Graphs, Tables, and Figures in Scientific Publications: The Good, the Bad, and How Not to Be the Latter

Lauren E. Franzblau, Kevin C. Chung, MD, MS

Graphs, figures, and tables can save readers time and energy, aid their understanding of an article, and reduce the word count of the main text. However, many graphics submitted to and published in scientific journals fail to meet their potential and include mistakes that jeopardize their clarity. Many formats are available for presenting data, as well as a variety of techniques for enhancing interpretability. When the appropriate format is used to depict data, it conveys the greatest amount of information in the clearest fashion, complements the text, and deepens readers' understanding. The aims of this article are to draw attention to the necessity of well-constructed graphs, tables, and figures in scientific publications, and to show how to create them. *(J Hand Surg 2012;37A:591–596. Copyright © 2012 by the American Society for Surgery of the Hand. All rights reserved.)*

Key words Graphs, tables, charts, figures, manuscripts.

IN THE HEART OF EVERY research article is its data, the information that gives rise to conclusions. However, without proper presentation, the data can be misinterpreted or, worse, ignored. To elude this trap, authors must use well-constructed graphs, figures, and tables to display data and trends, and to summarize information. Illustrations, or graphics, have 3 main purposes and advantages over text: first, they portray complex data and relationships in a way that is easier to interpret and understand[1–8]; second, they reduce reading time by summarizing and highlighting key findings or trends so that they are readily visible[5,6]; and third, they reduce the overall word count.[3,4,7]

From the Section of Plastic Surgery, Department of Surgery, University of Michigan Health System; Ann Arbor, MI.

Received for publication November 11, 2011; accepted in revised form December 27, 2011.

Supported in part by grants from the National Institute on Aging and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01 AR062066), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01 AR047328), and a Midcareer Investigator Award in Patient-Oriented Research (K24 AR053120) (to K.C.C.).

No benefits in any form have been received or will be received related directly or indirectly to the subject of this article.

Corresponding author: Kevin C. Chung, MD, MS, Section of Plastic Surgery, Department of Surgery, University of Michigan Health System, 2130 Taubman Center, SPC 5340, 1500 E. Medical Center Drive, Ann Arbor, MI 48109-0340; e-mail: kecchung@umich.edu.

0363-5023/12/37A03-0037$36.00/0
doi:10.1016/j.jhsa.2011.12.041

Well-designed graphics make data easier to understand by explicitly depicting trends and key points, and they are valuable assets to any paper or presentation. These visual aids enrich readers' understanding and improve the accuracy of their interpretations of the data.[9] On the other hand, poorly fashioned graphics slow down retrieval of data and lead to incorrect interpretations of findings.[4,6,8] Thus, it is imperative that the quality of graphs, figures, and tables that are submitted to and published in scientific journals be carefully considered and for data to be presented in such a way that strengthens the overall manuscript. The aims of this article are to bring to light common mistakes in graph, table, and figure construction, and to show how to create high-quality graphics that will enhance readers' acquisition and understanding of data.

WHAT NOT TO DO

Many submitted, and even published, manuscripts contain illustrations of "suboptimal quality."[8] Table 1 shows common errors that lead to this failure to fulfill graphical and tabular potential. The most glaring oversight is selecting the wrong format to present data. Suboptimal designs omit important information about the data that should be displayed.

Many computer programs offer graphing features that are easy to use and have a plethora of design elements. However, they tend to produce overcompli-

TABLE 1. Common Mistakes Found in Published Graphs, Figures, and Tables

Design Errors	Content Errors
Tables that are too large so that it is hard for readers to follow, or too simple so that the information should be included in the text[3,4]	Inclusion of nonessential data[4,6]
Failure to use shading and bordering in tables, when both techniques improve readability[1,4,7]	Redundancy with text[3,4,7]
Incorrect choice of graphical format or scale to portray data[1,3,4,6,8,10,12]	Excessive precision in tables (ie, including too many significant figures)[6–8]
Use of 3-dimensional graphs when 2 dimensions would suffice[6,10,11]	Not self-explanatory (ie, graphic cannot be fully interpreted when isolated from the main text)[1–3,7,8]
Design elements interfere with clarity of graph or figure[1–4,8]	Inadequate definition of symbols or abbreviations[2,3,8]

