Comment

Ned Glick

Professor Chatfield provides a compendium of *experiences* in statistical endeavors. He emphasizes tribulations—how to anticipate and, if possible, to avoid pitfalls—but also how to minimize damage in the event of complications or miscues, by oneself or others. An analogy in terms of traffic accidents would be to teach "defensive driving" that avoids accidents, but also to design cars, from basic frame to special safety equipment, that will protect occupants from injury when a car fails to avoid accidents. Yet driving can be fun, too.

Working well with data, and with other people, is quite different from the "theory and methods" (and computing) in most textbooks and courses for Ph.D. programs in statistics. I wish that someone had shared such perspectives with me when I was a graduate student 25 years ago. OK, I confess that wise and kind teachers *did* tell me the pitfalls of statistical consulting and collaboration. I just did not understand or heed my experienced teachers and colleagues, although they persevered with me, trying to make my education useful, even after I had begun to teach myself. At least I can appreciate my benefactors in hindsight; and I still can benefit from advice that several of them subsequently published.

Now, reading Professor Chatfield's advice, with dramatic descriptions and examples of statistical work, I have been tempted to rummage through my own experiences to find case histories that are as stunning or as ludicrous as his. Instead, I offer several responses that neither compete nor disagree with him.

First, exploring data is *fun*. Chatfield's litany is rather daunting. But I *enjoy* sharing the interests of "clients" or collaborators—and I bet that Professor Chatfield does, too.

Second, my own years of resistance to learning from experiences of other statisticians make me pessimistic about whether readers will be able to benefit from Chatfield's thorough and lucid exposition, or from my own observations, or from any "pitfalls" education.

Third, I believe that universities generally fail to reward statisticians for the breadth and depth of activities that Professor Chatfield advocates. In particular, university programs in statistics discourage faculty and students from absorbing and practicing his lessons.

My mentors 20 or 25 years ago had statistical experiences before they came to—or created—departments of statistics. They remained involved in projects or careers (academic and otherwise) in agriculture, anatomy, medicine, psychology, economics, geography, geology, engineering, law or other disciplines that use data.

In the past two decades, statisticians have developed elegant new probability theory—for point processes, record values, saddlepoint approximations, Chen-Stein methods for Poisson approximation, etc. Other inherently elegant statistical innovations now have been made practical by cheap, powerful computing: I note the advent of generalized linear models (for classical, logistic and contingency table analyses); bootstrap and re-sampling methods; density estimation and curve smoothing; interactive graphics; and so on.

Yet, during the same period, efforts to focus on "pitfalls" or to give "consulting" a distinct slot in the statistics curriculum implicitly acknowledge that *many statisticians no longer are immersed in substantive issues that require quantitative evidence and inference*.

Variants of "the ten commandments of statistical inference" can be stated in less than a hundred words (Driscoll, 1977). Other dozens or hundreds of longer articles incessantly admonish statisticians and others to do better with "real" data. Statistical consulting bibliographies have been compiled. Books have been published (Boen and Zahn, 1982; Hand and Everitt, 1987; Chatfield, 1988). And now also video tapes are used to teach consulting.

What is the problem? Why do we still need Chatfield's advice about avoiding pitfalls?

In contemporary statistics departments, many university professors and students ornament previous statistical theory or methods with baroque and rococo variations that may stretch the frontiers of mathematics or of computing, but not of real data analysis. Such work may involve data for illustrations—or to motivate funding—but not out of in-

Ned Glick is Associate Professor in the Department of Statistics and in the Department of Health Care and Epidemiology, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada. He also works as a consultant through Berkowitz and Associates and elsewhere.

trinsic interest. "The growing isolation of the statistics department is due in part to its mathematization," as described by Shafer (1990). Much of what statisticians now learn and do in their own departments may be never used elsewhere. Or worse, as Chatfield points out, techniques that are used may be *not useful*, especially when methods are applied inappropriately or are implemented incorrectly.

Many graduate students in departments of statistics are uncomfortable in their encounters with "real" data. They may not try to understand potential clients or collaborators who do not use statistics terminology (correctly). They may be not prepared to spend 90% of their time learning background material and *formulating* the statistics problems. (Often there is transparent resentment of the client who is not "smart enough" to state clear questions.)

Even when students do want to participate in the activities that Chatfield describes, their formal courses in statistics (theorems, models, books, computing) leave them less and less energy or opportunity. Undergraduate programs in statistics may further diminish the intellectual scope and liberal education of students in graduate programs. A university statistics department seldom is hostile to creative or useful intellectual ventures; but any academic bureaucracy is less inquisitive than acquisitive, wanting *more* courses (also more committees, publications, etc.) that can be *counted*.

