

Biostatistics in Public Health

Introduction

Nearly every day we see statistics used to support assertions about our health and what we can do to improve it. The press frequently quotes scientific articles assessing the roles of diet, exercise, the environment, and access to medical care in maintaining and improving our health. Because the effects are often small and vary greatly from person to person, an understanding of statistics and how it allows us to draw conclusions from data is essential for every person interested in Public Health. Statistics is also of paramount importance in determining which claims regarding factors affecting our health are not valid and which are not supported by the data, or are based on faulty experimental design and observation.

When an assertion is made such as "electro-magnetic fields are dangerous", or "smoking causes lung cancer," statistics plays a central role in determining the validity of such statements. Methods developed by statisticians are used to plan population surveys and to optimally design experiments, aimed at collecting data which allow valid conclusions to be drawn and thus either confirm or refute the assertions. Biostatisticians also develop the analytical tools necessary to derive the most appropriate conclusions based on the collected data.

In this article we discuss the role of Biostatistics in Public Health, and how Biostatistics is central to all functions of Public Health. Essential to appreciating that role is an understanding of variation in health data, and how this variation is quantified by statisticians. We next cover the standard sources of data used by Public Health professionals and list sources of publicly available data. We follow by a presentation of two fundamental and widely applied techniques of data analysis, namely chi square analysis and regression. We conclude with a brief list of useful references for those interested in further study of Biostatistics.

Role of Biostatistics in Public Health

We base our discussion on the general Public Health concepts that were summarized in the Institute of Medicine's Report on the future of Public Health (Committee, 1988). In that report, the mission of Public Health is defined as assuring conditions in which people can be healthy. To achieve this mission several functions must be undertaken:

1. Assessment. Identify problems related to the health of populations and determine their extent.
2. Policy Setting. Prioritize the identified problems, determine possible interventions and/or preventive measures, set regulations in an effort to achieve change, and predict the effect of those changes on the population.
3. Assurance. Make certain that necessary services are provided to reach the desired goals, as

determined by policy measures, and monitor how well the regulators and other sectors of the society are complying with policy.

An additional theme that cuts across all of the above functions is evaluation, that is, how well are the functions described above being performed.

Biostatistics plays a key role in each of these functions. In terms of assessment, the value of Biostatistics lies in deciding what information to gather and the identified health problems, in finding patterns in collected data, and in summarizing and presenting these in an effort to best describe the target population. In so doing it may be necessary to design general surveys of the population and its need, to plan experiments to supplement these surveys, and to assist scientists in estimating the extent of health problems and associated risk factors. Biostatisticians are adept at developing the necessary mathematical tools to measure the problems, to ascertain associations of risk factors with disease, and models to predict the effect of policy changes. They create the mathematical tools necessary to prioritize problems and to estimate costs, including monetary and undesirable side effects of preventive and curative measures.

As to assurance and policy setting, Biostatisticians use sampling and estimation methods to study the factors related to compliance and outcome. Questions that can be addressed include whether improvement is due to compliance or something else, how best to measure compliance, and how to increase the compliance level in the target population. In analyzing survey data, Biostatisticians take

into account possible inaccuracy in responses and measurements, both intentional and unintentional.

This effort includes how to design the survey instruments in a way that checks for inaccuracies, and the development of techniques which correct for nonresponse or for missing observations. Finally, Biostatisticians are directly involved in the evaluation of the effect of interventions and whether to attribute beneficial changes to policy.

Understanding Variation in Data

Nearly all observations in the health field show wide variation from person to person, making it difficult to identify the effect of a given factor or intervention on one's health. We have all heard the stories of someone who smoked every day of his life and lived to be 90, and of the death at age 30 of someone who never smoked. The key to sorting out seeming contradictions such as these is to study properly chosen groups of people (samples), and to look for the aggregate effect of something on one group as compared to another. Identifying a relationship, say between lung cancer and smoking, does not mean that everyone who smokes will get lung cancer, nor that if you refrain from smoking you will not die from lung cancer. It does mean, however, that the group of people who smoke are more likely than those who do not smoke to die from lung cancer.

