Gilbert Nestel Center for Human Resource Research The Ohio State University

A. Introduction

I must confess that at the time I agreed to accept Dr. Sagi's invitation to prepare a paper which would dwell on selected problems in the analysis of longitudinal data, I could not help but to think that this assignment was not unlike the one presented on the TV show, <u>Mission</u> <u>Impossible</u>. Recall for the moment the format of that show: an impossible task is described; the urgency of a solution is stressed; the relevant background data are given; the sponsoring agency disassociates itself in case of a failure; and the agent always accepts his assignment, however "impossible."

The remainder of this one-hour show, for those of you in the audience who may not have seen it, consists of a series of episodes, a climax, and a final resolution of the problem. The entire program is staged with such skill. ingenuity, and insight that the viewer is often dazzled at the ability of the writers and actors to describe an urgent contemporary problem and to provide a logical, minutely detailed, albeit fantasized solution. Despite its success as entertainment the program fails as a description of human behavior simply because it does not treat in a realistic way two of the more important dimensions in all decision making: namely, the role of information and the extent to which the individual can interpret interrelated and often conflicting data. The program is staged as if the participants do not have to weigh alternatives at critical crossroads. The various forks in the road are chosen inexorably. Questions of expected costs and future benefits of different actions are ignored. Uncertainties are rarely admitted. The story develops in much the same way as a game of chess might if one of the players knew in advance all of his opponent's moves and was also sufficiently skilled to take advantage of this information. Even the most intensive analysis of such a game by the most skilled of analysts would provide at best only limited insight to the more typical games of chess. Similarly, a Mission Impossible script is hardly a prototype description of decision-making under uncertainty.

The parallel scenario describing my research experiences with panel data would overlap the script of <u>Mission Impossible</u> but it would also deviate from it in several substantive ways. Both the agent in the TV show and I initially face very difficult assignments. In my case, however, there are only scattered references in the social science literature where authors have addressed themselves to questions which are directly relevant to the analysis of panel data. Moreover, I do not face an identifiable single protagonist, and thus the notion of a confrontation has no empirical content. In addition, the sponsoring agency's considerable monetary investment insures that this group will not dissociate itself from the project.

There is one additional difference which distinguishes the two scripts. Whereas the agent in <u>Mission</u> <u>Impossible</u> is always successful in completing his assignment, I unfortunately do not have a comparable script writer to assure me of this success.

What I will attempt to do in this paper is to paint some of the broad contours of the subject and to illustrate its terrain with data from an existing longitudinal study. An accurate painting of this canvas and the broad dissemination of the experiences of researchers working actively in this area is a necessary condition, I submit, for subsequent analytical progress in this field.

The data referred to in the illustrations were obtained from four National Longitudinal Surveys of the civilian mninstitutional population in the United States. The Bureau of the Census administered these surveys under a contract with the Department of Labor for The Center for Human Resource Research at The Ohio State University. A sample of approximately 5,000 people was selected by The Bureau from each of the following age-sex universes: young boys and young girls between the ages of 14-24; mature women 30-44; and older men who were between the ages of 45-59 when first interviewed. These surveys were designed so that nonwhites were overrepresented and comprised about 30 percent of the total sample. Each of the cohort groups are interviewed for five consecutive years after the initial interview. The specific groups that were chosen were done so because of the interests of the staffs of the Department of Labor and The Center for Human Resource Research in questions such as the following: (1) What are the occupational and educational aspirations of young boys and girls as they filter through the education system? Do these aspirations change in any systematic way over time and, if so, in what directions? To what extent are these aspirations realized and are there significant differences among the races? (2) To what extent is a woman's investment in her education and early job experiences fully depreciated when she decides some years later to return to the labor force? Does the depreciation of her human capital help to explain her subsequent experiences in the labor market? (3) Are the retirement expectations of the older men consistent with their actual behavior and to what extent are these expectations stable from one year to the next? What are the determinants of these expectations? The content of these questions and others that were identified strongly dictated that a prospective longitudinal design be used to collect the relevant data.

