Can an Index of Aging Be Constructed for Evaluating Treatments To Retard Aging Rates? A 2,462-Person Study

Richard Hochschild

Hoch Company, Corona del Mar, California.

Biomarkers of aging are needed to evaluate proposed treatments to retard aging rates. At present, the only validated biomarker of aging is maximum life span, which remains impractical for human use. Identification of other biomarkers awaits development of a method of biomarker validation. This paper outlines an approach for this purpose intended for selecting biomarkers usable in humans. Prospective biomarkers are validated and weighted according to their correlation with interventions that, in healthy individuals, influence life span, namely mortality risk factors. A general mathematical method is presented for combining biomarker scores into an index of aging rate. This method addresses problems encountered with the traditional (multiple regression) method of calculating biological age and develops an index termed standardized biological age, SBA. In applying the method to 2,462 office workers, SBA, based on 12 physiological tests under investigation as biomarkers of aging, was found to depend on most of 17 surveyed dietary, exercise, life style, and geographical risk factors for mortality or health, suggesting that many risk factors predict rates of common functional declines with age. The 12 candidate biomarkers of aging in this study differed widely in validity according to the criterion employed. The approach holds promise for assembling an experimentally useful battery of biomarkers of aging.

INTRODUCTION

BY various measures, individuals of the same species age at different rates. This observation underlies the growing interest in biomarkers of aging and their possible use to explore why aging rate differences exist and what influences them. Valid biomarkers of aging are needed to investigate mechanisms of aging and to evaluate treatments for retarding the aging process (Baker and Sprott, 1988; Bowden et al., 1989; Comfort, 1969; Harrison et al., 1982, 1984; Schneider et al., 1982).

Only one biomarker of aging has been validated so far, namely maximum life span, which remains impractical for human use (Masoro, 1989). Nevertheless, it appears possible that many functional, biochemical, or morphological parameters will be found to be effective biomarkers of aging. Hollingsworth et al. (1965), Comfort (1969), and others have suggested that aging rate assessments can be based on measurements of a sufficiently large and diverse test battery of such parameters.

Validating Biomarkers of Aging

The fundamental problem in developing an arsenal of biomarkers of aging has been the lack of a method of biomarker validation. Validation criteria have varied from author to author (e.g., Baker and Sprott, 1988; Harrison, 1982; Harrison et al., 1984; Ingram, 1983, 1984, 1988; Ingram and Reynolds, 1986). Selection of criteria of biomarker validity remains a controversial area. Lack of any independent method of measuring the rate of aging to use as a standard of comparison leaves only a number of indirect approaches, none of which is quite satisfactory.

Proposed biomarker validation criteria include correlation of biomarker scores with (a) mortality rate, (b) interventions that influence mortality rate, and (c) the quality of later life. Few attempts have been made to apply biomarker validation criteria (other than correlation with chronological age, which is an insufficient criterion). In mice, correlation with life span was used to assess the validity of candidate biomarkers of aging by Harrison and Archer (1983) and by Ingram and Reynolds (1986). Correlation with caloric restriction was used for this purpose by Harrison et al. (1984) and Ingram (1984). In human studies, Furukawa et al. (1975) investigated correlation with hypertension, and Webster and Logie (1976) checked correlation with health status as independent variables against which to evaluate a biological age index. Correlation with mortality rate, which has been frequently proposed as a biomarker validation criterion, presents problems in longer-lived species because of its time requirement.

This study tested the criterion that to qualify as valid, a biomarker of aging should be responsive to interventions that influence the natural life span (Bowden et al., 1989; Harrison et al., 1984; Ingram, 1984; Schneider et al., 1982). Examples of such interventions are caloric restriction in rodents, body temperature modulation in poikilotherms, and, in humans, a variety of dietary and life-style practices that are recognized as mortality risk factors. This criterion was tested by applying it, in a study of 2,462 office workers, to 12 physiological tests under investigation as potential biomarkers of aging. Associations were studied between each of the 12 tests individually and CVV, a composite index of risk based on 8 of 17 surveyed mortality and health risk factors. The 12 physiological tests are listed in Table 1,

Table 1. Physiological Tests Administered by H-SCAN
(j is test number)

j	Test	Unit
1	Vibrotactile sensitivity, finger tip	× 1.5 dB
2	Memory, sequence of lamps	# of jumps
3	Forced vital capacity	10 ml
4	Forced expiratory volume, 1-sec	10 ml
5	Alternate button tapping time, 30×	0.1 sec
6	Highest audible pitch	100 Hz
7	Visual accommodation	0.1 diopter
8	Auditory reaction time	msec
9	Visual reaction time without decision	msec
10	Movement time without decision	msec
11	Visual reaction time with decision	msec
12	Movement time with decision	msec

Table 2. Mortality Risk Factors Included in Composite
Validation Variable, CVV

	Risk factor	Weight
1	Cigarettes smoked per day	4
2	High-fat foods, servings per day	2
3	Emphasis on red meat in diet	3
4	Exercise intensity level	2
5	Combined exercise score based on hours/week + years + intensity	2
6	State life expectancy at birth	1
7	State life expectancy at age 45	1
8	Years of school completed	4

the 8 risk factors, in Table 2. All 17 surveyed risk factors are listed in the captions of Figures 13 to 31. In addition, associations were studied between the 17 risk factors individually and SBA, a composite index of aging based on the 12 tests.

The first study of the sensitivity of putative biomarkers of aging to life span interventions, albeit not for the purpose of validating the biomarkers, was reported by Hollingsworth et al. (1965), who studied survivors of Hiroshima some 15 years after radiation exposure. Other investigators have used various physiological tests to study aging effects of life span interventions in mice, such as hypophysectomy (Harrison et al., 1982) and caloric restriction (Harrison et al., 1984). Harrison and Archer (1987) demonstrated the value of this approach by observing that dietary restriction produced different aging effects in different strains of mice.

