Classification of Data in Research

Lawrence Garfinkel, M.A.

Between 1967 and 1968 the U.S. death rate for cancer of the lung and bronchus—which had been increasing at an average rate of 5.7 percent per year—suddenly increased by 9.6 percent. This caused concern among some cancer control workers, but not among vital statisticians who were aware that a new classification (the Eighth Revision) of the International Classification of Diseases (ICD) had gone into effect in 1968. They surmised that this sharp increase was probably not real, but resulted from classification changes in coding causes of death. A subsequent analysis of a sample of death certificates for persons dying with cancer coded by both the Seventh Revision and the Eighth Revision of the ICD proved that this hypothesis was correct. In the Eighth Revision of the ICD there was an increase of about 2.5 percent in the certificates coded to primary cancer of the lung and bronchus compared to the Seventh Revision.\(^1\) The major effect of the changes in the Eighth Revision with respect to lung cancer was to code many additional cases previously recorded as cancer, primary site unknown, to primary lung cancer.

Even when the same edition of the ICD is used, there can be substantial changes in classification of causes of death in international comparisons because of different interpretations of coding rules. In a recent study, a random sample of 1,246 death certificates which mentioned cancer were coded by experienced nosologists in seven different countries, including the United States. The analysis showed that there was some difference in coding of the underlying cause of death by at least one nosologist in 47 percent of the death certificates.\(^2\) Compared to the coding by the United States nosologist, the coders in the six other countries coded a different cause of death in 12 to 27 percent of the death certificates. The physician who completes a death certificate indicates what he believes to be the underlying cause of death in a majority of instances. Problems may arise either when more than one cause of death is entered on the death certificate and it is not clear which was the underlying cause of death, or when the physician uses ambiguous language. In these instances the coding clerk has to interpret what the physician had in mind and this can, of course, lead to discrepancies (see box, p. 4). In the Ninth Revision of the ICD—which will go into effect on January 1, 1979—rules for coding neoplasms have been greatly expanded and will be applied similarly by all countries, thus making international cancer mortality rates more comparable.

Problems of classification are present in a number of areas of medical and scien-

---

\(^1\) Garfinkel, M.A., Garfinkel, M.A., and Garfinkel, M.A. (1969). The Eighth Revision of the International Classification of Diseases, New York, N.Y.

\(^2\) Garfinkel, M.A., Garfinkel, M.A., and Garfinkel, M.A. (1970). The Ninth Revision of the International Classification of Diseases, New York, N.Y.
tific research. Investigators without full knowledge of definitions used should always be cautious in comparing results of different studies appearing in the literature. Although different investigators may put the same name to a diagnosis or a procedure, it is not always clear that they are referring to the same entity. For example, a study comparing the histologic type of lung cancer of the same slides read at different periods of time showed a 37 percent discrepancy. Definitions and concepts do change over time. There also may be considerable variability among pathologists reading the same slides. If one adds the effect of observer variability to changes in definition, the errors in classification become even more acute. In another study of histologic classification of lung cancer, intra-observer variability ranged from two to 20 percent when reading the same slide twice. Percentage disagreements from consensus reading varied among five pathologists, ranging from two percent for the most differentiated cancers to 42 percent for the most poorly differentiated types.

The same problems may exist in interpreting radiographs. In one series of evaluations physicians missed finding 32 percent of positive films, and incorrectly identified 1.7 percent as positive. In another study, there was unanimous agreement of the interpretation of positive or negative finding of pneumoconiosis in only slightly more than 50 percent of chest radiographs.

Potential errors in classification also apply to the comparison of results of different questionnaire studies. For example, consider the classification of “ex-smokers” in epidemiological studies of smoking and health. Ex-cigarette smokers are a motley group: they include both long-term and short-term former smokers, those who recently quit because of an acute illness but who will shortly resume smoking, former light, moderate and heavy smokers, etc. Mortality rates for this group as a whole will depend upon the proportions of the various subgroups represented in a particular study. Replies can also depend greatly on how the questionnaire is worded and how various subgroups of subjects classify themselves. Thus it is apparent that comparing results of groups called “ex-smokers” in several studies can be highly misleading if the definitions are not the same, or if the groups are not clearly described. In some studies “ex-smokers” are combined with those who never smoked. Consequently, relative mortality risks then have different meanings in different studies.

From the foregoing discussion it might be inferred that the results of any biomedical study may be suspect because of the problems of classification. This is not true. Fortunately, in many analyses the errors of classification are minimal and do not affect the results in any substantive way. But it does indicate that the publication of a properly conducted research investigation should always specify the definitions of the classification used, especially when the classification is based upon subjective recordings.

