Statistics 101 for Clinicians

Michael S. Smith, M.D., M.S.

We are awash in numbers.

Every day, we see percentages, correlations, trends, or graphs, drawing our own conclusions or accepting what others say. Many Americans suffer from "innumeracy," inability to deal effectively with numbers. Innumeracy allows so-called junk science to sway people and public policy.

Perhaps half the articles in the medical literature today have statistical flaws.[1] Retrospective observational results have been treated as clinical fact, leading, for example, to the unexpected discovery through experimentation that estrogen treatment did not decrease but rather increased the likelihood of dementia and heart disease in women.

The purpose of this article is to help busy clinicians enhance their reasoning skills.

Some Common Fallacies

Absolute v. Relative Numbers

Focusing only on a rate of increase or a proportion can give a very distorted picture about the meaning of a statistic.

For example, the statement that "there is a 25% increase in the cancer rates" may be interpreted by some to mean that the probability of being afflicted is 25%. In reality, the baseline incidence from which the increase is calculated may be very small. It is preferable to cite *both* the percentage increase and the absolute increase. If the initial cancer incidence is 1 in 10,000, a 25% increase changes it to 1 in 8,000. The absolute number of cancers per million increases from 100 to 125.

Stating "we operate on the correct side 99.9% of the time" may sound like a good record. Aside from the fact that the figure indicates about 12 wrong-side surgery cases a year in a moderate-sized hospital, the proper way to format those data is by counts, which should be zero, not percentages.

Predictive Value of Diagnostic Tests

People often believe that "if I have a disease, I have a high probability of testing positive. If I test positive I must therefore have the disease, right?"

Not necessarily. Suppose, for example, that a disease has prevalence of 0.5%, or 1 in 200 randomly selected people. Suppose further that a test for that disease is *perfectly sensitive* with a specificity of 99%. In other words, if you have the disease, you will always test positive, and if you don't have the disease, 99% of the time you will test negative.

Table 1. Test* Results in a Population with a 0.5% Prevalence of Disease

	Test Positive	Test Negative	*Disease Totals*
Disease Positive	50	0	50
Disease Negative	99	9851	9950
Test Totals	149	9851	10,000

* Sensitivity 100%; specificity 99%

Consider 10,000 people. Fifty would be expected to have the disease, all testing positive. Of the remainder, 99% would test negative. Table 1 shows that only a *third* of the patients with positive tests (50/149) actually have the disease.

"If It Happens to You, It's 100%."

While this statement may be true, probability is future-oriented, dealing with chance *before* the occurrence, not afterwards.

Understanding the Terminology

In any field, knowing the vocabulary is important. In our training, once we understood the vocabulary of medicine, we were well on our way to becoming physicians.

$$\begin{array}{ccc} \text{Population} & \rightarrow & \text{Sample} \\ \uparrow & & \downarrow \\ \text{Parameter} & \leftarrow & \text{Statistic} \end{array}$$

A *population* is a group of objects, people, or things about which we wish to learn. Populations may be small (people in a room), or large (the number of MRI procedures performed each year in the US). It is denoted by capital N.

A *sample* is a subset of a population. Samples, denoted by small n, may be of size 1, 2, … up to N. Random samples are one type of sample.

A *statistic* is a numerical measure of a sample, labeled with an Arabic letter. Examples include the sample mean (*X-bar*), standard deviation (s), correlation coefficient (r), and regression slope (b_1).

We use the sample statistic to make an *inference* about a population *parameter*, a numerical measure of the population, labeled with a Greek letter. We have the population mean (μ), standard deviation (σ), correlation coefficient (ρ), and regression slope (β_1). Parameters are numerical constants, usually unknown and unknowable. The exact percentage of obese Americans (a parameter) exists at a given time, but it is neither possible nor necessary to determine it, given an appropriate definition and a proper sample. What is quoted is a statistic that *estimates* that parameter.

