NORMAL HUMAN AGING:

The Baltimore Longitudinal Study of Aging

NATHAN W. SHOCK

Richard C. Greulich
Reubin Andres
David Arenberg
Paul T. Costa, Jr.
Edward G. Lakatta
Jordan D. Tobin

U. S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health
National Institute on Aging
Gerontology Research Center
Baltimore City Hospitals

NIH Publication No. 84-2450
November, 1984

CHAPTER VI

Longitudinal Studies of Aging

INTRODUCTION

There is not much evidence to support the concept of a general "aging process" that controls both physiological and psychological aspects of aging in an individual. It seems probable that aging, at least as we know it in humans, is rather the result of the interplay of many specific characteristics than a single process that regulates physiological and psychological functions. Aging is thus a highly individual phenomenon that can be characterized only by repeated observations in the same individual. This chapter summarizes the results of the Baltimore Longitudinal Study of Aging (BLSA) longitudinal analyses that have been completed and, in most instances, published. It must be looked upon as a progress report illustrating some of the findings about aging that have stemmed from longitudinal analysis.

Very few subjects have been followed over the full 25 years' existence of the BLSA, since new subjects have been added continually. Long periods of observation have, for example, produced five or more observation points on some variables for 667 subjects. Nevertheless, because of high variability of many of the measurements in different subjects, some longitudinal analyses have not yielded estimates of age trends in individual subjects with a desirable degree of precision.

The studies that are based on longitudinal analysis fall into five categories, for each of which a general summary has been prepared. The categories include a) studies that illustrate different conceptual approaches; b) analysis of anthropometric data on body size and composition; c) studies of physiological performance (basal metabolism, kidney function, blood cholesterol, and cardiovascular performance); d) studies of cognitive performance (vigilance, problem solving, learning, visual memory, the relation of temperament to visual retention, and the relation of hypertension to intellectual performance); and e) studies of personality characteristics (stability of personality, adjustment to aging, somatic complaints and neuroticism, hypertension and coronary disease in relation to personality, and personality and the life course).

Because of the many details that must be considered in longitudinal data analyses, we have chosen to reproduce in full all the publications from the Gerontology Research Center (GRC) that have used longitudinal analysis (see Appendix).

CONCEPTUAL CONSIDERATIONS IN THE IDENTIFICATION OF AGE CHANGES

By their very nature, longitudinal assessments of individual age changes commence with the acquisition of cross-sectional observations. Many of the studies reported in this chapter have been reported as cross-sectional investigations in Chapter V.

As Chapter I has noted, inferences of age changes from cross-sectional data fail to take into account birth-cohort effects. An example of the false impression that may arise from cross-sectional analysis of age differences is provided by our own experience with an assessment of the effect of age on body height and weight. Changes in these characteristics were examined in individual subjects over an interval of two to eight years during which three to eight measurements were made *(Shock, 1972).* The linear regression of height and weight on age was calculated for each subject on whom at least three measurements were available. Mean values for the slope of the individual regressions of height and weight on age were calculated for each subject within each age decade.

The average age decrement in body height, based on individually determined slopes, was similar to that derived by cross-sectional analysis. The results of cross-sectional analysis of body weight, however, differed from those based on longitudinal analysis of observations made on individual subjects. Cross-sectional analysis suggested a gradual decrement in body weight beginning with the 25-34 age group and continuing to age 80. Longitudinal analysis, on the other hand, demonstrated that, on the average, subjects 50 years of age or younger were gaining weight, while subjects 55 years and older were losing.

Early attempts to achieve an accurate longitudinal characterization of an individual's change in performance proved premature, since estimates of regressions on age based on only three observation points were highly variable and unreliable. Furthermore, practice or first-visit effects in some of the physiological variables greatly distorted the regressions on age when these first observations were included. For example, although three sessions were given to each subject in order to train him in performing the maximum breathing test *(Norris et al., 1964),* average values obtained on the second visit, 12 to 24 months after the first, were systematically higher than first-visit values, despite the fact that the average values fell with age and that values obtained on the third and subsequent visits decreased progressively. If the low initial value was used in the calculation of the age-regression coefficient, the resultant slopes did not accurately portray the changes in performance. None of the other pulmonary-function tests showed this practice effect.

The design of a longitudinal study requires the calculation or estimation of a number of statistical variables for each test. An extensive series of measurements of a variable serves as a data bank that can be used for determination of the sample size, the number of tests, the intervals between tests, and thus the total duration, required to characterize an individual's age changes in that variable with a specified reliability *(Schlesselman, 1973a,b).* These calculations require preliminary data to provide an estimate, based on cross-sectional data, of the average regression of the variable on age and of the total variance shown by the variable over the adult age span.

Schlesselman developed formulae and tables offering various combinations of study duration and number of measurements that yield the calculated value of "omega," a statistic calculated for each variable from estimated values of the slope of its regression on age, the standard error of estimate, and the selected degree of confidence in the computed individual slopes. For example, for an omega value of 0.20, three tests of kidney function distributed over a seven-year period provide as precise an estimate of the slope of individual age regressions as 30 observations made in three years. For a study of fixed duration, it is thus possible to establish the precision of the determination of the age regression that follows specified numbers of testings.

The longitudinal study of renal function *(Rowe et al., 1976b)* was based on the calculation of the regression of creatinine clearance on age in individual subjects. Although the *average* values for the individual slopes agreed surprisingly well with the average regression on age based on cross-sectional data, the study concluded that in order to characterize accurately the rate of age changes in creatinine clearance *in individuals,* serial observations over a period of at least 18 years may be required.

The study by *V.K. Elahi et al. (1983),* based on seven-day dietary diaries, illustrates the attempt to separate the effects of period and birth cohort from those of age. The nutrients considered were calories, protein, carbohydrate, fat, saturated fatty acids (SFA), polyunsaturated fatty acids (PFA), and cholesterol. Of the 12 indices derived from these variables, only three failed to show age effects: the percentage of calories derived from protein, intake of PFA, and P/S ratio. With the exception of carbohydrate, which rose, the intake of all of these nutrients declined with age. Birth-cohort effects were observed for none.

That period effects were also present over the duration of the study (1961–1975) was indicated by the systematic decrease in the absolute level of all the nutrients with the exception of PFA, which increased. The largest and most consistent period effects were an increase in the ratio of PFA to SFA and a decrease in cholesterol intake.

Douglas and Arenberg (1978) also attempted to separate age changes from cohort and period differences exhibited in responses to the Guilford-Zimmerman Temperament Survey (GZTS). Three hundred thirty-six subjects were measured on two occasions separated by an interval of seven years. To make possible cross-sequential and time-sequential comparisons, another 310 subjects were also tested for the first time when the original group were retested. The longitudinal analysis showed a significant age decline in Masculinity scores at all ages. The magnitude of the change increased with age. In contrast, General Activity scores declined only after age 50. Other scales (Thoughtfulness, Personal Relations, and Friendliness) showed declines from the first to the second administrations, but the magnitude of the change was not systematically related to age. The declines in General Activity and in Masculinity appeared to be age changes; the declines in Thoughtfulness and Personal Relations appeared to be related to changes specific to the period spanned by the two times of measurement as indicated by the differences within age groups in the time-sequential analysis; and the declines in Friendliness appeared to be attributable to period drift over a long time. In addition, later-born cohorts were lower in Restraint and higher in Ascendance than early-born cohorts.

Arenberg (1982a) devised a modification of the sequential strategies to explore further non-aging effects suggested by previous analyses of the Benton Visual Retention Test. Regression analysis, in which calendar time is a continuous variable, was used in an attempt to assess the magnitude of a period effect and to determine whether that effect could account for the within-cohort declines in performance. In four of the five oldest birth cohorts, declines were too large to be accounted for by the small period effects found for all age groups. Some portion of those declines was thus interpreted as representing the effect of aging.

BODY SIZE AND COMPOSITION

1. Fat Redistribution and Changing Body Dimensions

Cross-sectional population studies, as well as casual observation, indicate that the size, shape, and proportions of the human body change throughout adult life. Although all major tissues of the body change considerably with age, the changes in external size and body proportions occur because of changes in the relations among fat, lean tissue, and bone. The effect is most clearly evident in age-changes in subcutaneous fat distribution.

In a study that used both longitudinal and cross-sectional analysis, *Borkan and Norris (1977)* measured changes in subcutaneous fat distribution in relation to changes in other body compartments. The investigators compared various diameters and subcutaneous fat thickness to determine the role of fat, in relation to lean tissue and bone, in altering external dimensions.

The study population consisted of 699 BLSA participants aged 20 to 92 years. A total of 971 soft-tissue radiographs of these individuals, taken between 1958 and 1973, was available for analysis. Each radiograph contains views of seven body sites on the trunk and limbs. Because radiographic differentiation between skin and fat is difficult, measurements designated "fat" customarily include skin. A precision caliper was used to measure fat on radiographs of the trunk (at bony landmarks such as the top of the greater trochanter) and of the calf and forearm (at the widest parts). Reproducibility was tested by remeasuring a sample of 20 films; the average correlation between the paired measurements over the various sites was 0.95.

Other data used included body circumferences at sites corresponding to each radiographic fat measurement, height, weight, age, and biochemically assessed extracellular water and total body water. Values for extracellular and total body water were used to calculate the percentage of body fat. Total body water and extracellular water were assessed from the dilution curves for the distribution of standard intravenous doses of antipyrine and sodium thiocyanate. Extracellular water was subtracted from total body water to derive an estimate of intracellular water in the body.

Two separate data sets, one cross-sectional and one longitudinal, were used in the analysis. The cross-sectional sample, consisting of first-visit data for 699 individuals, was divided into ten-year age groups between the ages of 25 and 84. The longitudinal sample consisted of the first two of three radiographs of 234 participants (average interval between first and second films was 6.8 yr). Increments per year were calculated for all variables for each individual in the longitudinal sample. Each individual was placed in one of the ten-year categories of the cross-sectional sample based on his average age between the two test visits. Since third and fourth radiographs were available for very few individuals, they were not included in the longitudinal analysis.

Cross-sectional analysis showed that average values for both height and weight decline steadily after age 30. Total body water as a percentage of body weight declines in younger adults but becomes constant after age 55. Although the proportion of fat in relation to body weight increases with age, the calculated weight of fat changes little through the adult life span. The conclusion that the decline in body weight with age is due to loss of fat-free tissue is further supported by an age-dependent decline in percentage of intracellular water. The data on various body sites also gave evidence of a decline in fat-free tissue with age. Fat-thickness measurements of the trunk, particularly at the ilium and lower thorax, had the highest correlations with body weight. The marked reduction in forearm diameter after age 60 is due not to changes in fat thickness but to a loss of muscle tissue (Fig. V.5).

Longitudinal analysis showed weight increments in the three youngest age groups (25–34, 35–44, and 45–54 yr) and decrements between the ages of 55 and 74. The oldest individuals showed a marked increase in weight, possibly as a result of differential survival that results in unusually good health in very old individuals. Height decrements accelerated with age. Among measures of lean tissue, calculated fat-free weight declined in all age groups except the oldest. Body-fat increments, which were highly positive in the first age group, decreased with age, and the oldest group showed a marked decline. Decrements were observed in measures of arm and calf fat in all age groups. In the middle trunk, fat lateral to the lowest rib showed a decline through adulthood. In contrast to the cross-sectional data, which showed intervals of increase and decrease in fat lateral to the anterior-superior iliac spine, longitudinal data showed positive increments in the groups for which sufficient data were available (all groups except the youngest and the oldest). There were too few measurements at the greater trochanter to allow calculation of change.

The longitudinal data reveal that sites in close anatomical proximity behave differently during adulthood. While fat in the extremities remains relatively stable in the cross-sectional data, the longitudinal data show that it declines through adulthood, the middle trunk undergoing a net loss and the lower trunk a gain in fat. This shows that subcutaneous fat is in a dynamic state throughout adult life. The change is most striking in the longitudinal data, which show that subcutaneous fat in some sites may actually decline during the years in which overall body weight and fat increase.

These findings suggest that, at least during old age, when all subcutaneous fat in most sites is either declining or unchanged in thickness, total fat remains the same. The implication is that internal fat is accumulating while subcutaneous fat is decreasing with aging.

2. Height and Weight

Studies of the loss of height and weight with age have been described above, under "Conceptual Considerations in the Identification of Age Changes." It was found that height showed a decrement with age, both in age groups and in individuals. On the other hand, calculation of age regression of body weight in individual subjects showed that for the decades 25–34, 35–44, and 45–54 the regressions on age were positive even though the average regression calculated from cross-sectional data was negative.