When a panel of specialists or experts is involved in a study (pathologic diagnosis, evaluation of therapy, etc.) it is very important to test their agreement on definitions before the study begins. If they differ greatly, preliminary discussions can lead to mutually agreeable standards. Then the new protocol should be tested on a group of subjects not included in the main study to determine inter-observer variability. Obviously it would be impossible to get 100 percent agreement by panelists on subjective judgments, but if a consensus is to be used for a final score, or an evaluation or diagnosis, it must re-

"...if a consensus is to be used for a final score, or an evaluation or diagnosis, it must reflect the opinion of the great majority of all judgments—or it is of dubious value."

VOL. 29, NO. 1 JANUARY/FEBRUARY 1979
The standard international form for medical certification of cause of death designed by the World Health Organization (WHO) consists of Part I, used for reporting the disease or condition leading directly to death (a), and the antecedent conditions (b) and (c) that give rise to the cause reported in (a). The underlying condition is stated last. Part II is used for “other significant conditions” contributing to but not related to the disease or condition given in Part I. The underlying cause of death is used in national and international summaries.

WHO publishes the International Classification of Diseases (ICD) which includes a numerical and an alphabetic index to assist nosologists in selecting the correct code to apply for the underlying cause of death. In addition there is a set of coding rules published in each revision of the ICD. The U.S. National Center for Health Statistics (NCHS) annually publishes detailed instruction manuals based on these rules, and provides detailed coding rules and examples of coding procedures for cases where the underlying cause of death is not clear.

In 1968, NCHS developed a computer program called ACME, which in most cases relieves the coder of the decision of determining the underlying cause of death in ambiguous death certificates. The coder records the four-digit codes for each cause listed, as well as the line and order in which it appears, and the computer program for the majority of cases determines the underlying cause based on a series of decision tables, which have been developed to take into consideration the logical biologic relationships of the individual causes of death.

To illustrate how coding rules affect the site of cancer recorded as the underlying cause of death, consider the hypothetical example of how three physicians may have recorded the death of the same individual. These hypothetical death certificates have been adapted from actual examples.

The patient had an adenocarcinoma of the stomach which metastasized to the prostate. X-ray records also indicated some evidence of lung metastasis, and the patient died of hypostatic pneumonia.

The death certificates by three physicians may have read as follows:

Physician 1.
I(a) Hypostatic pneumonia
(b) Cancer of stomach and prostate
(c) Metastatic lung cancer

Physician 2.
I(a) Hypostatic pneumonia
(b) Cancer of stomach
(c) Cancer of prostate

Physician 3.
I(a) Metastatic cancer to prostate
(b) and lung
(c) Adenocarcinoma of stomach, primary

II Pneumonia, Terminal

On the first physician’s death certificate, cancer of the lung would be selected according to the rules as the underlying cause of death on the basis that lung cancer metastasized to the other two sites. On the second physician’s certificate, the underlying cause of death would be cancer of the prostate since this is on line I(c). On the third physician’s certificate, the underlying cause of death would be correctly ascribed to cancer of the stomach.

It is thus evident that the physician who certifies the cause of death should be careful to indicate the sequence of events leading up to death. If cancer is mentioned, the physician should indicate the primary site, or if it cannot be determined, state that the primary site was unknown.
reflect the opinion of the great majority of all judgments — or it is of dubious value. The same principles hold for testing the output of a new biomedical device that records electrical impulses, laboratory testing equipment, or a survey questionnaire. The limitations of a particular device and its random fluctuations must be tested under different conditions to assess the degree of experimental error. In a questionnaire, techniques have been developed to test how accurately questions have been designed to elicit “correct” answers.

Results of research studies and the interpretation of data are closely connected with the classification of the data, particularly when the differences observed are borderline. It is important, then, for the reader to know the definitions of the classifications used, the care that the investigator took to pretest concordance of evaluations of panelists, and the magnitude of the experimental error.

References

1. Percy, C.; Garfinkel, L.; Kreuger, D.E., and Dolman, A.B.: Apparent changes in cancer mortality, 1968: A study of the effects of the introduction of the Eighth Revision International Classification of Diseases. Public Health Rep. 89:418-428, 1974.
2. Percy, C., and Dolman, A.: Comparison of the coding of death certificates related to cancer in seven countries. Public Health Rep. 93:335-350, 1978.
3. Yesner, R.; Gelfman, N.A., and Feinstein, A.R.: A reappraisal of histopathology in lung cancer and correlation of cell types with antecedent cigarette smoking. Am. Rev. Resp. Dis. 107:790-797, 1973.
4. Feinstein, A.R. et al.: Observer variability in the histopathologic diagnosis of lung cancer. Am. Rev. Resp. Dis. 101:671-684, 1970.
5. Yerushalmy, J.: Reliability of chest radiography in the diagnosis of pulmonary lesions. Am. J. Surg. 89:231-240, 1955.
6. Reger, R.B., and Morgan, W.K.C.: On the factors influencing consistency in the radiologic diagnosis of pneumoconiosis. Am. Rev. Resp. Dis. 102:903-915, 1970.