How can we make statements about groups of people, but be unable to claim with any certainty that

these same statements apply to any given individual in the group? Statisticians do this through the use of models for the measurements, based on ideas of probability. For example, we can say that the probability that an adult American male dies from lung cancer during one year is 9 in 100,000 for a non-smoker, but is 190 in 100,000 for a smoker. We call dying from lung cancer during a year an "event", and probability is the science that describes the occurrence of such events. For a large group of people, we can make quite accurate statements about the occurrence of events, even though for specific individuals the occurrence is uncertain and unpredictable. A simple but useful model for the occurrence of the event, dying from lung cancer, can be made if we make two important assumptions: 1) for a group of individuals, the probability that an event occurs is the same for all members of the group; and 2) whether or not a given person experiences the event does not affect whether others do. These assumptions are known as 1) common distribution for events and 2) independence of events. It may be surprising to find that such a simple model can apply to all sorts of Public Health issues. Its wide applicability lies in the freedom it affords us in defining events and population groups to suit the situation being studied.

Consider a different example of brain injury and helmet use among bicycle riders. Here groups can be defined by helmet use (yes/no) and events become severe head injury resulting from a bicycle accident (Thompson et al., 1996). Of course more comprehensive models can be used, but the simple ones as described here are the basis for much of Public Health research.

Suppose we examine the following hypothetical data about bicycle accidents and helmet use in 30

cases, which could be gotten from a state registry.

	Wearing Helmet	Not Wearing Helmet
Severe Head Injury	1	2
Not Severe Head Injury	19	8

We can see that 20% (2 out of 10) of those not wearing a helmet sustained severe head injury, compared to only 5% (1 out of 20) among those wearing a helmet, for a relative risk of 4 to 1. Is this convincing evidence? An application of probability tells us that it is not, and the reason is that, with such a small number of cases, this difference in rates is just not that unusual. To better understand this concept, we must delve a little into the meaning of probability and what conclusions we can draw after setting up a model for our data.

Probability is the branch of mathematics that uses models to describe uncertainty in the occurrence of events. Let us suppose, for the moment, that the chance of severe head injury following a bicycle accident is 1 in 10. We will use a child's spinner, a disk with numbers "1" through "10" equally spaced around its edge, with a pointer in the center to be spun. When the pointer stops it will indicate a number from "1" to "10", and if the spinner is constructed properly, every number will be equally likely to show. Since a spinner has no memory, spins will be independent. Let the spin indicate severe head injury if a "1" shows up, and no severe head injury for "2" through "10". Now we spin the pointer ten

times to see what could happen among ten people not wearing a helmet. The theory of probability uses the Binomial distribution to tell you exactly what could happen with ten spins, and how likely each outcome is. For example, the probability that we would not see a "1" in ten spins is .349, the probability that we will see exactly one "1" in ten spins is .387, exactly two is .194, exactly three is .057, exactly four is .011, exactly five is .001, with negligible probability for six or more. So if this is a good model for head injury, the probability of 2 or more people experiencing severe head injury in ten accidents is .264 . Probability can also tell us just how likely it is to observe a risk ratio of 4 or more in samples of 20 people wearing a helmet and 10 people not wearing a helmet, assuming that the risk of injury is exactly the same, that is, 1 in 10, for both groups. The surprising answer here is that this happens quite often, about 16% of the time, which is far too large to give us confidence in asserting that wearing helmets prevents head injury. This is the essence of statistical hypothesis testing. We assume that there is no difference in the occurrences of events in our comparison groups, and then calculate the probabilities of various outcomes. If we observe something that has a low probability of happening under our assumption of no differences between groups, then we reject our hypothesis and conclude that there is a difference. To thoroughly test whether helmet use does reduce the risk of head injury, we need to observe a larger sample - large enough so that any observed differences between groups cannot be simply attributed to chance.