3. Joint Degeneration in Osteoarthritis of the Hand

Longitudinal anthropometric measurements based on radiographs were used by *Plato and Norris (1979b)* in their study of the rate of change in joint degeneration in osteoarthritis of the hand. The authors studied 478 BLSA participants between 21 and 97 years old who were grouped in four categories according to their ages at the time of their first x-ray. Group A included 107 individuals below the age of 40 ($\bar{x}$ age = 33.2 yr); group B 175 participants aged 40-54 ($\bar{x}$ age = 46.9 yr); group C 129 participants aged 55–69 ($\bar{x}$ age = 61.8 yr); and group D 67 individuals who were x-rayed for the first time when they were 70+ years old ($\bar{x}$ age = 74.1 yr).

The x-rays were graded according to published standards that recognize five grades of severity of osteoarthritis (grades 0 and 1 are considered normal, grade 4 severe). Osteoarthritis in the distal and proximal interphalangeal joints was evaluated separately by the DH, PH, and IH scoring system. The DH score represents the highest osteoarthritic grade among the five distal interphalangeal joints, the PH score the highest grade among the four proximal interphalangeal joints plus the metacarpo-phalangeal joint of the thumb. The IH is the highest osteoarthritic grade observed in any of the interphalangeal or metacarpophalangeal joints of the hand. All radiographs were evaluated by the same investigator to eliminate variability.

The results showed that joint degeneration from osteoarthritis is a relatively slow process. The maximum rate of degeneration was seen in the distal interphalangeal joints, where the average increase was about one grade per individual in the interval of 12 to 16 years between x-rays. The rate of degeneration in the proximal interphalangeal joints was much slower.

The progress of degeneration in distal interphalangeal joints of individuals (longitudinally evaluated) was very similar to that seen in an earlier cross-sectional study by the same authors (Plato and Norris, 1979a).

The rate of change in the osteoarthritic grade of individual hands agreed closely with that of their distal interphalangeal joints. This further supports the conclusion of the authors' earlier cross-sectional study (Plato and Norris, 1979a) that what has been referred to as an osteoarthritic grade of the hand of an individual may actually be the highest grade among the distal interphalangeal joints.

PHYSIOLOGICAL PERFORMANCE

1. Basal Metabolism

Basal metabolic rate (BMR) has traditionally been expressed in relation to body-surface area, which is calculated from height and weight by a standard formula. The use of body size as a reference, however, assumes that all tissues contribute equally to body metabolism. In reality, adipose, bone, and connective tissues have very low oxygen demands and contribute very little in comparison with other tissues to the bulk of the whole-body basal oxygen consumption. Early in this century, "active tissue mass"—the estimated weight of tissues actively consuming oxygen—was proposed as a more suitable reference for the BMR, but the difficulty of measuring it accurately has made the approach impractical.

A BLSA cross-sectional analysis of aging by Tzankoff and Norris (1977) confirmed results reported by Shock et al. (1963) showing that age-related decrements of whole-body basal oxygen consumption in adult men were largely attributable to concurrent decrements in the mass of metabolically active (creatinine-producing) skeletal muscle. The approach was advantageous in that it was not necessary to account for either the deposition or the loss of fat, and the questionable reference to body size was avoided. The authors later extended these findings in a longitudinal study of 355 BLSA participants in whom five or more paired determinations of whole-body oxygen consumption and muscle mass were obtained over a mean period of 10.7 years *(Tzankoff and Norris, 1978).*

Nonmuscle oxygen demand and muscle oxygen demand in the basal state were calculated from paired values of whole-body oxygen consumption and 24-hour creatinine excretion by the method of Tzankoff and Norris (1977). Each subject's data were represented by a slope obtained by least-squares linear regression of each variable on age. They were then grouped in age decades. The overall mean rate of change, –0.82 ml O_2/min per year, was similar to that calculated from cross-sectional data. Mean slopes of whole-body oxygen consumption, summarized by age-decade groups, were all negative, not very different from one another, and consistent with the cross-sectional trend. Nonmuscle oxygen consumption increased for the six older groups, in contrast with the lack of a cross-sectional age effect in skeletal muscle.

Among these subjects, 48 had died shortly after their last measurements ($\bar{x}$ interval = 1.9 yr), 46 of them from cancer or cardiovascular disease. As a group, the decedents had had significantly higher mean slopes for nonmuscle oxygen requirement than the average. In the absence of these diseases, muscle-mass decrement and the concomitant decrease in oxygen requirement accounted for age decrements in whole-body oxygen requirement. In the subjects who died, however, the overall decline had been slowed or even reversed by gradual increases in nonmuscle oxygen requirement during nearly the entire decade before death. These findings led the authors to qualify a conclusion of the earlier cross-sectional study that loss of muscle mass is responsible for all the decline in whole-body oxygen requirement with age. Men nearing terminal age were similar to the survivors in that they lost skeletal muscle at comparable rates, but they differed from the survivors in that their decline in muscle oxygen requirement was offset by gradual increases in nonmuscle oxygen requirement. The result was a lower rate of decline in their whole-body oxygen consumption. In longitudinal studies that include only two points, one early and one very late, this effect could be mistaken for a slower rate of aging.

2. Renal Function

All participants in the BLSA studied between July 1, 1961, and June 30, 1971, were included in a longitudinal and cross-sectional study of creatinine clearance *(Rowe et al., 1976b)*. More than 3300 creatinine clearances were obtained in 884 volunteers ranging in age from 17 through 96 years. On the basis of clinical examinations, subjects were placed in categories indicating the presence of specific diseases or medications that might alter the glomerular filtration rate. Only subjects who were not in these categories, a total of 548 individuals, were included in the data analysis.

Cross-sectional analysis showed a highly significant reduction in creatinine clearance with advancing age in healthy subjects. The data when grouped by decades suggested that creatinine clearance remained stable until age 34 and thereafter declined, the rate of decline increasing with each decade after age 65. In the cross-sectional analysis, however, this tendency toward an increasing rate of decline in the very old subjects proved not to be statistically significant.

The longitudinal analysis consisted of the calculation of the regression of creatinine clearance on age for each subject on whom five or more tests had been completed. The means of the individual slopes in the longitudinal analysis were similar to the mean differences in the cross-sectional analysis. Regression analysis of individual slopes by age indicated a minimally significant acceleration in the rate of decline in renal function with age after age 65. In view of the detailed screening of the subjects to eliminate those who might possibly have suffered from overt or occult renal disease, the authors concluded that the decrements represent true renal aging rather than the development of renal disease.

Shock et al. (1979) analyzed the longitudinal data from 398 BLSA subjects, aged 25 to 100 years, who were tested for 24-hour creatinine clearance five or more times over a period of ten years. They found that the correlation of individual regression slopes with age was highly significant, i.e., the rate of fall in creatinine clearance increased with age over the range of 25 to 90 years. Average slopes for 20-year age groups became increasingly negative, from –0.26 ml/min·1.73 m^2 per year for the 20–39 age group to –1.51 for subjects aged 80 to 100 years. This relation could not be demonstrated from cross-sectional data.

Although the standard error of individual regression coefficients based on as few as five measurements may be large, eight subjects were identified with statistically significant positive coefficients (i.e., their creatinine clearance improved with age). Longitudinal observations can obviously identify individual subjects who deviate markedly from the average pattern of age changes derived from cross-sectional data.

The same study yielded five or more observations on 59 subjects who died after the age of 55. Mean values for creatinine clearance and individual regression slopes for this group were compared with values based on a group of 106 subjects over the age of 55 who were still living after ten years. The mean ages of the two groups did not differ significantly. Although examination of death certificates showed that renal disease was not recorded as a cause of death for any of the 59 subjects, the individuals in this group had exhibited lower creatinine clearances and faster rates of decline in clearance than subjects of comparable age who were still alive.

The longitudinal observations illustrated clearly the individual differences that exist in the rates of change of these variables. Some individuals follow the pattern predicted from cross-sectional analyses (i.e., a gradual fall in clearance with age). However, some of the normal subjects showed a pattern of increased creatinine clearance with advancing age—at least to age 70. This is an important finding that bears further longitudinal exploration.

3. Serum Cholesterol

A decline in cardiovascular mortality in the past decade has been noted by several authors. The decline has coincided with public-health efforts to control such risk factors for coronary artery disease (CAD) as blood pressure, cigarette smoking, physical-activity levels, obesity, and serum cholesterol. Independent cross-sectional studies have provided evidence that serum-cholesterol levels have been dropping in recent years, and similar changes have been suggested by preliminary longitudinal studies. The trend to lower cholesterol levels was confirmed in a detailed study by *Hershcopf et al. (1982)* of longitudinal changes in serum-cholesterol levels in BLSA participants over a period of 14 years.

Serum cholesterol, height, and weight are among the variables that have been followed in the BLSA since its inception. In addition, dietary and exercise histories have been maintained in subgroups of this well-characterized population. It was thus possible to measure not only changes in serum cholesterol but also the effects on cholesterol levels of body weight, diet, and physical activity.

A total of 1011 males ranging in age from 17 to 102 had 5127 cholesterol determinations between 1958 and 1977. After data exclusions, 3088 cholesterol determinations in 783 participants were available to determine changes over the 14-year period from 1963 to 1977. As a measure of obesity, weight was corrected for height by use of the body-mass index defined as weight in kilograms divided by the square of the height in meters. Dietary intake was assessed from seven-day diaries kept by each participant under the guidance of nutritionists. Estimates of physical activity were obtained by interviews or questionnaires covering specific activities at home, work, or recreation, as well as variations in activities such as trips and seasonal sports.

Total daily energy expenditures were calculated for each subject by use of predetermined values for each activity (McGandy et al., 1966).

In the analyses of serum-cholesterol changes, the annual rate of change was computed as the slope of the regression line for each subject with three or more data points. In the analysis of secular change, each participant was represented by the mean of available determinations within each period. Simple linear correlations and regressions were used to assess the relation between individual performances.

Cross-sectional analysis revealed age differences in cholesterol data obtained between 1971 and 1977. Higher cholesterol values were found with increasing age from 25 to 64 years, but lower values were found in the 65–84-year-old group. When the data were examined as successive cross-sectional studies, serum-cholesterol values were fairly constant in the periods between 1964 and 1970 but dropped 6% in the periods between 1970 and 1972. Since then there has been little change. Because of the drop in serum cholesterol, the total span of the study was separated into two overlapping periods, Era 1 (1963 to 1971) and Era 2 (1969 to 1977). Within-age-group differences in serum cholesterol between Era 1 and Era 2 were found in all age decades studied.

Longitudinal analysis of serum-cholesterol values obtained within Era 1 showed that changes in cholesterol levels in individuals grouped by age followed the pattern seen in the cross-sectional analysis, i.e., cholesterol increased in the younger adult years and decreased in the later years. The drop in cholesterol levels between Era 1 and Era 2 seen in the comparisons within age groups was strikingly evident in the longitudinal analysis as well.

A shift in methodology was ruled out as an explanation of the period change in cholesterol values between Era 1 and Era 2, since re-analysis of stored frozen and lyophilized samples yielded values parallel to the original ones. When the effects of obesity, selected dietary constituents, and physical activity were examined as potential explanations, serum-cholesterol levels were not significantly correlated with levels of weight or body-mass index. Changes in weight, however, were significantly and positively correlated with changes in serum cholesterol. The fact that the study population did not as a whole experience a significant decline rules out weight changes as a source of the period drop in cholesterol levels. There were no significant correlations between the absolute value of any dietary variables examined and the absolute level of serum cholesterol. Although there were small but significant changes in most dietary constituents, only changes in caloric intake had significant positive correlation with changes in serum cholesterol. The small overall change in caloric intake, however, could explain less than 1 mg/dl of the average 11 mg/dl drop. There was no overall change in physical activity, and no significant correlations were found between either the level of or the change in physical activity and the level of or the change in serum cholesterol.

Since changes in fatness, caloric intake, and physical activity cannot fully explain the change in cholesterol levels between Era 1 and Era 2, other factors that must be involved should be examined in future studies.

4. Cardiovascular Function

Right-bundle branch block in apparently healthy men. Although the prevalence and incidence of several electrocardiographic (ECG) abnormalities clearly increase with age, the long-term prognostic implications are controversial primarily because the effects of aging and disease on the ECG have not been adequately separated. One such ECG abnormality is right-bundle branch block (RBBB). Most of the studies examining the long-term prognosis of this conduction defect are derived from hospital-based or military populations, neither of which is representative of the general community. Men from the BLSA provided a more reasonable population for such a study, and an age-matched control population made possible the separation of aging and disease processes.

