Bars and Pies Make Better Desserts than Figures

Thomas M. Annesley*

In a previous article on figures *(1)* I discussed line graphs and scattergrams, 2 of the most widely used approaches for presenting data and results in scientific papers. In line graphs and scattergrams, each axis is of a continuous variable. For example, the *x* axis may show a continuous range of phenytoin doses, and the *y* axis may show the corresponding range of the resulting serum phenytoin concentrations. Or, the *x* axis may be a range of months after chemotherapy, and the *y* axis may be the percentage of surviving patients. Assay comparisons, chromatograms, ROC curves, and PCR amplification curves are all examples of line graphs or scattergrams.

There are, however, situations in which the variables are discontinuous (also called "discrete" or "nominal" variables), meaning that they are categorically different (e.g., eye color); combined within an interval (e.g., ages 21–30 years, 31–40 years, 41–50 years); or numerically scaled (ordinal variables) (e.g., tumor stages 1, 2, 3, and 4). When a visual representation of the results for discontinuous variables is desired, 2 commonly used approaches are bar graphs and pie charts. In this article, I discuss the pros and cons of these 2 types of figures.

Bar Graphs

A bar graph, also known as a column graph when the bars are plotted vertically *(2)*, is a 2-dimensional figure in which a set of discontinuous independent variables is plotted versus a continuous dependent variable. Fig. 1 is a plot of the dollars spent per capita on healthcare, which is a continuous variable, for 5 industrialized countries with medium-sized populations, each of which is a discontinuous variable. The length of the rectangular bar (or column) represents the dollars spent by each country.

If you are considering a bar graph, the use of certain style elements can help you create an effective figure. First, the space (gap) between the bars should be narrower than the width of the bars so that the gaps do not dominate the figure and pull the focus away from the bars. A good starting point is a gap that is 50% of the width of the bars. Second, the shading or pattern within the area of the bars should be pleasing to the eye and easy to distinguish from other bars when multiple sets of data are plotted in the same figure.

Third, it is important, as with other types of figures, to avoid the use of a suppressed-zero scale (i.e., a scale that does not include 0), because this practice can exaggerate differences among groups *(1)*. Although all of the bar graphs in Fig. 1 present the same data, 3 have presentation styles that detract from the graph. In Fig. 1A, the width of the gap is the same as the bars. The color within each bar is white. This combination produces a figure that not only is initially difficult to grasp but also looks more like jail bars than data bars.