So Chatfield's admonitions and lists may not suffice to accomplish change—although they are not difficult to understand. Much of his advice boils down to a few simple principles.

Don't be shocked—be thorough. My father taught that a good trial lawyer, by the time he or she takes a case into a public courtroom, should never have to ask a witness any question for which he or she (the lawyer) did not already know the answer. Data are to the statistician what witnesses are to litigation. So, if possible, don't just sit in an office, but instead go to the relevant business site or field or laboratory or hospital: Watch the process. Check the sampling or experimental design. Chatfield's distinction between *numbers* and *data* also has been expounded by Finney (1975). Ferret out the 999,999 and the alleged woman whose height is 15 cm or whose weight is 800 kg.

Questioning may be helped by a checklist, such as provided by Finney (1982). Communication in consulting is a two-way street, as described by Moses and Louis (1984). Sometimes it is necessary to translate terminology. For example, while serving as a member of an Animal Care Committee, I learned that "animal model" means a live animal. In other circumstances a "model" may be a toy, or a sociologist's arrows drawn on paper, or a human being—or, sometimes, "model" may mean just what the term connotes in a statistics text.

A collaborator might die in mid-study. (I cannot refrain from mentioning one pitfall experience that Chatfield neglects.) Is the statistician prepared to continue the study?

If possible, design the research and do the statistics; if necessary, do more. Perhaps this principle just extends the preceding one. If I am asked to do analysis A but I know (after checking, as above) that graphic G would be more informative, then I use it. If I am collaborating with a surgeon and a pharmacist in a clinical trial of new medication to relieve pain, then why should it be their responsibility rather than mine to compose (or to plagiarize) a careful "informed consent" letter for potential subjects? Or, when I have two or more collaborators in a study, I may have to mediate disputes between them. If I am a "co-author" of a research publication, then I expect to *write* at least some of the prose.

Recognize that few projects exist solely to employ a statistician. I am most familiar with statistics for research in medicine, epidemiology and social sciences. In these and other research contexts, the livelihood and prestige of the statistician (if any) may be not paramount. Instead, the most important issues may be the livelihood and prestige of the researchers who provide the subject-area credentials.

In 1947, the playwright Arthur Miller incorporated in *All My Sons* (Act Two) a popular, romantic view of research:

Sue: Jim's a successful doctor. But he's got an idea he'd like to do medical research. Discover things. You see?

Ann: Well, isn't that good?

Sue: Research pays twenty-five dollars a week minus laundering the hair shirt. You've got to give up your life to go into it.

Today, however, research involves huge amounts of money, and not only to feed the researcher and his or her family. Through assessments for "indirect costs" or "overhead," these funds support the universities and institutes that house research. *The New York Times* within a span of just a few weeks recently published front-page news stories and editorials alleging:

- that "cold fusion" experimenters who became stars at one U.S. university actually had "cooked" data in their claims that were the basis for further grants exceeding $5 million;

- that the president of a prominent U.S. university on the east coast scorned or obstructed investigations that ultimately found fraud in cell biology data produced by one of his coauthors;
- that a prominent university on the west coast improperly spent hundreds of thousands of dollars (at least) from research "overhead," including thousands of dollars to pay for flowers and furniture in the president's residence.

In these instances we *statisticians* were *not* the culprits (as far as I know). Again waxing theatrical, I twist a couplet from Shakespeare's drama, *Julius Caesar* (Act One):

The fault, dear Brutus, is not in ourselves,
But in our [universities'] stars, for we are underlings.

Statisticians may anticipate that our devotion to clear, complete and unbiased information sometimes may conflict with the interests of a business or agency or litigation that employs us. We also should be aware, however, that our approach may conflict with the interests of a researcher in science or medicine. There may be financial *dis*incentives for good statistical practice.

Statisticians are educated to design research and to analyze data in ways that produce more—and more accurate—information in fewer steps in a shorter time for less expense. On the other hand, a researcher's income and status (as well as the university's "overhead") may be maximized by publishing papers are frequently as possible, while commanding as large a "team" and as large a budget as possible for as long a time as possible.

For these researchers, and for statisticians who are are weary of advice on how to *avoid* pitfalls, I have prepared a 10-step guide to complement the writing of Professor Chatfield.

HOW TO MESS UP WITH DATA ANALYSIS: A 10-STEP HANDS-OFF GUIDE FOR RESEARCHERS

(1) There is no substitute for flawed data. "Garbage in, garbage out" is an old proverb, but still a guarantee.

