

Teaching Bits: A Resource for Teachers of Statistics

Journal of Statistics Education v.7, n.1 (1999)

Robert C. delMas
General College
University of Minnesota
333 Appleby Hall
Minneapolis, MN 55455
612-625-2076

delma001@maroon.tc.umn.edu

William P. Peterson
Department of Mathematics and Computer Science
Middlebury College
Middlebury, VT 05753-6145
802-443-5417

wpeterson@middlebury.edu

This column features "bits" of information sampled from a variety of sources that may be of interest to teachers of statistics. Bob abstracts information from the literature on teaching and learning statistics, while Bill summarizes articles from the news and other media that may be used with students to provoke discussions or serve as a basis for classroom activities or student projects. We realize that due to limitations in the literature we have access to and time to review, we may overlook some potential articles for this column, and therefore encourage you to send us your reviews and suggestions for abstracts.

From the Literature on Teaching and Learning Statistics

"Random Variables: Simulations and Surprising Connections"

by Robert J. Quinn and Stephen Tomlinson (1999). *The Mathematics Teacher*, 92(1), 4-9.

The article presents three activities designed to help students develop a working definition of a random variable. Students work with coins, a six-sided die, and a decahedral die. For each chance process, students are asked to consider theoretical probabilities, work in pairs to collect empirical data, and combine data as a class to calculate theoretical probabilities. The activities use tree diagrams to explore the behavior of the random variables and to encourage students to think about the geometric sequence that corresponds to the probabilities.

"Who's Winning at MONOPOLY Junior?"

by Donald E. Hooley (1999). *The Mathematics Teacher*, 92(3), 197-203.

This article illustrates how a popular children's game can be used to introduce mathematical modeling and simulation. MONOPOLY Junior is a simpler version of the original MONOPOLY game by Hasbro, Inc. The article provides information on how to model the game with a computer simulation, a three-step Markov process, and a random walk to estimate the probabilities of landing on board spaces and expected values for the various properties in the games. The estimates indicate that the probabilities are not equal and the calculated expected returns are not always what one expects based on the property values alone. Students can work in groups to play the games and record data to test out the accuracy of the estimates.

"Studying Proportions Using the Capture-Recapture Method"

by Gregory D. Bisbee and David M. Conway (1999). *The Mathematics Teacher*, 92(3), 215-218.

For anyone who is not familiar with the capture-recapture method for estimating the size of a population, this article presents a fun way to learn the procedure by using plain M&M; candies to represent insects.

"Statistics and the Academy Award"

by Thomas J. Pandolfini, Jr., and Joseph L. Alfano (1999). *The Mathematics Teacher*, 92(3), 219-221.

The authors present a table of data on sixty-nine films that have won the Oscar for Best Picture. The table includes data such as the total number of Oscar nominations, total number of wins, and whether or not the film had nominations and wins for Best Director, Best Actor, Best Actress, Best Supporting Actor, and Best Supporting Actress. The article illustrates how this dataset can be used to teach about empirical probabilities, especially conditional probability (e.g., what is the probability that a film won Best Picture given that it did not receive a nomination for Best Director?). References are provided for additional information on the Academy Awards.

"Teaching Statistics Using Humorous Anecdotes"

by Hershey H. Friedman, Noemi Halpern, and David Salb (1999). *The Mathematics Teacher*, 92(4), 305-308.

This article offers some humorous ways to promote discussion among your students as they learn about the design of experiments and surveys. The authors offer a variety of fictitious anecdotal vignettes, many of which revolve around a fictitious Professor Ima Klutz. The vignettes are short, but rich in detail, and cover topics such as the representativeness of samples, homogeneity versus heterogeneity within a sample, response bias due to interview effects and the wording of questions, the importance of measures of dispersion, levels of measurement, joint probabilities, reliability and validity, correlation versus causation, and the appropriate use of inferential tests.

"Analyzing and Making Sense of Statistics in Newspapers"

Richard S. Kitchen (1999). *The Mathematics Teacher*, 92(4), 318-322.

