AMERICAN JOURNAL OF
Public
Editorials Health
September 1980 Established 1911
Volume 70, Number 9

Assessing the Quality of Vital Statistics

Three articles in this issue of the Journal focus on the quality and completeness of vital statistics data, particularly data on live births and infant deaths. The quality and completeness of these data are especially important because they constitute what may be the most widely and continuously used indicators of health status at the local, state, and national levels. It is a service to the Journal readers to have these articles in the same issue.

The articles are also of interest because they reflect contrasting approaches to the evaluation of quality and completeness of vital statistics data. Two of the articles—by Frost and Shy,[1] and McCarthy, *et al*[2]—compared information from independent but linked data files (births and deaths), while the article by David[3] examines the internal consistency and completeness of a single file (births). In addition, the three articles represent vital statistics systems in different parts of the nation, namely North Carolina, Washington State, and Georgia.

David's paper deals with the quality and completeness of both the sociodemographic and the biomedical data for three years of births (1975-1977) from the computerized files of the North Carolina Division of Health Services. It emphasizes two data elements—birth weight and period of gestation. David found that the reporting of birth weight, which is critical in studies of infant health risk, is virtually complete. Period of gestation, however, is completely reported only about 80 per cent of the time.

From the point of view of interpreting natality data, an important finding is that those records with missing or erroneous information are not randomly distributed among all births. Instead, reporting problems appear to be associated with those groups considered to be at high risk of potential infant health problems—that is, births among the young, the unmarried, and races other than White.

David suggests a number of approaches that might be used to improve the quality of natality data. For example, he recommends that birth records with values that are judged to be unrealistic be checked for errors and that feedback be routinely provided to the reporting hospitals on the quality of their data.

While David examines a single file of vital statistics data for completeness and internal consistency, the paper by Frost and Shy compares information from two different files for the same items. Frost and Shy examined the race classification in the birth and death files of Washington State for infants who die within their first year of life. The quality of this information is, of course, crucial for the study of racial differentials in infant mortality.

Frost and Shy found discrepancies in the classification of race between the birth and death files for the same infant. In general, the results are consistent with those of previous studies on this subject.[4] Where discrepancies do occur between the race recorded at birth and at death, there is a tendency for the infant to be reported as White on the death certificate. In addition, discrepancies are proportionately greater for infant deaths in numerically small racial groups, such as American Indians, Chinese, Japanese, and Hawaiians.

Frost and Shy show how discrepancies in racial designation would affect infant mortality rates for American Indians, but they do not show "corrected" rates for other racial groups. For the racial groups that constitute the largest proportion of infant deaths, namely White and Black infants, the statistical consequences of discrepant racial assignment are negligible.

Frost and Shy's finding that the greatest per cent of discrepancies were for those infants of small racial groups underscores two problems:

- Where small numbers of vital events are concerned, regardless of the quality of reporting, vital statistics data may have limitations for analytical use. Because small frequencies are affected by random variation, caution should be exercised in their use and interpretation, whether they are statistics for small demographic groups or for small geographic areas.
- The second problem is that reporting or classification problems can usually be expected to have the greatest impact on data for small groups.

The difference in reporting of race between the birth and death certificates for the same infant involves two factors. One of these factors is the difference in the way in which race is determined from the birth and the death certificates. In the case of births, the race of the infant is not reported on the certificate but is *derived* with a necessarily arbitrary set of rules using information on the race of *each* parent. Where parents are of different races, the rules assign a single race to the infant. For infant deaths, in contrast, the race of the infant is reported directly on the certificate. Because of this difference between birth and death files, some differences in assigning race can be expected, regardless of the quality of reporting.

The second factor is the possibility of errors in the *reporting* of race of the parents on the birth certificate and race of the infant on the death certificate. For births, race is usually reported by hospital personnel on the basis of information obtained from parents or by observation. For deaths, race is usually reported by funeral directors on the basis of information supplied by informants or by observation. A useful next step in Frost and Shy's analysis would be to determine why perceptions of race may differ between the filing of the birth certificate and the death certificate.

The study by McCarthy, *et al,* made use of linked birth and infant death records to evaluate the quality of vital statistics data. Following a cohort of low-birth-weight infants, they found significant underregistration of infant deaths in Georgia during the 1974-1977 period. The authors also present suggestive evidence that similar underreporting may have existed in a number of states in 1960. The McCarthy team also found, as did David, that problems of registration completeness appear to be concentrated among certain population groups—those with the highest incidence of infant health problems and infant mortality.

McCarthy and his colleagues have made a valuable contribution in a number of respects. Beginning with a clue to the problem of underreporting based on birth-weight-specific mortality rates, they developed a systematic procedure for following a cohort of low-birth-weight infants. Confirming that the deaths of some of these infants were not registered, they further examined the procedural factors that may have accounted for the incomplete registration, and then made appropriate recommendations. As a result, the State of Georgia instituted modified procedures effective in January 1978 to correct the problems.

The papers by Frost and Shy and by McCarthy demonstrate the usefulness of creating files of linked birth and infant death records. Frost and Shy show that such files can provide data of greater racial specificity—in terms of the race of each parent—than the death file alone. It is only with linked files that one can develop infant mortality rates that are completely consistent with respect to racial designation for the numerator and the denominator. McCarthy and his colleagues show that these linked files can also provide important infant mortality statistics, such as birth-weight-specific rates, and that birth-weight-specific rates can be important clues to the completeness of such data.

Studies like those reported in this issue of the Journal are invaluable because of the importance of continuously assessing the quality of the data that we use in research and in program planning, administration, and evaluation. The quality of vital statistics data is generally quite good inasmuch as states adhere to high quality control standards. Constant vigilance is essential, however, to ensure good communication between the personnel in state vital statistics offices where data are processed and the personnel in hospitals and funeral homes charged with responsibilities for completing and filing the death certificates.

While each of the papers in this Journal is concerned with a specific problem area, such studies provide model methodologies for evaluating quality and completeness of data, and—more importantly—provide some guidance on steps to improve data collection that can be widely implemented. It is through such evaluations that we can continue to improve the quality of vital statistics in the United States.

JOHN E. PATTERSON

Address reprint requests to John E. Patterson, Director, Division of Vital Statistics, National Center for Health Statistics, DHHS, 3700 East-West Highway, Hyattsville, MD 20782.

REFERENCES

1. Frost F and Shy KK: Racial differences between linked birth and infant death records in Washington state. Am J Public Health 1980; 70:974-976.
2. McCarthy BJ, Terry J, Rochat RW, Quave S, and Tyler CW Jr: The underregistration of neonatal deaths: Georgia 1974- 77. Am J Public Health 1980; 70:977-982.
3. David RJ: The quality and completeness of birthweight and gestational age data in computerized birth files. Am J Public Health 1980; 70:964-973.
4. Norris FD and Shipley PW: A closer look at race differentials in California's infant mortality, 1965-1967. HSMHA Health Reports 1971; 86:810-814.