Rationale for Using Correlation With Risk Factors for Validating Biomarkers

In humans, many proposed risk factors are unreliably correlated with mortality. Even for risk factors that are significantly related to mortality rate, correlations tend to be low, due in part to the large number of risk factors at work and in part to the high variability of age at death, death tending to be triggered by a chance event. A composite index of several risk factors can be expected to have a somewhat higher, but still not particularly high, correlation with mortality. That being the situation, is correlation with risk

factors a useful criterion for purposes of validating biomarkers of aging?

It is difficult to make a case for the use of risk factor data solely as a substitute for mortality data. There is, however, an additional rationale for using correlation with risk factors, and with mortality interventions in general, as a validating criterion. This is that many interventions that, in healthy individuals, biologically alter the natural life span may do so by their primary action on the rate of aging, thereby altering disease probability. Interventions that operate this way will be more closely related to the rate of aging than to the rate of death.

To illustrate this point, suppose that the task is to evaluate respiratory quotient, RQ, as a potential biomarker of aging in rats. RQ, or specifically its 48-hour pattern, is highly correlated with a mortality intervention, caloric restriction, CR (Duffy et al., 1989). Changes in RQ signal changes in main metabolic pathway (carbohydrate vs fatty acid metabolism). The 48-hour range of RQ can be used reliably to separate restricted from ad lib fed animals (Duffy, 1990) except for a small proportion of borderline animals such as those that are self-restricting in the ad lib group (as indicated by low fat-to-lean-body-mass ratio) or overfeeders in the CR group. Much evidence indicates that CR retards the rate of decline of age-associated physiological parameters and delays the onset of degenerative diseases (reviewed by Weindruch and Walford, 1988). If CR's effects on mortality rate are secondary to its effects on aging rate, CR would be a more appropriate variable against which to validate a biomarker of aging than is mortality rate. Correlation with CR offers an additional advantage. According to the data from the above example, the correlation between RQ and CR is higher than the correlation between RQ and mortality, partly because of the high variability of the latter. In this example, correlation with a mortality intervention is probably the more appropriate, and evidently the more sensitive, validation criterion.

The point is not quite so clear cut when applied to human studies, with risk factors in the role of interventions. No risk factor in humans — with some possible exceptions such as the composite of risk factors at work in being a practicing Seventh Day Adventist — has nearly the dramatic relationship with life expectancy that caloric restriction has in rodents. However, that does not invalidate the proposition that correlation with an intervention that biologically alters the natural life span can be a more appropriate and sensitive validation criterion for a biomarker of aging than is correlation with mortality, and that the former does more than serve as an estimate of the latter.

Correlation With Mortality Rate as a Validation Criterion

Mortality rate differences between groups are generally considered to be closely related to aging rate differences, at least in the laboratory, where the variable influences of predatorial activity, accidents, natural disasters, and (with exceptions) infectious diseases are not likely to be important factors. Accordingly, mortality rate, when available and accurate, remains a viable validating variable for biomarkers of aging. In the above example, if RQ were found to be associated with mortality rate among the ad lib fed animals

(or among the CR animals), this would further support the validity of RQ as a biomarker of aging.

Because, in the case of humans, collecting mortality data tends to be a 20-year or longer undertaking, little data exist that can be used to evaluate human prospective biomarkers of aging for correlation with mortality rate. Exceptions to this are the lung function parameters, forced vital capacity, FVC, and forced expiratory volume-1 sec., FEV-1. Much evidence has accumulated from various studies that lower-than-normal values of FVC and FEV-1 during times of health are associated with increased deaths from most major causes, not just lung disease (Anderson, 1979; Ashley et al., 1975; Beaty et al., 1982, 1985; Borkan and Norris, 1980; Higgins and Keller, 1970; Kannel et al., 1980; Kannel and Hubert, 1982; Menkes et al., 1984; Pardee et al., 1981; Petty et al., 1976). Lower-than-normal FVC was found to be associated with earlier mortality in the elderly as well as the young, in both sexes, and in nonsmokers as well as smokers. In some studies, lung function was more predictive of mortality than were smoking status, history of ischemic heart disease, blood pressure, or gender, and was second only to chronological age as a predictor of survival.

Correlation With Quality of Later Life as a Validation Criterion

A third proposed biomarker validation criterion is correlation of candidate biomarker scores with the quality of later life in healthy persons (Harrison, 1988). Humans perceive aging as the decline of characteristics that contribute to life quality. That perception may be sufficient to validate as human biomarkers of aging those physiological functions and appearance features that (a) decline with age in healthy individuals and (b) are commonly (though not necessarily) linked to life quality, such as acuity of hearing, memory, lung function, the speed of reactions and movement, and, to the extent that they undermine self-image, skin laxity, hair graying, and muscle sagging. Such functions and features probably integrate the effects of multiple aging mechanisms.

The 12 physiological functions in Table 1 were selected in part because of their evident role in determining the quality of later life. Age decrements in these functions are associated with some of the more noticeable and troublesome intellectual and physical deficits that commonly accompany human aging. The 12 functions are involved in physically demanding activities such as sports, travel, some occupations, and self-care in old age, and mentally demanding activities such as work, social interactions, and intellectual pursuits.

Critical Evaluation of Prior Mathematical Methods of Calculating Biological Age

A more powerful approach for estimating aging rates than examining data from individual biomarkers is to combine scores from a battery of biomarkers into a single index of aging. When in units of time, such an index is conventionally (not necessarily appropriately) called biological, physiological, or functional age. Attempts to estimate human biological age from physiological tests include those of Dirken (1972), Dubina et al. (1984), Furukawa et al. (1975), Hollingsworth et al. (1965), Nakamura et al. (1982), Voitenko and Tokar (1983), and Webster and Logie (1976).