The author presents two projects that help students to become more aware of how statistics are used in mainstream news articles. The projects are based on articles collected over several months from a local newspaper. Articles from the business section are excluded and only articles that include visual aids such as charts, tables, or graphs are selected. In the first project, students work in small groups to determine how

descriptive statistics are typically used in the news articles. The assignment asks them to answer questions such as what type of visual aids are used most often, how the use of visual aids affects the visibility of the article, and what topics or themes usually use statistics. The second project uses specific articles that are of high interest to students (e.g., university application and admissions rates as a function of ethnic background). Students are asked to analyze and interpret statistics and to engage in further research to resolve inferences they draw from the statistics. Both projects are designed to help students develop an awareness of the role statistics plays in their everyday lives.

The American Statistician: Teacher's Corner

"Developing Case-Based Business Statistics Courses"

by William C. Parr and Marlene A. Smith (1998). *The American Statistician*, 52(4), 330-337.

We provide guidelines for developing case-based business statistics courses. Specifically, we describe both the benefits and pitfalls of case-based courses, and list resources available for course development. We describe the characteristics of the instructor (and the classroom!) which augur well for case-based teaching.

[Download this article in .pdf format \(100K\).](#)

"Designing a First Experiment: A Project for Design of Experiment Courses"

by C. M. Anderson-Cook (1998). *The American Statistician*, 52(4), 338-342.

A project suitable for use as a first and last assignment given in an introductory experimental design course is outlined, and its implementation discussed. Phase 1 of the project is designed to be given at the beginning of a first course in experimental design and involves designing an experiment within instructor-specified parameters based initially on intuition and common sense. Phase 2, at the conclusion of the course, represents the final assignment of the course and encourages the student to critique their first design, suggest a "new and improved" design based on their newly acquired knowledge, and finally conduct an analysis of computer generated data (here using MINITAB, although choice of package is flexible) from an actual run of their experiment.

[Download this article in .pdf format \(103K\).](#)

"Correlation, Regression Lines, and Moments of Inertia"

by Roger B. Nelson (1998). *The American Statistician*, 52(4), 343-345.

A student of engineering or physics discussing a random variable X with mean μ and variance σ^2 might refer to μ as the *center of gravity* of the (probability) mass distribution of X , and to σ^2 as the *moment of inertia* of X about μ . Is there a similar "physical" interpretation of ρ , Pearson's product-moment correlation coefficient for pairs of random variables? We answer this question affirmatively by showing that ρ is equal to the ratio of a difference and sum of two moments of inertia about certain lines in the plane. From this observation it is easy to derive familiar important properties of ρ . Similar results hold for the population version of the nonparametric correlation coefficient Spearman's rho. These ideas are readily accessible by students in an undergraduate mathematical statistics course.

[Download this article in .pdf format \(134K\).](#)

Teaching Statistics

A regular component of the Teaching Bits Department is a list of articles from *Teaching Statistics*, an international journal based in England. Brief summaries of the articles are included. In addition to these articles, *Teaching Statistics* features several regular departments that may be of interest, including Computing Corner, Curriculum Matters, Data Bank, Historical Perspective, Practical Activities, Problem Page, Project Parade, Research Report, Book Reviews, and News and Notes.

The Circulation Manager of *Teaching Statistics* is Peter Holmes, ph@maths.nott.ac.uk, RSS Centre for Statistical Education, University of Nottingham, Nottingham NG7 2RD, England. *Teaching Statistics* has a website at <http://www.maths.nott.ac.uk/rsscse/TS/>.

Teaching Statistics, Autumn 1999 Volume 21, Number 1

"Exploring Visual Displays Involving Beanie Baby Data" by Bob Perry, Graham A. Jones, Carol A. Thornton, Cynthia W. Langrall, Ian J. Putt, and Cris Kraft, 11-13.

The authors describe how a third grade teacher reversed the usual didactic of instruction to help students understand stem-and-leaf diagrams. The data used came from the homepage of the popular Beanie Baby stuffed animals. According to the article, "Instead of taking a set of data and showing the children how to construct a stem-and-leaf diagram, the diagram was presented and the children were challenged to analyse how the data were displayed. To make sense of this unfamiliar display, the children made conjectures, justified their reasoning, and validated their ideas against the on-going thinking in class." Examples of the children's discussion are presented.