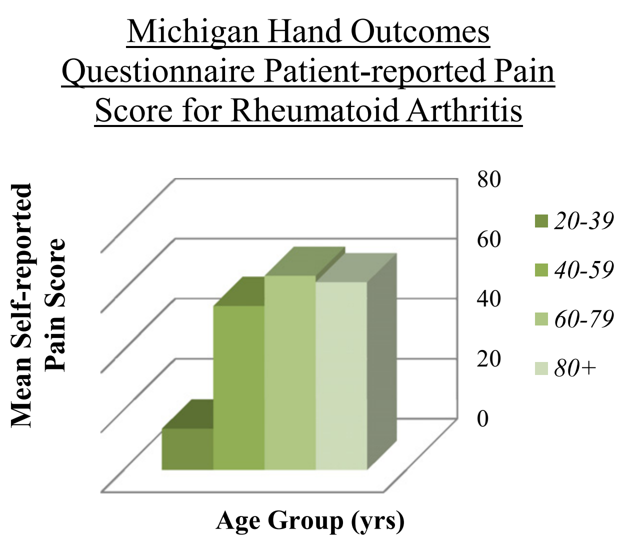

FIGURE 1: Three-dimensional bar graph.

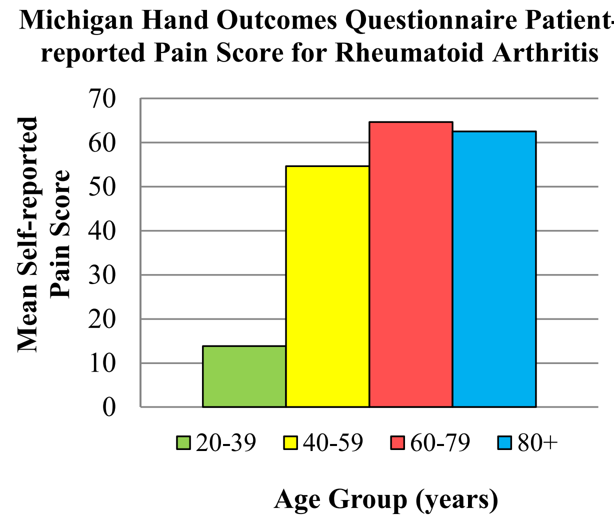

FIGURE 2: Two-dimensional bar graph.

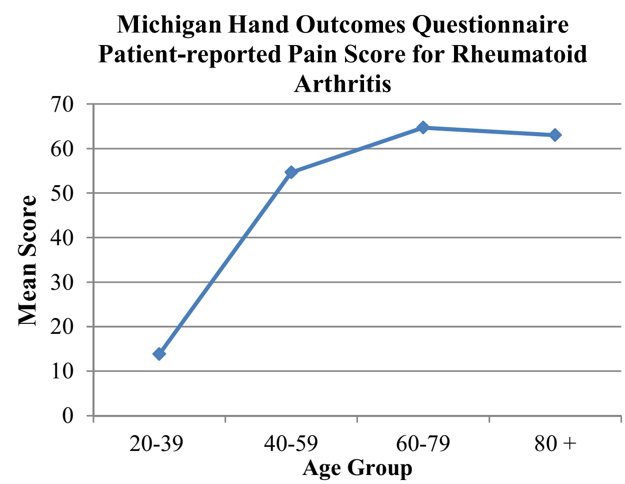

FIGURE 3: Connecting discrete data points.

cated, confusing graphics: in particular, the dreaded 3-dimensional graph with a skewed perspective (Fig. 1).[6,10,11] This flashy graph upstages the information it contains and distorts values as a result of its strange perspective. The color scheme of this graph, chosen by the computer, is monochromatic, which some readers would find challenging to distinguish. Figure 2 presents the same data in a 2-dimensional graph and is a better representation for the following reasons: the values are not distorted by the skewed perspective; the category labels are more space-efficient; the graph, not its title, occupies the most space; and the colors can be distinguished, even by a color-blind reader. Two dimensions, unless otherwise necessitated by the data, with a carefully chosen scale should be used to maintain simplicity.