Fig. 1B illustrates the opposite extreme. The gap here is only 15% of the width of the bars, so that when the ratio of the horizontal and vertical axes is set at 1, the bars become so wide that they more closely resemble a histogram than a bar graph. The use of lines or crosshatches within the bars to distinguish the data for the 5 countries makes the figure look busy. Now compare Fig. 1, A and B, with Fig. 1C, which has the 50% gap width mentioned above and a gray-scale shading that works well and is easy to distinguish from the white background. This bar graph has maximal clarity and would make a good published figure or slide in a presentation. The same data look much different (Fig. 1D) when a suppressed-zero scale is used. Note how much larger the differences in healthcare expenditures appear when the scale does not include 0. Look at newspaper or television reports, and you will find examples of the use of a suppressed-zero bar graph to magnify changes that might not otherwise appear so large.
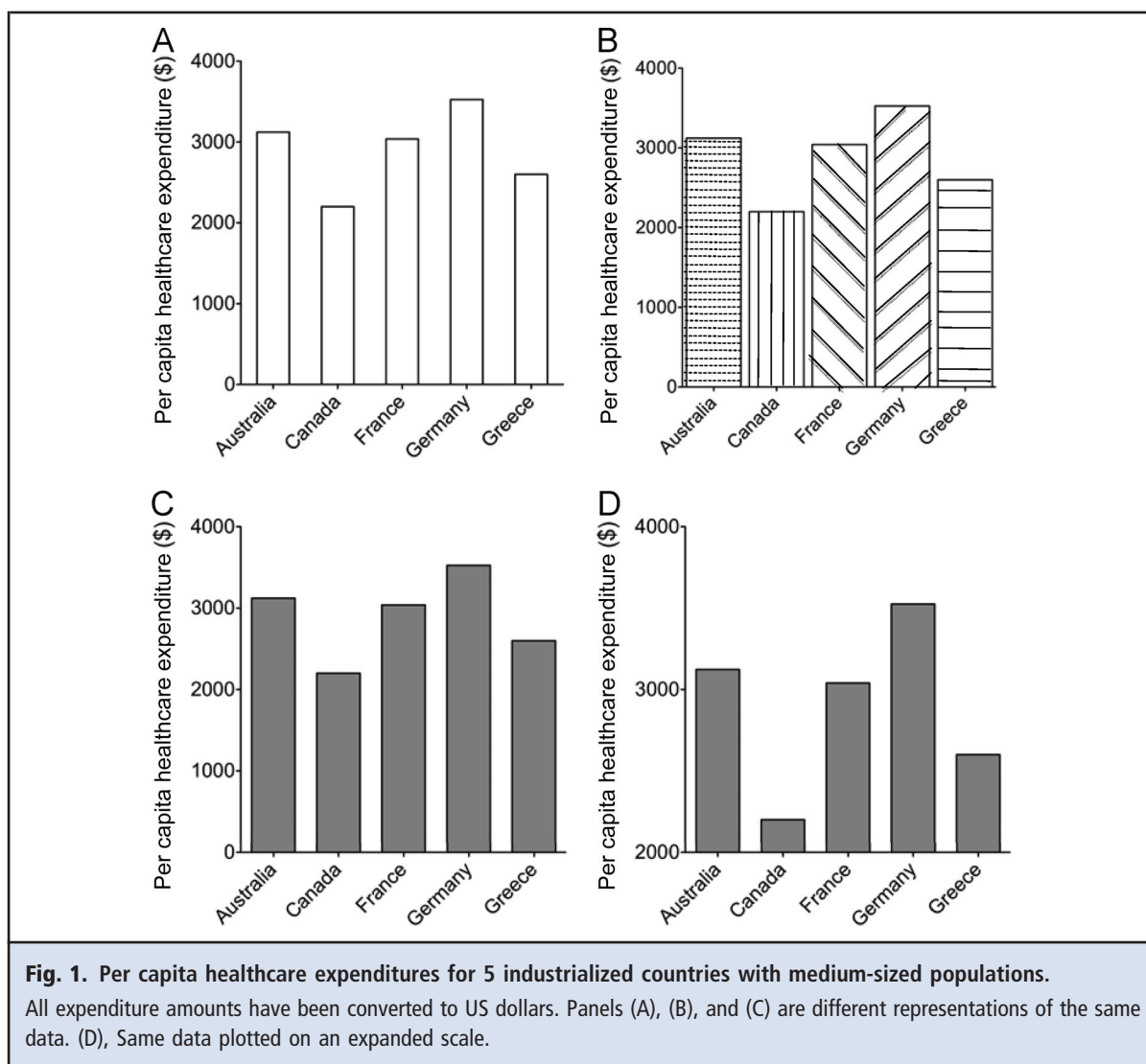
Current graphics software packages allow scientists to plot the same data in many different formats, some quite fancy. Three-dimensional bar graphs (Fig. 2, A and B), can look impressive, but they rarely add value and can actually be less clear to the reader. Note how the mortality differences in age groups 21–30 years and 31–40 years are more difficult to assess in Fig. 2A than in the other 2 panels of this figure. You will find, however, that presenting the same results as a 2-dimensional graph (Fig. 2C) will nearly always be easier to read. Note, too, that all 3 panels present the 2

University of Michigan Health System, Ann Arbor, MI.

* Address correspondence to the author at: University of Michigan Health System, Rm. UH2G332, 1500 East Medical Center Dr., Ann Arbor, MI 48109-5054. Fax 734-763-4095; e-mail annesley@umich.edu.