Sources of Data

Data used for Public Health studies come from observational studies (as the helmet use example above), from planned experiments, and from carefully designed surveys of population groups. An example of a planned experiment is the use of a clinical trial to evaluate a new treatment for cancer. In these experiments, patients are randomly assigned to one of two groups, treatment or placebo (a mock treatment) and then followed to ascertain whether the treatment affects clinical outcome. An example of a survey is the NHANES -- interviews conducted by the National Center for Health Statistics of a carefully chosen subset of the population to determine their health status, but chosen so that the conclusions apply to the entire U.S. population. Both planned experiments and surveys of populations can give very good data and conclusions, partly because the assumptions necessary for the underlying probability calculations are more likely to be true than for observational studies. Nonetheless, much of our knowledge about Public Health issues comes from observational studies, and as long as care is taken in the choice of subjects and in the analysis of the data, the conclusions can be valid. The biggest problem arising from observational studies is inferring a cause and effect relationship between the variables studied. The original studies relating lung cancer to smoking showed a striking difference in smoking rates between lung cancer patients and other patients in the hospitals studied, but they did not prove that smoking was the cause of lung cancer (Doll and Hill, 1950). Indeed, some of the original arguments put forth by the tobacco companies followed this logic, stating that a significant association between factors does not by itself prove a causal relationship. Although statistical inference can point out interesting associations that could have significant influence on Public Health policy and decision

making, these statistical conclusions require further study to substantiate a cause and effect relationship, as has been done convincingly in the case of smoking.

A tremendous amount of recent data is readily available through the internet. These include both already tabulated observations and reports, as well as access to the raw data themselves. Some of these data cannot be accessed through the internet because of confidentiality requirements, but even in this case it is sometimes possible to get permission to analyze the data at a secure site under the supervision of employees of the agency. We describe some of the key places to go for data on our health, and rely on the reader's ability to follow links to find further sources of data. Full internet addresses for the sites below can be found in the References. A comprehensive source can be found at Fedstats, which provides a gateway to over 100 federal government agencies which compile publicly available data. The links here are rather comprehensive and include many not directly related to our health. The key government agency providing statistics and data on the extent of the health, illness, and disability of the U.S. population is the National Center for Health Statistics (NCHS) which is one of the centers of the Centers for Disease Control (CDC). The CDC provide data on morbidity, infectious and chronic diseases, occupational diseases and injuries, vaccine efficacy, and safety studies. All the centers of the CDC maintain online lists of their thousands of publications related to our health, many of which are now available in electronic form. Other major governmental sources for health data are the National Cancer Institute which is part of the National Institutes of Health, the U.S. Bureau of the Census, and the Bureau of Labor Statistics. The Agency for Healthcare Research and Quality is an excellent source for data relating to the quality, access, and medical effectiveness of health care in the

United States. The National Highway and Traffic Safety Administration, in addition to publishing research reports on highway safety, provides data on traffic fatalities in the Fatality Analysis Reporting System which can be queried to provide data on traffic fatalities in the United States.

A number of nongovernmental agencies share data or provide links to online data related to health. The American Public Health Association provides links to dozens of data bases and research summaries. The American Cancer Society provides many links to data sources related to cancer. The Research Forum on Children Families and the New Federalism lists links to numerous studies and data on children's health, including the National Survey of America's Families.

Much of public health is concerned with international health and the World Health Organization makes available a large volume of data on international health issues, as well as providing links to its publications. The Center for International Earth Science Information Network provides data on world population, and its goals are provide support for scientists engaged in international research.

The internet provides an opportunity for health research unequalled in the history of Public Health. The accessibility, quality and quantity of data are increasing so rapidly that anyone with an understanding of statistical methodology will soon be able to access them to answer questions relating to our health.

