DISCUSSION

Gilbert W. Beebe, National Cancer Institute

These papers bear witness to a deepening interest in the use of existing statistical information to learn more about socioeconomic differentials in mortality. Despite its obvious importance for public policy, this has been a relatively neglected topic in the U.S., and I welcome the evidence of these papers that there are people and agencies working to devise record linkage patterns that may, in time, provide solid National data on socioeconomic differentials. As Kitagawa and Hauser have said, in their 1973 American Public Health Association monograph:

> "The importance of socioeconomic differentials in mortality is that they point to the possibility of reducing mortality through the betterment of socioeconomic conditions in the population."

If we extend the meaning of socioeconomic differentials to include those for detailed occupational categories, then such information also has potential value in studies of the etiology of disease, e.g., cancer, and in regulatory efforts to reduce the hazards of the workplace. Much of the needed information is already in government hands, but it is fragmented and needs to be brought together. One very favorable development, referred to by Dr. Rosenberg, is the Congressional mandate recently given NCHS in Public Law 95-623 to "develop a plan for the collection and coordination of statistical and epidemiological data on the effects of the environment on health."

The first two papers exemplify somewhat parallel approaches to the estimation of socioeconomic differentials in mortality. Caldwell and Diamond follow a 1969 IRS cohort through SSA files to develop mortality gradients with income. Information on other demographic variables, especially education and occupation, and on cause of death, would add greatly to the value of such information.

For me the significance of the paper lies in its illustration of certain aspects of method. Especially important is their suggested use of the Statistics of Income (SOI) sample:

1. The SOI provides a ready-made cohort suitable for longitudinal mortality follow-up through SSA records.
2. Since a different SOI sample is, I believe, selected each year, a cohort of substantial size could be defined from the SOI samples within a few years.
3. Although no costs are cited, it is noteworthy that 98 percent were acceptably matched -- both records bear the social security number or SSN. Furthermore, the IRS polices the SSN entry on Forms 1040 and 1040A, as I found this year when I transposed two digits of my wife's SSN on our joint return.

Rosen and Taubman provide a demonstration of another potentially useful linkage, the Current Population Survey (CPS) demographic data and SSA information on mortality. The CPS provides a stronger data base for exploring socioeconomic differentials in mortality than does the IRS sample, but it is small and less readily expanded to adequate size. Rosen and Taubman compare their socioeconomic differentials in mortality with those found by Kitagawa and Hauser in their 1960 study of death certificates manually matched to Census records. Whether their sample of 3,000 deaths is large enough to permit conclusions to be drawn about changes in these differentials between 1960 and 1973-1976 seems doubtful to me. But, with reference to the difference between the two studies as to mortality gradients with educational level of white males 65 or older, I should note that Kitagawa and Hauser did find such differentials among white females in 1960, and also that there have been great changes in mortality since 1960, especially, but not exclusively, from cardiovascular disease. Whereas cancer mortality has held firm over the period 1960-1976, cardiovascular mortality has dropped considerably in this interval and so has mortality from all other diseases. It would be interesting to know if socioeconomic factors played any role in the dynamics of these changes which have proven very puzzling to students of cardiovascular disease.

The third paper, by Alvey and Aziz, addresses the completeness of SSA mortality information. If the SSA file is to be used for death clearance, it is important to know how complete it is and how the degree of completeness varies with at least key demographic variables. Theirs is a very preliminary report on the results of matching a subsample of 2,000 deaths in 1975 selected from their entire test sample of 23,000. Not surprisingly, 11 percent of the death certificates lacked SSNs, and 1 percent had invalid SSNs. For only 80 percent did the initial work, collating on SSN, yield "probably good" matches. However, as supplementary nominal searches are made, the percentage of good matches will surely rise, perhaps to 90 percent or more, and a more definitive estimate of the completeness of the SSA information on mortality will become available.

Of course, the completeness of the SSA file in regard to deaths in 1975 may be better than that of earlier years. Epidemiologists have been celebrating the decision by NCHS to organize a National Death Index to which study samples may be taken for death clearance. However, we often forget that the Index will, at least initially, start with 1979 deaths. For some time, therefore, the SSA file will be the only general source for mortality in earlier years, and it is important that we have information on its completeness and the subsequent distribution of missing deaths.