The remainder of this paper is divided into two sections. In the first, I discuss the major differences between a longitudinal and a crosssection design and evaluate some of the advantages of the prospective longitudinal approach. The second section is restricted to a discussion of some of the problems that are typically present in a longitudinal study. Both sections provide an accounting framework within which this subject and some of its dimensions can be discussed.

B. Definition and Advantages

The unique properties of a longitudinal design help to differentiate it from still another survey instrument -- the cross-section sample. In the latter design the sample members are interviewed once and their responses provide a basis to characterize the universe or to extablish behavioral relationships among various variables at a single point in time. Any intertemporal changes in behavior can only be inferred from these data and only then on the basis of average relationships under ceteris paribus conditions. The selection of a prospective longitudinal design presupposes that the researcher is primarily interested in the intertemporal changes in behavior and that he has some theoretical reasons to believe that this behavior has a temporal ordering which can only be established or tested with data from a replicated experiment. The availability of repeated observations does not distinguish the two designs, however. The composition of the samples in the two designs is of far greater significance. In a prospective longitudinal survey the individuals who are selected at the time of the first sample comprise the sample of observations in all subsequent surveys. This control of the selection process is the essential characteristic of all longitudinal designs, and it provides a significant contrast with cross-section studies even when replicated.

It is also worth mentioning that a longitudinal survey does not necessarily require that the data be collected by repeated interviews. Individuals in the sample could be asked to answer questions about their past as well as current behavior. The major reservation about this procedure remains whether these people can accurately recall their past behavior particularly when asked about their plans, attitudes, or values. To the extent that they are unable to do this, errors of measurement are introduced which can seriously bias the statistical findings. These errors have very serious implications particularly in a study where the intent of the analysis is to measure change and to study its determinants. We shall return to this matter shortly.

The data from the young girls and women surveys provide an illustration of how longitudinal data from two different cohort groups can be used to establish possible causal paths. Consider the question of whether a woman's attitude about the propriety of employment of mothers with young children is related to her own labor force participation. The cross-section data from the women's survey show that the more favorable a woman's attitude about the world of work the greater the likelihood that she will be a member of the labor force. This association is not sufficient, however, to establish whether it is the attitude that determines the behavior or whether the behavior has conditioned the attitude.

The periodic monitoring of these attitudes and the subsequent behavior of the young girls over time could very well shed some light on this question. If it could be established that the girls who have the most positive attitudes toward working mothers prior to their marriage and entry into the labor force turn out also to have a very high participation rate after marriage and the birth of their children then the direction of the relation would be established. Unfortunately, this hypothesis cannot be tested at this time since we have not yet received the requisite data. Nevertheless we intend ultimately to test this hypothesis for its empirical content.

The data from a longitudinal study can be used for purposes other than to record and measure changes in behavior, or to assess the adequacy of alternative causal models. The measurements over time on selected characteristics of the cohort members provide one basis to study the distributional properties of these variables, and to identify the extent of the possible measurement errors whatever their origin. In a cross-section analysis a response is considered an outlier if the likelihood of the variable assuming this value is very small. However, the identification of the critical regions of the distribution is somewhat arbitrary at best. In any event the decision to include or exclude an observation is typically made without the aid of any additional measurements of the respondent's behavior in other time periods.

One is in much better position in a longitudinal study to evaluate the legitimacy of the various measurements. Thus if a variable takes on a value which deviates considerably from past behavior and if the magnitude of the response cannot be explained by changes in factors that are known or assumed to be associated with this variable, then these are two strong conditions for questioning the accuracy of the response. Thus the repetitive measurements on the same respondent that is typical of a longitudinal design also provide a mechanism for identifying extreme values and assessing their probable legitimacy.

The unique features of the longitudinal design are also of help when the interviewee fails to provide an answer to a specific question.

