RETROSPECTIVE LONGITUDINAL RESEARCH:
METHODOLOGICAL CONSIDERATIONS*

David L. Featherman

CDE Working Paper 79-19

*Work on this topic was made possible by the Agricultural Experiment Station of the University of Wisconsin-Madison, by the privileges of the author's Fellowship at the Center for Advanced Study in the Behavioral Sciences, Stanford, California during 1978-79, and by financial support from the Andrew W. Mellon Foundation and the National Institute of Mental Health (Grant No. 32 MH14581-03).

(Forthcoming in Journal of Economics and Business, May 1979)

April 1979
INTRODUCTION

Longitudinal research designs in the social and behavioral sciences are frequently cited as ideal frameworks within which the complex dynamic processes of human development and behavior are best studied. Cross-sectional or correlational designs are faulted for their inherently static approach. Yet longitudinal studies are costly scientific enterprises -- both in terms of their requisite funding and logistical support and in terms of man-years consumed in bringing panel studies to full term. It is perhaps these more practical limitations of longitudinal research designs that prompt both practicing scientists and research administrators to continually question the comparative utilities and "cost-benefits" of alternative designs. The question might be posed as follows: "When are longitudinal studies essential to derive descriptive and analytic data; conversely, when can cross-sectional designs yield trustworthy longitudinal data for the study of dynamic social and behavioral processes?" This paper addresses itself to that question.

The reader will notice that the question is phrased to distinguish between longitudinal designs and longitudinal data. The former are instanced by panel studies and by repeated cross-sectional studies on members of a common cohort (but the same subjects/respondents do not necessarily appear in each successive wave, as contrasted with the panel design). In these cases, a research program often extends over several years and entails repeated contacts with the same or statistically equivalent sample elements. Longitudinal data, on the other hand, need
not necessarily be generated solely from over-time research designs. Retrospective reports and prospectively focused life aspirations elicited from cross-sectional (one-wave) studies both provide dynamic data for the study of stability and change in human behavior.

This paper considers the use of retrospective reports, gleaned from one-time or cross-sectional research designs, as an alternative to the very costly, inefficient, and often statistically flawed data from panel designs for longitudinal research. It searches for an answer to the question of how trustworthy retrospective data are as substitutes for direct observations on the past (e.g., concurrent respondent reports in longitudinal panels, independent records). In suggesting a provisional answer, the paper reviews a sample (albeit an accidental one) of empirical assessments from a variety of disciplines about retrospective information across several life domains -- health, employment, migration, fertility, as examples. Fundamentally, the consideration of alternative methods reduces to an assessment of comparative validity and reliability. That is, can it be shown that retrospective reports provide valid representations of the past and that concurrent reports are reliable statements about the present? Further, do response errors about the past compare favorably with response errors about the present? That is, within the limits set by the reliability of concurrent reports, does the validity of retrospective reports suggest that such data are trustworthy materials for longitudinal research?

LONGITUDINAL DATA AND LONGITUDINAL RESEARCH: WHEN WON'T CROSS-SECTIONAL DESIGNS SUFFICE?

The substantive impetus to the collection of longitudinal data is the
1) **Change in interindividual differences in intraindividual differences.**

Let us assume that we are interested in relationship of fluid to crystallized intelligence within and between persons over time. Further, we assume that some test of each dimension of intelligence is available such that IQ\(_F\) is an individual's score for the fluid component and that IQ\(_C\) is the score for the crystallized component. In this first developmental study, we are concerned with change in the correlation of IQ\(_F\) and IQ\(_C\), calculated over the bivariate arrays of individuals' scores at \(t_1, t_2, \ldots t_n\). Such a comparison is signified in Fig. 2 as row "a" from the front face of the data cube as it passes as a plane through the "occasions" dimension. In the example of intelligence, the first developmental study seeks to understand the developmental course of the simultaneity of fluid and crystallized intelligence; it allows for individual differences in the pattern of such development to be captured by the correlation coefficient (or some other measure of statistical association as estimated over individuals).

This developmental study does not require a panel design. If the analyst is able to draw independent samples from the same population (e.g., birth cohort or school grade cohort at \(t_1\)) at each occasion, then the correlation \(r_{IQ_F, IQ_C}\) can be computed at each \(t_i\) independently of the computation and sample elements at \(t_{i-1}\). The condition of a constant or closed population across the interval of interest is apparent, and it would not be met were there selective losses or gains in the population vis-à-vis the distribution of intelligence. However, this condition is not the same as saying that no additions or losses to the universe can take place; those having no effect on the distribution of intelligence
(specifically, IQ₇ and IQ₉) are of no account.

2) Change in intervariable differences in interindividual differences. The second developmental study compares the distributions (e.g., variances) of individual differences in measures of two or more traits (e.g., IQₓ and IQᵧ) at two or more occasions. Thus, one might note that at age 6 the population variance in IQᵧ was nearly identical to the variance in IQₓ, but by age 18 the variance of the latter was twice that of the former. As in the first developmental study, this type of longitudinal research does not require a panel design if the conditions on a closed population hold and equivalent samples from a cohort of interest can be drawn at each measurement occasion. The second developmental study is signified in Fig. 2 by the plane of "b" in the data cube.

3) Interindividual differences in interoccasion differences (change) in intraindividual differences. This third study is signified by plane "c" in Fig. 2. It examines interpersonal differences in the stabilities and changes of intrapersonal traits or behaviors across various temporal segments of the life span. Specifically, the analyst would measure an array of variables for each individual at each occasion. In terms of the example of intelligence assessment used here, these variables would be IQₓ and IQᵧ, or subscales indexing these dimensions of intelligence. In fact, the variables in this study are really the occasions of measurement -- t₁, t₂,...tₙ -- and the IQ scores are the cases or observations over which correlations are computed. Thus, for each individual the analyst could compute correlations between t₁ and t₂, t₂ and t₃ and so on which express the extent of intertemporal consistency of intraindividual differences (in
research has been to lend some concreteness to the contention that longitudinal studies can be conducted with alternative research designs. In this and subsequent sections I wish to examine some of the positive and negative attributes of several designs appropriate to longitudinal research, beginning with prospective or panel studies.