Computer-made graphs can also contain errors that mislead readers: for instance, connecting discrete data points (eg, a series of average measurements taken from a group of patients) with a continuous line, as shown in Figure 3. The connecting segments suggest that there are values between age groups that fall on the lines, when in fact the author cannot know this.[4] 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, as in Figure 2. A similar graphical

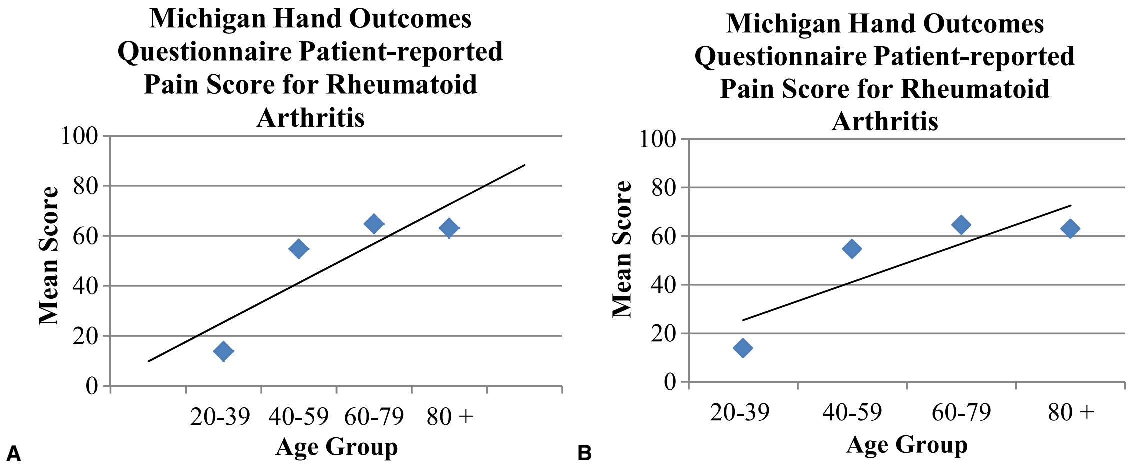

FIGURE 4: A Overextension of a linear regression line. **B** Appropriate linear regression line.

mistake is extending a regression line beyond the domain of the data. Figure 4 shows this mistake in the graph on the left and an example of a correctly fitted line to the right. Regression lines should not be extrapolated beyond the set of measured values.[4]

Unfortunately, mistakes can also be made outside the realm of actual data portrayal. When choosing a typeface, it is easy to get carried away with stylizing, overusing the bold, italic, and underline options, or selecting unconventional fonts and colors.[4] Although these techniques can distinguish important features, using too many in 1 graphic can make nonsignificant differences or values appear significant, and cause confusion.[4] The use of underlining, italicizing, and boldface type in Figure 1 is overwhelming, whereas the simple bolding of titles in Figure 2 does not detract from the graph. Moreover, using small typeface makes graphics hard to decipher and is not an appropriate way to save space. Critiquing the visual aids in other papers and journals can help authors learn to identify these mistakes and avoid making them in their own manuscripts.

WHAT TO DO

Before putting any data into a figure, graph, or table, ask yourself whether it *deserves* to be there, and how it will contribute to the article. Every illustration should enhance or supplement the text and provide necessary information.[1,7] Making tables or graphs that simply reiterate the text or contain extraneous data will only clutter the manuscript and burden readers. First decide what are essential data, and then select the best method to present them.