In a cross-section analysis where one does not want to exclude an observation for which there is not complete information, what is typically done is to stratify the universe according to several dimensions (thereby attempting to control for the factors known to be related to the variable where the nonresponse occurs) and then to impute a value for the missing datum that is computed from the responses provided. The quantity that is typically imputed is the mean or median value. With longitudinal data, however, one may not have to resort to a measure that is estimated from the responses of individuals who are assumed to share common characteristics. To the extent that the respondent has provided earlier or later measures for this variable then one can impute a mean or median value that is computed from his responses. Even if this is not possible, to the extent that there is a close and stable relation between the omitted variable and other factors for which we have complete information, an average relation among these variables could be estimated -- say, by the method of least squares -- and the reported characteristics of the individual then used to predict either the mean value of the omitted variable or the value itself.

The availability of longitudinal data also makes it possible to apply a more stringent test of whether a hypothesized behavioral relation is stable over time. Once again the superiority of this design can be traced to its unique properties. The repeated measurements on the same respondent mean that the data do not contain a source of variation that is introduced whenever a new sample is selected. This variation is present, however, in the case of the replicated cross-section sample, and it may well explain some portion of the observed intertemporal variation in the parameter estimates. In any event the magnitude of this additional variation needs to be assessed and tested for its statistical significance.

This concludes the discussion of some of the advantages of the longitudinal survey and how this design differs from a cross-section design. I turn next to a discussion of a selected number of problems.

C. Some Problem Areas

1. Panel Mortality

Perhaps the single most serious limitation of the longitudinal design is the loss of sample cases (panel members) over time, or what is referred to as the attrition or mortality problem. These noninterviews are a legitimate concern to any researcher particularly when the losses are large, or where there is some evidence that the characteristics of the members who disappear from the sample are significantly different from those who remain as cohort members.

Whether the loss of observations seriously biases the statistical analysis is a more subtle question and one that cannot be answered definitively once and for all. The longitudinal design does make it possible to use the data generated in one of the years, generally the first, to study the characteristics of the noninterviews in a subsequent survey. Although this comparison becomes a less useful test as the time interval between surveys increases it may provide suggestive evidence when used properly to allay some of the more lingering 'doubts. Our experience thus far in retaining the cohort members in the four age-sex groups have been very encouraging. The first surley of the men began in 1966 when 5,034 individials were questioned. by 1969 there were still 4,400 men, or 87.5 percent of the original group who were available to be interviewed for the fourth consecutive year. A study of the reasons for the loss of approximately 14 percent of the panel members indicated that 5 percent had either died or were institutionalized and the remaining 8 percent consisted of men who disappeared between surveys and could not be located.

The corresponding sample mortality rates for the young boys is not nearly as heartening, but part of the difference can be explained by the number of youths who entered the armed forces between survey dates. Interviews of the boys began in 1966 with 5,234 youths. Three years later (in 1969) the cohort consisted of 4,017 members or 76.7 percent of the original sample. Deaths and institutionalization accounted for less than 2 percent of these losses. Slightly more than one-half of the losses (13 percent) were to the military services. The residual classification--or, the losses that are considered to be "pure" attrition--comprised approximately 10 percent of the total sample.

We are not too surprised to find a greater attrition rate among the boys than among the men. Over time one would expect that the youths would show a greater degree of geographical mobility. Thus to the extent that these movements are positively associated with the number of disappearances, the greater the mobility potential the higher the likelihood of attrition.

At the time of this writing we have had three years of experience in interviewing the mature women and two years of experience with surveying the young girls. The initial survey of the women took place in 1967 when 5,083 women were interviewed. By 1969 the original cohort had been reduced to 4,547 members, or to 87.5 percent of its original size. Death or institutionalization of a panel member claimed less than 10 percent of these losses (or 1 percent of the total sample). Disappearances accounted for more than 90 percent of the total attrition, or 10 percent of the total sample. It might also be noted that the second interview of women was by a mail questionnaire and therefore the personal interviews were two years apart.

The final cohort group-the young girls--was first surveyed in 1968 when 5,197 female youths were interviewed. In the following year 4,971 members of the original panel were interviewed. Thus the Bureau of Census interviewers were able to locate and interview better than 95 percent of the original sample of females. The lack of any detail description of the sources of the noninterviews at this time prevent us from analyzing the reasons for these losses.