Analysis of Tabulated Data

One of the most commonly used statistical techniques in Public Health is the analysis of tabulated data which is generally referred to as contingency table analysis. In it, we compare observed proportions of adverse events in the columns of the table by a method known as a chi-square test. Such data can naturally arise from any of the three data collection schemes mentioned. In our helmet use example, observational data from bicycle accidents was used to create a table with helmet use defining the columns, and head injury, the rows. In a clinical trial, the columns are defined by the treatment/placebo groups, and the rows by outcome - disease remission or not for example. In a population survey, columns can be different populations surveyed, and rows indicators of health status - availability of health insurance for example. We can show mathematically that the chi-square test used for these data does not depend on whether we switch rows and columns, so the choice of columns for group definition is purely arbitrary. The chi-square test assumes that there is no difference between our groups, and calculates a statistic based on what we would expect to see if no difference truly existed, and on what we actually observe. The calculation of the test statistic and the conclusions proceed as follows:

1. We calculate the expected frequency E (assuming no difference) for each cell in the table by first adding to get the totals for each row, the totals for each column, and the grand total (equal to the total sample size); then for each cell we find the expected count:

$$E = \frac{(\text{rowtotal})(\text{columntotal})}{(\text{grandtotal})}.$$

2. For each cell in the table, let O be the observed count, and calculate: $\frac{(O - E)^2}{E}$

3. We sum up the values from step 2 over all cells in your table. This is the test statistic, X.

4. If X exceeds 3.84 for a table with 4 cells, we then declare the contingency table to be statistically "significant". If we repeat our sample and the resulting analysis a large number of times, this significant result will happen only 5% of the time when there truly is no difference between groups, and hence this is called a 5% significance test.

5. For tables with more rows and/or columns than 2, we need to use different comparison values. For a table with 6 cells, we check to see if X is larger than 5.99, with 8 cells we compare X to 7.81, and for 9 or 10 cells we compare X to 9.49.

6. This method has problems if there are too many rows and columns and not enough observations. In such cases, collapse rows or columns if any of the E values drop below 5.

Suppose now we found a larger data set for our helmet use/head injury example and tabulated it.

	Wearing Helmet	Not Wearing Helmet
Severe Head Injury	5	10
Not Severe Head Injury	95	41

It is worthwhile noting that the proportions showing head injury for these data are almost the same as before, but that the sample size is now considerably larger, 151. Following the steps outlined above we calculate the chi-square statistic and draw our conclusions:

1. The total for the first column is 100, for the second column 51, for the first row 15, and for the second row 136. The grand total is 151, the total sample size. The expected frequencies, that is the E's, assuming no difference in injury rates between the helmet group, and the no helmet group, are given below.

	Wearing Helmet	Not Wearing Helmet	row totals
Severe Head Injury	9.934	5.066	15
Not Severe Head Injury	90.066	45.933	136
column totals	100	51	151

It is quite possible for the E's to be non-integer, and if so, we keep the decimal part in all our calculations.

2. The values for $\frac{(O - E)^2}{E}$ are presented in the table below.

	Wearing Helmet	Not Wearing Helmet
Severe Head Injury	2.45	4.805
Not Severe Head Injury	0.270	0.530

3. $X = 8.055$

4. Because this exceeds 3.84, we have a significant association between the helmet use and head injury, at the 5% significance level.

The chi square procedure presented here is one of the most important analytic techniques used in Public Health research. Its simplicity allows it to be widely used and understood by nearly all professionals in the field, as well as by interested third parties. Much of what we know about what makes us healthy or what endangers our lives has been shown through the use of contingency tables.

