G.L. Maddox (Ed.)
Encyclopedia of Aging (3rd ed.). New York: Springer Publishing Co.

RESEARCH METHODS IN GERONTOLOGY

This article discusses selected methodological issues of particular concern to aging researchers. The focus is on the treatment of age and time as central variables, upon some characteristics of observations, and upon the specification or mis-specification of data collection strategies that include the age or time variable. More extended discussions of research methodology and aging may be found in Lawton and Herzog (1989), Schaie (1988, 1996), and Schaie, Campbell, Meredith, & Rawlings (1988).

Age and Time

Gerontological researchers frequently employ calendar age as their independent variable of prime interest. What remains frequently ignored, although crucial for the design of studies and the estimation of causal models, is the fact that while calendar age is a time-ordered process, it is not a given that time of onset or duration of a specific behavior must necessarily be correlated with age. For example, executive burnout could result as a function of length of time on the job or due to the half-life of professional knowledge. Since the burnout phenomenon is temporally indexed by time elapsed since initial training, it may be spuriously correlated with calendar age if there is a customary age at which one pursues initial training or enters a job. The same phenomenon, however, would be uncorrelated with age when observed in a population where career entry occurs over a wide age range; in that case, it is time-dependent, but not age-related. Many so-called aging phenomena are probably of the latter type. They may be equally interesting, but it would clarify design specifications if the term "age-related" is applied only to those processes that have an age-specific onset, asymptote or rate of change. By contrast, it would be more appropriate to label as "time-dependent" those processes that imply developmental change, but that have wide latitude in age of onset or progression (Schaie, 1994).

A bothersome feature of measurement in the developmental sciences is the fact that data collected over time and age are not normally distributed, but instead represent ordered sequences whose form may vary depending upon life stage or substantive content. The most common form hypothesized to fit age-related phenomena is the well-know Gompertz curve which assumes a steeply decelerating increment until a young adult asymptote is reached, with slowly accelerating linear decline thereafter. The implied irreversible decrement model is unduly restrictive. It does not allow for the recursive or lagged phenomena so characteristic of development and aging. The unidirectionality of time-ordered observations, however, presents a sound basis for causal inference, in that most consequents occurring later in time cannot be said to have "caused" antecedent behaviors and events. The unidirectionality of time-ordered behaviors, therefore, provides the major rationale for applying cross-lagged correlation and structural equations modeling (SEM) techniques to developmental data or for the informed choice of many of the fixed parameters required in model estimation.

Nevertheless, many investigators assume erroneously that the unidirectionality of time-ordered observations implies that time-ordered change must also be unidirectional in nature, even though there are many age-related processes that may be unidirectional at best over limited portions of the life-span, and others that may generally take cyclical or recursive form.

Characteristics of Observations

Except for demographic and/or quite specific physiological indices there are very few directly observable behaviors that directly contribute to our understanding of human development. Instead, observable behaviors typically serve as indicators of the respondent's standing on some theoretically defined or empirically abstracted latent construct of interest. This is fortunate, because the equivalence of single measures of a particular construct over wide age ranges and time periods may often be questionable. One of the major flaws of many aging studies, therefore, is the failure to apply the variety of techniques now available that allow us to study the measurement equivalence of multivariate systems across the developmental comparisons to be made (Schaie, Campbell, Meredith, & Rawlings, 1988; Schaie, Maitland, Willis, & Intrieri, 1998.)

Adequate demonstration of measurement equivalence across wide age intervals may also be limited by discontinuities in the development of behavior over time. Such discontinuities are most obvious in infancy, where many behaviors that have regular developmental progressions are nevertheless not predictive of later behavior. This is so because certain behaviors reach an asymptote required for survival such that the surviving populations becomes quite homogenous with respect to the predictor variable. In old age, similarly, behavioral decrement at earlier ages may not be predictive of later events, because the magnitude of decrement is below the threshold levels that must be attained before other behavioral domains are impacted.

