DISCUSSION

Daniel H. Garnick, Bureau of Economic Analysis

It is well known (1) that advancement in the empirical study of the processes of social change depends on the development and application of longitudinal microdata records, (2) that the Continuous Work History Sample (CWHS) represents a rich lode that has been increasingly mined over the past decade and a half, but (3) that the informational potential of the CWHS is constrained by problems with respect to (a) the quality of the records and (b) access to this record system and to other record systems with which it may be exactly matched or statistically associated. This session on research underway for improving and augmenting the CWHS is an impressive tour de force regarding all of the above. I am encouraged to observe further that, in spite of all the perversities afflicting the Federal statistical system, the studies reported by the participants of this session will ultimately succeed, because (1) the policy issues that can be illuminated by improving and augmenting the CWHS records are of great import, (2) the enrichment is cost effective and in harmony with the principle of reduced reporting burdens, and (3) an institutionalized learning curve is emerging through the efforts of this session's participants and others that transcends the turf of any single agency and that must prevail over the forces of obscurantism and inertia so long as rationalism retains any persuasive force at all. The methodical organization of this session, the excellence of the individual presentations, the joint cooperation and ongoing nature of the studies illustrate all of the above.

The Kilss-Scheuren-Buckler paper is an impressive overview that goes well beyond the usual technical description for matching the several files and the usual "wish list" of potential uses. It is grounded in a history of interagency joint cooperative efforts and in detailed knowledge of the potential uses, the strengths, and the limitations of the data. It meets head-on the problems of data quality and of access. It points to many research and public policy applications that an upgraded CWHS could accommodate. It reports on the Linked Administrative Statistical Sample project, which emerged as a result of interagency planning studies with respect to increasing the availability of data bases for epidemiological research, for which an augmented CWHS would be central.

I would like to point to one other potential application that would afford great illumination of sensitive regional economic and social issues: the Social Security Administration (SSA) employer file includes geographic and industry codes for county-level reporting units; but as the file has been maintained, it does not provide current payroll or employment data, and historical files have generally not been retained for more than two or three years after the file in question had been updated with additions of new reporting units and deletions of defunct units. Improvements in establishment reporting for SSA records and improved file maintenance procedures would enhance the longitudinal analytic capabilities of the SSA employer file for tracing individual business establishment location changes, thus complementing the CWHS capacity for tracing the patterns of earnings by industry and place of employment of individual workers over time. In a new employee-employer CWHS file, workers could be longitudinally traced as they change industry and geographic area of work in response to plant openings, closings, expansions, or contractions. While other data files are currently used for plant location, expansion and contraction analysis, the CWHS is the only large data source that could be used to directly analyze the dynamics of regional employment and income change and, correspondingly, workforce mobility, in terms of the longitudinal income and demographic characteristics of the workers. Moreover, the CWHS could be used to trace the individual characteristics of employers regionally over time: the age-sex-race and earnings characteristics of workers by size of employer, by industry and region of employer reporting units. Changes in the numbers of workers could be monitored over time to test hypotheses with respect to the relative growth and worker characteristics by small v. medium v. large size-class of employers. The Levine paper, at the end of this session, touches on some of these hypotheses. The linkage of an employee-employer file has great potential import.

The DelBene-Aziz paper is a straightforward manifestation of a learning curve: it carefully proceeds in establishing a records match between death reports on the SSA earnings file and State death certificates, and outlines approaches to improving matching among the record sets for the year 1975. It is a methodical continuation of the Alvey-Aziz study reported on at last year's ASA meetings.

The Aziz-Buckler paper examines the changing coverage of the several SSA files and external data from the National Center for Health Statistics over a 20-year period. It offers, perhaps, the best, most salient statement of the content of the files and the changes in coverage over time that I have ever seen. It carefully establishes a sampling of approximately 400 thousand decedents (with 15 combinations among the merged files) and describes the differential goodness of the death reporting over time by demographic characteristic. Even the charts illustrating improved death reporting in the earnings file over time appear to reflect a learning curve. (The demographic characteristics of the death reports, of course, reflect the greater coverage of white males v. other race-sex groups in the SSA files.)

The Sailer-Orcutt-Clark paper raises important questions with great clarity, in a very professional design for research, offering promising results for merging occupational information from IRS with the CWHS demographic and economic data. As an added bonus, it promises to feed back information to the Bureau of Labor Statistics for

improving the Standard Occupation Classification. Augmentation of the CWHS worker characteristics with occupation is clearly vital for longitudinal epidemiological studies. Its use in upward mobility and career ladder studies by demographic characteristic could be similarly vital.

The Levine paper also reports on a well-designed research effort. Not only does it evaluate the self-employed numbers and income changes by demographic characteristics, it demonstrates the unique capabilities of longitudinal employee-employer files for data validity checks. As an added bonus, it offers some interesting policy-illuminating findings and hypotheses concerning worker characteristics of small v. large employers. It has been established time and again that the CWHS files are central to empirical analysis of the processes of social and economic change, and, now, for epidemiological processes, as well; what a shame it would be to allow misreporting and deteriorating maintenance of the files to diminish this data base's usefulness. The Levine paper reporting on a joint cooperative effort by SSA and the Bureau of Economic Analysis for evaluating the CWHS files represents a strong, but only the first, step in needed improvement.

While working for change in the Federal statistical establishment is generally considered equivalent to submitting one self to the Chinese Water Torture, one must be encouraged by the work described in the papers presented in this session: the past development of the uses (and evaluations of the limitations) of the CWHS; the dedication of the current persons, in seeking to enrich and upgrade the files; and the efforts underway to improve access to those augmented data systems. Such an outlook is necessary: the Federal Government is too rife with burnt-out personnel whose cynicism too often drains the energy out of relatively large-scale and long- or intermediate-term statistical projects. Moreover, the (understandably) short-time horizons and attention spans of many of the appointees to whom we may directly or indirectly report, and the complexity and "nonsexiness" of most statistical projects, also, too often, impede long- or intermediate-term developmental efforts. So, while, individually, we must be responsive to brush fires, and on guard against burnt-out cynics, as well as wary of the overzealous protection of individual agency-perceived turf, we must also remain steadfast in pursuing our goals for improved statistics. Furthermore, we must renew ourselves, continually, through such step-by-step institutionalized learning curve processes as those so well illustrated by the workman-like reports in this session and other sessions centering on improvements in the statistical uses of administrative records. I recommend that those of you who have not already read the bounded volume of papers included in Statistical Uses of Administrative Records with Emphasis on Mortality and Disability Research, (SSA, October 1979), obtain copies and read them. A poignant note is sounded in the introduction to the papers on Linked Administrative Statistical Samples for Mortality Research by Iwao Moriyama:

> The importance of data linkage for mortality research has long been recognized. For example, Dr. Halbert L. Dunn, then Chief of the Federal Division of Vital Statistics, introduced the social security number into the 1940 revision of the standard death certificate to make possible the linking of death certificates to social security records. Dr. Dunn had a great interest in the general problem of record linkage, but he was regarded as something of a visionary and his interest was not shared broadly by his peers. A meeting such as this one would have been a source of great gratification to him.

A meeting such as this one would be a source of gratification to many others early in the game of conceiving and/or implementing studies utilizing longitudinal microdata sets for public policy analysis. It was foreseen and it remains in sight.