"Push-Penny: Are You a Random Player?" by Mike Perry, 17-19.

The author describes an activity that helps students develop an intuitive understanding for the consequences of randomness through data handling and the construction of graphs and tables. Students push a coin on a board trying to make the coin land on one of five lines. The lines are drawn perpendicular to the push direction, with lines spaced exactly two coin widths apart. The activity allows students to explore the behavior of binomial variables and introduces them to modeling and inference.

"Examples of the Use of Technology in Teaching Statistics" by Tom Scott and Susan Jackman, 20-23.

The authors describe the use of Microsoft Excel, graphing calculators, and a multimedia computer-based learning package (Statwise) in statistics courses at Napier University in Edinburgh, Scotland.

"THINK: A Game of Choice and Chance" by Ruma Falk and Maayan Tadmor-Troyanski, 24-27.

The authors present a modified version of a game (SKUNK) introduced by D. Brutlag that they have renamed THINK. THINK is a game of chance where players must make appropriate decisions to maximize their expected gain. On each round, players choose to either remain in the game (trying to earn more points, but risking the loss of earned points), or leave the game (keeping points already earned). After playing several games, the instructor can ask students to make suggestions for an optimal strategy. Students can then calculate probabilities and expected gains to arrive at the theoretical optimal strategy, which is then tested empirically.

Topics for Discussion from Current Newspapers and Journals

"Science and Technology: Trial and Error"

The Economist, 31 October 1998, pp. 87-88.

October 30, 1998, marked the 50th anniversary of the first clinical trial: the test of streptomycin to treat tuberculosis. This article salutes the tremendous medical advances that have resulted from clinical trials experiments, but also reminds readers of the shortcomings of this approach. Four major areas of concern are discussed.

First, while trials work well for drugs, they are not easily adapted to other kinds of treatments. For example, it is not clear how to use them to evaluate psychotherapy or minor surgery. For psychotherapy, there is no obvious placebo for comparison. For surgery, there are obvious difficulties with having a placebo group. However, there is actually a trial now underway to test arthroscopic surgery for knee injuries. Some doctors feel the surgery is no better than simply allowing the knee time to heal. Patients in the control group will be given a memory-blocking drug, and will undergo a minor incision designed to look like a surgical scar.

Second, since private firms are driven by profits, they are reluctant to test herbal treatments, which cannot be patented. Similarly, there is less incentive to investigate new uses of a drug like aspirin, which is now out of patent but has been shown beneficial for new uses such as treating heart attack patients. Even more ominous was the 1996 story of a University of California researcher who was forced by Boots Pharmaceuticals to withdraw a paper because it showed that the firm's drug for thyroid disease was no more effective than less expensive existing treatments.

Third, trials may be conducted under conditions that will not be reproduced in the outside world, which limits their relevance. The article cites a trial called ENRICHD, which is investigating the use of psychotherapy to improve the health of heart attack survivors who are suffering from depression. The concern is that the quality of care in the trial is much higher than what can be expected in the real world. Such concerns are compounded when studies are conducted in poor countries, whose typical standard of care is far below anything likely to be experienced in a drug trial. For example, AZT treatment for HIV infection is seen as prohibitively expensive for many other countries. And, as the article points out, 50 years after the trial of streptomycin there are still 3 million annual deaths from tuberculosis.

Finally, while research initiatives tend to focus on physiological measures of success, there is comparatively little emphasis on quality of the patients' lives. Hilda Bastian of the Consumers' Health Forum of Australia cites data indicating that fewer than 5% of trials published from 1980 to 1997 measured emotional well-being or social function of the participants.

"Sampling and Census 2000"

by Morris L. Eaton, David A. Freedman, Stephen P. Klein, Kenneth W. Wachter, Richard A. Olshen, and Donald Ylvisaker, *SIAM News*, November 1988, pp. 1 and 10.