In all of these studies, (a) an insufficient criterion, namely association with chronological age, was used to validate tests as biomarkers of aging; (b) a method inappropriate for the purpose, namely multiple regression, was used to calculate biological age from test scores; and (c) no attempt was made to reduce unwanted variance in calculating biological age. These three criticisms will now be examined in some detail.

Criticism #1: Association With Chronological Age Is Insufficient as a Validation Criterion

That association with chronological age, CA, is not a sufficient criterion for validating biomarkers of aging has been pointed out by Costa and McCrea (1980) and Ingram (1983, 1988). This follows from the consideration that a hypothetical biomarker that approaches perfect correlation with CA (and would therefore receive the highest validation) could be replaced by a calendar and would be insensitive to differences in aging rates between individuals. Many physiological parameters that are correlated with CA may be irrelevant to aging. For example, degree of baldness increases with age, especially in men, but men who grow bald early do not necessarily show signs of accelerated aging nor is baldness a risk factor for early death.

This does not mean that CA can be ignored in calculating an index of aging from physiological tests that depend on age (as do practically all functional tests that have been proposed as possible biomarkers of aging at present). For age-dependent tests, the relevant information is not the subject's absolute score but the deviation of that score from the norm for the subject's age and sex (based on some chosen norm group). These norms can be determined by regressing test scores on CA (as in this article). If CA were not used in combining scores from age-dependent tests, as was tried by Chodzko-Zajko and Ringel (1987), who substituted principal component analysis and weighted scores according to the resulting eigenvalues, it would constitute an unwelcome source of additional variance.

Criticism #2: Multiple Regression Leads to Errors in Calculating Biological Age

Multiple regression, used in all of the cited studies of human biological age and in many animal studies as well, leads to errors in calculating biological age from physiological test scores because of three, and under some conditions, four problems (Hochschild, 1989a).

(a) Multiple regression calls for an inappropriate step, equating BA to predicted-CA, causing regression to sample mean age. — In all of the cited prior biological age studies, biological age was calculated analogously to the following procedure. Let j be an integer identifying one of n physiological tests whose scores, Y_j, for a subject of chronological age CA are to be combined into biological age BA. The multiple linear regression equation is

$$CA = a + \Sigma b_j Y_j \qquad (1),$$

the summation being over test number j = 1 to n. Running the regression using norm group data produces coefficients a and the set of b_j from which predicted-CA can be calculated for an individual's set of scores as:

Predicted-CA $= a + \Sigma b_j Y_j$ (2).

The final step, and the one which is questioned here, is to define BA as follows:

BA $=$ predicted-CA   (3).

At least since 1972, a troublesome side effect of the above traditional approach has been recognized, namely regression to sample mean age (Costa and McCrea, 1980; Dirken, 1972; Dubina et al., 1984; Ingram, 1983; Nakamura et al., 1982; Webster and Logie, 1976). This effect causes biological age values calculated for persons younger than sample mean age to tend to be too old, and those for persons older than sample mean age, too young. The effect will be considered further under Methods and Results, and is illustrated in Figure 6, which shows that the sample's predicted-CA values cover less of an age range than does sample CA.

Corrections were proposed by Dirken (1972) and Dubina et al. (1984) in which artificial changes in slope and intercept were calculated, separately for each physiological test, to rotate the too-nearly-horizontal regression lines to 45° to the horizontal. This made the problem disappear cosmetically without, however, addressing its cause.

In the above procedure, CA is considered to be the dependent variable and scores the independent variables, on the notion that CA is to be predicted from scores. However, CA does not depend on test scores but on the calendar. Therefore it is meaningless to make CA the dependent variable and to use a regression equation designed to predict it. If predicted-CA has no meaning, it is equally meaningless to equate BA to it. This applies both to simple and multiple regression.

Reversing the direction of the regression should solve the problem and is the approach used in this study. Regressing scores on CA precludes the use of multiple regression, requiring the substitution of individual regressions, one per measured variable. Later in this article, results obtained using the two methods are compared and the issue of whether or not reversing the regression eliminates the problem of regression-to-the-mean is discussed.

(b) Multiple regression automatically and inappropriately weights the contributions of the included tests according to their correlation with CA. — If correlation with CA is not a sufficient validation criterion, then it is inappropriate to weight biomarker scores according to this criterion in combining them into a single index, as occurs with multiple regression. Reversing the regression avoids this by avoiding multiple regression and furthermore permits the optional use of alternative weighting criteria. In this study, weights are based on each test's validation rating.

(c) Multiple regression produces "wrong" signs for some coefficients. — Frequently, when multiple regression is used to calculate biological age, one or more of the resulting regression coefficients in Equation (1) carries a sign which is opposite to the sign of the coefficient for the same physiological test obtained from a simple regression of CA on scores for that test alone. For tests that contribute to the multiple regression with the "wrong" sign, scores char-

acteristic of older individuals in simple regressions correspond, inappropriately, to younger biological ages in multiple regressions, and vice versa, a consequence of applying the method of least squares fit. It doesn't make sense to use computational methods that equate poorer (adult) scores on any test to a younger biological age. The problem is removed by substituting individual regressions of scores on CA.

(d) Multicollinearity can confound results of multiple regression. — Multicollinearity occurs when two or more of the independent variables in a multiple regression are strongly (but not perfectly) related to one another. This condition can lead to unstable and misleading estimates of the multiple regression coefficients. Examples are given by forced vital capacity, FVC, and forced expiratory volume-1 sec., FEV-1. These tests are used in this study and have not infrequently been included in biological age assessments. Not only are FVC and FEV-1 strongly correlated with each other, but each is correlated with a third variable, height, which must also be included on the right side of Equation (1). (Thus multicollinearity would be a problem even if scores from only one lung function test plus height were included in a multiple regression.) Regressing scores on CA individually by test avoids this problem because scores and height can be placed on opposite sides of the equation.