It needs to be emphasized that the low attrition rates in the different cohort groups may be explained by the comprehensive procedures that the Bureau of the Census have adopted and followed in this study. Thus our experiences may not be representative of other logitudinal surveys.

It becomes increasingly more difficult with the passage of time to assume that the increase in the magnitude of the noninterviews has no effect on the findings of the statistical analysis. We have earlier documented that the size of the noninterviews in each of the age-sex cohort groups is not very large. Even though this in itself is very encouraging it is not evidence that the characteristics of the noninterviews are similar to those of the remaining members of the panel and that therefore no biases are introduced.

The 1970 Consuses of Population and Housing will provide an additional opportunity to evaluate implications of the noninterviews in the four age-sex cohort groups for selected characteristics. Admittedly the test will not be a perfect one since there will be the inevitable differences in the wording of the questions, who reports the information, the reference periods referred to, as well as when the interview takes place and how the variable is measured. Nevertheless this analysis will at least provide one benchmark for a potentially more intensive study of this problem area.

2. Measurement Errors

Response accuracy is a second aspect of the data analysis that has to be evaluated. Errors in measurement are not an inherent characteristic of a longitudinal design. However, if a panel study can be thought of as a series of crosssection surveys where the composition of the sample has been restricted, then the likelihood that a specific characteristic of the individual will be reported inaccurately at least once may be greater under this design than in a replicated cross-section survey where the composition of the sample changes over time. The presence of these errors thus depends on the length of the survey period, but it is also conditioned by the types of questions asked, the period of recall, and the evolving attitude of the respondent and his relationship to the person doing the interviewing.

The significance of these errors for the statistical analysis is also not <u>a priori</u> determined but is related to the focus of the analysis and to the distributional properties of the different error terms. For example, if the purpose of the cross-section analysis is to estimate a universe parameter (e.g., a mean value or total) and the aim of the longitudinal analysis is to study the intertemporal stability of these parameters, then the errors in measurement have one kind of implication. The same two sets of data could also be used to estimate a behavioral relation in which case any variables reported with error would have a different implication.

I can report in this context two of our many experiences which clearly highlight the potential seriousness of this problem.

A case in point has to do with the extent of intrafirm occupation mobility among blacks and whites between the first and second surveys of the men. The universe is restricted in this discussion to men who were employed in both years and a change in occupation is measured by the difference in their 3-digit occupation codes in the two time periods. The data as orginially tabulated by the Bureau of the Census show that 18.4 percent of the whites and 19.7 percent of the blacks appear to have made an occupational change. When, however, the sample is further restricted to include only men who provided a reason for their change we find that only 3.3 percent of the whites and 4.2 percent of the blacks who were eligible to change actually showed a change in their occupations.

Variation in rates of pay between 1966 and 1967 for the cohort of men 45-59 provide a second illustration of how response errors can affect the statistical findings. Between the first two years, 21 percent of the blacks and 23 percent of the whites reported a decrease of 10¢ per hour or more in their rate of pay. Even when the universe is restricted to men who have the same job status in each year the percentage of whites and blacks whose wage rate declined by 10 or more cents per hour did not change significantly. Thirty-one percent of the whites and 20 percent of the blacks reported a decline of at least 10ϕ per hour between the two years. Moreover, the decline was 30¢ per hour or more for 14 percent of the whites and 11 percent of the blacks.

The sharp decline in reported rates of pay cannot be explained by any identifiable changes in the characteristics of the respondents or by changes in their environment. We chose therefore to select a sample of these men and study how the wage rate variable was calculated in each of the two years. Forty-nine men were selected at random or 10 percent of the total number of sample cases and the rate of pay associated with those individuals identified. An hourly rate of pay was reported in both years by only two respondents. In another three cases a coding error or misplaced punch was found. And of the 46 cases where no errors in punching or coding was present slightly more than one-half of the respondents reported their earnings in different time periods. In addition, one fourth of the men also reported a different number of hours worked in the two periods.