Even the recent Supreme Court ruling has not given the final word on the use of statistical sampling in the Census, and the controversy is likely to be with us for some time. David Freedman is one prominent statistician who has consistently raised concerns about the sampling proposals. This article, based on one of his technical reports on the subject, does a nice job outlining the arguments.

The article begins by pointing out that there are two proposed uses of sampling for Census 2000. The first involves identified housing units from which responses are not obtained during the mail-in enumeration phase of the census. In the past, the bureau attempted follow-up visits to all of these. The new plan proposes using a sample of such residences to estimate the total number of non-respondents. The second use of sampling, called the Integrated Coverage Measure (ICM), is the analog of the Post-Enumeration Survey (PES) proposed for the 1990 Census. It would conduct a nationwide survey to check the accuracy of the enumeration phase and then statistically adjust the first phase for under- and over-counting within demographic groups.

The authors cite the following four general weaknesses with the proposals:

1. Many somewhat arbitrary technical decisions will have to be made. Some of them may have a substantial influence on the results.
2. Many of ICM's statistical assumptions are shaky.
3. There is ample opportunity for error.
4. The errors are hard to detect.

The article provides specific details on all of these. For example, on July 15, 1991, the PES estimated an undercount of 5 million. However, the authors assert that 50-80% of this figure reflects errors in the estimation process itself, not in the enumeration phase of the Census. Indeed, the Bureau later discovered an error in its PES analysis which had added 1 million people to the undercount and shifted a Congressional seat from Pennsylvania to Arizona! In any case, the net undercounts in 1980 and 1990 were in the 1-2% range. To improve the count, any new techniques would need to have errors of less than 1%. But the authors feel this level is not attainable with available survey techniques.

The proposed ICM would use a cluster sample of 60,000 Census blocks, representing 750,000 housing units and about 1.7 million people. Census officials would then attempt to match data for every residence in the sample blocks to data from the Census. An ICM record without a match may represent a "gross omission" in the Census, that is, a person who should have been counted but was not, whereas a Census record without a match may represent an "erroneous enumeration" in the Census. Finally, some people will not be counted either time. Their number is estimated by adapting the capture-recapture method, with Census records being the "captured and tagged" group, and ICM records being the recapture group. However, such estimates are flawed by "correlation bias" because the two groups are not really independent samples: people hard to find during the Census are also likely to be hard to find during the ICM. Moreover, even cases that are "resolved" through ICM field work will still create errors if the respondents do not provide accurate information. People who have moved, for example, may not give accurate information about their place of residence or household size on the official Census day (April 1).

The capture-recapture methodology alluded to above actually needs to be modified because the undercount is known to differ according to demographic and geographic variables. Therefore, the population is divided into "post strata," and a different adjustment factor is estimated for each such stratum. For example, one post strata might be Hispanic male renters age 30-49 in California. The authors object to the "homogeneity assumption," under which factors computed from the ICM are applied to all blocks in the country.

The authors also worry about how the first use of sampling -- namely sampling to follow up for non-response in the mail-in phase -- will interact with the ICM procedure. One immediate concern is that the ICM sample will find people who did not mail in their Census forms but were not chosen for the follow-up sample. Such people will be accounted for twice: once in the ICM adjustment, and once in the non-response adjustment. To avoid this, the Census Bureau proposes 100% follow-up for non-response in the blocks that will later comprise the ICM sample. The authors note that this makes two assumptions: "(1) census coverage will be the same whether follow-up is done on a sample basis or a 100% basis, and (2) residents of the ICM sample blocks do not change their behavior as a result of being interviewed more than once." Failure of these two assumptions is called "contamination error." The authors state that the magnitude of this error is not known.

"Count Monkeys Among the Numerate"

by Rick Weiss, *The Washington Post*, 23 October 1998, A1.