Fig. 1. Per capita healthcare expenditures for 5 industrialized countries with medium-sized populations.
All expenditure amounts have been converted to US dollars. Panels (A), (B), and (C) are different representations of the same data. (D), Same data plotted on an expanded scale.
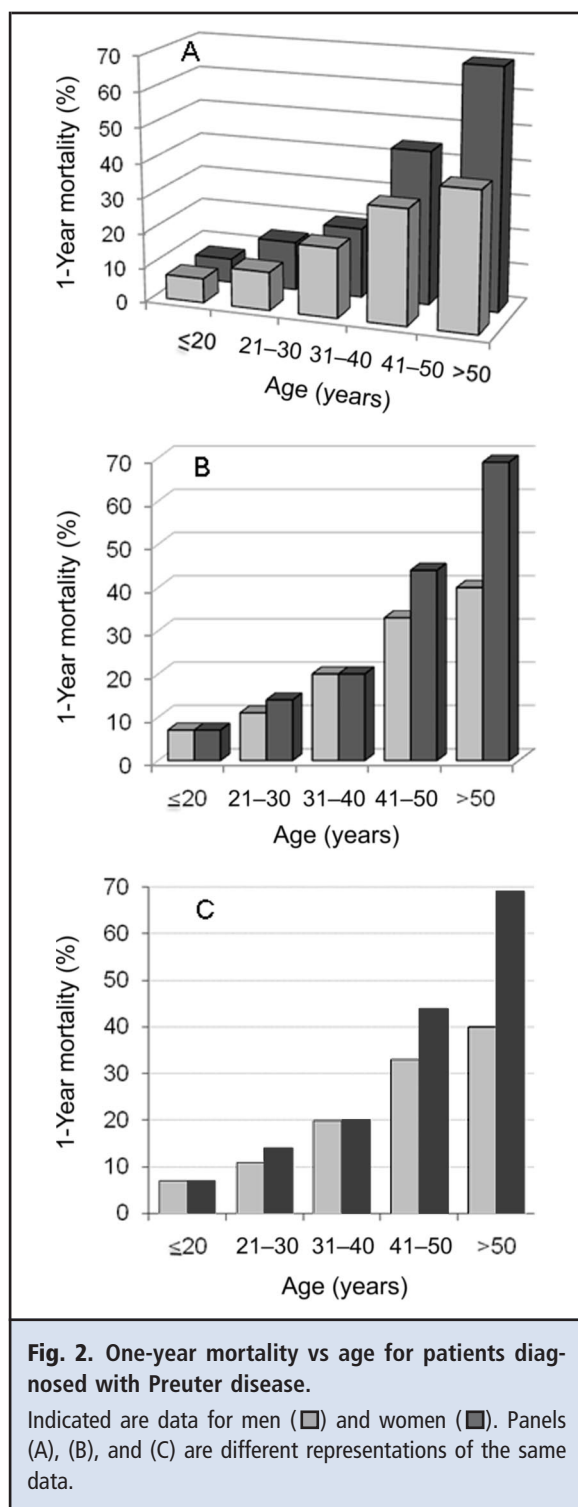
sets of results (women vs men) in complementary gray tones that stand out from the background. When multiple sets of data are plotted on the same graph, a gradation of shades from white to gray to black will be easier to read than patterns.
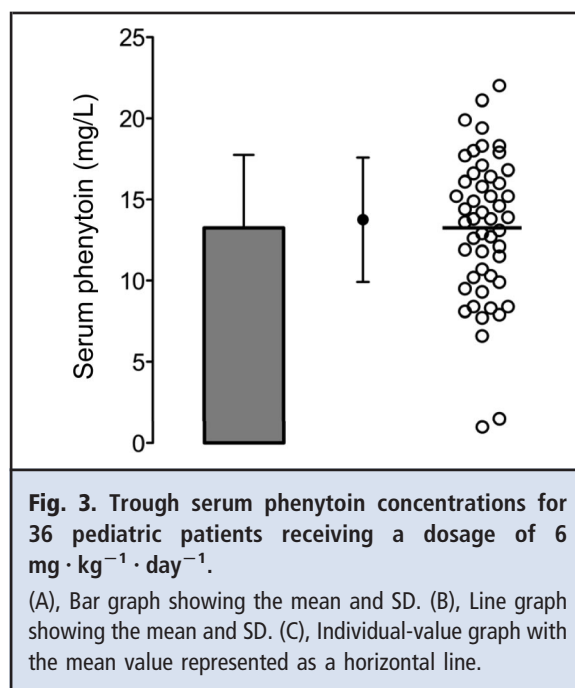
Bar graphs are useful for visual comparisons of data (Fig. 1) or for showing trends in the data (Fig. 2) and are most informative when you are more interested in the actual value of a variable than its CI *(3)*. This feature is why bar graphs are popular in slides for presentations. They focus the audience's attention on a single data value. In scientific publications, however, the distribution of the data also is critical for interpreting the data and results. The display of information regarding the data distribution is an area in which bar graphs have potential limitations. One can create a summary-data chart by adding the SD (Fig. 3, left), the