A favorite example comes from the 1950s. Epidemiologic studies of cancer suggested that a woman had greater risk for cancer of the cervix if her husband was not circumcised. Usually data regarding circumcision (yes or no?) came from asking the wife or husband. It became more difficult to interpret this research after one study (Lilienfeld and Graham, 1958) found that physical examination by a physician contradicted self-report of circumcision status for more than one third of all men in a consecutive series of 192 male patients. (Men misreported in both directions.)

Data should be collected with no purpose in mind, so that the study is "unbiased"; and it will be "double-blindfold" also if the data are not tailored to any particular type of analysis.

Select samples haphazardly; ignore biases or confounding factors; and choose inappropriate controls (or none). Also consider how to mess up practical matters as well as statistical theory: remember to set dose levels unrealistically; use temperatures or pressures at which measuring instruments and chemicals do not function; and so on.

If a study must have some design or protocol—for example, to obtain a research grant or contract—then alter the protocol whimsically after the money is in hand. Better yet, leave the project to paid assistants who are not allowed to read the protocol. Vary data collection as much as possible while the study is in progress. (Why else call observations "variables"?) But never document changes.

Throw away any data that you do not like.

The remaining data can be degraded by mistakes when numbers are keyed into computer files. Or, later, use incorrect definitions or formats for the stored data.

(2) Keep no back-up copies of data files. If assistants insist on producing *good* data, then lose the data.

(3) When the study is complete, do not look at the data. You might notice something.

Dataholics Anonymous advises that even a peek at data ("just a quick one") or some innocent "social" discussion of data can lead to heavy thinking.

For the researcher who is bored or impatient while waiting for assistants to compute the p-values, one way to avoid looking at the data is to keep busy writing new requests for more funding.

(4) Never let data analysis be influenced by the context and purpose of the research, nor by the design of the study.

If only medians are appropriate, then report mean values instead. Analyze any sequence of observations as independent, identically distributed random variables, no matter how strong the chronologic trend may be. Ignore interaction terms in analysis-of-variance or in logistic regression.

If the purpose of an epidemiologic study is to predict presence or absence of a particular disease that is prevalent in 50% of a population, then report all "risk factors" that have "significant" coefficients in a logistic regression, even if the corresponding predictions are only 51% correct. (Better yet, do not even estimate success and error rates; just concentrate on those logistic regression coefficients.)

Of "the three cardinal sins of statistical analysis" identified by Morgan (1984), perhaps the easiest to commit is "failure to use paired analysis [...] comparing two closely matched populations." For example, in 1986 and again in 1989 funds from several research grants sponsored questionnaire surveys of microcomputer use among "all full-time" members of the Faculty of Health Sciences at one Canadian university. Since the response rates were high (84% in 1989) and presumably the Faculty was comprised of mostly the same members in 1989 as in 1986, paired analyses would seem to be appropriate. Instead, however, changes were tested by computing the usual Pearson chi-square statistics (with continuity correction) to compare 1986 versus 1989 proportions—comparing the proportions of faculty who used microcomputers for statistical analyses (38% in 1986 versus 47% in 1989) or for word processing, etc.—as if the corresponding 1986 and 1989 proportions were independent. However, the authors did use a "Bonferroni adjustment for multiple comparisons."

(5) Avoid simple graphics. (Instead, insist on p-values, as discussed below.)

In particular, never plot X, Y data in a simple scatter diagram. The graph could reveal some problem with linear regression—when numeric results otherwise seem publishable. A few years ago, for instance, at the university where I work, a new teaching hospital installed software that mislabeled the slope and the intercept of fitted lines. I discovered the problem while consulting with a young microbiologist. The hospital had difficulty hushing the whole matter, so that other researchers would not be disturbed. (I do not know if anyone else noticed when the regression routine was corrected.)

When graphics cannot be avoided, follow the manual on "How to display data badly" (Wainer, 1984).

(6) Always describe *data analysis* as "computing" and refer to *statistics* as "number crunching" —or as "details."

It is beneath the dignity of any busy researcher to understand either statistics in general or the analysis of his or her own data in particular. For an important person who always pretends to know all about everything, it is OK to *pretend* to understand statistics.

A researcher who does know *some* statistics (contrary to advice above) should use new and unfamiliar number crunching techniques that he or she does not understand. Chances are good that colleagues or editors also will not understand; so the analysis is less likely to be criticized—and certainly it will not be *improved* by someone else.

Try multivariate methods *first*. In high-dimensional models for which the researcher has neither intuition nor grasp of mathematical theory, detection of any errors or oddities is less likely (especially without graphics).