Over the fall, the story of an experiment showing that monkeys were capable of counting was widely reported in the popular press. It is interesting to contrast the slant on the story that different newspapers presented. The Washington Post article attempted to describe the monkey experiment in the context of a long history of attempts to demonstrate animal behaviors that correspond to the use of mathematical concepts such as addition. It provided quotations from other experts in the field praising the design of the current study and testifying to the importance of its results. According to one of the experts, "This is still an eye-popper for most philosophers and mathematicians."

"No Monkeying About With Numbers"

by Andrew Derrington, *Financial Times* (London), 7 November 1998, p. 2.

Mr. Derrington is a psychologist. His article attempted to provide perspective by relating the achievements of the monkeys to those of children. He stated that while children can tell when two numbers are different, there is no evidence that they can tell whether one is bigger than the other until they have learned to talk.

By and large, the popular press did not give a very careful description of the actual experiment conducted with the monkeys, as originally reported in *Science*.

"Ordering of the Numerosities 1 to 9 by Monkeys"

by Elizabeth M. Brannon and Herbert S. Terrace, *Science*, 23 October 1998, pp. 746-749.

This research report shows that the experiments were more complicated than commonly reported in the news. The monkeys were first trained to order the numbers from 1 to 4. News reports consistently made it sound like the monkeys were then able to "count" from 5 to 9. However, it turns out that they were not asked to place the numbers from 5 to 9 in sequence. Instead they were presented with all $C(9,2) = 36$ pairs of numbers drawn from $\{1, 2, \dots, 9\}$. In each case, their task was to order the pair. The researchers divided these pairs into three groups: familiar-familiar, familiar-novel, and novel-novel, where familiar meant a number from $\{1, 2, 3, 4\}$ and novel a number from $\{5, 6, 7, 8, 9\}$. The monkeys performed better than chance on all three groups, but the novel-novel group was the most impressive achievement.

"Six Sigma Enlightenment"

by Claudia H. Deutsch, *The New York Times*, 7 December 1998, C1.

Ask statisticians or probabilists what the phrase "six sigma" means to them, and the answer will probably involve the tails of a normal distribution. In recent years, "six sigma" has been appropriated by the business community to refer variously to methodologies and goals in the area of quality control.

In the Six Sigma model, if one wants to increase the probability that a product is defect-free, one first determines all of the steps that are currently used to produce that product. The present article use the example of a diagnostic scanner, which is a device used in medicine to produce images of the inside of the human body. Such a scanner is of course very complicated to design and build. In an effort to increase the quality and longevity of the product,

engineers at General Electric Medical Systems (Gems), broke down their existing scanner into its component parts, and considered how to design each part so that it was more reliable.

The next step in the Six Sigma model consists of considering how much the improvement of each part's design contributes to the overall quality of the finished product. Using computers, one can simulate trial runs in which various possibilities for improving certain parts are assumed, and then the change in overall quality is computed. By making many such computer runs, one can determine which parts are most worthy of attention. In the case of the scanner, this revealed several different ways to obtain the same change in overall quality of the product. Some of these ways were more expensive and/or more time-consuming than others. Given such comparisons, the design engineers were able to make more informed choices about which parts to concentrate on.

The "six sigma" phrase is widely reported as corresponding to a production process that produces no more than 3.4 defects per million. But if one calculates the probability that a normal random variable takes on a value at least 6 standard deviations from the mean, one obtains the value 2 per billion, which is orders of magnitude smaller than the 3.4 out of one million figure. The following explanation seems to shed some light on the issue. The six sigma approach deals with the production process, and does not take into account the fact that raw materials used in this process are themselves subject to variation, which in many cases affects the degree of variability of the finished product. It is reported that a value of 1.5 sigma is a good one to assume for the degree of variability of many types of inputs. Accepting this means that one is really striving for 4.5 sigma quality ($4.5 = 6 - 1.5$) when one says six sigma. Calculating the probability that a normal random variable exceeds its mean by more than 4.5 standard deviations, we obtain the value 3.4 out of one million.

"College Board Study Shows Test Prep Courses Have Minimal Value"

by Ethan Bronner, *The New York Times*, 24 November 1998, A23.