The unit of measurement differences, or intertemporal metric, has obvious implications for the computation of the wage rate and may help to explain some of the observed intertemporal variations in these rates. In the cases where the respondent fails to provide an hourly rate this rate was estimated by dividing his reported earnings on his current job in whatever time unit he chose to express this amount by an estimate of the total number of hours worked in the same time period. Thus the computed wage rate can be in error if the numerator, denominator, or both of these measures are reported inaccurately. Before closing this discussion I might mention that the selection of the four age-sex cohort samples also makes it possible to study some aspects of the errors-in-variable question.

The Bureau of the Census has had to exercise considerable discretion in designing this study or run the risk of incurring prohibitive costs. In effect what they have done to contain these costs is to allow certain households to be represented in more than one cohort group. In the survey of the older men, for example, one out of every three households has at least one other member of the household represented in one of the four age-sex cohort samples. In approximately two out of every three households in the sample of women there is at least a second household member represented in one of the cohort groups. Only one out of every four households in the young boys and young girls samples can be classified as a single respondent household; the other households all share the common characteristic of having at least two household members represented. Finally, approximately 3 percent of the households in each cohort group were of sufficient size to have at least one member represented in every cohort group.

The overlapping of households (this may well explain the low attrition rates we observe over time) provides one way to study the frequency and magnitudes of the response errors for selected characteristics of the respondents. For example, these selected households can be studied to examine whether the family's total income in a calendar year or its income by source as reported by the wife in the women's survey is consistent with the responses reported by her husband in the men's survey. And to the extent that the household is represented in the surveys of young men and young women one can study the extent to which these members have accurate knowledge of the family's total income. The same kind of approach can also be followed in studying whether the number of weeks worked in a calendar year by the male head of the household, or the occupation he held for the longest period within the year are known by other family members in the household. The longitudinal nature of these studies also makes it possible to study these questions across a time domain as well as at a single point in time.

The research design that I have just described has its limitations, as one would expect. The different timing of the four interviews within the year means that only certain types of questions are amenable to a comparative analysis. Moreover, the differences in the content of the questions asked (particularly the time reference) also limit the kinds of comparisons which are possible. Finally, the distribution of response errors for different household members and their characteristics are likely to also vary over the lifetime of the longitudinal survey.

One suspects that the variations in responses among household members will be larger in the first year of the interviews but that these differences will be reduced in subsequent reinterviews. The repeated questioning of the same respondents could condition them to answer more consistently over time than they would have done in a single interview. In addition, the ease with which individual members of the same household can exchange information also suggests that they would be more apt in a later interview than in the first survey to give more consistent and accurate responses. Both of these factors point to smaller response errors with the passage of time. On the other hand, the respondents may lose interest in this survey and become annoyed at having to respond periodically to very similar types of questions. This may well lead to indifferent answers that are largely inaccurate.

3. Nonresponses

A third major problem area, and one that is again not unique to longitudinal studies has to do with the frequency of the nonresponses and the associated characteristics of the individuals who fail to provide answers to specific questions. As I suggested earlier what is typically done to neutralize this problem is to stratify the universe by various socioeconomic and demographic dimensions that are known to be associated with the characteristic that is to be imputed, and then to assign a value to the nonresponse that is based on the reported information of other people in the same stratum. The Bureau of the Census is currently following this policy in its March CPS survey which it uses to estimate the incomes of husband-wife families in the civilian noninstitutional population of the United States.

The Bureau has found that very often respondents fail to report all sources of their family's income. Rather than eliminating these families from the sample and readjusting the sample weights of the respondents who report complete information the Bureau has evolved a procedure which imputes a value to each source of income that is not reported. In practive the procedure followed is slightly different for the earnings component of income than for the "other sources." In the earnings imputation mutually exclusive classes are formed that are based on the sex, age, color, weeks worked last year, occupation, and class of worker of the respondent. The same set of controls and the earnings and employment status of the respondent are used to impute values for other sources of income. Each of the sample cases is then sequentially assigned to one of the available classes and when a member is found who does not provide information on a source of income he is assigned the value reported by the last person previously added to that class. The process continues until all assignments and imputations have been completed.