Criticism #3: No Attempt Was Made To Reduce Unwanted Variance in Calculating Biological Age

The physiological tests used in this and prior studies of biological age tend to differ markedly in variance. Table 5 illustrates this for the tests of this study. Standard deviations, *SD*, of the (transformed) test scores cover a 4-to-1 range. A problem arises when scores from tests that differ appreciably in variance are combined into a single index. An unknown part of each test's variance is "useful" in the sense that it represents genuine aging differences between individuals, while another part is "useless," originating from such sources as measurement error, baseline (e.g., inborn) differences between individuals, and short-term variations in test subject responses (e.g., differences in alertness, fatigue, etc.) (Hochschild, 1989b). The problem is that combining scores from tests differing in variance can lead to information loss.

If (transformed) scores from the various tests were simply averaged in calculating a composite index of aging, the tests would contribute to the variance of their mean in proportion to their individual variances, quite the opposite of what is desired if most of the variability is useless rather than useful. This leads to the undesired condition that the useless variance contributed to the mean by high-variance tests will swamp the useful variance contributed by all tests. The result is information loss and a reduction of the ratio of useful-to-useless variance of the mean. This matter seems to have received little prior attention.

Standardization and weighting are variance control techniques. In the approach used in this study, transformed test scores are standardized (i.e., transformed to z-scores) before combining them. Because standardized transforms are in units of standard deviations, they contribute equal (i.e., unity) variance to their mean. Secondly, the approach op-

tionally provides for weighting the contribution which each test makes to the mean according to any selected variable. In analyzing the study data, the weights used are the score each test earns on the biomarker validation criterion. The weighted mean, after being restandardized to put it into units of standard deviations, is designated "standardized biological age," SBA, the index of aging developed in this study. A method is added for calculating BA from SBA.

METHOD

Sample

Seventeen life insurance companies in various parts of the United States contributed a total of 2,462 headquarters office employees as study subjects. Of the 1,485 females, 1,344 were white, 86 black, 23 oriental and 29 other, or declined to say. Of the 977 males, 931 were white, 22 black, 15 oriental, and 9 other, or declined to say. Figures 1 and 2 show the age distributions of males and females, respectively. Test locations were selected to cover a range of different state life expectancies to allow investigation of the relationship between geographical factors and aging rate.

Participating companies are listed in Table 3 together with number of subjects by sex, the state in which testing took place, and male and female life expectancies at birth and at age 45 in each test state. Each company was asked to select randomly up to 200 subjects from headquarters office employees aged 35 and above and to avoid volunteers (who might constitute a self-selected, healthier-than-average group). Participation was not mandatory, but most companies reported that only a small proportion of those asked to participate declined to do so. Use of life insurance company headquarters employees had the advantage of providing a study sample that was widely distributed geographically while being rather well matched educationally, occupationally, and socioeconomically.

Physiological Tests and Instrumentation

Table 1 lists the 12 physiological tests that were evaluated in this study as prospective biomarkers of aging. All had been included in one or more of the cited prior human studies of biological age. They were selected for this study because they (a) met the criteria for age-sensitive tests according to the earlier studies, (b) were judged to meet the "quality of life" criterion, and (c) could be made to fit the format of the automatic instrument I developed to measure them. In addition, two of the selected functions, FVC and FEV-1, were judged to meet the "correlation with mortality rate" criterion.

The instrumentation used to administer the 12 physiological tests and the risk factor questionnaire was the H-SCAN (Hoch Company, Corona del Mar, CA). This instrument operates automatically and requires no operators to be in attendance. Thus data variance associated with differences in the quality of instructions and the degree of motivation provided by operators (important in performance-oriented tests) is eliminated. The instrument fits on a table top, is designed for field use, and can be set up quickly by untrained persons.

Subjects follow simple instructions which appear in large letters on a screen. Instructions are designed particularly to motivate subjects to perform maximally. Messages tend to be friendly, sometimes humorous, and have been revised many times as it was learned, over several thousand testing sessions conducted prior to this study, what mistakes participants were making. The program is designed to catch procedural errors or attempts to cheat and responds appropriately, prompting the participant to correct the error.
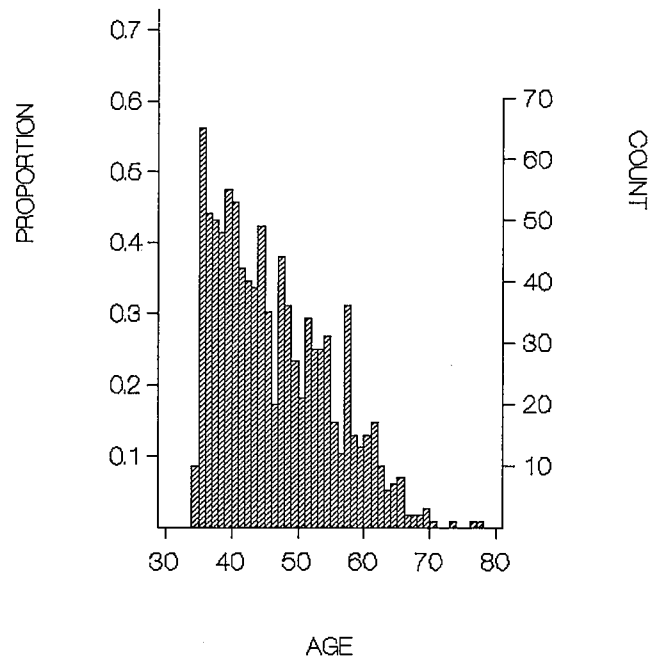