There has been a long-running debate over whether students can improve their SAT scores by taking preparation courses, such as those offered by Kaplan Educational Centers or Princeton Review. Kaplan advertises that the average increase in one's SAT scores after taking their course is 120 points (out of 1600 possible points); Princeton claims an average increase of 140 points. The College Board has long maintained that their tests are objective measures of a student's academic skills, and that preparation courses do not improve students' scores. (Nevertheless, the College Board itself publishes preparatory material for the tests, explaining that familiarity with the test styles is beneficial!)

It is not easy to design a study to settle the debate. First, the set of people who choose to take preparation courses is self-selected. Second, those who choose to enroll in such courses seem to be more likely to employ other strategies, such as studying on their own to help them get a better grade. Third, if one takes the SAT test several times, it is likely one's scores will vary to a certain extent.

The College Board's own study, undertaken by Donald E. Powers and Donald A. Rock, is the focus of this article. It found that students using one of the two major coaching programs were likely to experience a gain of 19 to 39 points more than those who were uncoached. This is considerably less than was claimed by these coaching services. Moreover, since the College Board estimates a standard error of 30 points for differences in tests taken at two different times, the study concludes that there was no significant improvement in scores due to the coaching.

"Madison Avenue and Violence Don't Mix"

by Sally Beatty, *The Wall Street Journal*, 1 December 1998, B9.

A study in the December issue of *The Journal of Experimental Psychology: Applied* reports that viewers who watched a violent film clip had poorer recall of advertising messages than viewers who watched nonviolent material. This suggests that violent TV shows may not be desirable for advertisers.

The study involved 720 undergraduate students at Iowa State University. Scenes from ten different movies were used. "The Karate Kid III," "Die Hard," "Cobra," "Single White Female," and "The Hand that Rocks the Cradle" provided violent scenes; "Gorillas in the Mist," "Awakenings," "Chariots of Fire," "Field of Dreams," and "Never Cry Wolf" were sources of nonviolent scenes. Included with each clip was one of three possible commercial messages, chosen at random. One was for Wisk laundry detergent, one for Plax mouth rinse, and one for Krazy Glue. When the students were later tested on brand-name recognition, brand-name recall, and advertising message, those who had watched the violent scenes scored lower.

Professor Brad Bushman, the Iowa State professor who directed the study, explained that the students who saw the violent scenes reported feeling angrier. He reported that such anger appears to be the major reason people had trouble remembering the ads. Critics of the study expressed concern about generalizing from college students to the population at large. Furthermore, Jim Spaeth, president of the Advertising Research Foundation, expressed concern about the use of movie clips. The research might say something about the reaction to feature films shown on TV, but this does not represent the majority of programming.

Bushman countered by saying that all the movies he used had in fact appeared on television. As to the first criticism, he says: "My question is, why would we expect the memory of college students to differ from the memory of others?"

"Ask Marilyn"

by Marilyn vos Savant. *Parade Magazine*, 3 January 1999, p. 16.

Marilyn's latest response to a probability puzzle could be used to discuss the distinction between sampling with or without replacement. In an earlier column (*Parade Magazine*, 29 November 1998, p. 26) she responded to the following question:

You're at a party with 199 other guests when robbers break in and announce they're going to rob one of you. They put 199 blank pieces of paper in a hat, plus one marked "you lose." Each guest must draw a piece, and the person who draws "you lose" gets robbed. The robbers think you're cute, so they offer you the option of drawing first, last or any time in between. When would you take your turn?

Marilyn said she would choose to draw first, explaining that "It would make no difference to my chances of losing -- any turn is the same -- but at least I'd get to leave this party as soon as possible." Not all of her readers agreed, and the present column contains responses from some of them.

One letter argues for drawing first: "You said any turn is the same, but I believe that would be true only if the partygoers all had to replace the papers they drew before another selection was made. But if they keep the papers (the scenario intended by the question), wouldn't the odds of losing increase as more blanks were drawn? If so drawing first is best."

Another reader argued for drawing last: "Though you have a 1-in-200 chance of getting a blank paper and not being robbed if you go first, the odds are 199 in 200 that the drawing will end with a loser (other than you) before you draw if you go last. You should go last."