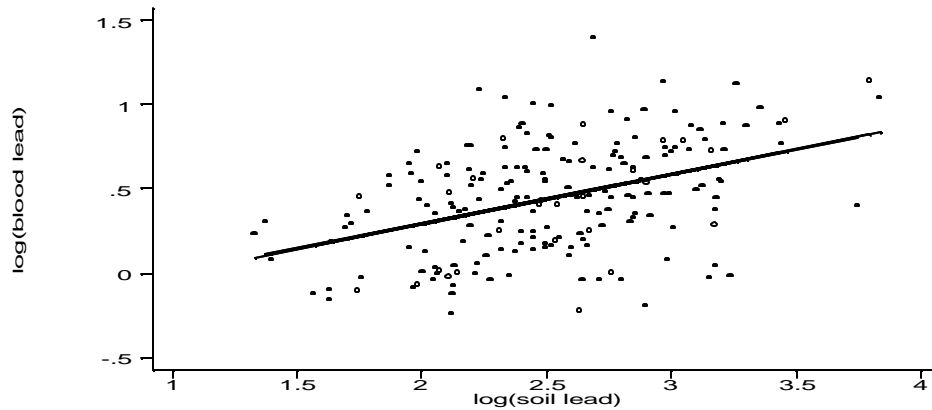
Studying Relationships among Variables

A major contribution to our knowledge of Public Health comes from understanding trends in disease rates and examining relationships among different predictors of health. Biostatisticians accomplish these analyses through the fitting of mathematical models to data. The models can vary from a simple straight line fit to a scatter plot of X-Y observations, all the way to models with a variety of non-linear multiple predictors, whose effects change over time. Before beginning the task of model fitting, the Biostatistician must first be thoroughly familiar with the science behind the measurements, be this biology, medicine, economics, or psychology. This is because the process must begin with an appropriate choice of a model. Major tools used in this process include graphics programs for personal computers which allow the Biostatistician to visually examine complex relationships among multiple

measurements on subjects.

The simplest graph is a two variable scatter plot, using the y-axis to represent a response variable of interest, the outcome measurement, and the x-axis for the predictor, or explanatory, variable.

Typically both the x and y measurements take on a whole range of values - commonly referred to as continuous measurements, or sometimes as quantitative variables. Consider, for example, the serious health problem of high blood lead levels in children, known to cause serious brain and neurologic damage at levels as low as 10 $\mu\text{g}/\text{dl}$. Since the removal of lead from gasoline, blood levels of lead in children in the United States have been steadily declining, but there is still a residual risk from environmental pollution. One way to assess this problem is to relate soil lead levels to blood levels in a survey of children - taking a measurement of the blood lead concentration on each child, and measuring the soil lead concentration from a sample of soil near their residences. As is often the case, a plot of the blood levels and soil concentrations shows some curvature, so transformations of the measurements are taken to make the relationship more nearly linear. Choices commonly used for transforming data include taking square roots, logarithms, and sometimes reciprocals of the measurements. For the case of lead, logarithms of both the blood levels and of the soil concentrations produce an approximately linear relationship. Of course, this is not a perfect relationship so, when plotted, the data will show a cloud of points as in the following example for 200 children.



This plot was produced by a statistical software program called Stata, using as input, values from a number of different studies on this subject. On the graph, we also asked the software program to plot the fitted straight line to the data, called the regression equation of y on x. For these data the equation is $y = .29x - .3$ which is linear with slope .29 and intercept -0.3 . We note that many calculators and all statistical software for personal computers can calculate the best line for a given data set.

Commercially available statistical software packages such as Stata, SAS, and SPSS can be purchased in versions for both IBM and Macintosh PC's. One comprehensive package, EpiInfo 2000, is free from CDC, and can be downloaded on the internet.

How does one interpret this relationship? First, it is not appropriate to interpret it for an individual; it applies to the population from which we took the sample. It says that an increase of 1 in log(soil lead) will lead, on average, to an increase in log(blood lead) of .29. The main point is that, from the Public Health viewpoint, there is a positive relationship between the level of lead in the soil and blood lead levels in the population. An alternative interpretation is to state that soil lead and blood lead levels are positively correlated.