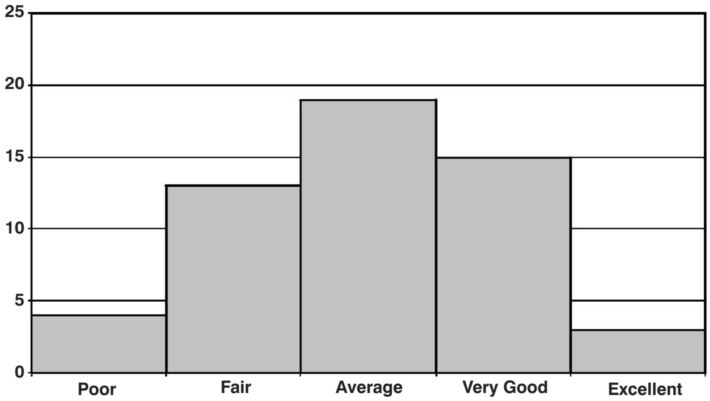

Figure 1. Patient Satisfaction with “Average”=3

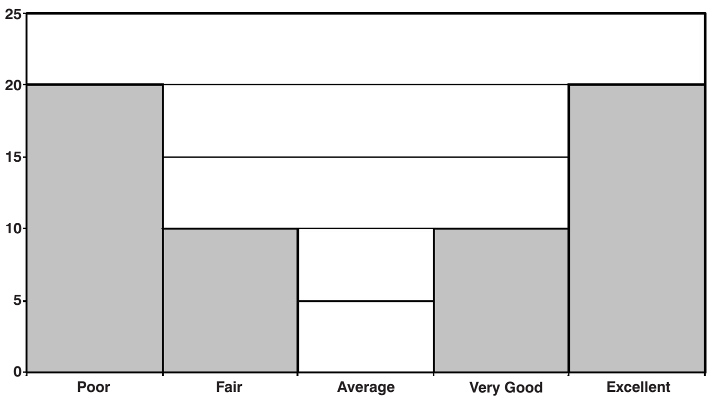

Figure 2. Patient Satisfaction with “Average”=3

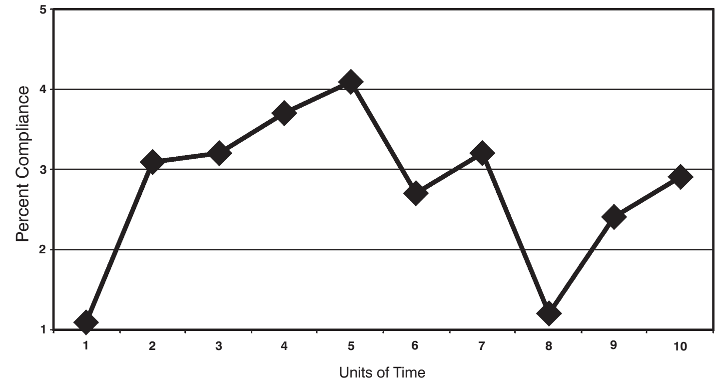

Figure 3. Line Graph. Range of data is 3.

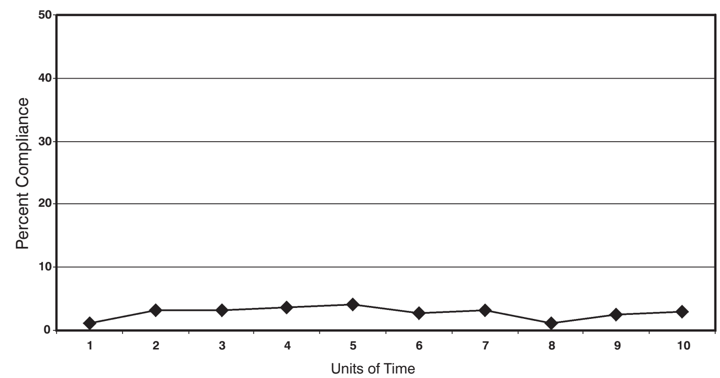

Figure 4. Same Data as Figure 3, Different Scale.

