

## DISCUSSION

Thomas J. Mason, National Cancer Institute

I would like to thank the authors for providing their papers in a timely manner, and will attempt to provide some comments from an epidemiologist's perspective. If we define epidemiology to be the study of the distribution of disease(s) among a population then this session plays to a central concern of epidemiology: that of the opportunistic utilization of data resources in an attempt to facilitate our understanding of the disease(s) under investigation. Epidemiologists are opportunists; and rightly so, for historically analyses of data such as those discussed today have improved our understanding of etiology, and are being used in many major preventive medicine programs.

In attempting to discuss these papers it appeared that one could dichotomize this area into attempts to facilitate counting (1) persons affected (i.e. exposed/diseased) and (2) persons at risk for the event (s) (i.e. exposure/disease). In this session we have heard about methods of acquiring these data.

As is true of any data acquisition effort there are a multitude of potential problems. They range from: missing data, non-specificity of the data for a subsequent study, biases in reporting due to any number of reasons, the cost incurred in collecting the data and the often forgotten "societal" costs of not collecting the data.

I would like to comment on the U.S.'s attempt to provide easy access to the mortality data collected by our States and maintained by our National Center for Health Statistics. Our National Death Index was a long time coming, and many of us who lobbied long and hard for its adoption are quite pleased. We've heard about the time periods, covered, the basic data elements which are maintained, and the cost of utilizing this file. We've also heard from our Canadian counterpart as well as from the National Heart Lung and Blood Institute that probabilistic matching criteria would enhance the yield for any of the potential users of the National Death Index.

Martha Smith and her colleagues are to be commended for their pioneering work in this field, and I was most pleased to see that their work with the Alberta Cancer Registry is going along so well. I must also admit to a certain amount of envy on my part for their having data which cover such a long span of time. Their concern for facilitating a successful search at minimal cost is commendable.

On our side of the border Gene Rogot has pointed to the need for the timely improvement of the methods for matching. None of us could bear the burden of a 66% rate of false positives (in terms of direct monetary costs, or costs in personnel time and the associated frustration). In some instances the matching went well. However, since we are not always blessed with

having all of these data elements available to us on historic cohorts, improvements need be made. The approach outlined by Gene and his colleagues should provide a mechanism for us to realize these improvements.

The other two papers in this session were concerned with the identification of populations at-risk. We've heard from Bill Crouse and his colleagues that the U.S. Bureau of the Census codes for Industry and Occupation can be used for mortality surveillance with our standard death certificate as a source. Quite a number of criteria for a system such as this were enumerated. Some rather optimistic results from a pre-test evaluation were quoted and the follow-up approaches would certainly appear appropriate.

We have also heard from our colleagues at the IRS. Patricia Crabbe and her colleagues are to be commended for their efforts. Although they have certain advantages over some of us with regard to having data accessible to them, their road has certainly been a rocky one. However, their findings to-date appear to support a sense of optimism for the future.

This does not mean that all is well with the world. It is not, but progress is being made in spite of many limitations. I would like to now mention some of the problems which remain after we've successfully identified a population at risk for some disease and attempted to quantify that risk.

Let's first consider the coding of occupation and industry. On the one hand, the use of the death certificate, there is the phenomenon of relatives reporting a job of higher status for the deceased than was actually performed. A systematic investigation of the source of the information might add to our understanding of the current dynamics of this practice. Using occupation data from tax returns has, as Patricia Crabbe mentioned, a related problem, in that a number of persons do not respond in a sensible way. The proportion which fall into this category is small and is not a major concern. The major limitation of any classification scheme is in its inherent lack of specificity for potential disease-causing exposures. The utility of time and cost effective descriptive epidemiologic studies is not in question as long as one recognizes that these studies are merely steps toward more analytic studies. An intermediate step which would appear to be appropriate would be linkage of job and industry classifications to the National Occupational Exposure Survey and the subsequent tests for consistency with known etiologic associations.

Our colleagues, who would provide for the timely ascertainment of vital status for persons under study, are not limited by technology for as you've heard from both Martha Smith and Gene Rogot, improvements are being made on probabilistic methods. I'm confident that these will be

incorporated into the U.S. National Death Index. The most critical piece of information is the underlying cause of death. In the U.S. and Canada we certainly have seen evidence that mortality from any number of causes of death is not distributed uniformly across large areas, but rather there are local aggregations. We interpret these distributions initially as suggesting areas where additional study would seem warranted. We have been successful in

using this approach in the United States for some causes of death. For other causes of death we found evidence of regional differences in death certification practices and determined that our efforts did not add to our understanding of etiology.

I am optimistic for the future of epidemiologic research-- thanks, in no small part to the work being done by the speakers at this session.