The way that data are presented can affect how accurately they will be interpreted.[9] Certain types of graphs, figures, or tables can be matched to the data at hand.[2,4,5,7,8,10,12] Tables, for instance, should be used when exact values are important, but graphs are better at depicting complex relationships.[7] As Cooper and Schriger[2] articulately state, "the distribution of the results should be the author's guides when choosing the graphic format that displays the optimal amount of detail necessary to accurately tell the story of the experiment." Table 2 shows when certain formats of data presentation are appropriate.

The following process outlines how to compare formats and find the optimal method to display a dataset. As seen in Figures 1 and 2, a 2-dimensional graph is a better fit for a univariate dataset. Despite outcompeting Figure 1, Figure 2 is still not the ideal format for portraying this information. Table 3 illustrates the same data in a more thorough, interpretable way than Figures 1 and 2. However, it, too, is flawed; it has too many significant figures, lacks units or references, and is a small, simple table that could easily be explained in text:

> Patients were divided into 4 age groups: 20 to 39 (n = 6), 40 to 59 (n = 11), 60 to 79 (n = 6), and 80 and older (n = 1). The mean patient-reported pain scores out of 100 were 13.8, 54.6, 64.7, and 63.0, respectively.

This text is clear and readily understandable, and it has greater data density than the table, but it still does not give a complete, satisfactory illustration of the data. The averages tell the reader very little about the dataset, sample, trends, and possible relationships among variables.

Table 4 and Figure 5 both show the entire dataset used in the previous example, explicitly pairing the patients' ages and pain scores. In this scenario, the scatterplot is the optimal representation of the data; it uses the least space, shows trends in the data, includes

TABLE 2. When and Why to Use Common Data Presentation Techniques

Type of Illustration	When to Use It	Advantages and Disadvantages
Tables		
Text table	Comparing different groups such as treatments or inclusion and exclusion criteria, or listing key points.	Summarize points and reduce the word count in the main text.[7]
Summary of findings	In any paper that analyzes large amounts of data or multiple datasets, such as a systematic review.	Further clarifies and summarizes the main points of a paper in 1 location. Should not be implemented for small amounts of data that can be synthesized easily by the reader.[5]
Data table	To compare multiple groups, show demographic information, or list raw data obtained in a study.	Beneficial when precise values or individual data points are important. Not appropriate for small amounts of data that can be stated succinctly in 1 or 2 sentences in the text.[7]
Figures		
Photograph	When the text is difficult to understand and can be clarified by a visual representation.	The value of photographs can be diminished by black-and-white printing or when they must be shrunk to fit the parameters of the paper.[3]
Line drawing	To show a key point in a procedure, characteristic signs of an illness, or other important points that can be clarified visually.	Simply depicts a situation or image. Useful when a photograph is not available or appears cluttered.[3]
Flowchart	To depict a process with multiple discrete steps, including an exclusion process, patient flow in a trial, or treatment protocol.	Minimal text is required, so flowcharts can shorten the main text. Flowcharts are not ideal for highly branched, complex processes.[4]
Decision tree	To show potential outcomes and decisions in a sequential manner (eg, outcomes of different treatment options).	Useful for cost-utility analyses and other outcomes studies. May not fit on 1 page and can be hard to follow when they depict extensive rounds of decision making.
Graphs		
Bar graph	For comparing categorical data or summary statistics from 1 or multiple groups. Stacked bars can be used to distinguish multiple contributors within a single category.	Easy to interpret. Bar graphs cannot show individual data points and do not convey more information than a table unless they compare multiple groups.[3,4,7,10,12]
Line graph	To depict how a single variable changes over time or compare the behavior of multiple variables over time.	Clearly shows data values and slope between them. Line graphs are not appropriate for representing averages of a group or other nonsingular measurements.[3,4,7,12]
Histogram	To portray a sampling distribution with a continuous independent variable	Histograms are best for showing the shape of the distribution of univariate data with a continuous variable. If the intervals are too large, the distribution will not have the correct shape.[3,10,12]
Box plot	To show the distribution of data of 1 or multiple groups. Box plots can be added outside the axes of scatterplots to show the univariate distributions.	Good for showing and comparing distributions of large datasets. Cannot show individual data points other than outliers. Not appropriate for small datasets, which can be represented in a histogram.[10,12]
Pie chart	To show relative frequencies or percentages, not precise values.	Good for presentations. Generally too simple to include in a scientific paper, although multiple pie charts in a single graphic can convey data more clearly than a table or text.[3,7,10]