Many aging studies fail to give adequate consideration to the scale properties of their measurement instruments. In particular, there seem to be many instances when ordinal scales would better fit the conceptual model employed than the interval scales that are proposed for reason of analytic convenience). Whenever hierarchial structures are investigated within groups or an individual, a strong case can be made for the application of ipsative measures that describe the relative position of an individual or particular variate within an individual with reference to a finite and specified set. In fact, one of the crucial, and often unaddressed, design questions that must be solved on theoretical and substantive grounds is concerned with the issue of whether developmental change should be considered to be an absolute or relative phenomenon.

An interesting recent development has been the increasing emphasis on the observation of older persons in naturalistic settings. One of the dilemmas in evaluating the adequacy of data collected in such settings has frequently been the discrepancy in the care with which settings are described, and the often quite causal specifications on the quality control for the recording of participant observer data and the reduction of such data to a form in which quantitative tests of hypotheses might be feasible.

Design Specifications and Sampling Issues

It is widely held that the highest quality of aging data arises from longitudinal studies. Investigators who are unsure about the conceptual and methodological reasons that would buttress such an opinion (cf. Baltes & Nesselroade, 1979; Schaie,1988,1996), consequently often provide an apology for their proposed use of cross-sectional data, without ever addressing the question whether their proposed design is or is not relevant to the research question asked. What is missing too often is an understanding of the considerations involved in determining whether a cross-sectional approach may indeed by the method of choice, or whether the question that is asked requires knowledge of the magnitude of intraindividual change, and thus cannot be answered without longitudinal data. Sometimes planned design misspecifications may be useful where assumptions that would otherwise be unduly strong can well be justified. For example, cohort effects are likely to be trivial when the range of ages compared is small, while period effects may be relatively trivial in many laboratory paradigms, except as expressions of reactivity or instrumentation effects (cf. Schaie, 1988, 1994, 1996).

The choice of appropriate age/cohort boundaries for one's comparison groups is not always readily apparent. Although it is easy to dismiss out of hand research paradigms that compare young adult college students with senior citizen center visitors, it is far more difficult to prescribe positive alternatives. If the investigator has no theoretical model that specifies the age at which reliable changes are to be expected, it might be quite sensible to sample individuals at random ages. Where cross-sectional studies are used to provide a first estimate of age-related change, however, it may be important to minimize cohort confounds. It would then be best to choose narrow age ranges centering about that point where age changes have been observed anecdotally in substantial numbers of persons.

Design specification issues often converge on problems of sampling, where requirements may differ markedly among different behavioral and social science disciplines. Many secondary data analyses proposed by economists, political scientists and sociologists require national probability samples; hence new data collections seek to obtain data sets that are as representative as possible of the total parent population. A critical question frequently asked in this context is whether hypotheses regarding change over age must be examined by follow-up studies of the same subjects, or whether drawing a new sample from the same population cohorts would suffice. The latter method, of course, would avoid the thorny issues of sample attrition and other effects that could threaten the internal validity of the study (see Schaie, 1996).

By contrast, few psychological studies are intended to obtain national or regional population parameter estimates. Representative samples for the psychologist or ethnographer, therefore, may mean no more than study populations that are broad enough to cover the full range of occurrence of the behavior of interest. Nevertheless, there are also a number of different options for these more limited sampling objectives. One may, for example, wish to sample the occurrence of behaviors roughly proportional to their incidence within a specified parent population. Alternatively, when the primary research interest is in the mechanism of a phenomenon rather than its population distribution, it may be quite legitimate to select subjects in terms of their suitability for the study of that mechanism. Psychological laboratory or ethnographic field studies, however, must also deal with the problems of external validity and thoughtful specification of subject characteristics are always mandatory. If meaningful comparisons are to be made across different ethnic groups, then the smaller minority group or groups must be oversampled to obtain sufficient statistical power for group comparisons.