The last two papers attack a deeper and more difficult problem, namely, the creation of mortality data by occupation and with occupation used not only to scale subjects as to socioeconomic status, but also as an index to environmental exposures of the workplace. Those of us who are interested in the etiology of disease, or in cleaning up the workplace, need desperately to have systematic, National data on occupational mortality to direct our attention to possible trouble-spots. Cancer epidemiologists believe that 80 to 90 percent of all cancer is environmentally induced. In fact, there have been estimates, not widely accepted, that, in forthcoming decades, 20 to 40 percent of all cancer mortality might be occupationally induced. Furthermore, problems of the workplace affect not only the workers, but, in some cases, their families, and even people living near the plant. Hence, the discovery of a cancer hazard in the workplace may signal a problem that is more widespread.

Koteen and Grayson report on the various efforts that have been made to code the brief occupational statement on the individual income tax return. It seems evident that this statement can be useful in scaling taxpayers as to socioeconomic status, but, by itself, it could not be expected to support coding at the three-digit level, which is what is needed by, for example, the National Institute for Occupational Safety and Health and the National Cancer Institute. The present coding experiment, therefore, also makes use of the SSA information on industry, and looks hopefully to the Post-Enumeration Survey for the 1980 Census as a way of testing the validity of this combination.

Finally, Rosenberg and his colleagues report on the codability of occupational entries on the death certificate. Their results are encouraging; perhaps 75 percent of the certificates are reliably codable at the 3-digit level. But this is codability, not validity. It represents a situation in which such information is not being used Nationally or even being coded by many States. With use, feedback, quality-control, and demonstrated interest, one would expect considerable improvement.

Taken together, what do these papers signify? First, that there is real interest in building a National data system for monitoring mortality differentials by occupation and other economic factors. Second, that there are options that need to be explored more fully, some of which are more promising than others. As I see it, these are, depending on source of information on occupation:

1. The death certificate alone, with no denominator, and relying on proportional mortality as the method of analysis. This option is not immediately feasible, as most States do not code occupation, and a considerable amount of developmental work would have to be done. But it warrants very serious attention. I would like to see it developed for a sample of adequate size.
2. The Census of population, with matching to death certificates or to the National Death Index, in the fashion of Kitagawa and Hauser. Without the SSN on the Census form, the matching would be difficult and expensive to whatever fraction of the Census population should have the long form. If the SSN could be added to the long form, this would surely be the way to go. (The earliest, however, this option could begin would be the mid-decade Census of 1985.)
3. The CPS has perhaps the very best information on occupation, other demographic information, and some clues to health status, but it is almost certainly too small unless accumulated over several years. Furthermore, the SSN is only routinely sought for a fraction of the sample. Presumably, however, it would be easier to extend the coverage of the SSN inquiry for the CPS than for the population Census. The CPS would then have to be linked with SSA records or with the National Death Index to obtain mortality.
4. The individual income tax return combined with SSA information on industry, with survival status obtained from SSA files, and cause of death from the States. Much of the feasibility of this approach depends on the outcome of the planned test of IRS occupational coding. Should that yield good results, then a drive could be mounted to routinely collect this information in both the SOI sample of the IRS and Social Security's 1% Continuous Work History Sample (CWHS). There are, in fact, certain advantages in working with the CWHS. These include, especially, its 20-year history, and the fact that it is larger than could be accumulated from the SOI sample in 5 years or so. Of course, the 1976 Tax Reform Act probably provides an immediate barrier to the transfer of IRS data on occupation, and this would have to be surmounted before a system could be put into place.

It should be noted, however, that before any of these approaches can be implemented on a wide enough scale, one more crucial question must be answered: Is there any hope for the Administration's proposed bill (Confidentiality of Federal Statistical Records) to establish a larger number of statistical enclaves, "protected statistical centers," entitled to share individually identifiable data for valid public purposes? I have just come back from a meeting on record linkage in Ottawa hosted by Statistics Canada and was quite impressed by the ability of Statistics Canada to link different kinds of official records in support of epidemiologic studies.