Continued

TABLE 2. Continued

Type of Illustration	When to Use It	Advantages and Disadvantages
Scatterplot	To show individual data points from bivariate data. May or may not include a regression line.	Preserves both dimensions of data and individual points, and shows the relationship between the variables. Points that fall on the same or close coordinates may not be distinguished.[4,8,10]
Survival curve	To show the cumulative changes in the population, such as deaths or cures.	Cannot show stratification of variables; this would require multiple survival curves and could reduce the data density.[10]

TABLE 3. Patient-Reported Pain Scores

Age Group	N	Mean Score
20–39	6	13.83333
40–59	11	54.63636
60–79	6	64.66667
≥80		63.00000

TABLE 4. Michigan Hand Outcomes Questionnaire Patient-Reported Pain Score for Rheumatoid Arthritis

Age (y)	Score (Out of 100)
23	18
28	3
32	9
35	3
36	34
37	16
44	36
45	46
49	41
49	39
52	45
52	76
53	47
55	64
57	61
58	52
58	94
65	83
69	63
69	46
69	67
72	75
78	54
81	63

every data point, and has greater data density than any of the techniques involving mean scores. Those formats suggested that the pain scores follow a bell curve and decline in the highest age group. In reality, there is a linear correlation between age and pain score, which is clearly depicted in the scatterplot. If the individual subjects are the focus of the author, he may choose to include the table with each of the scores in addition to a scatterplot.

No matter which type of graphic an author uses, his or her utmost concern should be reader understanding. Always keep this in mind when choosing format and design elements. Color coding and shading are wonderful techniques for distinguishing or highlighting datasets and categories, as long as they do not steal attention away from the information.[1,4,7] However, reds and greens should not be paired, to accommodate colorblind readers.[6] Authors must also consider the requirements, limits, and color-printing costs for illustrations set by journals.[4,6,7] For instance, do not spend time choosing colors for a graph that will be published in black and white.

In addition to being easy to read, good graphs, figures, and tables must be autonomous, meaning they should be fully understandable even outside of the context of the paper or abstract.[1–4,7,8] In other words, the reviewer or reader should not flip back and forth between the figure and the text to understand what the figure is showing. An independent graphic requires thorough, precise titling and labeling of all components and units. Judicious use of footnotes is a good way to expand on the various components of the illustrative material, such as explaining any superscripts, abbreviations, statistical tests, and missing values. A good way to test whether an illustration can truly stand alone is to ask someone who is unfamiliar with the data to interpret it without reading the text in the paper.

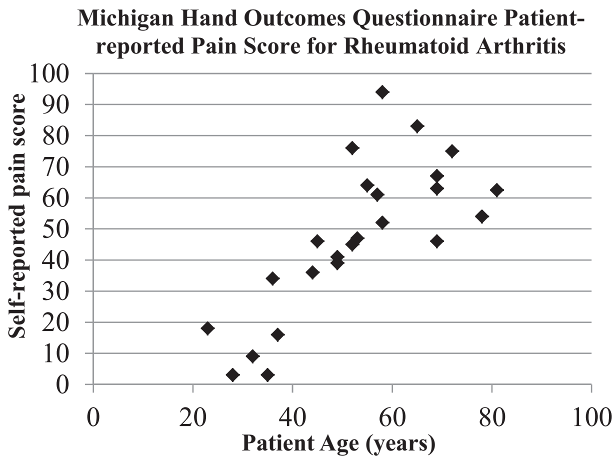

FIGURE 5: Scatterplot.