Prospective designs seem ideally suited to developmental research, whether they are conceived as essentially psychological studies (focused on the developmental course of intraindividual differences) or sociological studies (focused on interindividual variation in intraindividual differences). These prototypical prospective investigations begin with a carefully drawn sample of some well-defined population and follow it faithfully so as to minimize the impact of selective case loss. The fundamental criterion which dictates the use of a panel design is the necessity of repeated measurements on the same sample elements. This criterion has two related implications: 1) that careful attention will be given to replication of measurements (same instrumentation); 2) that measurements at "earlier" panels \( t_{1-k} \) be uniquely suited to that developmental state (that is, no substitutions of data collected in a later panel for information appropriate to an earlier panel are possible). Let me elaborate these two points in reverse order.

The sine qua non of prospective research is the contention that the developmental course of some behavior or attribute is rooted in (causally) antecedent phenomena that can only be measured concurrently -- at the time the antecedent phenomena are active or present. Fluid and crystallized intelligence at age 6, mastery behavior at age 10, parent-child patterns
serious practical and methodological limitations.

The major drawback to prospective longitudinal research appears to be a practical one -- the substantial cost in professional man-years and dollars that are required to follow-up panels of respondents, sometimes for decades. This practical consideration is compounded when the objectives of research require large samples, when the original sampling frame is improperly designed (e.g., such that temporary exits from the "at-risk" population selectively eliminate cases from the longitudinal phases of the research), and when the prospective nature of panel designs is compromised. Thus, the practical consideration is closely linked to the methods and, as I observe, the "natural history" of many prospective projects.

It is not uncommon at the outset of prospective developmental studies that the behavioral and personality outcomes of interest cannot be confidently forecasted. Neither is it usually possible to specify all the variables and processes that might be implicated as "causal" elements of behavioral and personality development throughout the often long period of the life cycle that is to be investigated. Three corollaries of this observation are 1) that large and heterogeneous samples must comprise the panels, 2) that the substantive foci at each successive panel change as the analyst's understandings and hypotheses are sharpened (and as the turnover in personnel and in the "state of the art" of social science research dictate new variables and new methods), and 3) that the cumulative longitudinal data file grows exponentially in relation to the much smaller file of true panel information (viz., that based on replicate measures across the panels). Thus, there is a tendency for prospective projects
to devolve into a series of repeated cross-sections.

I shall illustrate these points with observations about Project TALENT, the National Longitudinal Survey of Young Men aged 14-24 in 1967 (NLS), and the Income Dynamics Panel (DYNAMICS).

One of the primary purposes behind the initiation of Project TALENT was a desire to ascertain the processes by which American youth of varying interests and abilities were sorted into jobs and occupational careers. A goal in that connection has been the construction of the "talent" profiles and sociological "life histories" of jobs and career lines. That is, using the panel data for eventual incumbents of occupations, TALENT staff has aspired to construct a careers manual that suggests which unique kinds of personal and social "requisites" are associated with quite specific and detailed lines of work such as brain surgeons, corporate lawyers, nuclear physicists, lathe operators, insurance salesmen, janitors, and homemakers. Of course, TALENT could have taken a retrospective view of the life courses of present incumbents -- sampling within specific occupations -- along the lines of such vocational researchers as Strong, Holland, and their students. That research strategy would have assumed that the interests and abilities of adults are good proxies for those of adolescents, if not as stable indexes of age-specific mean levels of these traits then as acceptable indicators of the (stable) relative rank of individuals in their cohort at some earlier age. Instead, TALENT staff designed a prospective study with talent inventories in the initial wave.

It is a shortcoming of the overall design of the TALENT research program that the ability and interest inventories were not administered
subsequently, for in failing to do so it has constrained analysts to assume the stability of "talent" levels or ranks over successive waves. A decade or so ago psychologists who espoused a trait approach to personality might have been comfortable with such a decision. Today, the personality is viewed as substantially more plastic and situationally molded, the demonstration of which requires repeated measurements and the modeling of feedback processes between "talent" and "experience." Ironically, the absence of repeated measurements of the adolescent inventories renders the justification for the prospective versus retrospective design choice substantially less defensible after the fact. Of course, any future panels of TALENT could repair this shortcoming and, in my view, should do so. (The only retesting of abilities and interests involved 118 ninth graders as they matriculated into grade 12. The retesting was conducted on a modest scale using abridged -- nonreplicate -- instruments.)

The decision to initiate a prospective longitudinal study of eventual incumbents of detailed occupations and careers entailed the sampling of a very large number of American youth -- 400,000 high school students in 1960 -- so that enough observations at \( t_n \) would be available in each occupation to permit analysis. At the same time, the enormity of the undertaking has been telling, specifically in the achieved response rate across successive panels. Perhaps the tyranny of size is best seen by comparing TALENT's response rate with that of YOUTH IN TRANSITION (Bachman, O'Malley, and Johnson, 1978) -- a national panel of 2200 high school boys in 1966, some 74 percent of whom still were active members of the study in 1974. The losses in the TALENT panel have been staggering -- as high as
80 percent -- necessitating elaborate and ingenious reweighting of the remaining data to approximate the base sample (Wise, 1978).

It is doubtful that TALENT staff has been any less vigorous in pursuing their panels than have their colleagues at Michigan. The frequency of initiation of new waves -- virtually every year -- across the four grade cohorts, the task of maintaining an active address listing, and the sheer volume of fieldwork and of information processing associated with the hundreds of thousands of sample cases necessarily reduced the efficiency of the TALENT study. Something of a substantiation of this supposition about the effect of size appears in the rather high order of success achieved in following up modestly-sized samples of 2500 nonrespondents in the 11-year panel -- about 76 percent, or roughly equal to the YOUTH IN TRANSITION experience (Wise, McLaughlin, and Steel, 1977:16).