As we did in the case of contingency tables, we can test for the significance of the regression. In this case, as in all of statistics, statistical "significance" does not refer to the scientific importance of the relationship, but rather to a test of whether or not the observed relationship is the result of random association. Every statistical software package for personal computers includes the test of significance as part of its standard output. These packages and some hand calculators, along with the slope and intercept, will produce an estimate of the correlation between the two variables, called the correlation coefficient, or r . For the data in the scatter diagram above, r is 0.42. This number can very easily be used to test for the statistical significance of the regression through the following formula:

$$t = r \sqrt{\frac{n-2}{1-r^2}} = 0.42 \sqrt{\frac{198}{.8236}} = 6.51$$

which we compare to 2 to determine whether it is significant or not, at the 5% level. The number r

calculated this way must lie between -1 and +1, and is often interpreted as a measure of how close to a straight line the data lie. Values near ± 1 indicate a nearly perfect linear relationship, while values near 0 indicate no linear relationship. Do not make the mistake of interpreting r near 0 as meaning there is no relationship whatever -- a curved relation can lead to low values of r .

We could have studied the relationship between soil lead and blood lead using the contingency table analysis discussed earlier. For each child, we could have measured whether the soil lead was high or low, and similarly classified blood lead levels as high and low, choosing appropriate criteria for the definitions of high and low. This too would have shown a relationship, but it would not be as powerful, nor would it have quantified the relationship between the two measurements as the regression did.

Choosing a cutoff value for low and high on each measurement which divides each group into two equal size subgroups leads to the following table:

	low blood lead level	high blood lead level
low soil lead level	63	37
high soil lead level	37	63

The chi square statistic calculated from this table is 13.5 which also indicates an association between blood lead level and soil lead levels in children. The conclusion is not as compelling as in the linear regression analysis, and we have lost a lot of information in the data by simplifying them in this way.

One benefit, however, of this simpler analysis is that we do not have to take logarithms of our data and worry about the appropriate choice of a model.

Regression is a very powerful tool, and it is used for many different data analyses. It can be used to compare quantitative measurements on two groups by setting $x=1$ for each subject in group one, and setting $x=2$ for each subject in group two. The resulting analysis is equivalent to the two-sample t-test discussed in every elementary statistics text.

The most common application of regression analysis occurs when an investigator wishes to relate an outcome measurement y to several x variables -- multiple linear regression. For example regression can be used to relate blood lead to soil lead, environmental dust, income, education, and sex. Note that, as in this example, the x variables can be either quantitative, such as soil lead, or qualitative, such as sex, and they can be used together in the same equation. The statistical software will easily fit the regression equation and print out significance test for each explanatory variable and for the model as a whole. When we have more than one x variable, there is no simple way to perform the calculations (or to represent them) and one must rely on a statistical package to do the work. A recent article includes an analysis such as the one described here (Lewin et al., 1999).

Regression methods are easily extended to compare a continuous response measurement across several groups -- this is known as analysis of variance, also discussed in every elementary statistics text. It is done by choosing for the x variables indicators for the different groups -- so-called dummy variables.

An important special case of multiple linear regression occurs when the outcome measurement y is dichotomous - indicating presence or absence of an attribute. In fact, this technique, called logistic regression, is one of the most commonly used statistical techniques in Public Health research today, and every statistical software package includes one or more programs to perform the analysis. The predictor x-variables used for logistic regression are almost always a mixture of quantitative and qualitative variables. When only qualitative variables are used, the result is essentially equivalent to a complicated contingency table analysis.

Methods of regression and correlation are essential tools for Biostatisticians and Public Health researchers when studying complex relationships among different quantitative and qualitative measurements related to our health. Many of the studies widely quoted in the Public Health literature have relied on this powerful technique to reach their conclusions.