95% CI, or the interquartile range, but the bar usually remains the major visual element and therefore can mask the distribution of the data. A better alternative is either a line graph (mean and SD, or median and 95% CI), in which a symbol represents the mean (Fig. 3, middle), or an individual-value graph that shows the mean value as a horizontal line, as well as all of the data points and their spread (Fig. 3, right).

Fig. 4 illustrates another important point to remember: Bars do not have to be plotted vertically. This figure shows the per capita healthcare expenditures for 19 countries that participated in an economic survey. The format that many individuals would choose (Fig. 4A) is the placement of the country names horizontally and the bars vertically. If the number of variables is small or the category name (the country here) is short in length (e.g., Fig. 1), then the reader may be able to
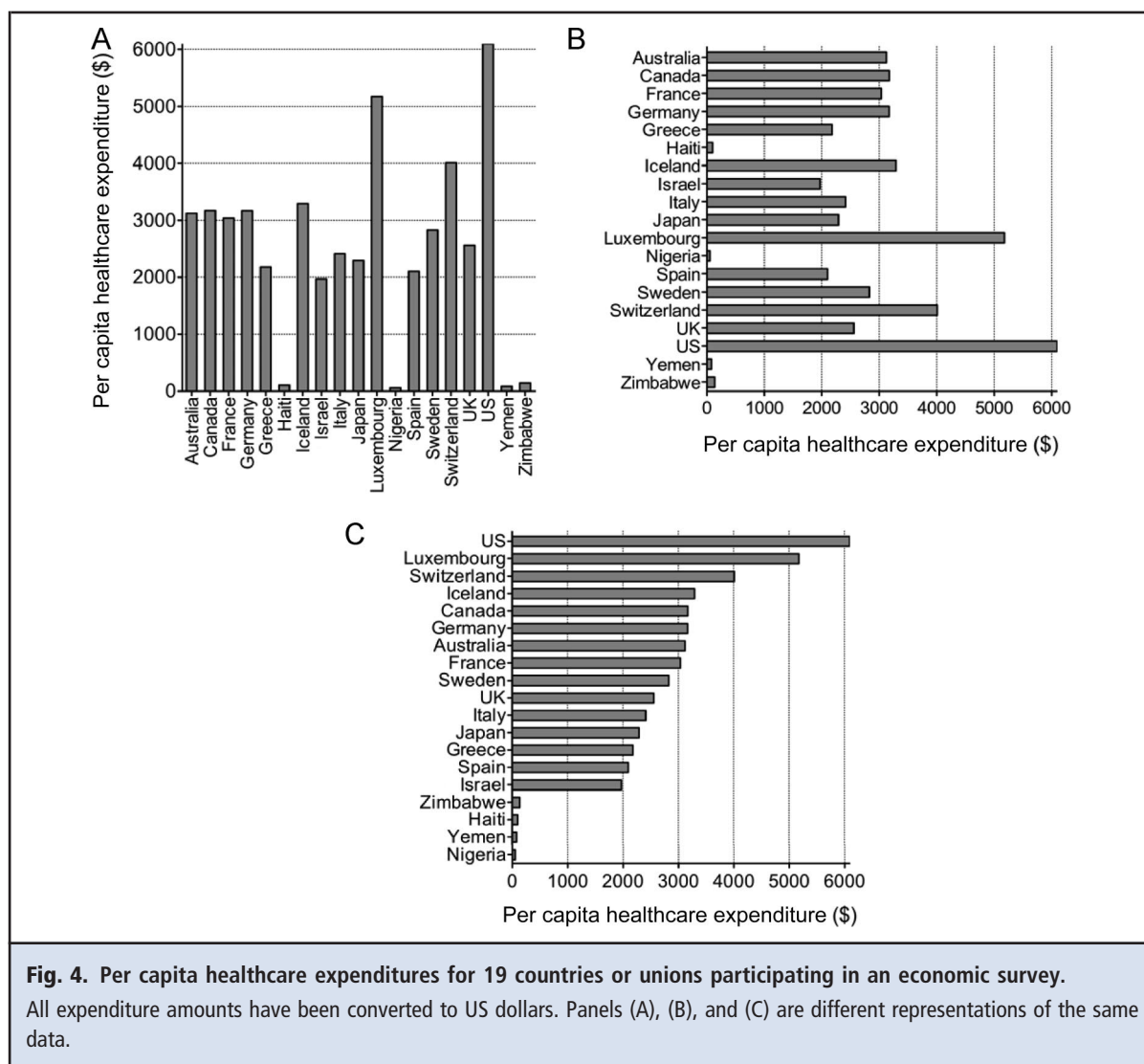
Fig. 2. One-year mortality vs age for patients diagnosed with Preuter disease.