The substantive efficiency and "cost-effectiveness" of prospective studies can be compromised by more than the ambitions of sample size and the rapidity of successive panels. The NLS (National Longitudinal Survey of Young Men aged 14-24 in 1966) provides an illustration. This is an annual prospective survey of originally 5000 young men designed to enlighten the relationships among schooling, labor force entry, employment through mid-life, and earnings histories; it was specifically intended to inform about the complex life-stage transition from school to work. The major shortcoming of the NLS study stems from its sampling frame which systematically excluded youth "at risk" to making the transition from school to work at t_{1+k} on the basis of their life circumstances at t_{1}. The failing illustrates the long-run vulnerability of prospective studies to errors
of design (as distinct from those of execution, however unavoidable, as in the case of TALENT).

The NLS was targeted for the civilian noninstitutional population of young men aged 14-24 in 1966. By definition it excluded students residing in institutional quarters or youth in the military in 1966. Once eliminated by their nonuniverse status from the initial panel, these noncivilian or student youth were not part of the prospective chronicle of occupational choice and labor force experience that the NLS data were intended to collect for an entire birth cohort. The whole impact of this population restriction is hard to estimate, but one example is informative. In 1966 roughly 5 million males were on active military duty -- up slightly from the post-Korean lows of 1960 but not as high as the peak of the Vietnam period (3.5 million in 1970). If one assumes that the most vulnerable ages for military service in 1966 were ages 17 to 20, then one can say that the birth cohorts of 1946-49 were at the greatest risk of being incompletely sampled by the NLS frame. Fully 46, 55, 51, and 42 percent of men in these birth cohorts ever served in the military. The median age by which men in all these cohorts would have entered the service was 18, although 25 percent prior to their eighteenth birthdays. Average duration of military service for each of these cohorts was over 4.5 years.

By inference, a large segment of the NLS birth cohort aged 17 to 20 in 1966 was and is missing from the prospective panel because of a unique historical (period) event and an ill-conceived sampling design. The impact on long-range analyses of transitions from school into first jobs -- as part of some general understanding of the life-course transition into adulthood for an entire male cohort--is significant. For instance, from retrospective questions in the 1973
Occupational Changes in a Generation Survey (Featherman and Hauser, 1978) one can estimate that the median ages of first full-time civilian jobs of the cohort born between 1946-49 ranged between 19.5 and 20.0 (Winsborough, 1975: Table 2); fewer than 25 percent of these cohorts took their first jobs prior to the average age of military induction. Thus, a large segment of the first-job taking process (Ornstein, 1976) is totally missing from the prospectively derived "history" of the birth cohort nominally represented by the NLS survey.

What is surprising about the NLS prospective design is that it apparently foundered on the same shoals as earlier and less extensively representative population samples such as the Oakland Growth Survey, the Berkeley Growth Survey, and the Fels Institute study. That is, despite the intention of collecting dynamic information for a representative sample of a historically interesting birth cohort, the NLS result is a flawed sampling frame, incomplete or non-comparable data from panel to panel, and essentially static approximations of life-course transitions. The latter point emphasizes the tendency in virtually all prospective studies to approximate the continuous flow of behavior by discontinuous information about statuses and conditions at the time of each successive interview or panel observation. This data collection strategy compromises the panel design, for it leaves gaps in the record of events -- namely, those that occur between the panels.

The analyst's difficulties in exploiting the longitudinal potentials of prospective research are therefore not unique to TALENT or the NLS (see Bielby, Hawley, and Bills, 1978, for summaries and critiques).
of the NLS studies and the uses of the data by the social science community). For example, the economist James Morgan and his colleagues have been following a nationally representative panel of some 5000 households annually for nearly a decade--the Panel Study of Income Dynamics. At each previous panel prior to the most recent, heads of household were interviewed about their current labor force attachment, wages and salaries, other income, and family statuses; in the most recent wave, spouses also were interviewed. Despite the apparent availability of panel information--at least on a year-to-year basis--for both male and female heads of households, the DYNAMICS staff has just written a monograph comparing males and female earnings in which 1) the earnings measure is static (annual 1975 earnings), 2) information on weeks of employment, unemployment, and nonwork activities in 1975 is elicited in the 1976 panel via retrospective report, and 3) information on cumulative job-training experiences, tenure with employers and in specific jobs, and other long-term experiences in the labor force are gleaned as in a cross-sectional interview rather than from whatever record has been compiled prospectively (see Duncan and Morgan, 1978).

One conclusion I wish to draw from these illustrations is that prospective designs for longitudinal research are not without serious methodological and procedural problems that are manifest in the histories of many if not all such projects. By no means has this commentary touched upon all such problems. Indeed one matter of central importance to the evaluation of the prospective method is the effect of continued contact with the same sample cases on the quality of information elicited from the person or household. For example, higher incidence of current unemployment
is reported by persons participating for the first time in the monthly
Current Population Survey than by those who are participating in their
fourth month; the latter, in turn, are more likely to report current
unemployment than are those in their eighth month of participation
(Taeuber, 1975). Since all respondents retrospectively report their
experiences of the same calendar period on the same measurement occasion,
and since there is no basis to expect true differences in unemployment by
length of participation in the panel, one can only surmise that the
differences are some effect of frequency of contact.

