):591–596