Resting 12-lead ECGs identified 24 men with RBBB who had shown no evidence of cardiac disease on initial presentation. To determine their long-term cardiovascular prognosis, they were compared *(Fleg et al., 1983a)* with an age-matched control group; mean age on presentation with RBBB was 63 years. No differences in the prevalence of major coronary risk factors or obstructive lung disease were noted initially. Over the 8.4-year average follow-up period, the incidence of angina pectoris, myocardial infarction, cardiomegaly, congestive heart failure, high-grade heart block, or cardiac death did not differ between RBBB and control groups. Likewise, no differences in cardiothoracic ratio, mean blood pressure, maximal aerobic performance, or maximal exercise heart rate were found between groups on most recent examination. Left-axis deviation, however, occurred more commonly in RBBB subjects (46% vs. 15%, $p < .01$), and their mean resting heart rate was lower than that of controls. In addition, PR-interval prolongation of 40 msec or more over time was more frequent in men with RBBB than in controls (21% vs. 6%, $p < .03$). These findings suggest that RBBB in clinically healthy men is found primarily in older individuals and has no effect on long-term cardiovascular morbidity or mortality. The frequent association of RBBB with left-axis deviation, as well as the slower heart rate and the fact that the tendency for prolongation of atrioventricular conduction with time is greater in men with RBBB than in controls, suggests a primary disorder of cardiac conduction in the former.

Longitudinal cardiopulmonary chest x-ray changes in normal men. The standard chest x-ray is the most common radiographic procedure employed in medicine. Despite the ubiquity of the examination, most of the information on radiographic changes seen with advancing age has been derived from cross-sectional studies in populations not carefully screened for the presence of cardiopulmonary disease.

To determine the changes attributable to the aging process, we evaluated cardiovascular and pulmonary structures on two standard postero-anterior chest x-rays taken at least ten years apart ($\bar{x}$ interval = 16.9 yr) in 67 carefully screened healthy men *(Ensor et al., 1983)*. Only normotensive individuals with negative maximal-exercise ECGs and normal pulmonary-function tests were included in the survey.

The mean aortic knob diameter increased from 3.4 ± 0.6 cm to 3.8 ± 0.5 cm, enlarging in 78% of subjects. Although mean cardiothoracic ratio (CTR) increased from .405 ± .04 to .427 ± .04 overall, only 3% of men developed a CTR greater than .50, and none exceeded .51. Pulmonary abnormalities on initial chest x-ray consisted mainly of hyper-inflation (27%) and increased markings (19%), both of which doubled in prevalence during follow-up. Kerley B lines and enlarged pulmonary arteries, which were rare initially, increased four- to fivefold over time. Commonly accepted x-ray criteria suggested chronic obstructive lung disease in 21% of the final chest films, despite the absence of clinical or spirometric abnormalities. These data provide a framework by which chest radiographic changes with age can be interpreted. The most significant finding is that the frequently quoted value of .50 for the normal upper limit of CTR remains valid even in advanced age.

Cardiothoracic ratio. CTR—the ratio of the cardiac diameter (CD) to the thoracic diameter (TD) as determined from x-ray measurements—has been used to assess heart size. CTR values exceeding 50% have been regarded as evidence of abnormal enlargement of the heart. Cross-sectional studies on other populations have reported values exceeding 50% in 10% of older men and in 20% of older women. Although these studies excluded individuals with overt heart disease, rigorous screening for the presence of heart disease was not conducted.

The availability of standard six-foot postero-anterior chest x-rays on BLSA subjects made it possible to examine the effect of age on this ratio *(Potter et al., 1982).* The study used 1124 chest x-rays from 243 men, aged 20 to 95 years, who had been followed for an average of 12.3 years (range of follow-up was 8 to 21 yr). Criteria for inclusion in the study were a) three or more technically adequate x-rays; b) eight or more years between first and last x-rays; c) age at time of last x-ray under 50 or over 60 years. Measurements of CD and TD, which were made blind to subject age and date of x-ray, were reproducible to 0.1 cm.

In the longitudinal analysis the rate of change in CD and TD was computed as the slope of the age-regression line for each subject estimated from three or more data points.

Cross-sectional analysis indicated a slight but significant rise in average values of CD up to age 75, whereas TD showed no change between the ages of 32–50 and 60–69 years but fell after age 70. The ratio CD/TD increased progressively from the 32–50-year group through the 80–95-year group.

Longitudinal analysis indicated that changes in CD, TD, and CTR in individuals generally followed the pattern suggested by the cross-sectional analysis: increasing CD and CTR after 50 years of age, and decreasing TD in the oldest groups.

Patients with diagnosed heart disease had higher values for CD and CD/TD ratio at all ages, but TD did not differ significantly from values for the normal controls.

For each of the 49 deceased subjects a surviving subject was selected who had been the same age at the time of first testing and whose cardiovascular status had been the same. Comparisons of mean slopes of CD and CTR in deceased individuals with those in survivors showed that individuals who had died with diagnosed heart disease had had significantly higher slopes than survivors. However, individuals who died without evidence of heart disease had slopes that did not differ from those of matched survivors.

COGNITIVE PERFORMANCE

1. Vigilance

"Vigilance" refers to a central process or state reflecting the individual's readiness to respond to specific infrequent and unpredictable stimuli. The primary index of vigilance is the proportion of the infrequent signals that is detected. The test consists of detecting and reporting instances when the pointer attached to a large blank circular clock face moves two intervals instead of the usual single interval. The full circle includes 100 intervals. In the course of the one-hour test, 23 double jumps occur at irregular intervals. A cross-sectional study (Surwillo and Quilter, 1964) showed that fewer signals were detected by an old than by a young group. The study also showed that middle-aged individuals had the fastest reaction time to infrequent signals, outperforming both old and young. The reduced proportion of stimuli detected and the slower reaction times demonstrated by elderly subjects suggest a reduction in the reactivity of the central nervous system late in life.

Thirty-three subjects from the original study were retested, by the same procedures, after an interval of 18 years *(Quilter et al., 1983)*. In addition to the cross-sectional comparisons of the second-time data and the longitudinal analysis of change, measures of men who were 51 to 69 years of age at the time of first testing were compared with measures of men who had reached ages of 51 to 69 by the time of their second testing.

The cross-sectional analysis of the number of double jumps that were detected showed a significant age effect. Subjects 70 to 88 years old detected 58% of the double stimuli, while 51–69-year-olds detected 71%.

A significant longitudinal effect occurred in those who were 70 to 88 years old in 1980–1981; they detected fewer targets (58%) than they had in 1962–1964, while those 51 to 69 years old in 1980–1981 detected one percent more targets than in 1962–1964. Longitudinal analysis thus confirmed the cross-sectional findings that young and middle-aged groups perform equally well on vigilance, while individuals approaching the age of 70 show a decline.

The longitudinal results also showed that in 1980–1981 the 51–69-year-olds had faster reaction times than 18 years previously, while the 70–88-year-olds were slower. The faster reaction times in the middle-aged group confirm the original cross-sectional study. It should be noted that the men who were 51 to 69 in 1980–1981 were similar to the men 51 to 69 in 1962–1964 in both proportion of signals detected and reaction time. Instead of increasing progressively over the entire age span, reaction time to infrequent signals increases primarily after age 70.

The results of the combined cross-sectional and longitudinal analysis and comparison support the conclusion that differences in vigilance, although they are not manifest in performance until after the age of 70, are true aging effects and are not due to cohort differences.

2. Problem Solving

Few studies have measured changes in problem-solving performance with age. Studies published before 1974 were cross-sectional. They showed that older adults had substantial difficulty in solving logical problems, but that some tests, which measure the ability to solve concept-type problems, did not consistently reveal differences between young and old adults. *Arenberg (1974)* was the first to publish the results of a longitudinal study of the effects of aging on problem-solving performance. The aims of the study were to determine whether performance on logical problems changes with age and whether old subjects decline more than younger subjects. A sample of 300 BLSA participants ranging from 24 to 87 years of age was studied.

All problems were administered during a single half-day session for each subject. After instructions, which included a sample problem and a practice problem, the subjects were given the main set consisting of up to three problems of increasing complexity. The second session was scheduled a minimum of six years after the first ($\bar{x}$ interval = 6.7 yr). Of the 300 initial participants, 224 were available to take part in the second session.

Each input prior to achievement of a solution sequence could be evaluated as potentially informative or uninformative, depending on the pool of information available to the subject at that point. The number of uninformative inputs was the primary dependent measure. These inputs were further classified as overtly redundant, directly inferable, or indirectly inferable, in order to elucidate the age deficits in reasoning that were hypothesized.

Performance on the tests was analyzed both cross-sectionally and longitudinally. Cross-sectional results for all problems showed that, with increasing age, the proportion of subjects who achieved solutions decreased and uninformative inputs increased. Although the pattern of age differences varied somewhat, all three problems showed a decrease in proportion of successful solutions and a decline in effectiveness of solutions with age.

Longitudinal measures of change in both Problem 1 (n = 193) and Problem 2 (n = 166) showed a mean decline in performance only for the group that was over 70 at the first testing. For that group, the number of uninformative inputs increased primarily because of the greater number of overtly redundant inputs, even though memory demands were minimized and the entire written record of input-outcome events was always available to the subject. Thus the oldest subjects repeated inputs frequently, and the increase in repeated inputs accounted for a major portion of the decline in the effectiveness of their reasoning. These subjects also made attempts at solution sequences that were variations of earlier attempts but were ineffective because they did not make use of the information available from previous input-outcome events.

A follow-up analysis of mortality and reasoning performance for the 49 men in the initial group over 70 showed a relation between survival and success in solving Problem 1. Of the 36 men ($\bar{x}$ age = 74.3 yr) who solved the problem, six had died; of the 13 ($\bar{x}$ age = 76.0 yr) who failed to solve the problem, six had died. There had been no difference in vocabulary scores between the survivors and nonsurvivors.

The true age changes in this study were probably reduced as a result of several positive biases: Individuals who returned for the second test were a select subsample of the original group; most subjects who reached Problems 2 and 3 were superior performers; and only those who successfully solved a particular problem in both sessions could be included in the analysis of change in reasoning effectiveness. Although a substantial mean decline was found for the group over 70, even beyond age 75 the performance of some subjects had not declined from that observed six years earlier.

In 1967, a study of concept problem solving was initiated in the BLSA. Cross-sectional and longitudinal data collected from 1967 through 1979 were analyzed *(Arenberg, 1982b)*. The problems were designed to measure not only whether correct solutions were reached but how effectively each was solved.

There were twelve problems: four one-attribute (simple), four two-attribute with high initial information, and four two-attribute with low initial information. Each problem comprised four dimensions, and each dimension had two attributes. To make the problems easier to understand, they were described as attempts to identify the "poisoned" food or foods in a problem. In the "language" of poisoned foods, the attributes are foods, the concepts (solutions) are poisoned foods, subjects select meals, and each meal is designated as having caused an imaginary diner to "Die" (positive) or "Live" (negative), depending on whether it included the poisoned food or foods.

To minimize the memory component of the task, subjects were required to write each selection and its designation (provided by the experimenter). Each concept rule (simple, conjunctive, disjunctive) was explained thoroughly and was prominently displayed during the problem. The entire procedure was subject-paced.

During the first 13 years of this study, complete first-time data were collected for 751 men ranging in age from 20 to 87 years. Cross-sectional analysis of the number of correct solutions found a monotonic decrease in means from 10.4 for the men in their 20s to 5.4 for the men in their 80s.

By the end of 1979, 327 men had returned for a repetition of all 12 problems at least six years after the first session ($\bar{x}$ interval = 6.9 yr). Mean improvements were found for the younger age groups, but the men who had been in their 60s or 70s when first measured showed small mean declines in the number of correct solutions.

Estimates of age changes were calculated within birth cohorts by regression analysis of number correct (first session) and calendar time (date of session). These within-cohort estimates showed no change for the six ten-year cohorts born between 1897 and 1956; but for the earliest-born cohort (1887–1896) the estimate was –0.25 problems per year. The estimates of age change thus confirm the direct (longitudinal) measures of change, i.e., declines in performance only late in life.

Results using the effectiveness measures were similar to the results for number of correct solutions.

3. Learning

BLSA longitudinal data on age changes in learning ability, although largely preliminary, constitute perhaps the most extensive body of information on the subject. The data on performance in paired-associate and serial learning at two different anticipation intervals for each task were described by Arenberg and Robertson-Tchabo (1977) and *Arenberg (1983)*. Arenberg (1967b) had previously shown by cross-sectional analysis that small age differences existed before age 60 and larger age differences thereafter, particularly when only a short anticipation interval was used.

The BLSA data on age changes in learning are of two kinds: conventional longitudinal data based on repeated measures at least six years apart, and independent subsamples of the same birth cohorts measured during two separate periods. Six birth cohorts, based on dates of birth between 1885 and 1932, were studied in the early sample, which was measured between 1960 and 1964. The second subsample of each cohort comprised men who entered the study after the first sample, but data for only five of the six cohorts were available for the later sample, which was measured between 1968 and 1974. Comparison of independent subsamples born in the same periods (cross-sequential analysis) is advantageous in that it eliminates retest effects and the effects of noncomparability of learning material, but is disadvantageous in that age becomes a between-subjects variable and any precision attributable to the within-subjects design is lost. The advantage of the design lies in the argument that if repeated-measures data are corroborated by intra-cohort differences, the evidence for age changes is compelling.

