Behavioural Science Section

Gerontology

Gerontology 2002;48:18–21

An Examination of the Hofer and Sliwinski Evaluation

Timothy A. Salthouse John R. Nesselroade

Department of Psychology, University of Virginia, Charlottesville, Va., USA

Key Words

Research design, elderly · Cross-sectional analysis · Longitudinal analysis

Abstract

Although we fully endorse the thesis of Hofer and Sliwinski that all research designs and analytical procedures should be critically examined before accepting conclusions based upon them, in this commentary we question whether their critical examination is complete. In particular, we suggest that they neglected some disadvantages of longitudinal and narrow age range cross-sectional designs and that they may have been overly pessimistic with respect to the potential usefulness of cross-sectional designs for certain research questions.

Hofer and Sliwinski [1] appear to make three major claims in their article. First, they claim that analyses of cross-sectional data can yield misleading estimates of the amount of shared variance associated with age because of mean-induced covariation among variables that are all related to age. They further argue that these types of analyses should not serve as the basis of evidence for theories and hypotheses of aging. Second, they propose that some of the problems of cross-sectional analyses can be eliminated by focusing on narrow age groups, because the mean age trend is no longer present when a single age group is examined. And third, they suggest that the strongest evidence for an association, or for the lack of an association, between age-related effects on different variables will come from longitudinal, and short-term 'bursts of measurement', designs. We agree with the authors that it is valuable to examine limitations of different designs and their attendant analytical procedures and to consider possible improvements on existing methods. However, we suggest that the current examination can be usefully extended, and the purpose of this commentary is to describe what we believe was incomplete in their article.

With respect to the third claim regarding longitudinal designs, we have little disagreement with the authors about the potential information value of longitudinal analyses when change is the focus. However, we believe that they have failed to acknowledge limitations of most existing data sets and analyses, and they have not sufficiently emphasized the importance of focusing on the process of change.

Among the limitations of most longitudinal data sets of which we are aware are the following. First, many are based on a single cohort, and thus they confound age and time of measurement. Second, the samples of participants and the sets of variables are seldom representative of the

KARGER
Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0304–324X/02/0481–0018$18.50/0

Accessible online at:
www.karger.com/journals/ger

Timothy A. Salthouse
Department of Psychology
University of Virginia
Charlottesville, VA 22904 (USA)
Tel. +1 434 243 5034, E-Mail salthouse@virginia.edu

Table 1. Range of absolute scores corresponding to the difference between scaled scores of 7 and 13 (–1 and +1 SD) from table A.1 of the WAIS III Administration and Scoring Manual [4]

Age years	Vocabulary	Comprehension	Information	Block design	Matrix reasoning	Digit symbol
20–24	25.5	12	11.5	27	9.5	32.5
25–29	24.5	11.5	11.5	27	10.5	31
30–34	25	12	12	26.5	10	32
35–44	25	11	11.5	26	10.5	33
45–54	24	10.5	11.5	24	11	30.5
55–64	27.5	12	12	24	12	30
65–69	27.5	12	12	23.5	11	30
70–74	27.5	11.5	11.5	20	9	29.5
75–79	27	11.5	12	18	8	29
80–84	27	12	11	16	7.5	30
85–89	26	11.5	11	16	6.5	28.5

population or variables that could be examined. Third, in many cases only a few variables are studied, and thus there is little opportunity to examine independence across a broad range of variables. Fourth, the correlation between the scores across measurement occasions is frequently quite high which suggests that the systematic variance in change is small relative to the variance in the scores at the two measurement occasions. Fifth, few attempts have been made to distinguish systematic age-related changes from practice effects, unsystematic fluctuation, and other changes attributable to factors other than endogenous maturation. That is, not all longitudinal change is due to aging, and the various possibilities should be separated to achieve valid inferences. And finally, the issue of the reliability of change scores has seldom been carefully examined in a manner that would allow meaningful interpretation of the correlations among changes in different variables. Some researchers have relied on latent constructs to minimize measurement error, but in most of the analyses of which we are aware, the constructs tend to be based on the assessment of variables at a given measurement occasion and not on aspects of the changes (e.g., rates and changes in rates) in the variables across occasions – but see McArdle and Nesselroade [2]. It is, therefore, possible that although the researchers may have been examining change in a reliable construct, they may not have been examining a reliable construct of change. In order to incorporate a truly dynamic focus, we would encourage the study of relations (and of the possible structure) among age-related *changes* in different variables, and not just the change in constructs that are defined on the basis of scores at a single measurement occasion. We believe that it is no longer premature to emphasize rates of change and changes in rates of change if research on adult development and aging is to progress beyond the static, equilibrium-oriented perspective that has dominated the field. Hofer and Sliwinski's [1] reference to designs involving 'bursts of measurements' indicates a desirable leaning toward capitalizing on interrelations defined over short intervals as a more promising 'base line' against which to model longer term changes [3].