In this column, Marilyn restates her original position that it makes no difference where in the process you draw. She argues that the answer would be the same as if everyone drew simultaneously, in which case it would be more intuitive that everyone has the same 1-in-200 chance. She offers another argument based on people buying

tickets for a church raffle, explaining that it makes no difference whether you buy your ticket immediately when you arrive or wait until just before the drawing.

"Getting it Right on the Facts of Death"

by Lawrence Altman, *The New York Times*, 22 December 1998, F7.

When a patient dies, the doctor is required to fill out a death certificate specifying the cause of death. These death certificates have many important uses. They provide information to families that can be important in the case of hereditary diseases. They may also affect insurance payments. From the point of view of research, data on cause of death influence how policy makers allocate money for future research, and how researchers themselves assess the results of medical studies.

Previous investigations of death certificates have found that a substantial proportion fail to give the correct cause of death. A number of reasons have been suggested for this. The death may be sudden and the doctor have little information to go on. Doctors may be influenced by family wishes to not have certain kinds of death recorded. Also, when several factors are involved, there is a natural tendency for specialists to see their own area of interest as the cause of death.

The present article reports on a study focusing on the accuracy of death certificates with respect to heart disease. The researchers considered 2,683 participants in the Framingham Heart Study who died between 1948 and 1988. They asked a panel of three experts to examine all the information available at the time of the deaths and classify the cause of death as (1) coronary heart disease, (2) stroke, (3) other cardiovascular disease, (4) cancer, (5) other, or (6) unknown.

The death certificate gave coronary heart disease as the cause of death for 942, or 35%, of the deceased; the corresponding figure for the expert panel was 758, or 28.3%. For patients aged 85 or older, the death certificate assigned coronary heart disease for twice as many patients as did the expert panel. The authors found indications that when doctors were uncertain of the cause of death, they listed the death as coronary heart disease. The discrepancy between the cause of death given on the death certificate and by the panel was significantly less for cancer, where there was a longer history to go on.

The authors assert that this level of error has serious implications for studies that are carried out to test drugs for treatment of heart disease. They argue that it produces a bias towards the hypothesis that the drug is not effective.

You can read more about the study in the *Annals of Internal Medicine* online (<http://www.acponline.org/>). See "Accuracy of Death Certificates for Coding Coronary Heart Disease as the Cause of Death," by Donald M. Lloyd-Jones et al. (15 December 1998) and also the accompanying editorial "Fifty Years of Death Certificates" by Claude Lenfant et al.

"Fiber Does Not Help Prevent Colon Cancer, Study Finds"

by Sheryl Gay Stolberg, *The New York Times*, 11 January 1999, A14.

"They Stand by Their Granola"

by Ginger Thompson, *The New York Times*, 24 January 1999, Section 1, p. 33.

Consumers often express dismay at the conflicting health advice that seems to appear almost daily in the news. But one message has seemed clear for many years: a high-fiber diet is good for you. One commonly accepted

benefit was a lower risk of colon cancer. This idea originated in a study done in Africa, where certain groups of people were observed to have high fiber intake and low rates of colon cancer. Still, in the years since this study, many other studies on this relationship have been undertaken, with mixed results.

The latest controversy arose from a cohort study involving more than 88,000 women nurses, who were tracked over a 16-year period. The subjects filled out questionnaires that included questions on such things as amount of physical exercise, smoking, aspirin use, fat intake, and family history of colon cancer. After these data were collected, the subjects were divided into five groups, depending upon their average daily intake of fiber. The null hypothesis -- that there is no relationship between amount of fiber intake and the rate of incidence of colon cancer -- could not be rejected for any of the five groups. In fact, looking just at the subset of subjects who ate the most vegetables, the risk of colon cancer was actually higher by 35% than the overall average risk. The authors believe that this last result, while statistically significant, was probably due to chance. They conclude that, although their study does not establish a connection between intake of fiber and a reduction of risk of colon cancer, there are numerous other studies that do show an inverse relationship between fiber intake and things such as heart disease.