RETROSPECTIVE LONGITUDINAL RESEARCH: LONGITUDINAL DATA ELICITED WITHIN
CROSS-SECTIONAL DESIGNS

Retrospectively-focused questions lodged within cross-sectional
surveys provide an alternative means of collecting longitudinal data.
While the retrospective method is not without potentially significant
limitations, it does avoid many disadvantages of the prospective method.
In addition, retrospective longitudinal data can be collected efficiently
for more than one (birth) cohort simultaneously -- through a cohort-
sequential design (Schaele, 1965; Baltes, 1968). Thus, developmental
studies based on cross-sectional designs and retrospectively elicited
longitudinal data can become cost-efficient alternatives to panel studies,
especially for (cohort) comparative investigations.

Nearly all social surveys elicit some small amount of information
retrospectively, even if these items are limited to questions about a
respondent's family of orientation and social origins -- parental education
and occupation during R's adolescence, for example. Other surveys, such
as the U.S. Bureau of the Census' Current Population Survey (CPS), rely
quite heavily upon reports both about the present and the past. For example, at least once a year the CPS collects information (for all persons over age 14 in a selected household) on length of schooling, current occupation, longest job in previous calendar year, income from detailed sources in last complete calendar year, current employment status and length of unemployment, and job search behavior in last year. (These data are subsequently used to estimate official employment and unemployment statistics.) Strictly speaking, nearly all of the CPS data are reports about the past, albeit the immediate past. Measures such as "current" occupation refer specifically to occupation during the full week preceding the survey. "Current" employment status also refers to that previous period and is indexed through a series of questions about "main activity last week" and about job-seeking behaviors. Even "current" education, while referring to an essentially fixed or stable characteristic of most adults is in fact a report about a series of events that terminated some time in the past -- perhaps as many as several decades in the past.

Thus, retrospective information is elicited to some degree in most cross-sectional designs, even those surveys like the CPS which are designed to portray current population conditions with as much precision as is feasible. In addition, retrospective reports have formed the core of data from which empirical assessments of social mobility and of education's role in achieving equality of economic opportunity have been made. The replicate social mobility studies, "Occupational Changes in a Generation (OOG)" in 1962 and 1973 illustrate this usage in longitudinal research (Featherman and Hauser, 1978).
The OCG studies were designed to address two main questions:

"How much opportunity is there for socioeconomic mobility between and
within generations?" "Have opportunities for achievement changed in
recent years and for younger cohorts?" Using the March CPS in 1962
and 1973 as sampling frames, the OCG studies supplemented the CPS in-
formation about "current" levels of schooling, employment and occupa-
tion, and income with retrospect reports about social background, jobs
held at earlier points in the work career, and life cycle events like
marriage and military service. The availability of retrospective in-
formation about the social origins and schooling of several birth
cohorts -- derived simultaneously from a single OCG survey (actually,
derived independently from two separate samples -- 1962 and 1973 -- from
a single OCG universe) permits the construction of a time-series on
access to education by socioeconomic background. Table 1 illustrates
a type of "cohort-sequential" design in which the association between
social "class" and educational achievement is traced across the ex-
periences of successive (birth) cohorts. The trend to greater access
to schooling irrespective of social background is evident in this table.
It permits the analyst to speculate about the temporal or historical
conditions that underlie such trends and to study cohort-specific
processes without the need to undertake separate prospective studies of
the individual cohorts.

Other studies of socioeconomic achievement over the life course
drawn from a combination of prospectively and retrospectively derived
longitudinal data. For example, Sewell and Hauser (1975) have followed a 1957 graduation cohort of Wisconsin high school students into young adulthood. Based on their 2-wave data (1957 and 1964), they construct a dynamic model of achievement as illustrated by Figure 3. The 1957 panel elicited information about current (senior year in high school) statuses as well as retrospective reports about social background; the 1964 wave collected information about early career attainments, including schooling, jobs, and income. While the Wisconsin panel study permits the construction of richer, more elaborated models of the longitudinal process of achievement than do the OCG studies, it is noteworthy that in identically specified, rudimentary models of that process the two types of designs yield estimates of relationships that are substantially the same (Sewell and Hauser, 1975).
The most extensive use of retrospective longitudinal designs appears in what might be termed "life history" studies. The tradition of such designs is longstanding, as instanced by the classic work of W. I. Thomas and Florian Znaniecki, *The Polish Peasant in Europe and America* (1918-20). More recent applications of this method included the "Life Circumstances" study, carried out by Coleman et al. (Blum, Karweit, and Sørensen, 1969), a replicate of that investigation in Poland (Pohoski, 1974), and quite parallel researches in Mexico (Balán, Browning, and Jelfn, 1973), and Hungary (Andorka, 1977). The Coleman et al. (HOPKINS) study is illustrative of the lot. Here, a respondent is asked to recall the history of events in a series of life domains -- schooling, work, marriage, residence, and parenting. Taking age 14 and schooling as a starting point, the respondent is led by an interviewer through the years, grades, and schools. Every time a respondent changes grades or schools, these events are recorded by month and year. (In HOPKINS, the life domains were represented by columns in a questionnaire and time -- years and months -- constituted the rows. The interviewer completed each column -- domain -- sequentially from age 14 through the present.) With the history of one domain fully accounted, the respondent moves to the second, sometimes beginning with the past and working forward -- as with schooling -- and sometimes starting with the present and sequencing backward -- as with the marital history. In either case, the histories of domains can be cross-referenced by interviewer and respondent alike in order to provide contextual markers for recall and for putatively greater sequential accuracy in any single domain.
When complete, life history investigations like HOPKINS permit extensive longitudinal research within the confines of a cross-sectional design; they provide a limited alternative to panel designs. Their advantages and disadvantages are as follows.

**Advantages of retrospective life history studies:**

1. Within the same measurement occasion -- a single cross-section -- more than one cohort can be studied. Clearly the histories of cohorts would be of unequal length, regardless of the cohort definition (e.g., birth, school, marital). Nevertheless, for the period of the life cycle (or of whatever developmental process is indexed by the cohort definition) that is common to all cohorts, the analyst obtains cohort-sequential data at a single measurement occasion (as in the OCG illustration). Such a feature is not unimportant, especially in view of the emerging awareness in a variety of social science disciplines of the life-span, cohort-specific nature of developmental processes (Riley, Johnson, and Foner, 1977; Baltes and Schaie, 1973; Elder, 1975; Vinovskis, 1977).