The mean period between tests was 6.8 years for the 102 men assigned to the short anticipation interval (fast-pace condition) who had two valid measures on the paired-associate task. The youngest group (initially 30–38 yr) showed a small mean decrease in errors. The other two groups of men initially less than 54 years of age showed moderate mean increases. The three oldest groups showed larger mean increases in errors, the largest in the oldest group (initially 69–76 yr). Cross-sequential analysis found the smallest increases within birth cohorts in the three latest-born cohorts (1909 to 1932) and the largest increases in the two earliest-born cohorts (1893 to 1908), i.e., the oldest groups. Mean raw scores on the Vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS) were similar in the early (1960–1964) and the later (1968–1974) samples.

The mean interval between administrations was 6.7 years for the 111 men assigned to the slower pace who had two valid measures on paired-associate learning. The three youngest groups (initially 30–55 yr) showed the smallest mean declines, the two oldest (initially 61–74 yr) the largest. The smallest difference within birth cohorts was found in the youngest cohort (1925–1932). Large differences favoring the younger subsamples within birth cohorts were found for the two earliest-born cohorts (1893–1908), the oldest groups. Again, the mean WAIS Vocabulary raw scores for the early and later samples were quite similar.

For the 104 men who had two valid measures of serial learning at the faster pace, errors decreased, although the two oldest groups (initially 63–76 yr) showed mean increases. Comparison of early and late subsamples within birth cohorts showed the largest increases in errors in the two earliest-born (oldest) cohorts.

For the 102 men who had two valid serial-learning measures at the slow pace, errors increased. Again the largest decline in performance was found in the oldest group (initially 68–76 yr). All birth cohorts showed age differences; mean errors were consistently greater for the older subsample within each cohort. Although no clear pattern emerged between the magnitude of the mean age differences and the birth periods of the cohorts, the earliest-born cohort showed the largest age difference.

Thus, for both paired-associate and serial learning, age decrements in conventional longitudinal comparisons were found in the oldest age groups. In addition, comparisons within birth cohorts showed that in early-born cohorts older subsamples performed less effectively than younger subsamples. The evidence indicates a deficit in verbal learning in the later years of life.

4. Visual Memory

In a longitudinal study of age changes in nonverbal memory published by *Arenberg (1978),* performance consisted of reproducing geometric designs from memory. In addition to cross-sectional data, the study included age changes based on conventional longitudinal data as well as on comparisons of independent samples from the same birth cohort measured at different times.

The subjects included every participant in the BLSA for whom a measurable initial performance during the period from 1960 to 1973 was available, a total of 857 men whose age range was 18 to 102 years. The total sample was divided into early, middle, and late subsamples based on the date of the initial measurement. The primary cross-sectional analysis was based on the early sample, the middle and late samples serving as replications. The primary longitudinal analysis was based on the re-examination of 268 men from the early sample, and a small longitudinal replication was based on the re-examination of 82 men from the middle sample. Age changes were also estimated in subsamples from the early and late samples on the basis of comparisons within birth cohorts (age differences between subgroups born at the same time but measured at different times).

All participants were given Form C of the Benton Revised Visual Retention Test, Administration A, usually on their first visit. Form E was administered on a subsequent visit at least six years after the first test. Errors were scored independently by two psychologists according to the test manual. At the same session the WAIS Vocabulary subtest was also administered.

For cross-sectional analysis, subjects from the early, middle, and late samples were classified in seven age groups. In all three samples the mean number of errors increased with age.

Although the magnitude of change increased with age in both longitudinal samples, statistical significance was not attained in the middle sample. Assessment of cumulative distributions of age changes for five age groups in the early sample demonstrated a striking difference between the oldest group and the other four: Fewer than half the men in each of the four youngest groups declined in performance, but more than two thirds of the group initially over 70 declined. The large mean increase in errors for the men in their 70s was attributable not to a few men who changed drastically but to a shift of the entire distribution in the direction of larger decrements in performance. Similarly, estimates of age changes based on comparisons within birth cohorts indicated age differences in all cohorts, the largest occurring in the earliest born.

In summary, increases in errors were small for the young groups, moderate for men in their 50s and 60s, and substantial for men over 70. WAIS Vocabulary measures for these same samples showed small cross-sectional differences favoring the older men, no overall longitudinal change but small deficits for the older participants, and small declines in estimates of age changes based on comparisons of independent samples born during the same period. In general, the results indicate age decrements in memory-for-design in men late in life, but small or no decrements in vocabulary for the same samples. No correlation was found in either longitudinal sample between change in memory performance and change in vocabulary score.

The picture is quite clear for memory for designs. Cross-sectional, longitudinal, cross-sequential, and time-sequential analyses provide consistent and strong evidence of age decrements late in life.

5. Temperament as a Predictor of Change in Visual Retention

Individual differences raise the important question whether predictors of an individual's decline in memory performance can be identified. In this case it might be possible to devise strategies for intervention in cognitive problems in the elderly. A preliminary BLSA study was therefore undertaken to examine temperamental traits as possible predictors of change in visual-memory performance *(Robertson-Tchabo et al., 1979).*

The 52 subjects in the analysis included every BLSA participant 70 to 79 years of age at the time of the initial administration of the Benton Test for whom there was a second performance measure, as well as both initial and second administrations of the GZTS, which assesses General Activity, Restraint, Ascendance, Sociability, Emotional Stability, Objectivity, Friendliness, Thoughtfulness, Personal Relations, and Masculinity. Multiple regression analyses were carried out with the ten scales of the GZTS as independent variables and residualized change in Benton total errors as the dependent measure. (Change in Benton total errors represents the part of the second measure that is uncorrelated with or independent of the initial level of total errors.)

The results of the multiple regression analyses, together with the simple correlations, showed that the variables Objectivity, General Activity, Restraint, Masculinity, and Ascendance were important correlates and predictors of Benton performance.

Although the consistency with which some of the scales emerged as important predictors in this exploratory study is encouraging, the small sample size and the need to control for the effects of other variables, such as health factors, clearly require replication of the findings. Furthermore, the observed relation between personality traits and visual-memory performance in this study permits no conclusions about cause and effect or the underlying mechanisms.

6. Blood Pressure and Intelligence

Although the prevalence of elevated blood pressure in older people has led to suggestions that it may be partly responsible for decline in cognitive function with age, the literature on the subject is contradictory (Wilkie and Eisdorfer, 1971). Uncertainty arises in part from the complexity of the variables under investigation. Hypertension is not a single clinical entity; there are many kinds, and this diversity may influence any relation with intellectual performance. Most research has been conducted in clinic patients, who are often under medication that could influence intellectual performance. Different kinds of cognitive tests show different age effects: Visual memory, as the preceding section has shown, decreases significantly after age 70, while vocabulary does not in general decline and may even increase. Other frequent sources of inconsistency are small sample size (which makes it difficult to measure hypertensive effects that are small in comparison with the range of individual differences in intelligence) and faulty sampling procedures.

Longitudinal studies are needed to assess the long-term effects of increased blood pressure on intellectual performance, since serial observations in the same individuals over time permit the determination of the relations between antecedents and consequences. A BLSA longitudinal study by *Costa and Shock (1980)* used the Army Alpha test, a general intelligence test that includes some components of intelligence that change with age and others that do not. The study had the further advantages of relatively large sample size and availability of information on health status and medication of the subjects.

The Army Alpha has two parallel forms, A and B, each of which consists of eight subtests: Following Directions, Arithmetic Problems, Practical Judgment, Synonym-Antonym, Disarranged Sentences, Number Series Completion, Analogies, and General Information.

All subjects were given Form A of the test on their first visit. Thereafter, they were treated differently according to age. After an interval of four to eight years, those under 70 were given Form B, which is parallel to Form A, to avoid practice effects. Subjects aged 70+ continued to be administered Form A at each annual visit. This provided two sets of longitudinal data: retest data after four to eight years on Form B for 350 men initially aged 17 to 65 years, and data from six administrations of Form A spanning five to 11 years for 51 men initially aged 66 to 84 years.

The two-point data for men under 65 were extensively analyzed. In the first set of analyses, subjects were classified in three age groups, 20–39, 40–49, and 50–65 years, and were cross-classified as low, average, or high in blood pressure on the basis of three sources of data: average basal systolic and diastolic blood pressure at the first visit (designated Basal), average of casual sitting pressures recorded from right and left arms at the first visit (designated First Casual), and average right and left casual sitting pressures at the second visit (designated Second Casual). In the Basal classification, subjects were divided into low, middle, and high thirds of the distribution. In the First and Second Casual classifications, subjects were classed as "low" if their systolic pressure was below 120 or their diastolic pressure was below 80, "high" if their systolic pressure was above 140 or their diastolic pressure above 90. The remaining subjects were classified in a "middle" group.

In the first set of analyses, scores on each of the eight timed subtests and the total Alpha score were examined in relation to systolic and diastolic blood pressures under Basal, First Casual, and Second Casual classifications, a total of six definitions yielding up to 18 groups. Age group and time were also used as classifying variables. There was no evidence in the data for any effect of hypertension (within the existing blood-pressure ranges) on changes in Army-Alpha performance among men under 65. Similar analyses were conducted on data from the Army Alpha test given under untimed conditions. Although three of 48 main effects and five of 144 interaction terms were significant, the investigators concluded that this finding could not be considered evidence of any effect of blood-pressure level on changes in performance.

Since this was an unscreened population in which the influence of chronic illness or drugs could have obscured a true effect, further analysis was undertaken after exclusion of such individuals. Additional analyses were performed in which subjects taking medication or diagnosed as having a serious chronic illness, other than hypertension, at any point during their first five visits to the GRC were excluded. These exceptionally stringent criteria left a smaller sample of 117 men, few of whom were in the "high" hypertension group. Since there was no evidence of an interaction between age and hypertension, subjects were collapsed across age groups, and age was used as a covariate in the analysis.

When timed subtests and totals were treated as dependent variables, the "high"-blood-pressure group performed least well on the Following Directions and Number Series Completion subtests. Although the effects were weak and inconsistently replicated, they gave at least mixed evidence that higher blood pressure may be detrimental to some kinds of cognitive performance among otherwise healthy men. Somewhat stronger evidence was found when the untimed condition of the Alpha was analyzed. Replicated main effects were seen for Arithmetic Problems, Synonym-Antonym, Disarranged Sentences, and Number Series Completion, as well as for the total untimed score. The "high" group showed the poorest performance in all these measures. None of the effects, however, was strong, and no finding was statistically significant in all six possible replications. In these data no support could be found for the prevalent notion that elevated blood pressure or hypertension causes cognitive loss or decline with aging.

In an analysis of men over 70, all subjects were classified as a) free of drugs and major chronic diseases other than hypertension, or b) "other." Since no significant differences were found in any of the eight subtests or total scores, it was possible to collapse the two categories in subsequent analyses. Repeated-measures analyses of variance were performed on the Alpha subtests and total score, using six levels of the repeated factor and classifying subjects by age and blood pressure. On only one of the subtests, Practical Judgment, was there a significant main effect for blood-pressure classification, and then only for systolic, on which hypertensive men were somewhat lower performers than normotensive men. It might be expected, from the literature, that normotensives would remain relatively constant or show only a modest decline over the six administrations, while hypertensives would show a more progressive and precipitous decline. On the contrary, the study suggests that elevated blood-pressure levels have little if any effect on intellectual performance for men ranging from 17 to 88 years of age. The conclusion must be interpreted cautiously because of several limitations, which include the modest elevation of blood pressure and the absence of cases of severe hypertension.

PERSONALITY CHARACTERISTICS

1. Age and Stability of Personality

The question whether personality remains stable or changes over the adult life span has become a major focus of attention in the last decade, precisely the period during which the field of personality psychology was undergoing fundamental revaluation. Neugarten (1977) pointed out that the attacks of social learning theorists (Mischel, 1968), humanistic psychologists (Maddi, 1976), and psychometrists (Fiske, 1974) had left both personality psychology and its subfield, the study of personality and aging, in disarray. The dominant theories of adult development, outgrowths of psychoanalytic approaches, had not proven their usefulness. The critiques of Schaie (1965) and Baltes (1968), who demonstrated that cross-sectional studies confound age changes with birth-cohort differences, had brought into question most of the evidence on the relation of personality to aging.