Indicated are data for men (☐) and women (■). Panels (A), (B), and (C) are different representations of the same data.

Fig. 3. Trough serum phenytoin concentrations for 36 pediatric patients receiving a dosage of 6 mg · kg^{-1} · day^{-1}.

(A), Bar graph showing the mean and SD. (B), Line graph showing the mean and SD. (C), Individual-value graph with the mean value represented as a horizontal line.

versed and the per capita expenditures are shown on the horizontal axis (Fig. 4B). This version is acceptable if you are interested more in an individual expenditure for a single country than a comparison with other countries. If the reader is going to see only a series of apparently unrelated bars, however, with each bar representing a single data point, you should consider whether the data might be more informative in a table, where the actual numerical information for each country could be listed. This argument reiterates the earlier point that bar graphs are effective for presentations but are not always so for scientific papers. If you are going to use a bar graph, the best representation of the data is that of Fig. 4C. The relationships among the countries are much more apparent, and the graph shows the differences and trends from the highest to the lowest values.

Pie Charts

A pie chart is a circular drawing that is divided into segments, with each segment representing a data category or group. The size of each segment reflects its percentage or proportion of the total area of the pie. Pie charts are popular but are not useful in most scientific papers. They are used more frequently in magazines and newspapers to illustrate a specific difference between selected groups, often to draw attention to large differences or to express a viewpoint. Pie charts are most understandable if the number of categories is limited to 6 or fewer. Presenting more than 6 categories

read and understand the graph readily. If the category names are long and many, however, they can become difficult to read unless the reader rotates the page. The same data become easier to read if the format is re-

Fig. 4. Per capita healthcare expenditures for 19 countries or unions participating in an economic survey.
All expenditure amounts have been converted to US dollars. Panels (A), (B), and (C) are different representations of the same data.
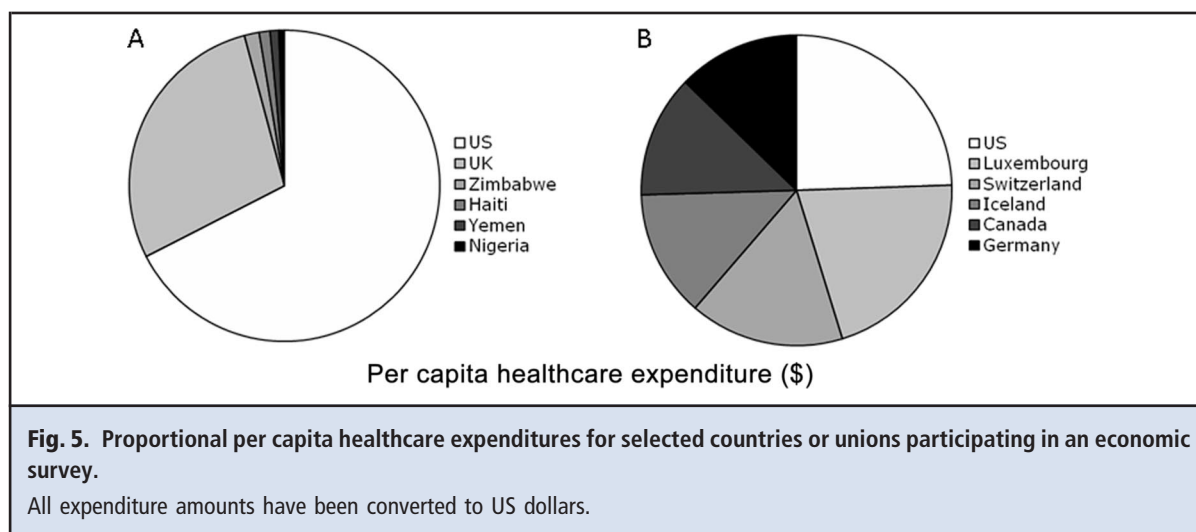
not only makes the pie chart appear busy and confusing but also makes finding usable, nonclashing color hues, shading, or background patterns more difficult. Although a pie chart is good for displaying the relative size or percentage contribution of each included piece of data, a potential problem with a pie chart is that readers may infer that the circle represents 100% of all possible data or all possible outcomes, which may not be the case.