Each type of graphic has its own conventions. For tables, these include aligning the entries, using the minimum significant figures, and including statistical values when applicable.[4,7] Rows and columns must be divided, especially in large tables, by either border lines or spaces.[4] This is an aesthetic choice, but it should be made consistently for all tables in an article. The typeface inside a table should ideally be the same as that of the main text, or within a few font sizes if multiple sizes are used.

In general, graphs should depict real data, not summary statistics, such as mean or median values.[4] Unlike true data, they cannot show trends, proportions, or relationships. The most prominent features of graphs should be their data points, regression lines, or bars; tick marks, gridlines, and labels are accessories to the data and should serve to clarify, not compete with them.[1,4]

Figures, which include flowcharts, photographs, diagrams, or line drawings, should only include simple text.[3] Filling them with long sentences defeats the purpose of using a figure, which is meant to provide an overview or simple example. Figures, especially photographs, have the unique concern of resolution, so it is important to know how they will be printed.

Good graphics are easily interpretable, shorten reading time, decrease word count in the main text, and save space. However, there is a limit on how many should be included; graphics should not account for more than one-third of the manuscript.[4,7] For instance, if a paper is 6 pages long, the illustrations should not take up more than 2 of those pages. Finally, aim for individual figures, tables, and graphs that fit onto 1 page, and be sure that they are designed on a scale that will be readable once they are printed.

Although this may seem like a lot of instructions, making effective graphs and tables is not hard as long as simplicity, accuracy, and clarity are the top priorities. For scientists who wish to communicate their results with a large body of readers, it is crucial to design graphics that maintain "accuracy in [their] data and clarity in [their] presentation."[11] Doing so will enable readers to glean important findings and relationships quickly and easily, and reduce the risk of misinterpretation.[2,3,5,6,8] Graphs and tables also take up less space and fewer words than text that explains the same information, making them valuable techniques for presented data.[3]

Being able to identify common mistakes and knowing how to achieve graphic and tabular excellence are stepping stones to improving the quality of graphs, tables, and figures in scientific journals. All are valuable modes of making complex information accessible to a variety of readers. The best way to communicate data and lead readers to the right conclusions is always the simplest.

REFERENCES

1. Annesley TM. Put your best figure forward: line graphs and scattergrams. Clin Chem 2010;56:1229–1233.
2. Cooper RJ, Schriger DL, Close RJH. Graphical literacy: the quality of graphs in a large-circulation journal. Ann Emerg Med 2002;40:317–322.
3. Currier DP. Let's reduce the communication gap: how to present data in figures and tables. Resp Care 1976;21:335–338.
4. Durbin CG. Effective use of tables and figures in abstracts, presentations, and papers. Resp Care 2004;49:1233–1237.
5. Knottnerus JA, Tugwell P. Better data presentation in graphs and tables is possible and needed. J Clin Epidemiol 2010;63:585–586.
6. McDonald JC. Charts, graphs and tables—reporting the data. Radiat Prot Dosim 2001;95:291–293.
7. Saver C. Tables and figures: adding vitality to your article. AORN J 2006;84:945–950.
8. Schriger DL, Sinha R, Schroter S, Liu PY, Altman DG. From submission to publication: a retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. Ann Emerg Med 2006;48:750–756.
9. Elting LS, Martin CG, Cantor SB, Rubenstein EB. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. BMJ 1999;318:1527–1531.
10. Schriger DL, Cooper RJ. Achieving graphical excellence: suggestions and methods for creating high-quality visual displays of experimental data. Ann Emerg Med 2001;37:75–87.
11. Shelly M. Three-dimensional graphs misleading. Infect Cont Hosp Epidemiol 1994;15:686–687.
12. Sonnad SS. Describing data: statistical and graphical methods. Radiology 2002;225:622–628.