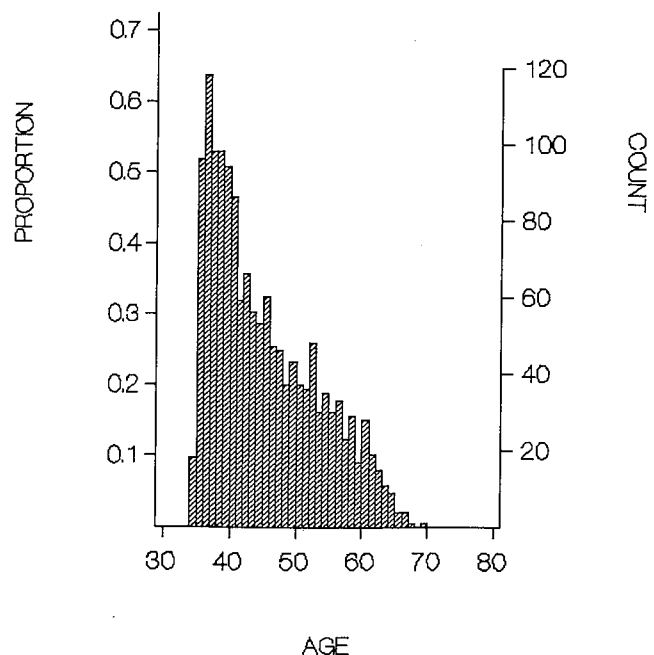Figure 1. Age distribution of 977 male participants.

Figure 2. Age distribution of 1,485 female participants.

Table 3. Employers, Number of Employees Participating by Sex, and State Life Expectancies at Birth and Age 45 by Sex

Company	Number of Subjects		State	State Life Expectancy[a]			
				At Birth		At Age 45	
	Males	Females		Males	Females	Males	Females
Allstate	53	110	Illinois	69.6	77.1	28.8	34.8
American General	58	139	Tennessee	69.2	77.5	28.4	35.1
American States	72	115	Indiana	70.2	77.5	29.0	34.8
American United	69	103	Indiana	70.2	77.5	29.0	34.8
Business Men's Assur.	67	86	Missouri	69.9	77.7	29.0	35.3
Chubb LifeAmerica	75	141	New Hampshire	71.4	78.4	29.7	35.5
Cuna Mutual	77	107	Wisconsin	71.9	78.9	30.2	36.0
E.F. Hutton Life	34	62	California	71.1	78.0	30.1	35.5
Gulf Life	84	112	Florida	70.1	78.0	29.9	36.0
Maccabees Mutual	35	40	Michigan	70.1	77.3	29.0	34.9
National Fidelity	20	43	Kansas	71.6	79.0	30.4	36.4
National Life of Vt.	50	68	Vermont	71.1	78.5	29.5	35.6
Pacific Mutual	82	128	California	71.1	78.0	30.1	35.5
Prudential	45	59	New Jersey	70.5	77.4	29.2	34.7
Southern Farm Bureau	96	106	Mississippi	67.6	76.4	27.9	34.6
United Insurance	39	38	Illinois	69.6	77.1	28.8	34.8
Unity Life	22	28	New York	70.0	77.2	29.1	34.7
TOTALS:	977	1,485					

[a] National Center for Health Statistics (1985)

About 40 minutes are required to complete the 12 physiological tests, 10 minutes more for the risk factor questionnaire. Results are printed out immediately after testing, all data also being stored on floppy disks from which the information was transferred to other computers for analysis.

For each of the 12 physiological tests, procedures, correlations with CA and their significance, variance, and other characteristics have been described by Hochschild (1989b). To gather the data for this study, H-SCANs were sent to the participating insurance companies for several weeks without accompanying personnel. Each company appointed an employee to the part-time task of overseeing the equipment and scheduling participants.

In our experience, the investment in automating the tests produced the following returns: (a) test operators were eliminated as a source of data variance; (b) repeatability of the procedure was improved over nonautomated tests, permitting data collected at different times and locations to be reliably compared and combined; (c) the test format facilitated testing large numbers of subjects efficiently; (d) there were almost no staff costs; (e) there was no need to train operators or send personnel to test sites; and (f) participants found the format attractive, resulting in high rates of participation among randomly selected subjects.

Missing Data

Scores for the physiological tests were listed as missing if a test was skipped at the election of the participant (usually for certain impairments), abandoned by the H-SCAN program due to excessive errors by the participant, or if the score was deleted for being sufficiently beyond expected limits to suggest a physical handicap, an improperly carried-out test, or a testing malfunction. Conditions for score deletions were as follows, with the total number of male/female scores missing given in parentheses: vibrotactile sensitivity score if under 6 dB (38/45), memory if less than 3 jumps (34/126), both lung function scores if vital capacity was under 1 liter or height was outside the working range of 140 to 220 cm (2/8), button tapping time if under 10 sec (24/61), highest audible pitch if under 5 kHz (91/60), and accommodation if under 1.3 diopter (224/475). The large number of missing accommodation scores was due to malfunctioning viewers at 3 of the 17 test sites. No reaction or movement time scores were missing. All 12 test scores were available for 645 males and 881 females. No other outliers were deleted, nor were deletions made for any reason not stated. To avoid bias, all deletions were made before data analysis began.

Risk Factor Questionnaire

Programmed into the H-SCAN to continuously follow the physiological tests, 33 multiple-choice questions appear, one at a time, on the H-SCAN screen, answerable by the push of one of six buttons. No questionnaire answers were missing because the computer does not go on until it gets an answer. Information on two additional factors, namely life expectancy at birth and at age 45 in the states of residence (Table 3), was obtained from the National Center for Health Statistics (1985).