Further Study

One of the best ways to learn Biostatistics there is to take a course designed to introduce the student to its concepts and practice. For those interested, however, in learning on their own, we have included in the Bibliography one self-sugdy electronic reference, as well as several textbooks treating more advanced topics. We selected a representative text in each of several areas that are particularly pertinent to the health sciences.

Word Count: 4886

Bibliography

Afifi AA, Lee M, and Cobb G, An Electronic Companion to Biostatistics, Cogito Learning Media Inc., San Francisco, 1998.

Armitage P and Berry G, Statistical Methods in Medical Research, Blackwell Science Ltd., Oxford, 1994.

Committee for the study of the future of public health. The Future of Public Health. Institute of Medicine, National Academy Press, Washington D.C., 1988.

Dean AG, Arner TG, Sangam S, Sunki GG, Friedman R, Lantinga M, Zubieta JC, Sullivan KM, Smith DC, EpiInfo 2000, a database and statistics program for public health professionals, for use on Windows 95, 98, NT, and 2000 computers. Centers for Disease Control and Prevention, Atlanta, Georgia, 2000.

Doll R and Hill AB, "Smoking and carcinoma of the lung. Preliminary report." British Medical Journal, 739-748, 1950.

Dunn OJ and Clark VA, Basic Statistics 3rd Ed., John Wiley and Sons, New York, 2001.

Glantz SA and Slinker BK, Primer of Applied Regression and Analysis of Variance, McGraw Hill, 2001.

Hosmer DW and Lemeshow S, Applied Logistic Regression, John Wiley & Sons, New York, 2000.

Kachigan SK, Multivariate Statistical Analysis :A Conceptual Introduction, Radius Press, New York, 1991.

Kheifets, L, Afifi AA, Buffler PA, Zhang ZW "Occupational electric and magnetic field exposure and

brain cancer: A Meta-analysis." Journal of Occupational and Environmental Medicine,

37:1327-1341, 1995

Lewin MD, Sarasua S, and Jones PA, "A multivariate linear regression model for predicting children's

blood lead levels based on soil lead levels: a study at four superfund sites." Environmental Research

Section A, 81:52-61, 1999.

NHANES: National Health and Nutrition Examination Survey, National Center for Health Statistics,

Hyattsville Maryland.1996.

The SAS System for Windows, Version 8.1. SAS Institute Inc., Cary, NC, USA, 2000

SPSS Version 10.1 for Windows. SPSS Inc. Chicago, Illinois, 2001.

Stata Statistical Software: Release 6.0. StataCorp.College Station, Texas, 1999.

Thompson DC, Rivara FP, and Thompson RS, "Effectiveness of bicycle safety helmets in preventing

head injuries - A case-control study." Journal of the American Medical Association, 276:1968-

1973, 1996.

Links for Data Sources

Agency for Healthcare Research and Quality <http://www.ahrq.gov/data/>

American Cancer Society <http://www.cancer.org/>

American Public Health Association http://www.apha.org/public_health/

Bureau of Labor Statistics <http://stats.bls.gov/datahome.htm>

Center for International Earth Science Information Network <http://www.ciesin.org/>

Centers for Disease Control <http://www.cdc.gov/scientific.htm>

Fedstats <http://www.fedstats.gov>

National Cancer Institute <http://cancernet.nci.nih.gov/statistics.shtml>

National Center for Health Statistics <http://www.cdc.gov/nchs/>

National Highway and Traffic Safety Administration <http://www-fars.nhtsa.dot.gov/>

National Institutes of Health <http://www.nih.gov>

National Survey of America's Families <http://newfederalism.urban.org/nsaf/cpuf/index.htm>

Research Forum on Children Families and the New Federalism <http://www.researchforum.org/>

U.S. Bureau of the Census <http://www.census.gov>

World Health Organization <http://www.who.int/whosis>

William G. Cumberland, Ph.D.

Professor and Chair of Biostatistics

Abdelmonem A. Afifi, Ph.D.

Dean Emeritus and Professor of Biostatistics

UCLA School of Public Health