Variables indicate a specific characteristic that may take on more than one value. Discrete variables have specific values: male-female, live-die, or scores on the Glasgow Coma Scale (GCS). There is no continuum between the values. Continuous variables may take on any value in their domain. If a variable may be sensibly subdivided, it is probably continuous (weight or dollars, for example). Continuous variables may be treated as discrete (greater than or less than a hundred pounds); discrete variables cannot be treated as continuous.

Variables may also be classified as nominal, ordinal, or interval/ratio. Nominal variables are discrete, without order, and fit into a single category (automobile make, hospitals in a region).

Ordinal data are ranked, with qualitative divisions among the ranks. Examples include the GCS and patient satisfaction on a poor-to-excellent scale. A patient with a GCS of 14 is more conscious than one with a 7 (but not twice as conscious), and a very good rating is more approving than a fair one (but not twice as approving).

Interval/ratio data have equal intervals between adjacent values. Ratio data have a useful zero; thus, 100 kg is twice 50 kg; and a serum glucose of 150 mg% is half of 300 mg%. Interval data do not: 50^{o}C is not half as warm as a 100^{o}C, and the year 2000 does not mark the passage of twice as many days since the onset of time as the year 1000.

The measurement scale is important because tests we use for ratio data may not be appropriate for either ordinal or nominal data. A t-test comparing patient satisfaction scores–a discrete, ordinal measure–is not an appropriate test, despite its wide use. Even if the data were interval and continuous, comparison of the two values means little.

Presenting the Data

Because averages may hide distributions, it is helpful to draw a histogram showing the distribution. For ordinal data, this approach is the easiest and often the most useful. For two distributions, one may also compare proportions of the top or bottom, test the two distributions for homogeneity, or compare medians.

Both distributions have the same “average,” 3, but Figure 1 shows values clustering around the most common response, and Figure 2 shows marked dissatisfaction or marked satisfaction.

Notice that the scales are identical in Figures 1 and 2, allowing the graphs to be directly compared. Changes in scale may produce different appearing graphs with the same data (Figures 3 and 4).

It may be useful to convert ratio data to ordinal data, even though the conversion removes information. Suppose you learn that the average wait in an emergency department is three hours with a standard deviation of two hours. Suppose the same data show that 20% of the patients wait more than seven hours. Which is easier to comprehend? From a data collection standpoint, it is easier to look at a chart and determine whether the patient waited more or less than seven hours than it is to calculate the hours and minutes.

These figures illustrate the importance of the *shape* of the distribution. Some may question how 20% of the patient times can be more than two standard deviations above the mean. Shouldn’t that be 2.5%? Not if the distribution is significantly skewed. The empiric rule many of us learned (68% of observations are within one standard deviation of the mean, 95% within two, and 99.7% within three) holds only for normal or Gaussian distributions. It is entirely possible for 25% of observations to be more than two standard deviations from the mean, 11% more than three, and 6%

more than four. (Consider a data set such that 19 of the observations are zero and the 20th is 20. The mean=1, standard deviation=4.5, and 5% of observations are more than four standard deviations above the mean.)

Outliers skew the distribution, so it may no longer be treated as Gaussian, a fact that those who profile physicians often forget. *Chebyshev's inequality* places an upper bound on what percentage of observations *may* be more than k standard deviations from the mean: $1/k^2$.

Interpreting the Results

Many varieties of statistical software are available, such as Microsoft Excel, Minitab (www.minitab.com), SAS (www.sas.com), or SPSS (www.spss.com). With each, the operator must know the appropriate test to use, whether the requirements or assumptions for the test have been met, and how to interpret the results, statistically and practically.

Consider this example, modified from the *Barrow Quarterly.*[2] The ordinate or Y-axis displays the mortality associated with severe head injury at Virginia Commonwealth University from 1980 through 1997. An accompanying comment was that there had been a "steady decrease" in the mortality.

What do you think? Has mortality really fallen? What else would you like to know?