Selection of Mortality Risk Factors In Table 2

Table 2 lists the 8 surveyed risk factors that were combined into composite validation variable, CVV. Their selection and the listed assigned weights were based on evidence of the association of these factors in healthy individuals with subsequent mortality. Weights represent subjective judgment of the strength of such evidence, based largely on literature cited under Results in separate "Background"

subsections for each risk factor. When aspects of the same risk factor were represented twice in the list, the assigned weights (4 for exercise and 2 for state life expectancy) were split 50-50 between the two similar factors. To avoid bias, weights were assigned before data analysis began.

Evidence of relationships to mortality in healthy individuals exists for four other of the surveyed factors which were, however, excluded from CVV for the following reasons: Parental longevity was surveyed by asking whether a subject's natural mother or father or both had died before age 75. For most subjects, one or both parents remained alive and under 75. To have included these two factors would have reduced sample size by about 60% and would have introduced a bias favoring older subjects. A relatively weak predictor of mortality, parental longevity was further weakened in our survey by being limited to a simple determination of whether parents had reached age 75. Race (black/white) is a strong predictor of mortality, but the representation of blacks in our sample (4.5%) was too low for the effective use of that factor. Nevertheless, data for all races were retained for all analyses. The relationship between alcohol consumption and mortality remains ambiguous and apparently disease-specific, being positively associated with cancer, negatively (in moderate quantities) with coronary heart disease, according to the cited evidence.

Composite Validation Variable, CVV

In the description of the mathematical method to follow, virtually all variables are sex-dependent. For this reason, all steps of the analysis are carried out separately by sex, whether or not this is explicitly stated.

To evaluate the 12 physiological tests of Table 1 as biomarkers of aging, the criterion in this study was correlation with composite validation variable, CVV, based on the 8 mortality risk factors of Table 2. The following mathematical method, however, can be generally applied to any selected set of validation variables. Let m be an integer identifying one of N validation variables F_m (such as risk factor scores) which are to be combined into a single variable, CVV. Standardization (transformation to z-scores) is a useful method for putting differently scaled variables all on the same scale.

Designating the standardized transform of F_m as SF_m,

$$SF_m = \frac{F_m - \bar{F}_m}{\sigma_m} \qquad (4),$$

where $\bar{F}_m$ is the mean and σ_m is the standard deviation of F_m for the norm group (in this study, the sample as a whole). Because a distribution of standardized scores always has a mean of 0 and a standard deviation of 1, SF_m indicates how far F_m deviates from the norm group mean in units of standard deviations.

Weighting the SF_m in calculating their mean is appropriate when the included factors differ in quality as validators or when two closely related variables each deserve half-weight. Designated as MVV, the weighted mean is

$$MVV = \frac{\Sigma W_m SF_m}{\Sigma W_m} \qquad (5).$$

The weighting coefficients, W_m, used in this study are listed in Table 2. Summations are over the N values of m.

CVV can then be defined as the standardized transform of MVV:

$$CVV = \frac{MVV - \overline{MVV}}{\sigma_{MVV}} \qquad (6).$$

Having CVV in standardized form makes it possible, for example, to use normal curve scales to express results in percentiles of the norm group when results are nearly normally distributed. This is useful if CVV has additional applications such as estimating relative risk in a Health Risk Appraisal.

Plotting SBA (calculated below) against CVV results in a scatterplot. One way to make such a plot more intelligible is to graph mean SBA by CVV decile (as in Figures 13 and 14).

$$CVV\text{-decile} = INT\left(10 \frac{R}{S} + 1\right) \qquad (7).$$

Here INT(x) is the largest integer that is less than or equal to x, R is the rank of each CVV score, and S is the total number of sample cases. Higher rank and higher deciles correspond to higher risk. The number of cases per decile may not be exactly S/10 because cases having identical CVV scores will carry the same rank and all cases of equal rank are sorted into one decile.

Age/Sex Norms of Physiological Test Scores
In Terms of CA

Standardized biological age, SBA, is the composite index of aging to be developed. It is calculated from a subject's *n* physiological test scores, Y_j, where j is test number. For the 12 tests and over the age range 35 to 65+ of this study, eyeball examination indicates that the scatterplots of the Y_j vs CA do not depart perceptibly from linear. Hence *n* linear regression equations of scores on CA replace the multiple regression of CA on scores of Equation (1):

$$Y_j = d_j + e_j CA \qquad (+ f_j H) \qquad (8),$$

where j = 1 to *n*, H is height, and the term in parentheses is added only when j designates a height-dependent test (FVC and FEV-1 in this study). No multicollinearity results because H is not on the same side of the equation as Y_j.

Test Age, TA_j

Running the regressions of Equation (8) with the norm group data produces regression coefficients d_j, e_j, and as needed, f_j that can be used to calculate predicted-Y_j for age CA on test j:

$$\text{Predicted-}Y_j = d_j + e_j CA \qquad (+ f_j H) \qquad (9).$$

Equation (9) can be used in reverse to estimate the age-equivalent of a participant's score on test, j. This estimate will be termed "test age," TA_j, defined as follows:

$$TA_j = CA \text{ for which the norm group's predicted-}Y_j$$
$$\text{equals obtained score } Y_j \qquad (10).$$

In effect, TA_j is the norm group age that is typical of score Y_j. Substituting TA_j for CA and Y_j for predicted-Y_j in Equation (9) yields

$$TA_j = -\frac{d_j}{e_j} + \frac{1}{e_j}Y_j \quad \left(-\frac{f_j}{e_j}H\right) \tag{11}$$

with the term in parentheses again applying only when j is a height-dependent parameter. Equation (11) can be rewritten

$$TA_j = a_j + b_j Y_j \quad (+ c_j H) \tag{12}$$

using the definitions

$$a_j = -d_j / e_j \tag{13}$$

$$b_j = 1 / e_j \tag{14}$$

$$c_j = -f_j / e_j \quad \text{(lung function only)} \tag{15}$$

where the coefficients d_j, e_j and, as needed, f_j, are generated by the n regressions of Equation (8). Equation (12) represents n equations defining a participant's set of test ages, TA_j. For the 12 tests of Table 1 and the study group, the coefficients a_j, b_j, and c_j are listed by sex in Table 4 and can be used to plot score vs age by test. These coefficients are based on the units of measurement listed in Table 1.