For example, one could take the per capita healthcare expenditure data of Fig. 4, select 6 countries from the bar graph, and plot them as a pie chart (Fig. 5). Because the pie chart in Fig. 5A contains a selected subset of the data, it shows that the per capita healthcare expenditure in the US is roughly 45 times that of Zimbabwe. Is the intended message that the US spends too much? That Zimbabwe

spends too little? One might even conclude that the US spends 75% of all per capita healthcare dollars, which is not true. Readers might be drawn to a different conclusion if 6 other countries, all with higher per capita expenditures, were compared in a pie chart (Fig. 5B). In this case, US expenditures do not look so out of proportion.

A pie chart is most accurate when all available data or possible outcomes are included. For example, in a hypothetical clinical study of the efficacy of a new chemotherapeutic regimen, the 4 possible outcomes or end points could be (*a*) complete remission, (*b*) partial remission, (*c*) no improvement, or (*d*) death due to treatment complications. All patients in the study should fit into one of these categories. A pie chart representation of the results might look similar to Fig. 6. All of the available data

Fig. 5. Proportional per capita healthcare expenditures for selected countries or unions participating in an economic survey.

All expenditure amounts have been converted to US dollars.

are included, and the total number in each group is provided.

The common convention in creating pie charts is to consider the circle as a clock face *(4)*, starting with filling the largest section (wedge) in at 12:00, plotting subsequent sections in clockwise manner, and ending with the smallest section approaching 12:00 (Fig. 5). In some cases, the categories have a natural order or association, as in Fig. 6, that is best understood if they are plotted in a specified order (e.g., the best outcome to the worst outcome).

Of course, a pie chart is not necessary in most scientific papers. The same data can be presented in a table or even in the text. For example, the data in Fig. 6 could be stated in the text as: "In the men's group, 21

patients achieved complete remission, 52 patients achieved partial remission, 40 patients experienced no change, and 6 patients died from complications believed to have been due to the chemotherapy. In the women's group, 44 patients achieved complete remission, 41 patients achieved partial remission, 22 patients experienced no change, and 4 patients died from complications believed to have been due to the chemotherapy." Authors must decide on the best use of page space and word count.

Learning Exercise

With the data provided in Fig. 6, transform this figure from a pie chart to a bar/column graph. Be sure to add an appropriate legend to the new figure as well. After you have finished this exercise, compare your graph with the examples shown after the list of resources and additional reading materials.

Final Thoughts

Bar graphs and pie charts can be effective for summarizing data in a slide presentation or poster. They serve as a visual anchor for the audience while you explain the data and can highlight important differences or trends that might be missed if the data were presented only in text or a table. In scientific papers, however, a bar graph or pie chart must not only present the data but also be easily understood without having to refer repeatedly back to the main text. Authors can easily confuse readers with graphs that are unnecessarily complicated or that potentially misrepresent or underrepresent the data. In many cases, bars and pies make better desserts than figures.