It might have been expected that the resurgence of the field would require the development of a new conceptual model. Instead, the findings of stability argued for one of the oldest models of personality: trait theory. Previous trait theory (Eysenck, 1960; Cattell, 1973; Guilford et al., 1976), despite its discouraging theoretical and technical differences, had converged empirically on two basic dimensions of temperament, which may be labeled Neuroticism and Extraversion. A new trait model, the NEO schema (Fig. 1) (Costa and McCrae, 1980c), which added the dimension of Openness to Experience, provided support for earlier BLSA studies that had consistently found that adult personality remains stable with advancing age.

Statistical definitions of stability and change. The term "stability" has three different and largely independent meanings. A trait may be considered stable for a group if the mean level of the trait in a group of individuals is constant over time. This will occur if all individuals remain at the same level, or if equal numbers increase and decrease over the interval. Analysis of variance on age groups for repeated administrations is the usual way to determine whether significant changes in mean level have occurred. Another approach compares a person's standing on a trait in relation to that of others in the group. Stability of this kind is usually assessed by a test-retest coefficient, which will be higher if individuals maintain the same relative ordering on the trait over time, regardless of the level of the trait. If some developmental process leads to a uniform increase or decrease of a variable over time, it will have no effect on the retest coefficient.

The implication of these considerations is that, except in the artificial case in which all individuals score identically on repeated administrations of a test, the issues of mean-level stability and retest stability must be addressed separately.

There is a third way in which stability or change in personality might be seen across the life span. Personality is often conceptualized in terms of the relation between discrete variables. The pattern of intercorrelations among a group of traits might alter with maturation; and although this kind of change is least familiar, it is logically prior to the other kinds of stability or change. The most common method of comparing patterns of intercorrelations is factor analysis of the battery of tests showing that the same (or different) factors emerge in different age groups or administrations. "Age-comparative

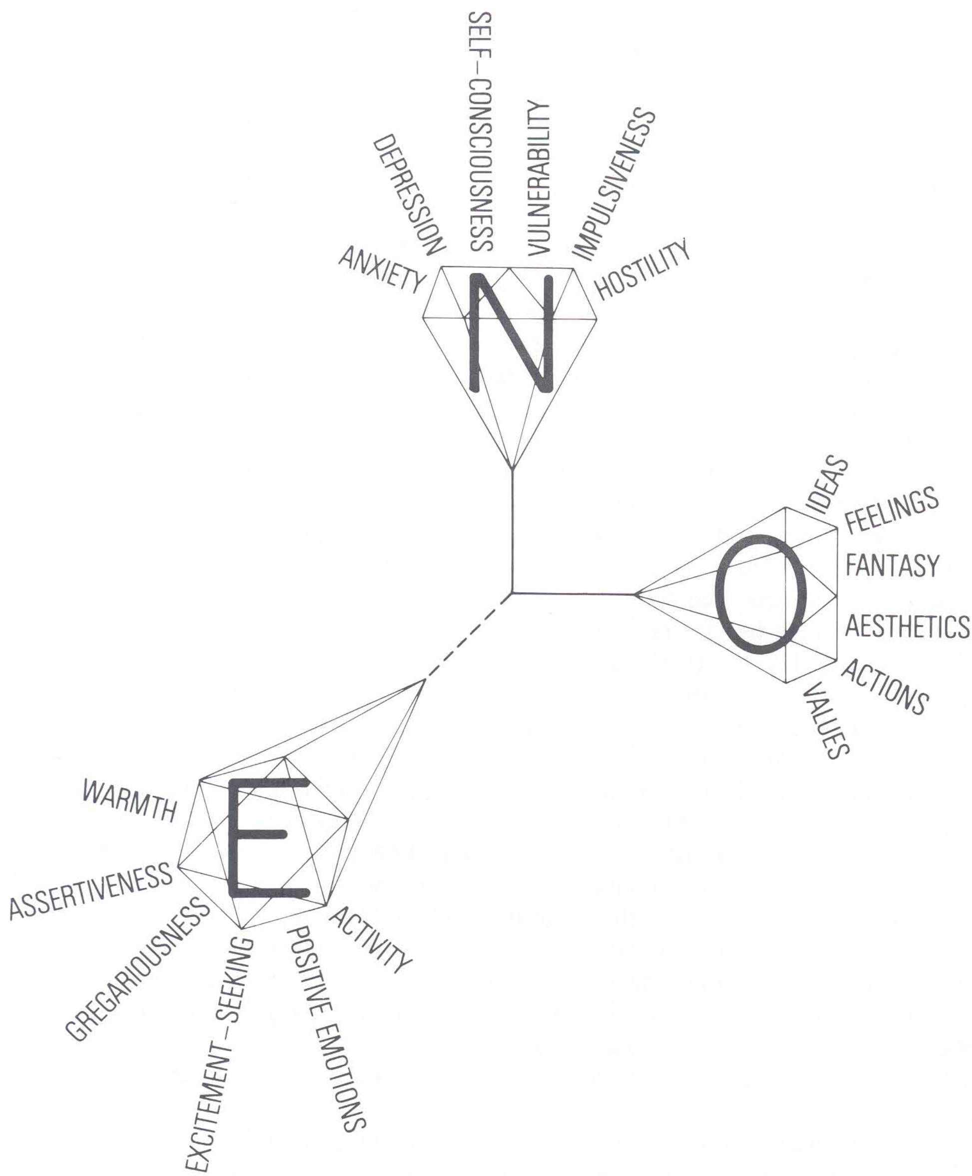

Figure VI.1. Schematic representation of the 18-facet neuroticism-extraversion-openness (NEO) model.
From McCrae and Costa (1980).

factor analysis'' is a term occasionally used to designate this kind of analysis, while the type of stability is usually referred to as ''structural stability.''

The three definitions of stability were applied in a series of BLSA studies *(Douglas and Arenberg, 1978; Costa et al., 1980b; McCrae et al., 1980)* that investigated the question whether personality remains stable or changes with age.

Stability of group mean level. Evidence of the constancy of group mean level is provided by a study by *Douglas and Arenberg (1978),* which included cross-sequential and time-sequential designs in the analysis of a large sample on a standard personality instrument over a considerable span of time. The authors of the study investigated the relation between age and the ten scales of the GZTS in light of the following questions: a) Does the GZTS demonstrate age differences? b) Does it demonstrate age changes within individuals? c) Are age changes in repeated measures of the same subjects confirmed in cross-sequential analyses of the same birth cohort measured at different times? d) Are the changes found in repeated measures of the same individuals also found in time-sequential analyses (i.e., are the changes due to period effects rather than to aging)?

Between 1958 and 1974, the GZTS was administered to 915 BLSA participants from 17 to 98 years of age. Repeated measures were obtained for 336 men from 5.6 to 9.9 years after initial testing. Each GZTS scale was analyzed cross-sectionally and longitudinally. Time- and cross-sequential analyses of independent samples were also performed in an attempt to separate the effects of aging from those of birth cohort and period.

Only two scales (General Activity and Masculinity) were interpreted as showing aging effects: Preference for fast-paced activity started to decline at age 50; masculine interests declined at all ages. In general, results provide strong evidence for the stability view of adult personality. Only two of the ten scales showed evidence of age-related change, and the amount of change was very small for both. Personality differences among individuals at all ages are far more pronounced than the effect of age itself.

Stability of individual rank order. Retest correlations or stability coefficients assess the magnitude of personality consistency or change in the relative ordering of individuals, regardless of absolute level. These are among the most important analyses for longitudinal studies, for although different samples at different times, or different birth cohorts at one time, can be used to estimate age changes in trait levels, only repeated testing of the same individuals can speak to the degree of stability of individual differences.

The stability of individual rank order on the scales of the GZTS was examined longitudinally in BLSA participants who were tested on three occasions over a period of about 12 years *(Costa et al., 1980b).* Of primary interest was the degree of stability over such extended intervals. In addition, two subsidiary hypotheses were tested. The first hypothesis was that certain socially desirable traits, like sociability, assertiveness, and others that define the broad domain of Extraversion, are particularly stable, and that change is more characteristic of undesirable traits that might be interpreted as elements of the Neuroticism domain. The second hypothesis asserts that stability coefficients will increase with age. This hypothesis was based on evidence that stability coefficients increase for children (Nesselroade and Baltes, 1974), and suggestions that personality is increasingly consolidated in old age (Neugarten, 1964). The subjects were 460 BLSA participants ranging in age from 17 to 85 at the time of first testing. Subjects were readministered the GZTS at intervals of approximately six years over a

Table VI.1. Six- and 12-Year Retest Coefficients for GZTS Scales for Total Samples and 12-Year Retest Coefficients for Three Age Groups

	6-Yr Retest	12-Yr Retest	12-Yr Retest		
Scale	Total Sample (17–85 yr)[a]	Total Sample (20–76 yr)	Young Group (20–44 yr)	Middle Group (45–59 yr)	Old Group (60–76 yr)
General Activity	.83 (410)[b]	.77 (192)	.77 (60)	.82 (93)	.78 (39)
Restraint	.71 (418)	.72 (193)	.61 (62)	.74 (94)	.76 (37)
Ascendance	.82 (401)	.83 (194)	.85 (62)	.85 (95)	.77 (37)
Sociability	.81 (393)	.74 (182)	.64 (62)[c]	.81 (88)	.66 (32)
Emotional Stability	.74 (427)	.70 (203)	.63 (68)	.76 (96)	.71 (39)
Objectivity	.71 (405)	.69 (191)	.66 (64)	.76 (87)	.59 (40)
Friendliness	.77 (418)	.74 (193)	.74 (64)[d]	.68 (88)[e]	.87 (41)
Thoughtfulness	.72 (418)	.73 (199)	.78 (64)	.71 (94)	.71 (41)
Personal Relations	.73 (385)	.68 (188)	.70 (62)	.64 (89)	.73 (37)
Masculinity	.75 (417)	.72 (200)	.73 (66)	.71 (94)	.70 (40)
Mean stability	.77	.73	.72	.75	.73

[a]Numbers in parentheses in column headings are age ranges at Time I. All correlations are significant at $p < .001$.
[b]Ns are given in parentheses.
[c]Difference between young and middle groups is significant at $p < .05$.
[d]Difference between young and old groups is significant at $p < .05$.
[e]Difference between middle and old groups is significant at $p < .01$.
From Costa et al. (1980b)

12-year period. Three age groups, based on the date of first testing, were formed: young (17–44 yr, $\bar{x}$ age = 36.7 yr, 145 men); middle (45–59 yr, $\bar{x}$ age = 51.5 yr, 183 men); and old (60–85 yr, $\bar{x}$ age = 67.9, 132 men).

Table 1 gives the six-year and the 12-year stability coefficients for the total samples, and the 12-year stability coefficients within each of the three age groups.

Under the hypothesis that stability should be greater in older age groups, one-tailed tests of the significance of differences between correlations were computed for each pair of age groups on each scale. Six- and 12-year stability coefficients were exceptionally high, comparable to the short-term retest reliability. Coefficients were quite similar for the three age groups. Of the 30 comparisons of stability coefficients in different age groups at each interval, only six were significant for the six-year interval and three for the 12-year interval. Of the nine significant differences, four were in the predicted direction, five in the opposite direction. None of the specific findings for scales at the six-year interval was replicated at the 12-year interval. The hypothesis that younger men would show lower stability of personality traits than older men was not supported by the data. When corrected for unreliability, data were also inconsistent with the hypothesis of differential stability for different traits. These data provide strong evidence for the stability of personality traits in adulthood. Increased stability with age may be found among children and adolescents, but by young adulthood stability in these dimensions of temperament is so high—near the limits of reliability of the instrument—that a ceiling effect diminishes the likelihood of any further increase in stability. Personality stability across time and instruments provides support for one implication of the general stability position: that the time at which personality is measured is irrelevant. Once the individual has reached full adulthood, by age 30, a single measurement of personality would suffice for a lifetime. In other words, personality might be considered a constant throughout the individual's adult life span, including old age.

There are, of course, a number of reasons to stop short of so sweeping a stand. It assumes a perfect correlation of tests over time, a condition rarely observed even when corrections for unreliability are made. It also assumes that there are no circumstances under which personality might regularly be expected to change. Evidence to date shows only that the events encountered over the course of a lifetime by volunteer subjects do not systematically produce change. But other circumstances, such as therapeutic interventions, cataclysmic events, or severe illnesses, might affect traits. The longitudinal study of personality in adulthood is simply too young a field to rule out these possibilities.

On the other hand, if we turn from the individual to the group, much stronger arguments can be made. On the aggregate level, there appear to be good grounds for claiming that the time at which measurements are made should not affect the results; in particular, the relations between different tests ought not to depend on the times at which they are administered. Data collected in the 1960s might be used to validate tests created in the 1970s, just as if the administration had been contemporaneous.