2. A related feature is the replicative nature of measurement and concept definition within the cross-sectional, cohort-sequential design. Since measurement across all reference periods (dated events in the historical record of domains for a single individual) and cohorts occurs simultaneously, instruments (questions, tests, scales) are consistently applied and have nominally equivalent meanings. In the absence of multiple measurement occasions, investigators do not have a chance to change or refine questionnaires as reflections of conceptual "sharpening,"
development. Without prior knowledge of the likely temporal boundaries of such changes, discrete-time panels may well miss the critical points, especially so if there is high intra-individual variability in their timing.

6. The last advantage is entirely practical. To the extent that retrospective life histories elicited through a cross-sectional design can approximate cohort-sequential panel data, the costs of longitudinal research can be reduced. For the one-time cost of a single survey, longitudinal data are acquired for multiple cohorts. Such studies are conceivable and can be brought to completion within the lifetime of a single research team which would not be facing the risks of failing interests of potential funding sources for successive panels.

Disadvantages of the retrospective method:

1. While the risks of heavy sample attrition in longitudinal research are substantially reduced by retrospective designs, life history studies are vulnerable to the selective effects of mortality. In any population cross-section, the surviving members of older birth cohorts are less fully representative of their cohorts at some earlier time -- say 10 years previously -- than are survivors of younger cohorts. Thus, cross-sectional life history studies tend to underrepresent the earlier experiences of older cohorts as a function of differential mortality. This kind of bias, whether within a given cohort over time or at a given time across cohorts, can be remedied in part through multiple decrement cohort life table techniques (Shyrock and Siegal, 1976: Ch. 15).
2. In an age cross-section, the life histories of younger cohorts are truncated or incomplete. For some phenomena, such as fertility behaviors (e.g., Potter, 1966), life table techniques can be employed to fill in incomplete histories. (See also Heckman, 1979, for a statistical treatment of truncation issues.)

3. The major shortcoming of the retrospective approach to developmental research arises from the irretrievability of certain psychological states and sociological conditions. This may occur for two reasons: 1) because the datum cannot be supplied by the respondent or by some other source -- adults cannot report their grade 6 intelligence battery scores, although such scores might be available from school records; 2) memory decay -- deterioration of the neural traces for events -- may obviate a capacity to recall (Bach, 1952). Thus, for most purposes, the retrospective method is not a satisfactory alternative to panel studies of personality development; however, there are some exceptions (e.g., Rosenthal, 1963; Mikesell and Tesser, 1971; Back and Bourque, 1970).

4. Decay of neural traces and the bias of recall are related. In the absence of memory, recollections become fabrications, but recall can be self-serving distortions even in the presence of an active memory (Cahalan, 1968-69). In either case, retrospective reports are subject to anamnestic errors -- errors of recall and bias. Because the retrospective approach appears so vulnerable to these biases, it is essential that the trustworthiness of retrospective data -- their validity and reliability -- be examined in some detail.
ARE RETROSPECTIVE REPORTS ACCEPTABLE MATERIALS FOR LONGITUDINAL RESEARCH?

Because of the interpersonal aspect of childhood memories of adults and because of their functional role in achieving adult tension reduction, it is systematically impossible to reconstruct the childhood life space from retrospective reports of adults, for such retrospections are geared to the social adjustment requirements of the situations in which the retrospection is made. And even when these situational factors are such as to facilitate the most uncontaminated view of the past, implications from neurology make it highly implausible that an adult brain can mediate 30 years later the exact reproduction of childhood experience. (Bach, 1952:97.)

This rather negative view of a clinical therapist expresses a good many of the substantial reservations one might have about the acceptability of retrospectively reported data for developmental or longitudinal research, irrespective of a specific interest in a subject's childhood. With Freud, one might suppose that certain materials of revealed memory are contaminated by denial and ego defense functions. Or, one could concur with Lewin that retrospection is shaped by situational influences -- perhaps even just in the interview setting -- as adaptive reconstructions to personal and social needs and situational demands (i.e., social desirability). Others might agree with Adler that memory is self-serving, tending to perpetuate neurotic themes and fictions about a consistent, temporally constant Self. All of these views cast memory as a functional mechanism for ego, and these functions can be played without reference to what might be regarded as the more fundamental question: "How long do traces of events remain neurologically viable?"

But these reservations about retrospective reports might also be raised
as objections to any interview or social survey materials; they are perhaps ones of degree and not of kind that distinguish concurrent from retrospective data. If so, then in answering the question that heads this section one must compare biases in retrospective reports with those in reports about concurrent states. Put another way, will a person's public construction of some event in his or her life be as valid ("true") in the long run as in the short? Or phrased even more precisely, will retrospective reports be no less valid than concurrent ones? (One can cite instances where the validity of some datum might actually increase over time, e.g., Mason et al., 1976. Additionally, Powers et al., 1978, report no essential difference in the set of relationships between several personality measures and either contemporaneous or retrospective (ten-year recall period) reports of employment and economic states.)