Relative Test Age, RTA_j

Differences in variance by test can be examined by dividing TA_j by CA to remove CA as a variable. Calling the result RTA_j for "relative test age,"

$$RTA_j = \frac{TA_j}{CA} \tag{16}$$

RTA_j expresses the deviation of a subject's TA_j from its age/sex norm, which is 1. For example, $RTA_j = .9$ indicates a score on test j that corresponds, in the norm group, to a test age that is 10% younger than the participant's CA.

Value of Standardization and Weighting in Calculating a Composite Index

An age-independent composite index of aging could be produced by averaging a subject's set of RTA_j. However, this will lead to information loss when the RTA_j differ widely in variance, as they do in this and most other studies of biological age (see Introduction and Table 5). The RTA_j will contribute to the variance of their mean in proportion to their individual variances with the result that the useful variance (variance related to aging rate differences between subjects) contributed by each test will be swamped by the useless variance contributed by high-variance tests.

This potential for information loss can be limited by standardization and weighting. The standardized transforms, STA_j, of the RTA_j will each contribute equal (unity) variance to their unweighted mean.

$$STA_j = \frac{RTA_j - \overline{RTA_j}}{\sigma_j} \tag{17}$$

As usual, $\overline{RTA_j}$ and σ_j are the norm group mean and standard deviation of RTA_j for test j.

Even though the STA_j all have the same variance, they still differ in ratio of useful-to-useless variance. Variance in the composite index can be further reduced by weighting the STA_j according to a selected biomarker validation criterion, allowing relatively more valid biomarkers to contribute more heavily. Designating the weighted mean as MSTA,

$$MSTA = \frac{\Sigma \, w_j \, STA_j}{\Sigma \, w_j} \tag{18}$$

The summations are over the n values of j and the weights w_j are "biomarker validation scores" earned by the individual tests. In this study, these weights were the Pearson correlations, r, between STA_j and CVV as listed in Table 6. (It might have been as appropriate to use r^2, the F-ratios, or some other parameter describing the relationship.)

Standardized Biological Age, SBA

Standardizing MSTA produces the composite index of aging employed in this study, termed standardized biological age, SBA. Thus

$$SBA = \frac{MSTA - \overline{MSTA}}{\sigma_{MSTA}} \tag{19}$$

Although it is not in units of years, SBA is a particularly useful way to express an individual's biological age or comparative aging rate. SBA expresses, in units of standard deviations, how far a subject's composite test score deviates from the norm for the subject's age and sex. At any age, SBA has a norm group mean of 0 and standard deviation of 1. Positive values of SBA indicate subjects who are aging more rapidly than average by the included tests, negative values corresponding to slower than average aging. Contrary to the traditional concept of biological age in units of time, SBA is age-independent. Thus it facilitates comparisons

Table 4. Norm Coefficients, by Sex, for Calculating Test Age, TA_j, Using Equation (12).
Applicable Units Are Given in Table 1. Height is in cm.

j:	1	2	3	4	5	6	7	8	9	10	11	12
Males												
a_j	176.6	141.9	−66.54	−34.71	−144.1	113.8	109.8	−209.0	−327.1	−64.44	−273.7	−79.31
b_j	−6.894	−9.437	−0.223	−0.275	0.927	−0.530	−1.475	1.947	1.833	0.713	1.298	1.034
c_j			1.239	1.058								
Females												
a_j	170.0	149.4	−55.44	−38.99	−140.6	127.7	111.0	−139.8	−358.3	−75.21	−227.5	−58.65
b_j	−6.31	−11.74	−0.280	−0.329	0.834	−0.619	−1.656	1.286	1.925	0.651	1.129	0.688
c_j			1.201	1.075								

between experimental groups comprising individuals of mixed ages.

Our data indicate that SBA distributions are nearly normal over the age range of the study sample, essentially 35 to 65. Therefore, to good approximation, the usual normal curve percentile scales apply (e.g., Jensen, 1980), whereby 68.26% of individuals selected randomly from the norm population will have SBA scores between -1 and $+1$ *SD*, and 95.44% of individuals will score between SBA $= -2$ and $+2$ *SD*. SBA < -1.28 *SD* identifies individuals testing in the youngest (conventionally designated as the highest) decile for their age. The other decile boundaries occur at SBA $= -.84, -.53, -.25, 0, .25, .53, .84$, and 1.28 *SD*.

Mean standardized biological age, $\overline{\text{SBA}}$, can be calculated for mixed-age groups characterized by different levels of a factor of interest. For example, in Figures 13 through 31, SBA is the variable used to illustrate the relationship between 12-function aging rate and 17 surveyed health risk factors.