This approach was adopted in a study relating archival GZTS scores to recently collected personality scores *(Costa and McCrae, in press)*. The criteria of primary interest were scales from the GZTS (Guilford et al., 1976), administered from 1959 to 1979 to subjects on their first or second visit to the GRC. Although the instrument was subsequently readministered, for purposes of simplicity only first-administration data are analyzed here.

Over the same period, data were also collected on the Cornell Medical Index (CMI). Twelve sections deal with physical symptoms, six with psychiatric complaints. Scores from the WAIS Vocabulary scale and the Army Alpha total collected in the same two-decade period were also used.

All these measures were used to consider the convergent and discriminant validity of three personality instruments administered in the past four years. Form A of the Eysenck Personality Inventory (EPI), with scales for Extraversion, Neuroticism, and Lie, was mailed to subjects to be completed at home in September 1979. The NEO Inventory was mailed to subjects in February 1980, and the NEO Rating Form was completed at home in August 1980 by spouses of subjects whose husbands or wives were also participants in the BLSA. Because some spouses did not participate, rating data are available for only a subset of subjects.

The NEO Inventory measures six facets, or aspects, of three global domains of personality. Neuroticism is represented by anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability. Extraversion includes scales for warmth, gregariousness, assertiveness, activity, excitement-seeking, and positive emotions. Openness to Experience is measured in the areas of fantasy, aesthetics, feelings, actions, ideas, and values. Total scores for the three domains are obtained by summing the scores of the six facets in each. The analyses provided evidence of validity against new criteria; in addition, since the construct validity of the NEO scales is already fairly well established, they provided a demonstration of the feasibility of using data collected several years previously in validation studies.

The correlations between GZTS scales administered between 1966 and 1979 and

NEO facets administered in 1980, at least one year after the last GZTS, are presented in Table 2. The average interval between the two tests was 9.1 years. Despite this considerable lapse of time, construct validity is very much in evidence. General activity from the GZTS is most strongly correlated with NEO activity; GZTS ascendance with NEO assertiveness; GZTS sociability with NEO warmth; GZTS emotional stability and objectivity (negatively) with NEO anxiety; GZTS friendliness (negatively) with NEO hostility; GZTS thoughtfulness with NEO openness to ideas. At a more global level, the GZTS Extraversion scales, (G, A, and S) are correlated chiefly with NEO Extraversion scales; GZTS Emotional Health scales (E,O,F,P, and M) show consistent negative correlations with NEO Neuroticism scales. Finally, the magnitude of the correlations requires comment: Almost all exceed .30, and several reach .60. In view of the somewhat different conceptions underlying the two instruments, these are high correlations.

Correlations with NEO ratings by spouses are generally smaller than correlations with self-reports and occasionally fail to reach statistical significance, in part because they are based on considerably fewer cases. But in view of the fact that the correlations cross instruments, methods of measurement, and two to 20 years, the pattern of results is convincing. A correlation of .59 between GZTS Extraversion and spouse-rated NEO Extraversion is particularly remarkable.

2. Historical Effects on Test Data

As the value of longitudinal studies becomes more apparent, we can expect more researchers to collect and archive personality data, or to retrieve and retest subject populations measured some time before. Longitudinal methodologists have pointed out that there are potential problems in the use of such data, since the meaning of the test may have changed in the intervening years. This phenomenon can readily be envisioned in the case of attitudinal research: Attitudes toward women that were radical in the 1950s might be middle-of-the-road today.

In the case of personality measures, this concern appears to be less well founded. Over the past 20 years, scales measuring the major dimensions of personality do not appear to have altered substantially in meaning. Subjects today appear to respond to test items in much the same way as they did in the 1950s and 1960s. Basic aspects of temperament are relatively impervious to historical changes, and longitudinal research can use older personality data with some confidence that the meaning of test items has not changed.

Indeed, it seems more likely that what has changed is our conceptualization of what older tests measure. For example, the Bernreuter (1933) inventory contained a scale labeled "introversion" that we might now conceptualize instead as social anxiety, a part of Neuroticism. Such hypotheses are testable if we readminister older instruments and revalidate them against measures that are better conceptualized and psychometrically more sophisticated. Since the relations between tests appear to change little over an interval of years, we could use the results of contemporary studies to reinterpret older findings on the basis of our current understanding of the nature of the constructs they measure.

Stability of factor structure. Constancy and change in adult personality organization can also be assessed by comparison at different ages of the factor structures of personality instruments. This approach was used in another BLSA study by *McCrae et al. (1980),* which compared the factor structure of the GZTS longitudinally in three

Table VI.2. Correlations between GZTS Scales Administered 1959-1966 and NEO Facets Administered in 1980

NEO Facets	GZTS Scales[a] G	R	A	S	E
Neuroticism:					
Anxiety	–14	–01	–25**	–25***	–67***
Hostility	23**	–24**	07	–12	–47***
Depression	–04	–02	–38***	–38***	–58***
Self-Consciousness	–12	–04	–37***	–33***	–55***
Impulsiveness	12	–35***	16	11	–44***
Vulnerability	–31***	08	–44***	–34***	–39***
Extraversion:					
Warmth	11	–08	31***	56***	18*
Gregariousness	31***	–25***	41***	52***	17
Assertiveness	34***	–14	61***	41***	20*
Activity	61***	–18*	20*	10	03
Excitement-Seeking	41***	–34***	42***	21***	–13
Positive Emotions	32***	–29***	27**	30***	10
Openness:					
Fantasy	07	–12	10	–06	–30***
Aesthetics	19	10	26**	26**	19*
Feelings	22*	00	31***	21**	–09
Actions	22**	–10	31***	20*	03
Ideas	12	13	29***	08	07
Values	10	–12	23**	–05	–14
	O	F	T	P	M
Neuroticism:					
Anxiety	–55***	–31***	17	–31**	–28***
Hostility	–44***	–53***	04	–26**	–24**
Depression	–52***	–26**	11	–29***	–26**
Self-Consciousness	–46***	–25**	02	–22**	–26**
Impulsiveness	–31***	–46***	–06	–27**	–13
Vulnerability	–38***	–03	01	–10	–26***
Extraversion					
Warmth	15	08	10	12	–03
Gregariousness	16	–05	–23**	16	00
Assertiveness	15	–20	01	01	00
Activity	01	–16	01	–08	06
Excitement-Seeking	–08	–33***	–05	01	02
Positive Emotions	07	–06	10	–07	–05
Openness:					
Fantasy	–20*	–22**	22**	–22**	–02
Aesthetics	15	13	31***	06	00
Feelings	–06	–24**	31***	–12	–15
Actions	05	00	09	08	24**
Ideas	08	05	35***	09	11
Values	–13	–11	19*	–11	00

[a]GZTS scales are General Activity (G), Restraint (R), Ascendance (A), Sociability (S), Emotional Stability (E), Objectivity (O), Friendliness (F), Thoughtfulness (T), Personal Relations (P), and Masculinity (M). N = 140 to 152. Decimal points omitted. *p < .05; **p < .01; ***p < .001.

From Costa and McCrae (in press)

administrations, three age groups, and two times of measurement. Factor analyses were performed on correlation matrices computed for eight groups. Longitudinal comparisons were based on analyses of first-administration data from 769 men, second-administration data from a subset of the first group consisting of 346 men, and third-administration data from a further subset of 171 men. Cross-sectional comparisons were based on first-administration data from three age groups: young (17–44 yr, $\bar{x}$ age = 34.4 yr, 314 men); middle (45–59 yr, $\bar{x}$ age = 51.6 yr, 242 men); and old (60–97 yr, $\bar{x}$ age = 70.4 yr, 213 men). Possible structural differences resulting from period effects were assessed by dividing the sample into two groups: 455 men who first took the GZTS before July 1968 (age range 17–83 yr, $\bar{x}$ age = 52.1 yr), and 314 men who first took the GZTS after that date (age range 18–96 yr, $\bar{x}$ age = 45.6 yr); these analyses were limited to data obtained at the first administration of the test. All factor analyses were restricted to subjects with valid scores on all ten GZTS scales.

Factor structures in the GZTS in this sample were clearly invariant. Despite aging, attrition, and possible practice effects, the same pattern was seen at each administration. The data revealed no meaningful differences when GZTS measurements before and after 1968 were compared, despite great social differences between the 1960s and 1970s; the implication is that the basic structure of personality is not greatly influenced by social and historical change in the life of the individual.

3. Eliminating Response Bias as an Explanation for Stability

Response sets and age. Most research on the stability of personality has been conducted with self-report measures, which have in general shown a better record of internal consistency, retest reliability, and construct validity than ratings or projective tests.

But self-report inventories are also prone to certain problems. The transparency of the items makes it possible for individuals so motivated to present themselves favorably or unfavorably. The use of a standard format for answering questions (yes-no, Likert scale, a rating bar) makes it possible for consistencies in the style of response to distort the scores obtained from the instrument; the result may be spuriously high consistency or correlations. The effects of social desirability, acquiescence, and extreme response have been the source of interminable debate among personality psychologists, and no definitive resolution has been reached.

In addition, there has been relatively little research on age trends in response sets. Some writers (Schaie and Schaie, 1977) have argued persuasively that there have been enormous changes in the amounts and kinds of testing to which individuals of different generations have been exposed, and that this may introduce unwanted sources of variance in tests. Increased cautiousness may alter the responses of older subjects (Botwinick, 1969), or standards of social desirability may change with age, bringing shifts in the influences of that set. Age changes or cohort differences in response could account for observed differences in scale scores or could mask real changes that are occurring. To date, these possibilities remain speculative, with little empirical foundation. Clearly, before any conclusions about aging and personality are drawn from objective personality measures, some information on these issues would be useful. Recent analyses of data from the Baltimore study *(Costa et al., 1983)* contribute to a resolution.

Over the past 20 years, subjects in the BLSA have been given the GZTS every six years. Since recruitment into the study has been continuous, new samples of individuals, ranging in age from the 20s to the 90s, have been tested successively. If their ages, birth cohorts, and times of measurement are divided into six-year intervals, a variety of analytic designs may be applied to aid in the interpretation of changes or differences.

The GZTS was given to all subjects with standard instructions that provide three response options—"yes," "no," and "?"— but subjects were instructed to use the "?" option only if they were completely unable to select "yes" or "no." In conformance with the suggestion of Guilford and Zimmerman (1949), scales containing more than three "?" responses were invalidated in all previously reported applications of the GZTS conducted by the Baltimore study. This exclusionary principle may, however, distort results; in particular, if age produces caution in responding, a disproportionate number of older subjects may be excluded, perhaps especially the most cautious. Because of that possibility, a new approach was adopted in the analyses reported here. All the GZTS answer sheets were keypunched, so that responses to individual items could be analyzed. The original scoring system, in which "?" responses were not scored, tended to lower the score of the individual. In the new system, responses were assigned a value of –1, 0, or +1, the "?" being represented as a neutral, rather than negative, value.

The handbook for the GZTS (Guilford et al., 1976) lists three scales that have been developed to estimate the influence of certain response sets: the Gross Falsification (GF), the Subtle Falsification (SF), and the Careless Deviancy (CD) scales. The first two are intended to screen individuals who may be attempting to present an unduly favorable impression; the third is composed of relatively rare responses, a high score on which is interpretable either as careless responding or as deviancy in personality. In addition, it is possible to measure at least three other response sets on the GZTS. The number of blanks was summed across all scales to give an index, as was the number of "?" responses. Since items in the GZTS are roughly balanced on most scales, it is possible to interpret the sum of "yes" responses as an index, not of any substantive personality trait, but of the tendency to acquiesce indiscriminately.

For each of these variables, cross-sectional, longitudinal, cross-sequential, and time-sequential analyses were performed. In the repeated-measures analysis, the subjects were 348 men ranging in initial age from 32 to 74 who were retested after four to eight years ($\bar{x}$ interval = 6.6 yr). They were classified in seven age groups, each spanning a six-year interval. In the time-sequential analyses, 328 men who were tested in the period from 1958 to 1964 were compared with 278 men tested between 1965 and 1971. They were cross-classified by the same seven age categories used in the repeated-measures design. In the cross-sequential analysis, 345 men tested between 1958 and 1964 were compared with 285 men tested between 1965 and 1971. In this design, however, they were cross-classified in seven cohorts born during six-year intervals from 1896 to 1932.

Repeated-measures analyses showed significant ($p < .05$) effect on the repeated factor for number of question marks, which increased from 6.2 to 9.6; on acquiescence, which decreased from 132.1 to 128.9; on the GF scale, which increased from 11.8 to 12.1; and on the SF scale, which decreased from 21.2 to 20.8. Age-group differences were seen for the number of question marks, which was highest in the 68–74-year-old group and lowest in the 50–55-year-old group. In addition, there were two interactions: Men aged 38 to 43 at first testing showed a decrease instead of an increase in question marks, and men aged 55 to 61 showed an increase instead of a decrease in acquiescence.

