COMMON STATISTICAL ERRORS EVEN **YOU** CAN FIND*
PART 3: ERRORS IN DATA DISPLAYS

By Tom Lang, MA
Tom Lang Medical Communications

This article is the third in a series in which I describe several of the more common statistical errors in the biomedical literature. The first article focused on 10 errors in descriptive statistics and in interpreting probability, or P values (*AMWA J.* 2003;18: 67-71), and the second article described 9 errors in interpreting differences between groups (*AMWA J.* 2003;18:103-106). This article addresses 5 errors in presenting statistical information in figures and tables.

Not surprisingly, errors in figures and tables can confuse the interpretation of data. For example, we tend to recall the visual impression of a figure better than the actual message presented by the data. We also tend to compare things that are side-by-side, including data in adjacent columns of a table. As the examples here illustrate, the most effective way (and the only ethical way) to use a figure is to make the visual impression correspond to the message of the data. The best way to use a table is to put the columns to be compared side-by-side.

ERROR #20. Visually distorting relationships on a column chart by starting columns at a baseline value other than zero.

This common error (Figure 1) is sometimes called the *suppressed zero problem.* Unless otherwise informed, readers assume that the baseline of any chart is zero on the Y axis. To read the chart, readers are supposed to see the values at the tops of the columns.

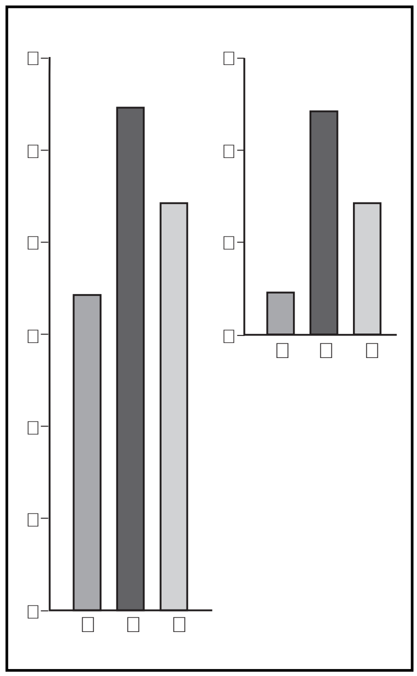

Figure 1. The Suppressed Zero Problem. *The suppressed zero visually distorts the relationships among quantities. Here, A is actually two thirds as large as B, but the suppressed zero makes A appear to be less than one quarter the size of B.*

Visually, however, readers actually compare the *heights* of the columns. In the suppressed zero problem, the length of the column is no longer proportional to the value it represents. Thus, in the graph on the right in Figure 1, column A appears to be less than one quarter the size of column B when, if fact, the value of column A is actually two thirds as large as that in column B. To prevent this distortion, the scale and the columns should be "broken" above the expected baseline of zero to indicate clearly that part of the scale has been omitted.

ERROR #21. Visually distorting relationships among data by manipulating the relative scales on the X and Y axes.

Sometimes called the *elastic scale problem*, this error is related to the relationship between the width and height of a graph (Figure 2). By expanding or compressing the scale on the Y axis, differences can be made to look larger or smaller. Expanding or compressing the scale on the X axis can make changes over time to look more sudden or more gradual.

There is no easy way to prevent this

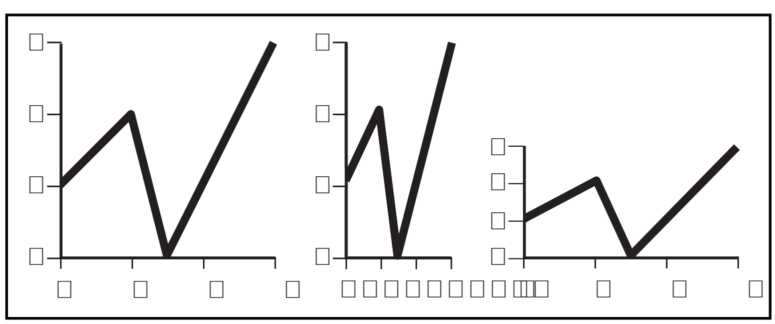

Figure 2. The Elastic Scale Problem. *Uneven scales visually distort relationships among trends. Compressing the scale of the X axis (representing time here) makes changes seem more sudden. Compressing the scale of the Y axis makes the changes seem more gradual. Scales with equal intervals are preferred.*

**This series is based on 10 articles first translated and published in Japanese by Yamada Medical Information, Inc. (YMI, Inc.), of Tokyo, Japan. Copyright for the Japanese articles is held by YMI, Inc. The* AMWA Journal *gratefully acknowledges the role of YMI in making these articles available to English-speaking audiences.*