One can conceptualize the comparative study of validity as in Figure 4. One would seek subject's reports about a variety of life circumstances or domains; these could be about one's self or one's own circumstances, or they could be proxy reports about one's parents or some other, or they could be some other's reports about ego. Such information would comprise the rows of Figure 4. The columns of the figure indicate the reference period for which the subject is asked to report. That is, irrespective of how far in the past some event began or terminated, the interval over which the event or state occurred can vary and (presumably) affect validity (e.g., "telescoping"). Finally, the recall period -- the third dimension of Figure 3 -- denotes lapsed time from the event of interest (and its reference interval) to the measurement occasion. Concurrent reports would have a short recall period while retrospective reports would have longer ones. Presumably, the scale for the "recall" and "reference" dimensions is some unit of clock time. Of course, the cells of this classification would contain data on validity. For example, one might like to know if the
Using this framework, one derives a rather complex answer to the question about comparative validity. Retrospective data about some life domains may be more trustworthy than about others, conditioned by the length of the reference and recall periods, respondent's socioeconomic attributes, and the methods of interviewing or elicitation. Obversely, retrospective longitudinal designs -- such as HOPKINS and similar life history cross-sections -- might be acceptable or preferable alternatives to panel designs when the analyst's substantive interests are of one kind but might not be as acceptable in the case of others.

Unfortunately, my own efforts to classify extant validation studies into the likes of Figure 4 are at an early stage. In the remainder of this section I shall review an illustrative subset of the small collection of studies of which I am now aware.

Validation studies: demographics

As part of an ongoing assessment of quality of measurement, the U.S. Bureau of the Census undertakes the validation of demographic information collected as part of the decennial censuses and the monthly Current Population Survey (CPS). If only because of this institutional commitment, more information seems to be available about accuracy of reporting demographic characteristics than of other sociological or psychological ones. The Census studies are of three basic kinds: 1) reliability studies in which reinterviews are conducted with sample elements after periods of a few days to 2 or 3 weeks (e.g., U.S. Bureau of the Census, 1974); 2) validation studies in which some interview response is matched to a record or some independent criterion (e.g., U.S. Bureau of the Census, 1970); and 3) consistency studies in which comparison of independent measurements
across the same subject population are made without a criterion or "true" record (U.S. Bureau of the Census, 1964; U.S. Bureau of the Census, 1965). In instances as the latter -- where an error-free criterion cannot be assumed or is not available -- the analyst often is constrained to use estimates of short-term stability (e.g., test-retest reliability) as the limit of validity and to estimate validity through split-half reliabilities and similar internal consistency checks. That is, validation is attempted out of a combination of studies 1) and 3) above, yielding a multiple-occasion, multiple-indicator design. Illustrations of the latter are Bielby, Hauser, and Featherman (1977) and Bielby and Hauser (1977) using data about one's concurrent demographic statuses and those (in retrospect) characteristic of one's family of orientation during adolescence. (See Gutek, 1978, for a different approach to estimating validity in the absence of a criterion.)

Table 2 cites data pertinent to the relative validity of concurrent and retrospective reports as drawn from the studies of Bielby and associates. In brief, the data were culled from a reinterview of several hundred adult participants (aged 20-65) in a large study of social mobility carried out as an adjunct to the March CPS in 1973 (Featherman and Hauser, 1978). Interviews and reinterviews covered information about social origins (parental characteristics at approximately respondent's age 16) and current occupation, education, and earnings as of the survey week (actually the week preceding the survey or, in the case of income, the preceding year). Test-retest correlations indicate a high order of similarity in the temporal stability of reports of both concurrent statuses (about self) and retrospective statuses (about family head). That is, across the "recall" dimension
of Figure 4, the accuracy of reporting nominally the same "type" of information is roughly the same even though the "reference" periods are dissimilar. (Parental statuses refer to states held "most of the time up to" or "about" age 16.)

But test-retest correlations are not measures of validity, since either (or both in this instance) the original or the remeasurement are subject to bias. Indeed, errors of reporting may be correlated across measurement occasions, adding to the size of the within-variable test-retest correlation. Correlation of errors among reports within the original and within the re-interview might also occur. Using the test-retest covariances as a basis for estimating common variance reflecting an unmeasured "true" score for each nominal variable, Bielby et al., derive bias estimates for concurrent and retrospective reports. This is accomplished by estimating a structural equation model of the "stratification process" in which measured variables are assumed to be less than perfectly valid and errors in variables are correlated within occasions (e.g., the errors in original report of head's education are correlated with those in original report of own education) or between occasions (e.g., the errors in the separate indicators of a common factor "education" -- embedded in the test-retest correlation -- are correlated). Because a model which specifies uncorrelated errors of both kinds fits the data for nonblack men, the "reliability coefficients" in Table 2 are nearly identical to the test-retest correlations. (The reliability coefficients are estimated from the model and are subject to its specifications about error structure; they are computed as the ratio of "true score" variance to total or observed variance under the conditions
of the model.) Given this set of conditions, one estimate of relative validity among variables is achieved by taking the square root of the "reliability coefficient." Validity is both high and comparatively indifferent to the retrospective-concurrent distinction (viz. the "recall" dimension of Figure 4).

Another set of validity estimates for sons' reports about their parental characteristics appears in Table 3. These data are drawn from companion work to Bielby et al., based on a record match between Wisconsin adult sons' reports (from the same nominal population of nonblacks as in Bielby et al.) and decennial census records of their family heads between 1920 and 1960 (Featherman and Hauser, 1975). This is a validation study as in 2) above where some criterion is regarded as a "true" score -- in this case, the family head's census data. So construed, the proxy-census correlations estimate a degree of validity in retrospective reports that is slightly lower than in Bielby et al. -- closer to .8 than in excess of .9, excluding for the moment information on parental income. Of course, it is doubtful that the census data are error free, on exactly the same argument that led Bielby et al., to assume that CPS reports of adult sons about their own jobs or schooling were less than perfectly valid. Accordingly, the correlations in Table 3 are biased estimates of validity, and further modeling -- along the lines of Bielby et al. -- needs to be carried out. What is important for this context, however, is that memories of adults about their demographic childhood appear to be more robust than the quotation from Bach (at the head of this section) would lead us to expect. Apparently, some aspects of childhood can be reconstructed retrospectively within acceptable degrees of error.
What about family income? The correlations of proxy and census reports in Table 3 is very low -- 0.28. This is misleading for the following reasons. Wisconsin men were asked to report the average annual income of their families about the time they were age 16. The census report is for family head's personal total income for a specified census year -- close to the son's sixteenth birthday. Sons reported income in various ways -- dollars per month, week, year -- and were converted mechanically into annual estimates. Heads (or their family proxies) reported components of income, making distinctions between net and gross receipts in ways that sons probably were not. Consequently, a census may record a zero or negative income for a head, whereas a son's may not make such calculations. Roughly 30% of Wisconsin sons retrospectively reported a nonzero (also non-negative) "family" income in cases where the family head was classified by the census as "loss or zero." Removing these cases and recomputing the correlation yields a "validity" estimate of 0.50 for these crudely corresponding definitions of parental income. On the argument that such estimates are more accurate when dollars are converted to natural logs (Bielby and Hauser, 1977), the correlation was again computed on the restricted sample of positive income reports, and the estimate rose to .57.