These results are somewhat puzzling; certainly they did not show a monotonic change in any of the response sets with age. Most of the changes were extremely small in magnitude, and if we require a significance level of $p < .01$, only two effects are significant: the increase in question marks and the decrease in acquiescence. The fact that neither of these longitudinal changes was mirrored in cross-sectional differences suggests that the changes are due either to time-of-measurement effects (a cultural change during the testing period) or to "practice" (repeated exposure to the test).

Examination of the cross- and time-sequential analyses, conducted on samples of more than 600 men, is revealing. Analyses of number of blanks, number of question marks, acquiescence, GF, SF, and careless deviancy show no significant ($p < .05$) effects for aging/cohort, aging/time, cohort/time, or cohort/aging, nor were there any significant interactions. These data suggest that the marginal cross-sectional differences and interactions in the repeated-measures analyses are best regarded as unreplicable error and that the longitudinal changes in acquiescence and use of question marks are attributable to practice effects. That acquiescence decreased while use of question marks increased by three items is suggestive: Subjects may have felt pressured on the first administration to avoid question marks at all costs and may thus have agreed to a few items of which they were uncertain. Some years later, as experienced subjects no longer so hesitant to assert themselves, they may have used the question marks when they felt they needed to.

In any case, these data imply that response sets are not ordered by age. Although longitudinal research may want to consider the effects of repeated administration of the same instrument, the particular effect seen here is small in its overall influence on scale scores and is probably unique to instruments, like the GZTS, that provide a question-mark option but fail to score it.

4. The Stability Model: Implications for Gerontology and Geriatrics

These results favor the stability model for objectively measured personality traits throughout adulthood, at least for males. The implications for gerontology and geriatric practice of these consistent findings are radical. Neugarten's classic formulation that adult development is "the changing basis within an individual for adaptation to life" may well have to be changed to read "the *stable* basis." Students of the relation of personality to aging, seeking the basis of the impact of major life events, may well have to shift the focus of their efforts from the vectors of personality change to the mechanisms by which personality preserves equilibrium. Not least significant is the corollary that a clinician who finds true personality change in an aged patient is likely dealing not with a normal event but with evidence of disease that can often be treated.

These findings do not prove that personality is unchangeable. Some individuals do change in one or more characteristics, for reasons not yet understood. It is reasonable to suppose that psychotherapeutic intervention can make real changes in personality, and a host of techniques, from cognitive behavior modification to biochemical intervention, may have profound effects as yet undocumented. What we can say is that such changes, for better or worse, are not likely to happen to anyone simply as a result of growing older.

5. Longitudinal Changes in Adjustment to Aging

Adjustment to aging and the effect of age on personal adjustment among the aged have long been of concern to gerontologists. A recent study *(Costa and McCrae, 1982)* used cross-sequential and time-sequential as well as traditional longitudinal designs to examine the stability or change in the Chicago Attitude Inventory (CAI). The eight sections of the CAI—Health, Friends, Work, Economic Security, Religion, Usefulness, Happiness, and Family—were supplemented by two global items concerning assessment of life happiness and satisfaction with accomplishments in life. Analyses were conducted on a sample of 425 men, aged 17 to 97, on whom first-administration CAI data had been collected before 1970. Results of cross-sequential, time-sequential, and repeated-measures analyses led to the conclusion that attitudes toward usefulness and work showed small age-related declines that were interpreted as non-maturational. Assessment of Life, Satisfaction with Accomplishments, and attitudes toward Health, Friends, Economic Security, Religion, Happiness, and Family showed no consistent changes with age. Personal adjustment was thus found to be quite stable across the adult and later years.

6. Personality and Adjustment to Aging

Enduring dispositions as predictors of successful personal adjustment to aging were examined in a BLSA longitudinal study by *Costa et al. (1981)*. Successful aging can be studied by focusing a) on outer social adjustment defined in terms of an individual's activities and social roles, b) on the inner subjective experience of personal adjustment, or c) on subjective well-being. From the second perspective, satisfaction with life's accomplishments, retention of high morale, or simple happiness is considered evidence of successful aging, while hypochondriasis, fear of death, and a sense of uselessness, loneliness, and depression are considered signs of poor adjustment to aging.

The second definition was used in this BLSA study, which examined the ability of longitudinal measures of Neuroticism and Extraversion to predict successful personal adjustment to aging. The study used the GZTS to measure personality characteristics and the CAI to measure subjective well-being. The ten GZTS scales yield scores on three factors: Neuroticism, Extraversion, and Thinking Introversion (a factor thought to represent meditative thinking or introspective tendencies). It was hypothesized that Neuroticism would have a negative relation to CAI measures of subjective well-being, that Extraversion would have a positive relation, and that Thinking Introversion would have no relation to those measures.

The CAI and the GZTS were administered to each of the men in the sample (aged 17–97 yr) on the first or second visit to the GRC. The CAI was readministered on the fifth and ninth visits. In the first analyses, correlations between the CAI scales and the contemporaneous GZTS factors were calculated for two subsamples: 418 men aged 18 to 49 and 391 men aged 50 to 97 years. In the second analysis predictive relations between personality and adjustment were sought by correlation of GZTS Neuroticism and Extraversion scores from the first administration with scales from the second and third administration of the CAI for the 577 men with repeat administration data.

As hypothesized, Neuroticism was related negatively and Extraversion positively to most concurrent measures of well-being in both younger and older subsamples, and Thinking Introversion was unrelated to the well-being measures. Predictive correlations between personality and subjective well-being over intervals of two to ten years ($\bar{x}$ interval = 5.3 yr) and ten to 17 years ($\bar{x}$ interval = 12.6 yr) confirmed earlier research (Costa and McCrae, 1980b) and showed that enduring personality dispositions precede and predict measures of personal adjustment to aging.

Using more sophisticated measures of well-being, *Costa and McCrae (in press)* have shown that an individual's affect balance and life satisfaction can be predicted years in advance by assessment of personality. They used GZTS scores collected in the period from 1959 to 1969 with well-being data collected ten to 23 years later ($\bar{x}$ predictive interval = 15.6 yr for first and 17.7 yr for second interval). As had been predicted by a model of psychological well-being (Costa and McCrae, 1980b), three of the four Extraversion scales were significantly related to well-being, especially to positive affect, while four of the five Neuroticism scales were significantly related to well-being, especially to negative affect. These correlations, which were not markedly different in size from contemporaneous measures, provide a strong longitudinal replication of the model.

7. Somatic Complaints as a Function of Age and Neuroticism

Self-perceptions of health are key components of health maintenance, since they influence efforts at self-medication as well as decisions to seek medical treatment. Studies have consistently shown that global self-ratings, which are moderately correlated with medical determinations of health, are also related to such psychological characteristics as health attitudes, morale, adjustment, and psychological distress. Despite the demonstrable increase in many kinds of illness with age, global self-ratings often fail to show any marked association with age.

Symptom checklists like the CMI offer certain advantages over global ratings. Since they ask specific questions about conditions and symptoms, they may be less influenced by general health attitudes and can be used to analyze medical conditions of body systems separately, as well as to determine whether age is differentially associated with complaints in particular somatic systems. The CMI is a self-report symptom checklist with 195 items divided into 12 somatic sections (A–L) and six psychiatric sections (M–R). Somatic sections I (Frequency of Illness) and J (Fatigue) had extremely low endorsement in this sample and were therefore combined in all analyses. The sum of sections A–L yields a measure of total physical complaints, while the sum of sections M–R provides a measure of psychiatric complaints. Like global ratings, checklists are influenced by both objective health and psychological factors. Hypochondriasis, neurosis, general anxiety, poor marital adjustment, and psychological problems have all been found to be linked to extremely high endorsement of CMI physical complaints.

Many of the psychological factors identified in previous research can be hypothesized to be related to Neuroticism. The hypothesis was tested in a BLSA longitudinal study that examined the relative influences of age and Neuroticism on self-perception of health or illness and on total physical complaints *(Costa and McCrae, 1980a)*.

At the time of their first administration of the CMI, the 1038 subjects ranged in age from 20 to 97 years. Although some analyses used data from a second or third administration of the CMI, the number of subjects was smaller because of insufficient length of participation in the study, death, or withdrawal. Endorsement of psychiatric items was low in this population, about 40% of the subjects endorsing none and 20% only one of the 51 items. The Emotional Stability Scale of the GZTS was used as an alternate measure of neuroticism. This is a 30-item scale covering Evenness vs. Fluctuation of Mood; Perseveration of Ideas; Composure vs. Excitability; Daydreaming; Feelings of Guilt, Loneliness, or Worry; and Cheerfulness vs. Gloominess. Two items explicitly refer to feelings of Good vs. Ill Health. Correlation between the GZTS Emotional Stability Scale and the CMI psychiatric score was found to be 0.52 in a sample of 915 subjects.

Longitudinal analyses were restricted to subjects whose second and third administrations of the CMI were, respectively, five to eight years and ten to 17 years later than the first. Each subject had also completed the GZTS Emotional Stability Scale at either the first or second visit to the GRC. Four sets of repeated-measures analysis of variance were conducted, with age and Neuroticism as classifying variables. In all analyses, subjects (N = 248) were classified in three groups of equal size as young (20–44 yr), middle-aged (45–56 yr), or old (57+ yr) on the basis of their ages at the first administration. In two of the sets of analyses, Neuroticism was measured by the GZTS Emotional Stability Scale and subjects were classified as unstable or stable. In the other two sets, Neuroticism was measured by the CMI psychiatric (M–R) score, and subjects were classified as high or low (subjects endorsing only one item in the M–R section were omitted from these analyses).

Supplementary cross-sequential analyses using different subsets of the same cohort (551 subjects), and time-sequential analyses (637 subjects) were also conducted on data collected at the first administration. The data showed a consistent pattern across different methods of analysis and different measures of Neuroticism. Age had a selective effect on physical complaints, while Neuroticism appeared to produce a more general and diffuse effect. Problems in sensory, cardiovascular, musculoskeletal, and genito-urinary systems increased with age, while health habits improved. The fact that only certain systems showed age-related increases in complaints may in part account for the finding that global health ratings are only weakly related to age. The finding of significant effects of Neuroticism despite its limited range in this healthy and well-adjusted BLSA sample argues that the relation between Neuroticism and physical complaints must be quite strong. The influence of Neuroticism on health perception may be even more pronounced in the general population.

It was concluded that research in the influence of physical health on morale, sick-role behavior, or adjustment should either use objective measures of health or supplement self-ratings with measures of Neuroticism in order to control for its effects. Most important, the findings convincingly contradict the conception that the aged are typically hypochondriacs obsessed with their bodily functions.

8. Hypertension, Somatic Complaints, CAD, and Personality

Hypertension, Hypochondriasis, and Neuroticism. Reports that hypertensives score higher than normotensives on measures of maladjustment, hypochondriasis, and neuroticism have led some writers to conclude that hypertension is related to personality (Sainsbury, 1960). Specifically, the chronic internal stress of repressed rage or anger is thought to lead to the development of hypertension. However, two alternative hypotheses are also consistent with the data. First, the association may be artefactual, the result of self-selection among clinical populations. Under this hypothesis, neuroticism leads only to the discovery or detection of hypertension. Second, the causal relation may be reversed; hypertension itself, the patient's awareness of illness, or medical treatment may lead to poorer psychological adjustment.

In an attempt to decide between these alternatives, *Costa et al. (1980a)* examined longitudinal data on systolic and diastolic blood pressure and data from the CMI and

GZTS from more than 700 BLSA subjects. Of the volunteers, 101 who were taking medication for hypertension either when they entered the study or at the succeeding four visits were eliminated from all analyses because of the possibility that their medication might increase somatic concern while decreasing blood pressure, and thereby obscure any real positive association between the two variables.

Each subject was given the GZTS and CMI during his first or second visit to the GRC. The GZTS was readministered approximately every six years, the CMI on the fifth visit. Longitudinal analyses were limited to subjects who took their second GZTS and second CMI four to 12 years after their first medical examinations. The average predictive intervals were 7.4 years for the GZTS and 6.5 years for the CMI. A total of 12 personality and physical-complaint variables were examined.

In cross-sectional analyses, higher somatic complaints, Restraint, Friendliness, and Good Personal Relations, as well as lower General Activity and Masculinity, showed small associations with higher blood pressure. However, when the effect of age (which is related to blood pressure) was partialled out, no statistically significant association between personality characteristics and blood pressure was found.