With respect to the authors' advocacy of the value of the narrow age cohort design, we are less optimistic than they, because it requires the very strong assumption that individual differences in change are large relative to individual differences in initial values. That is, the single narrow age group design assumes that performance at a given age is determined by one's initial level of ability and by age-related changes in that ability. However, without multiple age groups it is impossible to separate these two sources of variance. Furthermore, this model seems to lead to an expectation of larger variability at older ages as the age-related change accumulates, an expectation that can be questioned on the basis of existing data.

Among the best sources of normative age trends is the standardization data from commercial test batteries, because they include a wide variety of psychometrically sound variables with nationally representative samples. The Wechsler Adult Intelligence Scale III [4] is one such test battery that can be used to examine the relation between age and interindividual variability. Table 1 contains the range of raw scores corresponding to the difference between a scaled score of 13 (representing 1 SD above the mean) and a scaled score of 7 (representing 1 SD below the mean) for eleven age groups in the WAIS III standardization data. Each age group was composed of

200 individuals except for the two oldest groups which were based on 150 and 100 individuals, respectively. Inspection of table 1 reveals that variability increases across age only for the vocabulary variable, and then only to about age 60. To the extent that the samples are truly representative of their age groups, the data appear to be inconsistent with a simple view of additive variances of initial ability and of change because that would imply increased variability that does not occur.

Another expectation from the assumption that the variability in change is large relative to the variability in initial level is that retest (stability) correlations should be relatively low. The most interesting retest or stability correlations are those across relatively long intervals, because they provide the greatest opportunity for cumulative change. Among the studies that have reported retest correlations for cognitive ability test across intervals of 20 years or more are those by Schaie [5], Owens [6], and Deary et al. [7]. In each case the correlations were computed between a composite score of intellectual ability. The correlation reported by Schaie across 21 years was 0.86, that reported by Owens across 42 years was 0.78, and that reported by Deary and coworkers across an interval of 66 years was 0.63. These moderately high to high values are relevant to the authors' argument for a narrow age cohort design because as the retest correlation increases from 0, more of the variance in the latter score is predictable from the original score, and hence less is available to be associated with age-related change mechanisms.

Because very strong assumptions are needed to justify the narrow age group design for the study of aging, and because at least some potentially relevant evidence appears to be inconsistent with those assumptions, we believe that multiple age cohorts are needed to separate initial level and effects of change. Note that the existence of multiple age cohorts is a defining characteristic of a cross-sectional design.

One of the reasons the authors proposed a narrow age group design is their claim that the association between two variables (e.g., X and Y) could be due to the effects of age on each variable (i.e., mean-induced covariation). However, this situation is similar to the classic 3rd-variable problem in correlational research that has an equally classic solution, namely, control the variation in the 3rd variable to determine whether the critical relation is spurious. In this case, age is the 3rd variable that can be controlled or partialled when examining the relation between X and Y.

The authors suggest that it may not be informative to partial age from cross-sectional data because the relation between the two variables of interest might change as a function of age. However, it is possible to examine whether the relations between two variables (e.g., X and Y) change as a function of age by using multiple regression procedures to test the interaction of age and X in predicting Y or the interaction of age and Y in predicting X. If the X-Y relation varies as a function of age, then one of the interactions should be significantly different from zero. If neither interaction is significant, then it is meaningful to examine the relation between X and Y at the average age which is accomplished by partialling age. Of course, the researcher needs to be sensitive to statistical power in interpreting the results of this type of analysis, but statistical power is a concern with any analysis and is not specific to this application.

