Possible Progress and Unresolved Conflicts Resulting from Guidelines on Good Epidemiologic Practices

C. Ralph Buncher, ScD; James J. Collins, PhD; and William Halperin, MD, MPH

The papers in this issue of the Journal have illuminated a number of points within this topic of good epidemiologic practices and how to put these into effect. This paper is one place to reflect on these issues and to see what has been accomplished. In particular, is this new set of guidelines going to improve occupational epidemiology or will it be another annoying set of rules? Will these guidelines improve the general quality of studies or will they ultimately be a useless addition to paperwork? Finally, we discuss briefly some unresolved dilemmas in this field, particularly those that epidemiologists have yet to reach a consensus on the underlying issues; further discussion is needed.

The good epidemiologic practices (GEP) guidelines require a great deal more paperwork, quality control, quality assurance, documentation, and attention to archiving the data base and results than was previously the case. One estimate is that 15% of the resources currently available will be devoted to these tasks, although the actual cost in any particular setting is arguable. Few anticipate additional funding to pay for the costs of GEP. Therefore, there may be 15% less occupational epidemiology conducted, resulting in what might be called a "practice tax" on epidemiology. Thus the specific question is whether the GEPs will overcome this practice tax with studies demonstrating quality, excellence, and conviction.

From the University of Cincinnati Medical Center, Cincinnati, Ohio (Dr Buncher); the Monsanto Company, St. Louis, MO (Dr. Collins); and the National Institute for Occupational Safety and Health (Dr. Halperin).

Address correspondence to Dr Buncher, Professor and Director of Epidemiology and Biostatistics, University of Cincinnati Medical Center, Cincinnati, OH 45267-0183.

0096-1736/91/3312-1261$03.00/0

The answer to this question depends a lot on one's view of the state of epidemiologic practice prior to GEP. For example, no one seriously doubts that a protocol should be written and peer-reviewed prior to beginning the study and that a more complete protocol leads to better and more productive studies. Few contest the concept of archiving the data from a completed study. If these activities are already routine in an organization, then additional costs will be minimal.

Some of us have seen how a set of rules can be used as a tool to destroy a truly good study. Guidelines always become rules; those who ignore a guideline are asked to explain their deviation. If the study ends up in litigation, then the GEPs can be used by some to measure how well the study was accomplished. Unfortunately, the rules can become the focus of disputes that do not really judge the quality of the study. For example if the dispute is over the issue of whether quality control was adequately done, the disputants may lose sight of the fact that quality might have been excellent in the study even if there was no formal quality control program.

Quality control is not a goal unto itself; a quality study is the goal. Yet the epidemiologist must remember that often the legal dispute is over form rather than substance. This type of checklist mentality could result in GEPs becoming more of a problem than an accomplishment. Only the future will give us the answers to this concern. In the next 5 to 10 years, as GEPs become part of the environment in occupational epidemiology, let us all watch carefully to be sure that the new guidelines are useful rather than simply another hurdle to overcome.

Exceptions must also be considered. For example, "response epidemiology" will not permit time for written protocols. By response epidemiology, we mean those instances in which enough of a problem is known before the study is started that immediate remediation is required rather than the more familiar goal of establishing

that a hazard exists. If someone notes that six cases of bladder cancer have occurred among 34 workers, then it is clear than an epidemic already exists. If 22 workers out of 73 have a trauma disorder, then changing the situation as soon as possible is the concern. In establishing the facts of the situation, one also does most of the study. Thus the facts of the situation are frequently completed before the protocol is written.

Release of Data Sets

Data from federal government researchers are now available in public use tapes or through Freedom of Information Act requests within the limits established by the Privacy Act. Data from industry, academia, and consultants have occasionally been shared. The data set is usually desired because someone or some organization wishes to see an alternative analysis. Most understand that an alternative analysis can illuminate an issue; in many cases, it can make the readers more sure of the initial analysis. Thus, making the data from epidemiologic studies available has many advocates.

There are also controversies in the issue of release of data from occupational studies. In a world of open science, it would be wonderful to make available all data sets to any other person who has a use for the information recorded in a study. Therefore, open access to data for additional study, for testing alternative hypotheses, and for educating the next investigators are all very positive goals. At the same time, most of us want particular information about us to remain private. We do not want our medical status, our clinical history, our psychological comments, our work evaluations, and numerous other characteristics to become part of the public record. Privacy of patient information is thought to be essential for the patient and essential to obtain the cooperation of patients in the future.

These opposite goals of access to data banks and privacy produce conflicts in how and to what extent data should be made available for research purposes by others. The simplistic view is to remove personal identifiers and thus the data set can be released. Those who work in the field know that removing a name does not prevent identifying the individuals in a data base. There are few men born on March 1, 1944 who weigh 100 kilograms (220 pounds), began working on June 20, 1967, were exposed to products A, B, and F, and have a diagnosis of kidney cancer.