Does hypertension affect personality? If elevated blood pressure itself affects personality, the effect should have been evident in the correlations discussed above. It is possible, however, that the interval between a blood-pressure increase and the measurement of personality was not long enough for the hypertension to influence personality. To test this possibility, predictive analyses using average pressures at first measurement were correlated with GZTS and CMI scores collected four to 12 years later. Systolic pressure was correlated with greater Ascendance and Sociability and poorer Personal Relations, but the correlations were only marginally significant.

It can be argued that the GZTS, which is based on self-reporting, is inherently incapable of accurately detecting repressed rage and that clinical judgment is needed. An attempt was therefore made to approximate clinical judgment by examining combinations of personality scales. By cross-classification on scales F (Friendly vs. Hostile), R (Restrained vs. Impulsive), A (Ascendant vs. Submissive), and M (Masculine vs. Feminine), six hypotheses were tested: that higher average pressure might be found in individuals who were a) hostile but restrained, b) ascendant but restrained, c) masculine but restrained, d) hostile but submissive, e) masculine but submissive, or f) hostile but feminine. Neither main effects nor interactions proved significant when age was used as a covariate.

In summary, when the influence of age was controlled, no association was found between blood pressure and any of 12 personality and psychosomatic concern measures; nor did patterned combinations of traits show any relation to blood pressure. Attempts to predict personality scores four to 12 years later from blood-pressure levels also failed; the inference is that any changes in anxiety or somatic concern due to hypertension or its treatment are short-lived.

Does coronary disease affect personality? Examination of the relations between CAD and personality revealed that BLSA subjects who complained of angina were lower than average on the GZTS scales Emotional Stability and Masculinity. Was this an example of personality change resulting from medical illness, or were personality variables diagnostic of and perhaps causally involved in the development of the disease? To test out these possibilities, it was necessary to separate subjects into different groups and to take into consideration the temporal course of the disease *(Costa et al., 1982)*. Eighty-eight subjects were selected who were measured on the GZTS during their first or second visit to the GRC, and were free from both anginal complaints and certain ECG signs on these visits. Over the next twenty years, four groups emerged: those who developed CAD as evidenced by both anginal complaints and ECG signs; those who showed ECG signs of CAD, but did not report angina in a follow-up period of from five to 15 years; those who reported angina but showed no ECG signs of CAD in the same follow-up period; and a control group (age-matched to the total of the three groups) that showed neither ECG signs nor angina in a follow-up period of ten to 20 years. (All subjects were classified according to their status at last examination; the variation in the follow-up intervals reflects the fact that subjects entered and left the study at different times.)

The first group is easily diagnosed as having CAD. The second and third groups are more ambiguous: Those with only angina seem to be overly sensitive; those with only ECG signs appear under-sensitive. Do any of these distinctions show up in personality measures taken before the development of disease?

Results showed that there were no differences between the first and fourth groups, that is, between those who definitely did and did not develop CAD. Thus none of the traits measured by the GZTS appears to have etiological significance. But there were pre-existing differences between the two intermediate groups. Individuals who complained of anginal pains, but who gave no ECG evidence of disease, were less emotionally stable than those who reported no angina despite ischemic ECG signs.

It is well known (Hurst et al., 1978; Froelicher, 1977) that resting and stress ECG signs in themselves are far-from-perfect indicators of CAD, and the number of subjects in this study (88) is too small to be conclusive. Nevertheless, some interesting interpretations are suggested by the data. Men high in Emotional Stability may be more likely to minimize minor chest pains, so that without routine medical examinations those who have CAD are likely to remain undiagnosed and untreated (cf. Berglund et al., 1975). On the other hand, men who are low in Emotional Stability are very sensitive to chest pains, may report anginal symptoms even in the absence of organic pathology, and may request medical attention that is not required. Taken together, these analyses suggest that although the personality variables measured by the GZTS do not affect the development of CAD they seem to affect the presentation of symptoms.

If these personality differences were found at the same time as the symptoms, several alternate interpretations could be offered. We might argue that the experience of angina was sufficiently traumatic to lower the individual's emotional stability or we might argue that self-selection played the crucial role: Those individuals who were aware of chest pains and who were also predisposed to worry about their health (i.e., the more neurotic) would be most likely to join and remain in a longitudinal study that promised periodic monitoring of their health. Better-adjusted men with angina, some of whom could be expected to show no ECG signs and would thus be classified in the third group, would be less concerned with their health and less likely to volunteer for the study. These arguments have been supported in studies of the relationship between hypertension and personality *(Costa et al., 1980a)*.

In this case, however, we can rule out those alternatives. Since measurement of personality preceded the development of angina and ECG signs, it could not have been influenced by them. Self-selection on this basis is likewise impossible, since the individuals were presumably unaware that they would experience angina in the next few years. Retrospective accounts of personality difference could not be trusted in this context; only the archives of a longitudinal study permit the kinds of inferences drawn here.

These rather simple examples of the kinds of inferences that can be ruled out with longitudinal data are useful in part because they illustrate the underlying logic. Path-analytic techniques, however, offer far more sophisticated statistical models, which estimate not only the direction of causal influence but its degree, expressed as a regression weight (see Kenny, 1979). Causal inferences based on these statistical techniques are subject to a number of restrictions, many of them related to the assumptions necessary to construction of the model. In general, fewer assumptions are necessary and more reliable results can be obtained if more measurement points are used and more of the variables that might plausibly influence elements in the model are measured. Clearly, longitudinal studies are ideal for this type of analysis.

9. Research on Personality and the Life Course

The major drawback to the use of a stable individual-difference approach to life-span development is that current conceptions and methods of research on aging and personality are designed primarily for the study of change. The elegant models intended to separate true maturational changes from cohort differences and cultural changes are not very useful if there is no meaningful maturational change. The widespread attempt to chart the developmental course of personality in adulthood no longer seems profitable, and it may not be immediately clear what direction future research should take.

We have argued elsewhere *(McCrae and Costa, 1982)* that personality dimensions may be more usefully construed as causes than as effects. Age, and its attendant social, cognitive, and biological changes, should be considered in conjunction with personality as joint shapers of the life course. Personality can help explain the choices (educational, career, familial) that must be made at specific life transitions. The stability of personality dispositions may contribute to the individual's sense of identity and to the continuity and coherence of the life course. Finally, adaptation to life at all ages is likely to be powerfully influenced by personality.

Prospective life histories collected by longitudinal studies can help to answer such questions as the following: How do the lives of introverts differ from the lives of extraverts? Does openness to experience lead to a more fluid and unpredictable life course? How is neuroticism typically manifest at each stage of life? Answers to these questions may help integrate the insights of life-span developmental psychology with those of personality in the study of lives.

OUTCOME STUDIES

1. Obesity and Longevity

"Desirable" or "ideal" weight goals for adult men and women have been recommended since the turn of the century by the insurance industry on the basis of its experience with "insured lives." The tables have been and continue to be used widely—by the lay public, physicians, public-health officials, and clinical investigators. The most recent tables (Metropolitan Life Insurance Co., 1959) are based upon data subsequently published in a report of the Society of Actuaries (1960).

Since that publication, however, there have been reports of the obesity-mortality association in more than 40 other diverse populations (Andres, 1980a, 1981), most of them in the United States and western Europe with isolated reports from Japan, Australia, and Israel. There is a strong consensus that minimal mortality does not occur in the leanest segments of the population (as is commonly averred), but in individuals who range from the middle of the "desirable" weight range in the 1959 insurance tables to a value at least 20% over the mid-point. Indeed, one of the recent reports is an update of the experience of the insurance industry (Society of Actuaries and Association of Life Insurance Medical Directors of America, 1979); its data indicate that updated "desirable" weight tables will be required and that weights in them will be considerably increased over those of the 1959 tables.

Experience in the BLSA also indicates that mild or moderate overweight has been over-emphasized as a risk factor for mortality (Andres, 1980b). One of the possible confounding variables in the obesity-mortality link is cigarette smoking. Since smokers as a group weigh less than nonsmokers, it is theoretically possible that what appears to be the benefit of mild overweight may, in fact, be due to the fact that the percentage of smokers in that group is lower than that in the lean segment of the population. This hypothesis was tested in the BLSA subjects and was rejected *(D. Elahi et al., in press)*. In both current smokers and nonsmokers, there was a U-shaped relationship between obesity and mortality; mortality was highest in the leanest subjects in both groups, reached a nadir at moderately obese levels, and increased again in the more severely obese subjects. It must be emphasized that the BLSA population includes very few severely obese subjects. There is no question of the deleterious effects of severe obesity on longevity as well as on many specific diseases. Further analyses of the BLSA data will examine the impact of mild and moderate obesity on such end-points as hypertension, diabetes, and CAD.

2. Age and Some Physiologic Variables

A study by *Tobin (1981)* examined four physiologic variables clinically related to health. The variables included tests of the respiratory system (forced expiratory volume in one second), cardiovascular system (systolic blood pressure), renal system (standard creatinine clearance), and metabolism (glucose tolerance). Each is influenced by age and, at its extreme, is also indicative of a disease (chronic obstructive pulmonary disease, renal failure, hypertension, and diabetes). Normative data were obtained from individuals who were free of disease, poor health habits such as smoking or inactivity, or medications known to influence the organ system under study. This was done to ascertain the effect of age as such, in distinction from disease or other factors. To judge the importance of each variable, the age-adjusted T-scores for individuals who had died were compared with those individuals who had survived a ten-year interval.

There was no effect of age on the fasting plasma-glucose level. At two hours of a standard oral glucose tolerance test, there was a clear effect of age on blood-glucose concentration, with higher levels in each successive age group, although there was large variance at every age. Results of the tests of forced expiratory volume, systolic blood pressure, and creatinine clearance likewise showed a progressive impairment of performance across the age span, subjects in each successive age decade performing less well than the younger ones.

T-scores calculated within age decades for each of the 162 volunteers who had died since the BLSA was initiated were compared with the T-scores of those who were still alive (matched for age). T-scores for those variables whose values increased with

Table VI.3. T-Scores for Live and Dead Groups

		BP	$FEV_{1.0}$	C_{cr}	Glucose
Live					
	Mean	50.2	49.4	49.6	47.8
	SEM	0.33	0.37	0.35	0.46
	N	860	813	772	676
Dead					
	Mean	45.9	44.9	46.8	46.9
	SEM	0.77	1.08	1.01	1.20
	N	162	135	155	100
P		< 0.001	< 0.001	< 0.01	NS

From *Tobin (1981)*

age (blood pressure and glucose) reflected poorer performance. Measurements of blood pressure, forced expiratory volume, and creatinine clearance of those who had died had all been significantly poorer at the time of first testing than the scores of those who lived, and represented on the average a poorer age-adjusted performance for these values (Tab. 3). There was no significant difference between the two groups in glucose concentration at two hours after oral ingestion of glucose, although both were significantly lower than the normative levels.

Most of the deaths in this population have, as was expected, resulted from cardiovascular disease and cancer. As the study proceeds, more specific mortality data will be available to allow comparison of age changes in function commonly occurring with disease entities.

3. Predicting Coronary Events in Asymptomatic Subjects

Despite the much-publicized decline in mortality from CAD over the last decade, it remains the most common cause of death in the United States, especially in the elderly. Although the presence of standard risk factors (smoking, hypertension, hypercholesterolemia) or an abnormal ECG response to exercise is associated with an enhanced risk of future coronary events (angina pectoris, myocardial infarction, or sudden death), these findings have low predictive value in asymptomatic subjects.

Thallium myocardial perfusion scanning is a relatively new procedure for detecting CAD without cardiac catheterization. We examined whether combining ECG and thallium-scanning (TS) during maximal treadmill exercise would improve the predictive accuracy for detecting CAD *(Fleg et al., 1983b)*. Because cardiac catheterization cannot routinely be performed in asymptomatic individuals, especially in those with a normal response to these non-invasive tests for CAD, we monitored these subjects for the subsequent development of cardiac events (CE) indicative of CAD (angina pectoris, myocardial infarction, and sudden death).

Accordingly, 235 clinically healthy BLSA volunteers aged 40 to 92 years received ECGs and TS in conjunction with maximal treadmill exercise. The incidence of subsequent CE in each of four groups defined by the results of exercise ECGs (EE) and TS is shown in the table below.

Groups	*N*	*CE*	*Age*	*Follow-Up (Yr)*
−EE/−TS	179	3	59.1	2.5
−EE/+TS	22	0	58.3	2.8
+EE/−TS	20	1	65.4	2.7
+EE/+TS	14	5	69.3	3.2

The group with +EE/+TS was significantly older than the other three groups and had a markedly higher rate of CEs during the follow-up period. Both age and the combination of +EE/+TS were independent predictors of CEs. Coronary events consisted of angina pectoris in seven subjects and myocardial infarction and sudden death in one subject each.

In this asymptomatic population, the combination of an abnormal EE and an abnormal TS thus identified a group of elderly subjects with a strikingly high incidence of subsequent CE.