The major issue raised by the authors concerns the practice of analyzing covariation of age-related variance in cross-sectional data. Although the authors are clearly critical of variance partitioning with cross-sectional data, we suggest that there are situations where this is both appropriate and informative.

For example, if a researcher is interested in predicting cross-sectional age-related effects on many variables from a smaller set of variables, then variance partitioning analytical procedures can be quite useful. The vast majority of studies investigating age-related effects on cognitive functioning are cross-sectional in nature. Hence it is valuable to know the dimensionality of the phenomenon or how many different parameters are needed to characterize the cross-sectional differences that have been observed in many different variables. Only if it were claimed that no cross-sectional research is relevant to theories of age-related influences would these analyses be irrelevant to theories of aging.

Another circumstance where these types of analyses can be informative is when a researcher is interested in determining the extent to which cross-sectional age-related effects on variable X are independent of cross-sectional age effects on variable Y. To the extent that cross-sectional age effects on the two variables are independent, then the magnitude of the relation should be very similar when the age-Y relation is examined after controlling the variation in X. Analyses of this type have been used to investigate the role of health status (as assessed by self-ratings) and amount of formal education (obtained from self reports) in age-related declines in cognitive functioning. If these factors play a causal role in the cross-sectional age trends on cognitive variables, one would expect that control of health or of education would substantially attenuate the age relations. However, this has not been the case because the results have revealed that the age relations on many cognitive variables are only slightly reduced when either health or education was statistically controlled [e.g., 8]. Results such as these have led to the inference that cross-sectional age trends in health and in education are largely independent of the cross-sectional age trends in many cognitive variables. Not all variables exhibit this pattern of independence, however, and a major challenge is to identify the degree to which age-related effects on different variables are independent of one another.

Finally, we agree that cross-sectional findings involving between-person comparisons provide a weak basis for inferences about age-related changes at the level of individuals. However, that is not the only interesting question in adult development, and as long as the researcher is sensitive to possible methodological artifacts, such as a mean-induced covariation of the type described by the authors, we believe that multivariate analyses of cross-sectional data can be valuable and informative.

Conclusions

The authors have initiated a very useful examination of analytical methods of developmental data, but we suggest that their examination was incomplete in several respects. In brief, we believe that the limitations of cross-sectional analyses were somewhat overstated and that relative weaknesses of current longitudinal studies were somewhat neglected. We also question the usefulness of the narrow age cohort design, because it seems to be based on assumptions of increased variability with age and/or relatively low retest correlations that are inconsistent with relevant evidence. All analytical methods rest on assumptions, and the strongest conclusions will come when results from different methods, each based on different sets of assumptions, converge. Although we do not agree with all of their arguments, we believe that Hofer and Sliwinski have made a valuable contribution by stimulating careful examination of the assumptions and limitations of different designs and analytical methods often associated with them.

References

1 Hofer SM, Sliwinski MJ: Understanding ageing: An evaluation of research designs for assessing the interdependence of ageing-related changes. Gerontology 2001;47:341–352.

2 McArdle JJ, Nesselroade JR: Using multivariate data to structure developmental change; in Cohen SH, Reese HW (eds): Life-Span Developmental Psychology: Methodological Contributions. Hillsdale, Erlbaum, 1994, pp 223–267.

3 Nesselroade JR, Featherman DL: Establishing a reference frame against which to chart age-related changes; in Hardy MA (ed): Studying Aging and Social Change: Conceptual and Methodological Issues. Newbury Park, Sage, 1997, pp 191–205.

4 Wechsler D: Manual for the Wechsler Adult Intelligence Scale III. San Antonio, Psychological Corporation, 1997.

5 Schaie KW: Manual for the Schaie-Thurstone Adult Mental Abilities Test (STAMAT). Palo Alto, Consulting Psychologists Press, 1985.

6 Owens WA: Age and mental abilities: A second adult follow-up. J Educ Psychol 1966;57:311–325.

7 Deary IJ, Whalley LJ, Lemmon H, Crawford JR, Starr JM: The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. Intelligence 2000;28:49–55.

8 Salthouse TA: Mechanisms of Age-Cognition Relations in Adulthood. Hillsdale, Erlbaum, 1992.