Omitting personal identifiers is not sufficient to prevent identifying individuals if there is a whole array of characteristics in the data set. Just two alternatives to each of 10 characteristics results in over a thousand different categories; therefore, several hundred people in such a data set essentially could be identified on an individual basis. Three alternatives for each of 12 characteristics produces an array of over half-a-million different possibilities. Thus, any extensive data set could be used to identify individuals and thereby thwart desires for privacy.

There are many approaches to dealing with the problem of identifiability. These range from censoring or expunging of identifiable data through introduction of random errors into the data set. One alternative that deserves research is random perturbations to the data. By introducing a small degree of error, it should be possible to prevent individual identification and at the same time not lose the essential characteristics of a data set. Thus, one could add a small randomly determined error to the date of birth, the date of hire, the weight, and even to exposures and diagnoses. Random changes to exposure variables will on the average reduce power, but random changes to extra variables will have no effect other than enhancing privacy. Randomly changing some of the persons but not changing all also implies that persons trying to identify individuals can not be sure that a putative match is not a random coincidence.

Research is needed to investigate whether the basic characteristics of the data set could be retained while privacy is enhanced by adding small errors to the data items, thus preventing exact matches. Could some level of random error hide the identity of individuals while retaining the essential occupational interrelationships of the variables? If such an algorithm can be produced, then data sets could be released more easily with individual privacy protected and group information intact for research purposes.

The statistical community may wish to investigate methods to protect privacy, perhaps including the introduction of random errors and their effect on the ability to match. Epidemiologists have wrestled with the other side of this problem in trying to match, for example, persons in a cohort with persons in a vital statistics mortality file. The question is to find the optimum scheme for randomizing data to maintain the relationships and at the same time to reduce the ability to match individuals.

Data Sharing

Another issue is making the basic data set available to others. Most epidemiologists believe that data sets should be available to be analyzed by responsible members of the profession. A few practical considerations, however, give pause to this professed goal. First is the issue of privacy as discussed above. In the federal government, there currently is a balance between access to information provided by the Freedom of Information Act and protection of individual privacy provided by the Privacy Act. This balance is not yet developed for epidemiologists working in other venues.

Another practical concern is whether the second analyst can read and understand the data set that has been made available. Only the naive or inexperienced think that one can get a data set and a code book and then at once do productive work with another person's data set. There are always idiosyncracies in the data set that produce confusion to those who have not mastered the nuances of the coding scheme. Why are there codes both for "not done" and "incomplete" if they have the same meaning? Answer: at one time we expected to record partial results but sufficient information was not available, so both codes now mean the same thing. The

response to a secondary analysis is frequently a valid statement that the second investigator did not understand the data set properly and has misinterpreted key variables.

An important consideration is that data may be made available to those who are perceived to have a malicious intent, ie, some one with entrenched interests who is paying these individuals to obtain your data and then impugn your conclusions. Reanalysts dream that they will find a fatal flaw in the original analysis and, therefore, the reanalysis will produce great enlightenment. A nightmare for researchers is that some critic will obtain your data, reanalyze it, announce that your analysis was faulty, and then persist on this path even after you have properly defended yourself. See J. Palca,[1] for an example of the potential conflict. Years of productive life can be spent in defending oneself and sometimes it is the attacker who is completely wrong. Any data set made public enables one's valid and invalid critics alike to comment on what you have done.

The epidemiologic community has yet to sufficiently weigh the advantages and disadvantages of making data available to others. It would also be useful to have empiric data on how often a reanalysis leads to enlightenment and how often it leads to lots of extra work but no additional insights.

Archiving Data

Archiving data sets sounds like it should be simple and worthwhile. One difficulty is the medium by which the material is archived. Should it be on computer tape, video disks, hard disk, floppy disk, cartridges, or what? Recent history shows that we have changed our storage devices every decade or so, and a decade after that the only machines available to retrieve that format for data are in a museum. In this changing environment, how can one archive for 20 years into the future and, preferably, for 50 years into the future? Archiving can be done on paper or a paper substitute such as microfiche, which is effective but inefficient.

One answer is that the archived data set must be routinely upgraded, which will cost additional time and effort, much of which will be wasted since many such data sets will never be queried again. Researchers have lost data sets because they were on ordinary computer tape and after 7 to 15 years without any use, the tapes were no longer readable. Who pays for the upgrades to the archive if funding for that project has ceased? Who is responsible for archiving the study? If one epidemiologist does the study, works for the company involved, and is responsible in the future, the answer is obvious. What about the instance of a company that is responsible but an outside epidemiologist, whether an academic or a consultant, has done the work? In a number of instances, only the epidemiologist who does the study retains the archived copy of the results.

A consultant for a company may obtain data that are to remain private from the company, eg, particular clinical information not available to the company. How are such data archived and by whom? Who is to enforce the responsibility for a contractor to archive the data for the next ten years, especially in those instances in which the consultant goes out of business or retires from the field? Do we need some unit to serve as the repository for archived data sets? A great deal of thought needs to be devoted to the issue of practical considerations in how to archive data sets.