Fig. 6. One-year follow-up status of patients with acute lymphoblastic leukemia treated with a combination of vincristine and cytarabine.
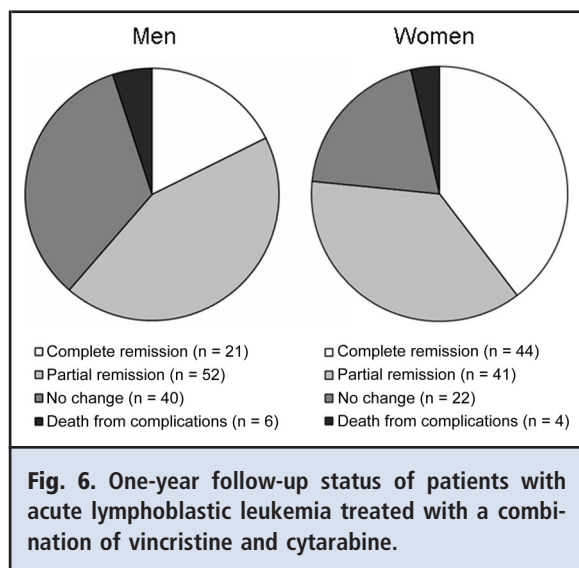
Additional Reading

Day RA, Gastel B. How to write and publish a scientific paper. Westport, CT: Greenwood Press; 2006.

Lang TA. How to write, publish, and present in the health sciences. Philadelphia: ACP Press; 2010.

Zeiger M. Essentials of writing biomedical research papers. New York: McGraw-Hill; 2000.

References

1. Annesley TM. Put your best figure forward: line graphs and scattergrams. Clin Chem 2010;56:1229–33.
2. Gustavii B. How to write and illustrate a scientific paper. New York: Cambridge University Press; 2008. p 24.
3. Browner WS. Publishing and presenting clinical research. Philadelphia: Lippincott Williams & Wilkins; 2008. p 101.
4. Freeman JV, Walters SJ, Campbell MJ. How to display data. Malden, MA: Blackwell Publishing; 2008. p 17.
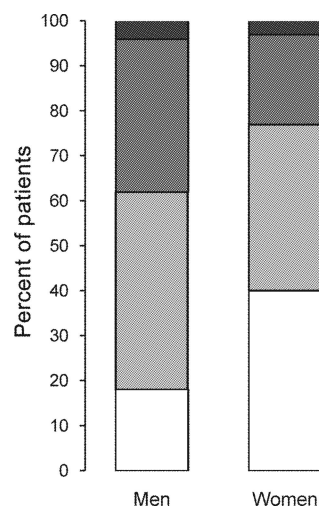
Answer to Learning Exercise
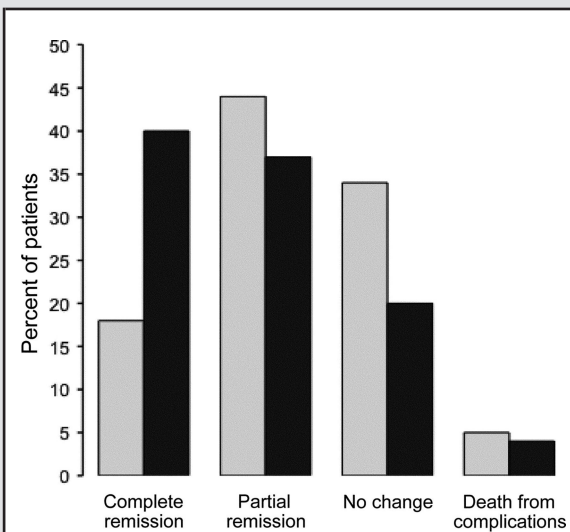
Example 1. One-year status of patients with acute lymphoblastic leukemia treated with vincristine plus cytarabine.

Men (▣), n = 119; women (■), n = 111.

Example 2. One-year status of patients with acute lymphoblastic leukemia treated with vincristine plus cytarabine.

□, complete remission; ▨, partial remission; ▧, no change; ■, death from complications. Men, n = 119; women, n = 111.

Commentary

Because the numbers of men (n = 119) and women (n = 111) differ, the best way to compare outcomes is to plot the percentages of men or women in each response category. It is also important to include the number of patients on the graph or in the legend. Example 1 is a clustered bar graph, in which the *categories* are plotted on the horizontal axis. The pattern of response rates is easy to see and compare for both sexes. When fewer than 3 groups are included, clustered bar graphs are better for showing trends and allow group comparisons. Example 2 is a stacked bar graph, in which the *groups* are plotted on the horizontal axis. Because stacked bar graphs must add up to 100%, they have the same characteristics as pie charts. When >3 groups are compared, a stacked bar graph may be easier to understand, especially if there is a natural order to the categories. This consideration is a good reason to plot your data several ways and then decide on the format that most clearly presents your message.