Structural Equation Models in aging research

Structural equation models are particularly useful in aging studies because the unidirectionality of time permits sounder guides for the specification of causal paths than is possible in studies using single observation points only. Longitudinal factor analysis is a useful approach for modeling individual differences in intraindividual change, the central focus of any individual differences approach to aging (cf. Schaie, 1996).. An equally important role for such models, however, and one that requires attention by those who do cross-sectional work as well, is the applicability of structural equation models to the demonstration of measurement equivalence (Schaie, Maitland, Willis, & Intrieri, 1998). Utilization of the same questionnaire or test apparatus does not guarantee measurement equivalence over time or different subject populations. Aging researchers face the dilemma that no two individuals, or groups of individuals have identical characteristics at the same point in time, nor does a group or individuals retain identical characteristics over different points in time.

Two fundamentally different aspects are involved: the first concerns the traditional problem of reliability of measures across occasion and regression to the mean when using fallible measures. The second issue involves the fact that measurement equivalence would not be guaranteed, even if only perfectly reliable measures were used because of systematic but nonuniform changes occurring among individuals over time.

A major concern in cross-sectional studies is the question whether a task that may be a good estimate of one construct in young adulthood remains so in late life or in fact becomes a measure of some other construct. What must be demonstrated then is the invariance of factor structure across multiple groups or subpopulations (cf. Schaie, Maitland, Willis, & Intrieri, 1998). Similar requirements pertain when intervention studies propose multivariate effect models. Here it is desirable to test hypotheses regarding the structural equivalence of experimental and control groups prior and subsequent to intervention. In longitudinal studies, moreover, it is important to demonstrate factorial invariance for the same individuals over time. Structural equation analysis seems to be the approach of choice to assess measurement equivalence issues involving multiple groups and occasions It is also beginning to be applied to decompose sources of aging effects in experimental studies (see Salthouse, 1992). Other recent structural equation applications in aging research include latent growth models (McArdle & Anderson, 1990), and survival analysis (event history).

Assessment of Individual Change

While structural equation models seem to be the appropriate avenue to the assessment of measurement equivalence and the modeling of structural changes, including mean structures, there remain some very practical reasons why many investigators may wish to rely on more simplistic and direct descriptors of observed change. The furor over problems interpreting change scores in developmental research started by Cronbach and Furby seems to have led quantitatively unsophisticated investigators to shun the use of change scores and for some journal editors to resist publication of change score results. Recent discussions in the methodological literature, however, suggest that there may still be a role for simple change scores, particularly in two-occasion studies. More complex paradigms for the study of change may be found in Collins and Horn (1991).

References

Collins, L. M., & Horn, J. L. (Eds.) (1991). *Best methods for the analysis of change.* Washington, DC: American Psychological Association..

Lawton, M. P, & Herzog, A. R. (1989). *Special research methods for gerontology.* Amityville, NY : Baywood Publishing Co.

McArdle, J. J., & Anderson, E. (1990). Latent variable growth models for research on aging. In J. E. Birren & K, W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 21-64). New York: Academic Press.

Salthouse, T. A. (1992). Shifting levels of analysis in the investigation of cognitive aging. *Human Development, 35,* 321-360.

Schaie, K. W. (1988). Internal validity threats in studies of adult cognitive development. In M. L. Howe & C. J. Brainard (Eds.), *Cognitive development in adulthood: Progress in cognitive development research* (pp. 241-272). New York: Springer-Verlag.

Schaie, K. W. (1989). The hazards of cognitive aging. *Gerontologist, 29,* 484-493.

Schaie, K. W. (1994). Developmental designs revisited. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Theoretical issues revisited* (pp. 45-64). Hillsdale, NJ: Erlbaum.

Schaie, K. W. (1996). *Intellectual Development in Adulthood: The Seattle Longitudinal Study.* New York: Cambridge University Press.Schaie,

Schaie, K. W., Campbell, R. T., Meredith, W. M., & Rawlings, S. C. (Eds.). (1988). *Methodological issues in aging research.* New York: Springer.

Schaie, K. W., Maitland, S. B., Willis, S. L., & Intrieri, R. L. (1998). Longitudinal invariance of adult psychometric ability factor structures across seven years. *Psychology and Aging, 13*, 8-20.