Validation studies: miscellany

(Owing to the early stage of my excursion through extant validation studies, those which were conducted on primarily nondemographic subjects are less systematically examined in this section. This is both a reflection of my impressions of the literature to date as well as of the status of my review program.)
It surely might be possible that reflections of childhood -- at least images of more qualitative dimensions of family background than the more demographic ones -- are distorted in retrospect (Vaillant, 1977; Yarrow, Campbell, and Burton, 1970). Distortion could entail low validity as well as bias among the reports of current and past events (e.g., correlated errors as in the Bielby et al., example). However, some studies demonstrate at least a prima facie case for the ability of adult subjects to relate detailed interpersonal and other early life experiences in connection with later life diagnosis of authoritarianism (Mikesell and Tesser, 1971). And, at least males seem able to offer valid retrospective accounts of their adolescent experiences and parental relationships (Rosenthal, 1963).

Whether the degree of distortion is worse than that within concurrent reports is (as this paper would have it) the real question. Without some multimethod, multioccasion design (again the Bielby et al. example is illustrative) these comparative error evaluations are hard to make. It seems unlikely, however, that the "sensitivity" of some datum is alone a very diagnostic dimension for what gets distorted and what does not. For example, adult women are able to report with apparently high accuracy the age at menarche; at least cohort distributions created with retrospective reports are nearly identical to concurrent reports of adjacent cohorts (USDHEW Series 11 No. 133, 1973). Former drug arrestees and patients readily report subsequent incarcerations and treatment readmissions when compared against criminal justice and clinical records, over a 3-to 4-year recall period (Simpson, Lloyd, and Gent, 1976). The 2 to 7% error rate with such information is consistent with other assessments of retrospective reports of illicit drug use, even when a "true" criterion is unavailable, indicating high reliability and validity (Single, Kandel, and Johnson, 1975). On the other
hand, victims of crimes are likely to underreport their incidence when occurring beyond a three-month recall period, to "telescope" criminal experiences into a reference period (tending to distort time or sequence analysis), and to differentially report exposure as a function of the type of crime (Turner, 1976). Perhaps personal salience and degree of desired detail are important components in the ability to reconstruct the past. For example, evaluations of consumer expenditures by length of recall and of reference periods suggests that major purchases (e.g., residential repairs and alterations) are subject to less "telescoping" and memory erosion than minor ones (Neter and Waksberg, 1965). In addition, "conditioning" (training) of respondents in an expenditures panel, coupled with "bounded recall periods" (reminding respondents of previously reported events during a specific reference interval), tended to enhance apparent validity and overcome differential salience (Ibid.).

Efforts to validate retrospective life histories (more precisely, the data collected via the retrospective method) are still incomplete. For instance, the HOPKINS staff and other analysts of retrospectively elicited histories have yet to report evidence about data quality. Such work as is underway (unpublished materials kindly provided by Dr. Nancy Karweit) shows mixed levels of acceptability across subjects or domains. However, such studies are perforce limited to mainly demographic comparisons of cohort and distributions in HOPKINS/independently drawn samples from census and special surveys. This limitation is not unique to HOPKINS, but it does not substantiate the degree to which individuals are able to reconstruct their own pasts. Obversely, it does not aid future users of the retrospective
method by suggesting improvements in the methods of assisting respondents (methods that are molded to the social circumstances of individuals) to recall with greater accuracy. (HOPKINS does permit a rough match of personal earnings with individual social security records.)

CONCLUSION

A concluding assertion about the comparative utility of the retrospective longitudinal method would be premature. This paper has described an agenda for the empirical assessment of the method, based on the relative biases that arise in concurrent reports (and which are not unique to retrospective studies) and that characterize retrospective reports of various types, over different recall periods, and with differing reference intervals. Since the major reservation about the retrospective method is its ability to elicit (comparatively) acceptable data for developmental research, it is unlikely that a convincing judgment on this question can or will be made from a thorough review of extant validation studies in the behavioral sciences. Included in the latter are still incomplete efforts to assess the quality of data elicited by retrospective life history studies.

Consequently, it is the judgment of this writer that specific studies need to be designed to assess the "robustness" of the retrospective method of longitudinal research (e.g., Powers et al., 1978). In that regard, the attractiveness of prospective panel studies like TALENT is very high. Using the panel as a sampling frame, the analyst can compare a retrospectively elicited life history with successive concurrent reports in the panel at the individual level. In addition, experiments with methods of elicitation could be applied in retrospectively focused interviews of panel members (e.g., U.S. Department of Health, Education, and Welfare, 1977). If bias is
Footnotes

1 Other efforts to address this question include the very thorough review monograph sponsored by the British Social Science Research Council (Wall and Williams, 1970) and a synoptic paper by Campbell (1978). Baltes (1968) offers systematic comparisons of longitudinal and cross-sectional designs for human developmental research. Some of his ideas are visibly reflected in this paper.