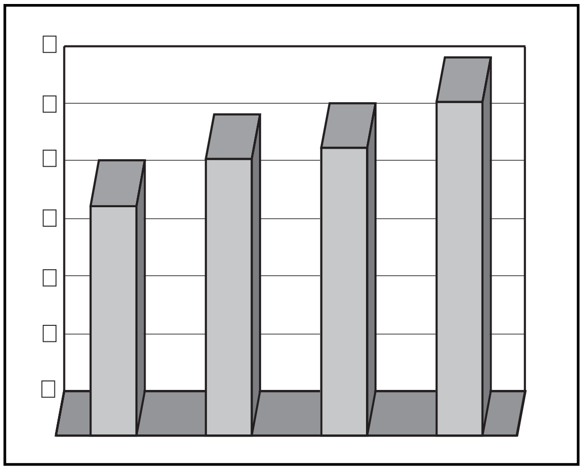

Figure 3. **The Double Perspective Problem.** *The double perspective problem confuses the reader by shifting the visual reference point, in this case from the back of the column to the front.*

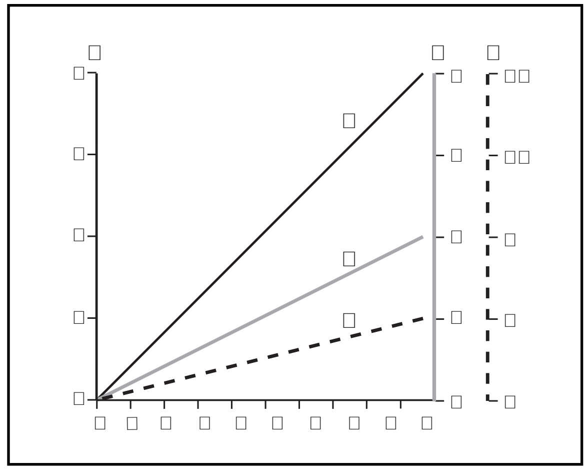

Figure 4. **The Double Scale Problem.** *Charts with 2 scales, each for a different line of data, can imply a false relationship between the lines, depending on how the scales are presented. Lines A, B, and C represent the same data, but their visual relationships depend on how their respective scales are drawn. Here, Line B seems to increase at half the rate of Line A, whereas Line C seems to increase at one quarter of the rate. Unless the vertical scales are mathematically related, the relationship between the lines can be distorted simply by changing one of the scales.*

error because the scales are likely to be dictated by the data. However, a scale that seems to be unduly compressed or expanded may be a clue that the authors, intentionally or otherwise, are trying to minimize large differences or maximize small differences in the data.

ERROR #22. Visually confusing relationships among data by adding unnecessary dimensions to the chart.
Three-dimensional charts are rarely necessary in biomedical research. The extra dimension, usually added only to make the figure more "attractive," can mislead readers by directing them to focus on the wrong part of the chart. In Figure 3, it is not clear whether the values should be read from the front or back of the columns. Again, 3-dimensional charts should be examined closely to determine whether the implied visual relationship is actually supported by the data.

ERROR #23. Visually distorting relationships among data by graphing two variables on a single graph using two unrelated vertical scales of measurement.
This *double scale problem* is a problem when the vertical scales can be expanded or compressed *independently* of one another. In Figure 4, lines A, B, and C rise from a value of zero to a value of 4, yet visually, line B appears to be increasing at half the rate of line A and line C appears to be increasing at a quarter of the rate of line A. Thus, the visual comparison between line A and line B can be manipulated by expanding or contracting the vertical scale on which line B is graphed. Graphs with 2 variables graphed on 2 *independent* scales should always be examined closely to determine whether the implied visual relationship is actually supported by the data. (Graphs with related scales, such as gallons on one side and liters on the other have a fixed relationship to one another, so the distortion does not occur.)

ERROR #24. Using tables to store data rather than to communicate information.
The figures and tables used to record data during a study are not necessarily the same ones that should be used to communicate the study's results. The sample tables show 4 of the 8 possible forms a table might take to compare 3 variables: age, sex, and nationality. All 8 forms of the table would contain the same data. However, the best table for communicating results is the one in which the columns to be compared are placed side-by-side. Thus, Table 1 is preferred for comparing the values of nationality by sex; Table 2 is preferred for comparing the values of men with those of women by nationality; Table 3 of sex by age group; and Table 4, of age groups by sex.

Table 1. Preferred Table Format for Comparing Values in the United States with Those in Japan according to Sex

	Men		Women	
	United States	Japan	United States	Japan
0-21 y				
22-49 y				
≥50 y				

Table 2. Preferred Table Format for Comparing Values of Men with Those of Women according to Nationality

	Japan		United States	
	Men	Women	Men	Women
0-21 y				
22-49 y				
≥50 y				

Table 3. Preferred Table Format for Comparing Values of Men with Those of Women according to Age Group

	0-21 y		22-49 y		≥50 y	
	Men	Women	Men	Women	Men	Women
United States						
Japan						

Table 4. Preferred Table Format for Comparing Values for Age Groups according to Sex

	Men			Women		
	0-21 y	22-49 y	≥50 y	0-21 y	22-49 y	≥50 y
United States						
Japan						