The same kind of reservations discussed earlier in the context of measurement errors holds here too. I suspect that the problem is further compounded if the variable is usually reported inaccurately. It might be true, for example, that there is a high degree of positive association between the rate of refusal and the likelihood of a measurement error. In these cases the imputated values may be in error for two reasons. First, the controls by which the various mutually exclusive classes are formed may not be sufficiently correlated with the characteristic that is being measured. But second, even if one can adequately control for this variation, a further error is introduced because of the inaccurate responses of those in the stratum who reply. In summary, the consumer of longitudinal data is faced with the dilemma of having to live with a sample which had been substantially reduced because of refusals or having to impute values for these nonresponses which may be seriously inaccurate. In either case a bias may be introduced into the analysis but the relative advantages and disadvantages of the two alternatives remain still to be evaluated.

The need for such an evaluation is very apparent from the results of the first round interviews of the various age-sex cohort groups. In the men's survey, for example, one out of every three whites and almost two out of every five blacks failed to provide complete information on his family's resources (net assets). The same set of questions was also asked in women's survey and here the responses of three out of every ten white females and one out of every five blacks were incomplete. In the young boys' survey the questions related to the net assets of a family were asked only of respondents who were heads of households. However, approximately one out of every five white male youths and slightly more than one out of every ten blacks who were not enrolled in school at the time of the interview and who also aspired to an occupation at age 30 which was different from their current occupation failed to provide a reason why they believed they will not attain their goal. These examples, and others which could be introduced, fully illustrate the potential range of the refusals.

4. Metric

A fourth problem area that has to at least be mentioned concerns the metric that is assigned to the changes in behavior that one observes intertemporally. Here again the longitudinal survey is not the only design where this question arises. Nevertheless its emphasis on measuring the same person's characteristics at several points in time means that the potential paths that this behavior may take are much more numerous.

The responses to the expected age at retirement (ERA) question is a case in point. In the first interview a panel member could have reported a specific age; he could have answered that he was already retired; he could have responded that he never expects to retire; he could have claimed that he did not know when he was retiring; or he simply could have refused to answer the question.

The same options were also available when he was reinterviewed in the second year. Thus with five alternatives in each of two years a total of 25 paths can be uniquely identified. These paths can be most easily summarized in a 5x5 matrix where the rows represent the various 1966 responses and the columns the responses in 1967. The cell frequencies when transformed to a relative base can be viewed as estimates of the universe transition probabilities, particularly if the cell frequencies are not very small.

Fortunately, the 25 transition probabilities are not of equal interest. Nevertheless, a metric may have to be assigned to one or more of the paths if the extent of changes in the expected retirement ages is the focus of the analysis. In studying the retirement decision how does one assign a measure to a man who provides an age in one year but when reinterviewed reports that he will never retire? Or how does one classify the behavior of a man who changes his expectation from a never retire response to an already retired?

5. Other Selected Problems

There are additional problem areas that are peculiar to a longitudinal design which should also be discussed. I will forego this temptation in the interest of keeping what is already a long paper within manageable proportions and simply provide a listing with only minimal comments. First, the intertemporal influence of the interviewer on the reported responses of the interviewee needs to be evaluated. Second, the extent to which the respondent recalls and reports an earlier response even though his circumstances or his environment has not changed also needs to be studied. Third, there is a need to study the extent to which the interviewing process itself conditions the change in behavior. Fourth, the use of a longitudinal design requires that the questions asked be comparable over time, that there be consistent coding of the responses, and that when subjective rules are adopted (e.g., in coding open-ended questions) that these rules be followed consistently. Finally, the prospective longitudinal design loses some of its significance if the environment which conditions the responses does not show considerable variation over the lifetime of the interviews.

D. Summary

My occasional references to the retirement decision of the older men was suggested to me by the strategy of the writers of the TV and movie previews--provide a setting and backdrop so that the viewer has an incentive to return at a later date. I had hoped to report on some of these findings at this session but I have already exceeded my allotted time and besides that study is another <u>Mission Impossible</u>. It might be mentioned in passing, however, that the problems discussed in this paper are general in content and carry over to that study with only minor changes.

ί,