The first *New York Times* article discusses the study, and the second, the reaction of the cereal industry and shoppers to the news. The industry rushed to remind customers that other studies have found that fiber-rich foods are important for reducing the risk of many other health problems, for example heart disease and high blood pressure. Customers generally indicated their mixed feelings about science and studies. For example, one woman remarked:

I feel like most scientific studies are done in a vacuum. I only listen to the ones that make sense to me. There have been many other studies that show high fiber is good for you for many reasons.

"The Cancer-Cluster Myth"

by Atul Gawande, *The New Yorker*, 8 February 1999, pp. 34-37.

Here is one more article on the challenges of epidemiological research. A "cancer-cluster" is a geographical area exhibiting an above-average rate of some cancer. In the currently popular movie "A Civil Action," John Travolta plays the real-life lawyer who brought suit against W. R. Grace, claiming that the company's contamination of ground water in Woburn, Massachusetts, was responsible for the elevated rate of childhood leukemia there. But how high above normal does the cancer rate have to be for us to sound an alarm? And do neighborhood clusters of cancer always indicate environmental problems are to blame?

Over the last twenty years, the number of identified cancer clusters has been increasing dramatically. In the late 1980s, about 1500 clusters a year were being reported to public health officials! A famous example from the 1980s concerned the farming town of McFarland, California, where a woman whose child developed cancer found four other cases within a few blocks of her home. After doctors found six more cases in the town (population 6400), people began to fear that groundwater wells had been contaminated by pesticides. This led to lawsuits against the manufacturers of the pesticides. Nevertheless, despite extensive investigations of hundreds of such clusters in the US, there are no cases in which environmental causes have been conclusively established.

Of course, the failure to identify a cause is profoundly frustrating for the "stricken" communities. Historically, there have been many success stories where disease clusters were used to identify causes. Think of John Snow's famous identification in 1854 of London's Broad Street pump as the culprit in a cholera outbreak. In recent times, AIDS first came to light through cases of an unusual form of pneumonia. Moreover, certain occupational clusters of cancer have led to the successful identification of carcinogens such as asbestos and vinyl chloride. But neighborhood cancer clusters are different. One reason is that many known carcinogens require exposure over an extended period of time before they trigger cancer. With today's mobile population, it is unlikely that residents of a community have lived together long enough for there to be a local cause for their cancers.

What then can be said about the neighborhood clusters? They may reflect nothing more than people's tendency to seek causes for patterns that are perfectly well-explained by chance variation. The article cites Kahneman and Tversky's psychological research into people's belief in "the law of small numbers." For example, the sequence of red-black roulette outcomes RRRRRR is perceived to be less random than RRBRBB. Similar misconceptions lead basketball fans to believe in the phenomenon of "streak shooting," even though statistical analysis fails to find more runs of hits and misses than would be expected by chance. The article also cites probabilist William Feller's famous analysis of bomb hits in London during WWII. Because the hits appeared to cluster, residents suspected that German spies were picking the targets. In fact, Feller showed that a simple Poisson model fit the data -- there was nothing in the patterns to suggest that the hits were non-random.

The article calls the tendency to focus attention on clusters the "Texas-sharpshooter fallacy," named for the self-proclaimed marksman who shoots at the side of a barn and then draws bulls-eyes around the holes. With cancer clusters, we first observe the cases, and then circle the at-risk population around them. The article quotes California's chief environmental health investigator as saying that "given a typical registry of eighty different cancers, you could expect twenty seven hundred and fifty of California's five thousand census tracts to have statistically significant but perfectly random elevations of cancer. So if you check to see whether your neighborhood has an elevated rate of a specific cancer, chances are better than even that it does -- and it almost certainly won't mean a thing."

[JSE Homepage](#) | [Subscription Information](#) | [Current Issue](#) | [JSE Archive \(1993-1998\)](#) | [Data Archive](#) | [Index](#) | Search JSE | [JSE Information Service](#) | [Editorial Board](#) | [Information for Authors](#) | [Contact JSE](#) | [ASA Publications](#)