2 This view of the necessity of standardized and replicated instrumentation in connection with longitudinal research may require qualification. Wall and Williams (1970:15) distinguish between predictive and genetic longitudinal studies. The former, insofar as they monitor stability in some specific characteristic with known ontogenetic properties in a population (assessing the effects of environmental change or manipulation), are consistent with the principle of standardization and replication at fixed temporal intervals. The latter, however, in their search for developmental sequences and the discovery of causal relationships, may neither profit from standardization nor from large and representative samples. The present writer is not convinced by this distinction, but admits to the desirability of an experimental attitude on the part of the longitudinal research, so long as he or she does not sacrifice replicability and the integrity of population representation. That is, one can experiment with new instruments even as one repeats the old and in so doing can examine the concurrent and cross-time relationships among indicants of nominally the same construct.
Bibliography

Andorka, Rudolph

Bach, George R.

Bachman, Jerald G., Patrick M. O'Malley, and Jerome Johnston.

Back, Kurt and Linda Bourque

Balán, J., H. O. Browning and E. Jelfn
1973 Man in a Developing Society: Geographic and Social Mobility in Monterey, Mexico. Austin: University of Texas Press.

Baltes, Paul B.
Balters, Paul B., and K. Warner Schaie (eds.)
1973   Life-Span Developmental Psychology: Personality and
Bielby, William, Clifford B. Hawley, and David Bills
1978   Research Uses of the National Longitudinal Surveys.
        Social Science Research Council. Washington, D.C:
        Center for Coordination of Research on Social Indicators.
Bielby, William T., and Robert M. Hauser
1977   "Response error in earnings functions for nonblack males."
        Sociological Methods and Research. 6(2):241-260.
Bielby, William T., Robert M. Hauser and David L. Featherman
1977   "Response errors of nonblack males in models of the
        stratification process." Pp. 227-251 in A.S. Goldberger
        and D. J. Aigner (eds.) Latent Variables in Socioeconomic
Blum, Zahava, Nancy L. Karweit, and Aage B. Sørensen
1969   "A method for the collection and analysis of retrospective
        life histories." Report No. 48, Center for the Organization
Buss, Allan
1974   "A general developmental model for interindividual
        differences, intrapersonal differences, and intraindividual
        changes." Developmental Psychology. 10(1):70-78.
Campbell, Richard T.

Cahalan, Don

Cattell, R. B.

Cattell, R. B.

Duncan, Beverly and Mark Evers

Duncan, Greg and James Morgan (eds.)

Elder, Glen H., Jr.
Featherman, David L. and Robert M. Hauser


Gutek, Barbara A.

"On the accuracy of retrospective attitudinal data."

Heckman, James J.

1979   "Sample bias as a specification error."

Kodlin, Dankward and Donovan J. Thompson

1958   "An appraisal of the longitudinal approach to studies in
       growth and development." Monographs of the Society for
       Research in Child Development, Inc. Vol. 23. Serial 67,
       No. 1:7-47.

Mason, William M. et al.

1976   "Models of response error in student reports of parental
       socioeconomic characteristics." Pp. 443-484 in William
       H. Sewell, Robert M. Hauser, and David L. Featherman (eds.)
       Schooling and Achievement in American Society. New York:
       Academic Press.

Mikesell, Richard H. and Abraham Tesser

1971   "Life history antecedents of authoritarianism: a quasi-
       longitudinal approach." Proceedings of the 79th Annual
       Convention, American Psychological Association, pp. 369-370.
Morgenstern, Richard D. and Nancy S. Barrett

Meter, John and Joseph Waksberg

Pohoski, Michal

Ornstein, Michael

Potter, Robert G.

Powers, Edward A., Willis J. Goudy and Pat M. Keith
Ramsøy, Natalie Rogoff


Riley, Matilda W., Marilyn Johnson, and Anne Fonner (eds.)

1977 Aging and Society: a Sociology of Age Stratification.
New York: Russell Sage

Rosenthal, Irene

1963 "Reliability of retrospective reports of adolescence."
Journal of Consulting Psychology. 27(3):189-198.

Sewell, William H., Robert M. Hauser, and David L. Featherman (eds.)

1976 Schooling and Achievement in American Society.

Sewell, William H., and Robert M. Hauser


Schaie, K. Warner

1965 "A general model for the study of developmental problems."

Shyrock, Henry S. and Jacob Siegel

1976 The Methods and Materials of Demography.
New York: Academic

Simpson, D. Dwayne, Michael R. Lloyd and Michael J. Bent

Singer, Burton and Seymour Spilerman


Singer, Burton and Seymour Spilerman


Single, Eric, Denis Kandel, and Bruce D. Johnson


Taueber, Conrad


Turner, Anthony G.


U.S. Bureau of the Census

U. S. Bureau of the Census

1970  "Record check study of accuracy of income reporting."
      Evaluation and Research Program of the 1960 U.S. Censuses
      of Population and Housing. Series ER 60 No. 8. U. S.

1965  "Response errors in collection of expenditures data of
      household interviews: an experimental study." Technical
      D. C.

1964  "Accuracy of data on population characteristics as
      measured by CPS-Census match." Evaluation and Research
      Program of the U.S. Censuses of Population and Housing,
      1960. Series ER 60 No. 5. U. S. Government Printing
      Office, Washington, D. C.

U. S. Department of Health, Education and Welfare

1977  "A summary of studies of interviewing methodology."
      Data Evaluation and Methods Research Series 2, No. 69.

1973  "Age at menarche, United States." Vital Health
      Statistics Series 11, No. 133. U. S. Government
      Printing Office, Washington, D. C.

Vaillant, George E.

1977  Adaptation to Life: How the Best and the Brightest Came
Vincovskis, Maris A.


Wall, W.D., and H. L. Williams


Wise, Laurrell


Wise, Laurrell, D. G. McLaughlin and L. Steel


Yarrow, Marian Radke, John D. Campbell, and Roger V. Burton