Study Hypotheses

Another conflict revolves around the issue of multiple comparisons. What does one do when there are numerous clinical conditions and numerous exposures? One common way to resolve this problem is to list one, two, or three key hypotheses in the protocol. These are the hypotheses that generated the current study and which are to be resolved by accumulating the data to provide an answer to the truth of the hypothesis. The other hypotheses are not formulated at the beginning of the study or are of a secondary nature in the hierarchy of hypotheses to be tested. These "other hypotheses" can have *P*-values or confidence limits calculated but they shall remain unproven. It is only through new and independent studies that one can really investigate these other hypotheses, say some epidemiologists. Others will make judgements based on available data and criteria for causation.

Study Power

A final topic of interest is whether to do a study with very low power, ie, a study with such a small sample size that it is very unlikely to be able to find a relationship between exposure and disease even if it is actually there. Another example would be a study of a potential carcinogen done before the latent period has elapsed. Thus a study of a chemical suspected to cause cancer is doomed to a finding of no relationship if it is done only 7 years after the exposure and the average latent period is 20 years.

Some would advocate that any study that cannot find a positive relation should not be done since it only produces a false sense of security. Others feel that the study must be done in many instances but reported with the caveat that only a very large risk would have been found in this study. One example is a company that produces a chemical suspected to be a risk factor. The company and its epidemiologists may feel required to study the chemical now, even though they know there is little chance of finding anything positive. The consensus in the field seems to be that a study and even the small addition to knowledge that its results produce is better than no study at all. Many small studies viewed together can overcome the lack of precision in a single study, but not the problem of insufficient latent period. Additional thought on this subject is warranted.

Some argue that study power is a double-edged sword when seen from the public health point of view. A large study can detect a more modest risk than a small study but still ignore virtually the same number of disease

occurrences as in a small study as not being meaningful for the testing of the null hypothesis. Study power is a useful planning tool when the investigator wants to draw study subjects from a larger pool of candidates. The concept of study power. however, is of more limited value in most occupational settings since the investigator chooses to include all persons from the limited number of those who were exposed.

Summary

In summary, producing guidelines also produces conflict. The purpose of this paper was to delineate some of the conflicts that result from, or are involved in, having a set of guidelines for good epidemiologic practices. The goal was to present briefly both sides of the conflicts without trying to resolve them. The hope is that others will cogently argue one side or the other of these conflicts and convince the community to resolve the conflict, or alternatively, perhaps that the current guideline is somewhere in the middle of the beliefs in the community and thus is correctly positioned.

References

1. Palca, J. Get-the-lead-out Guru challenged. *Science.* 1991;253:842–844.

Get Hooked on Seafood Safety

To get a truer picture of the safety of seafood, FDA's Center for Food Safety and Applied Nutrition, in cooperation with CDC, did a risk assessment study. It showed about one illness per million servings of seafood when raw or partially cooked molluscan shellfish (mussels, scallops, clams, and oysters) were excluded. . . . (In comparison, the risk assessment for chicken is about one illness for every 25,000 servings). (Adding) raw mollusks . . . to the FDA risk assessment, the chance of illness for seafood overall . . . jumps to something like 1 in 250,000 servings. That's still 10 times safer than eating chicken, but the agency figures that those raw oysters, clams, and mussels . . . account for a whopping 85% of all the illnesses caused by eating seafood.

Mollusks are troublemakers because most cannot move and have to feed by filtering water through their systems, pulling out nutrients in the process. In so doing, they also can pick up and store harmful bacteria and viruses that can cause a string of illnesses. When people eat these pathogen-packed shellfish raw they ingest the viruses and bacteria. . . . As Anthony Guarino, director of FDA's fishery research branch on Dauphin Island, AL, asks: "What other animal do we eat, digestive tract and all, without cooking it first?"

FDA's risk assessment study concluded that 1 out of every 1,000 to 2,000 servings of raw mollusks is likely to make someone ill . . . Not enough people realize the danger in eating them uncooked, particularly when they are taken from contaminated warmer waters or held and shipped without adequate refrigeration. . . . Two states—Louisiana and California—now require warning notices about eating raw shellfish at places where they are sold. In Louisiana, the following notice is required: *WARNING; Raw oysters, raw clams, and raw mussels can cause serious illness in persons with liver, stomach, blood or immune disorders.* The California notice requires a similar tag on the sack or container of oysters from the Gulf of Mexico.

Oysters taken from the Gulf of Mexico, particularly from March through October, may contain a naturally occurring pathogen called *Vibrio vulnificus* which is particularly pernicious to persons with liver disease. . . . Cancer patients, people with iron metabolism disorders, and those with weakened immune systems . . . may also be vulnerable. The risks are high. The fatality rate for at-risk individuals who become infected is more than fifty percent, with death usually occurring within two days.

—From "Get Hooked on Seafood Safety" by RW Miller.
FDA Consumer 1991; 25:5:6