Ideally, use incomprehensible procedures that are currently *fashionable*. Imitate what was done in someone else's recent study, but perhaps include some complication. For example, if the previous study used a difficult analysis-of-variance with "fixed effect" factors, then use a similar design, but with one "random" factor. (Just let some random sample of individuals play the role of factor levels.) Wonderfully incorrect p-values might be achieved by crunching through the same F-tests that were applied for the fixed effects model.

Dr. Donald Mainland has described himself as "one who graduated in medicine in 1925, the year of publication of R. A. Fisher's *Statistical Methods for Research Workers*." While noting many changes in medical research during his own career, Mainland (1982) recognized that some problems have persisted in data analysis:

> ... when I look through the articles [in medical journals] and when I see the recent eager acceptance by medical investigators of statistical cookbook recipes I wonder if conditions are much better than [...] when the experimental biologist Lancelot Hogben [1950] affirmed that "less than 1% of research workers clearly comprehend the rationale of the statistical techniques they commonly invoke."

(7) Restrict number crunching to "package" computer programs.

Although many good statistical software packages are available, it is best to select one (at random) and to use this same software for all future number crunching. Force all analyses to fit whatever options are found in this statistical package. (Some researchers prefer to choose a different statistical package at random for each new analysis; but this approach may be too time consuming.)

> There are people who have to do statistics who can't possibly understand it. And this is not going to get any better with the computer. We now have these packages, and people will understand even less of what they are doing than they did before. [Erich L. Lehmann, quoted by DeGroot, 1986a]

> I want to be absolutely sure that I know what is being done, so I don't like packages. [Charles Stein, quoted by DeGroot, 1986b]

> ... anything that looks like the production of 'package addicts' or like a switch from the teaching of statistics to the teaching of packages is to be deplored... [Preece, 1986]

> ... maximize the role played by computers

> [. . .] with a minimum of human tutorial contact. [Head of my department in a Faculty of Medicine, 1986]

(8) Delegate all number crunching to assistants (or, if necessary, to co-authors).

> Any important researcher will usually have a large number of subordinates (called graduate students, interns, or fellows). At least one of those subordinates can often be induced to study elementary statistics textbooks or learn how to run packages like SAS or BMDP. Graduate students have usually taken a vow of poverty and penitence, and some of them are so fond of the hair shirt that they will eagerly take to such tasks [Salsburg, 1985].

Keep subordinates busy enough to discourage them from any reading or any independent contacts relevant to new developments in statistical methodology.

(9) Display computer output and p-values as ritual offerings to placate gods of "significance" testing. For specifics, see essays on "Clinical trials as a religion" (Rimm and Bortin, 1978) and on "The religion of statistics as practiced in medical journals" (Salsburg, 1985).

The mechanical "significance" testing process sometimes makes data analysis a meaningless mess, but with results that usually look neat. As Preece (1986) says, "... so much statistical practice is not good [. . .] because so much is ritual gone blind and deaf. . . ."

When computer output is piled on the desk and there is some doubt about which numbers to report (in addition to p-values) then include *all* of them (means, standard errors, kurtosises, correlations, kappas, chi-squares, gammas . . .). Friends will be impressed; and critics will have a bigger job. The more numbers are reported, the better the chances of including at least a few statistics that are inappropriate or misleading.

> . . . repentance . . . may indeed be appropriate for one who in the 1930s started to propagandize statistical methods in medicine and now is depressed by the abundance of P's, t's, χ^2's, r's, SD's, SE's, and NS's in medical journal articles [Mainland, 1982].

(10) Avoid statisticians. Their views are meddlesome and contrary.

In hindsight, field trials of the Salk polio vaccine in 1954 are remembered as one of the largest (and most influential) experiments in the history of medicine or public health (Meier, 1989). These vaccine trials led directly to reduced incidence of polio. They also publicized and established the methodology of the randomized, placebo control, double-blindfold clinical trial as an epidemiologic paradigm.

But in 1953 Dr. Jonas Salk had other plans: "Let us [. . .] plan the recipe that will bake the cake that tastes the way we want it to"; and he wrote that "The use of a placebo control [. . .] would make the humanitarian shudder and would make Hippocrates turn over in his grave."

> In talks with many people in our own group, in Pittsburgh, and others as well, I found but one person who rigidly adhered to the idea of a placebo control and he is a bio-statistician who, if he did not adhere to this view, would have had to admit his own purposelessness in life [Jonas Salk, quoted by Carter, 1967, pages 186–187].

If statisticians must be involved, then the researcher at least should procrastinate.

> To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of [Ronald A. Fisher, 1938, quoted by Bibby, 1986, page 35].