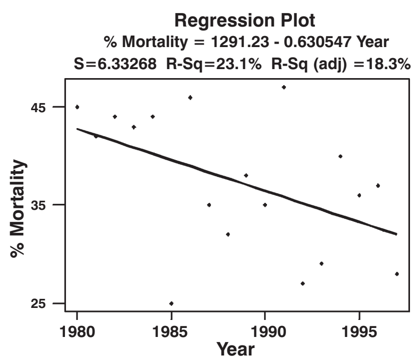

Figure 5. Mortality in Patients with Closed-Head Injury

The first question to ask is this: Are we comparing like patients? Severe head injury is not a uniform concept. Perhaps today more severely injured patients are seen because of improvements in early trauma care. The report does not address the issue, and we'll assume the patients were comparable.

Another question concerns the total number of cases treated each year. Percentages have numerators and denominators; if the denominators aren't equal, it is inappropriate to compare the quotients, unless they are weighted. If there were four patient deaths out of nine in 1982, and 40 out of 140 in 1993, we cannot treat each of those equally. We'll assume equal denominators.

What about the "steady decrease"? If one looks at the data and ignores the line, it should be quite clear that there is a lot of scatter. The mortality percentages vary by a factor of nearly two. The least-squares line shown is an artificial construct, minimizing squared deviations between the data and the predicted value for any possible line drawn through these points. The fact that one can draw a line through the data–which is always possible–does not mean that the line is the best model for the data. Software can place trend lines, but one must know how to test for statistical significance, what significance means, whether the assumptions for linear regression are met, or whether a curved line might fit better.

Is mortality decreasing? Perhaps a better way to know is to make a line graph and show the total average mortality (37%) concurrently. Figure 6 is a *run chart*. We determine trends by the number of times the data cross the average line as well as the length of each run of data on either side of that line.

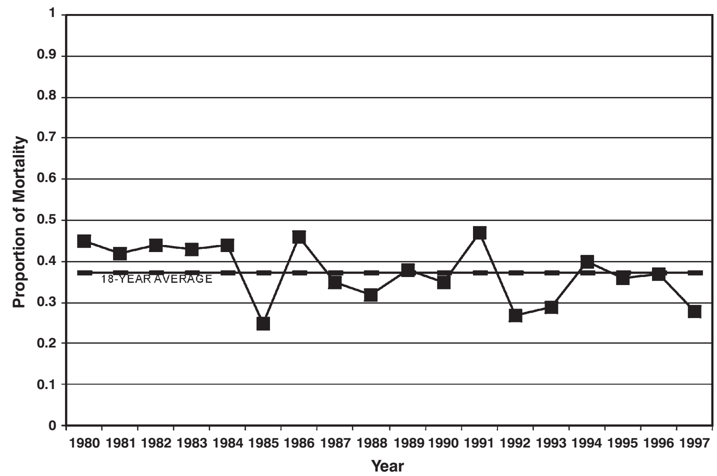

Figure 6. Run Chart Showing Percent Mortality

Here we have the same data, but a different picture! The run chart shows no statistical significance, but issues with patient selection and unequal numbers each year may affect the conclusion.

Evaluating a Research Paper

The conclusions in a research paper should not necessarily be accepted as stated. These are the steps to follow in assessing the findings for yourself:[3]

1. Determine the main question.
2. Determine the originating issue and who is raising it.
3. Ascertain the author's conclusion.
4. Evaluate the evidence offered in support of the conclusion.
5. Decide whether you accept the evidence.

Michael S. Smith, M.D., M.S., practiced neurology for 17 years and was medical director of St. Mary's Hospital (Tucson) from 1992-1998. He earned a Master's degree in Experimental Statistics from New Mexico State University in 2000 and is an independent consultant in process improvement in health care. He also teaches statistics. E-mail address: mssq@comcast.net. Telephone: (520) 410-7917.

REFERENCES

1 Altman DG, Bland JM. Improving doctors' understanding of statistics. *J R Stat Soc* 1991;154:223-267.

2 Bullock MR, Adelson PD, Marion DW, Orrico K, Ghajar J. The case for designated neuro-trauma referral centers in the United States. *Barrow Quarterly* 2003;19(3):7-8.

3 Paul, C. Finding the findings behind the news. *Stats* 2002;(35):3-6.