Biological Age, BA

Clarity is gained in describing $\overline{\text{SBA}}$ differences by expressing these differences, additionally, as differences in biological age, BA, in units of years at some given age. Therefore a method of calculating BA is needed. In deriving BA from SBA and CA, we are free to select any desired norm group mean, $\overline{\text{BA}}$, and standard deviation, σ_{BA}, for the following reason. Let z be the standardized transform of X:

$$z = \frac{X - \overline{X}}{\sigma} \qquad (20).$$

This transformation can also be run in reverse. Any standardized variable, z, can be transformed into a new variable, X, having any selected mean, $\overline{X}$, and standard deviation, σ,

$$X = \sigma z + \overline{X} \qquad (21),$$

or, in terms of SBA and BA,

$$BA = \sigma_{\text{BA}} \, SBA + \overline{BA} \qquad (22).$$

Here $\overline{BA}$ and σ_{BA} can be given any appropriate values. The situation is similar to the one encountered in defining the Stanford-Binet IQ scale, which also required the selection of an arbitrary mean and standard deviation. In that case, raw test scores, X, obtained on an IQ test by a norm population were first standardized to z-scores by Equation (20). IQ was defined in terms of these z-scores to have a mean of 100 and standard deviation of 16 (Jensen, 1980),

$$IQ = 16 z + 100 \qquad (23).$$

The choice of $\overline{BA}$ is not an issue. The concept of biological age implies that the average-scoring individual in the norm group should test his or her chronological age, i.e.,

$$\overline{BA} = CA \qquad (24).$$

However, there is no such obvious choice in selecting σ_{BA}. The preceding standardization and weighting steps wiped out any guidelines the data might have provided for determining σ_{BA}. As a first condition on σ_{BA}, it seems reasonable that it should be a function of CA because no postulated mechanism of aging can account for groups of young and old

individuals having the same σ_{BA}. The simplest expression of this condition is to make σ_{BA} a linear function of CA with a zero intercept at CA $= 0$, i.e.,

$$\sigma_{\text{BA}} = c \, CA \qquad (25).$$

As a second condition, the constant c remains to be selected. There is nothing critical about the choice of c, any more than there was about the standard deviation of IQ scores. But it seems reasonable that the distribution of biological ages around the mean age at death of the population under study should be similar to the distribution of deaths around the mean age at death. The latter is reproduced in Figure 3, which shows deaths at each age per 100,000 live births from U.S. Life Tables for 1979–81 for males and females combined (Vital Statistics of the United States, 1980).

The curve deviates appreciably from normal, having a marked negative skewness, and offers no more than a starting point for selecting c. Mean age at death is 73.4 years and the distribution has a standard deviation of 18.1 years. Substituting these values, respectively, for CA and σ_{BA} in Equation (25) gives c $= 0.25$. To compensate for the skewness, the contribution of deaths below age 35 and the greater homogeneity of typical study populations, I revised this value downward, setting c $= 0.16$ for human populations similar to the one of this study. Using Equations (22), (24), and (25), the definition of biological age for human studies becomes

$$BA = CA \, (0.16 \, SBA + 1) \qquad (26).$$

How this choice of c works out in practice is illustrated in Figure 4. Here the frequency distribution of BA values obtained by all females in the study aged 45 $\pm$ 1 year, and for whom all 12 scores were available, is shown superimposed on the distribution of U.S. deaths. (For this illustration, all w_j in Equation (18) were set $= 1$.)

By selecting other values of c, custom tailoring of BA distributions is possible. Because the BA values are nearly normally distributed, normal curve percentile scales apply, and c can be selected so that a specified percentage of norm population members whose age is CA will obtain BA values
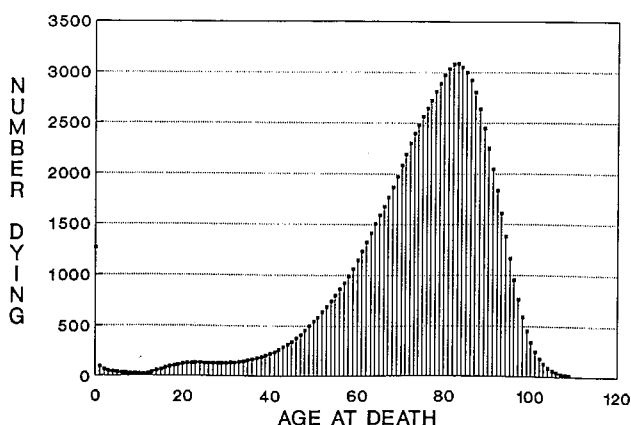

Figure 3. Distribution of U.S. total deaths per 100,000 live births, by age, males and females combined, 1979–81 (Vital Statistics of the United States, 1980). Mean age at death is 73.4 years. The distribution has a standard deviation of 18.1 years and a marked negative skewness.
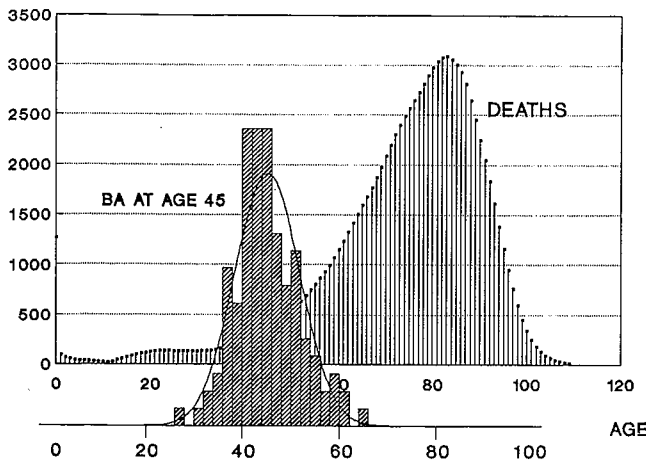
Figure 4. Frequency distribution of biological age, BA, scored by all females in the study aged 45 ± 1 year for whom all 12 test scores were available, superimposed on the distribution of U.S. total deaths by age.
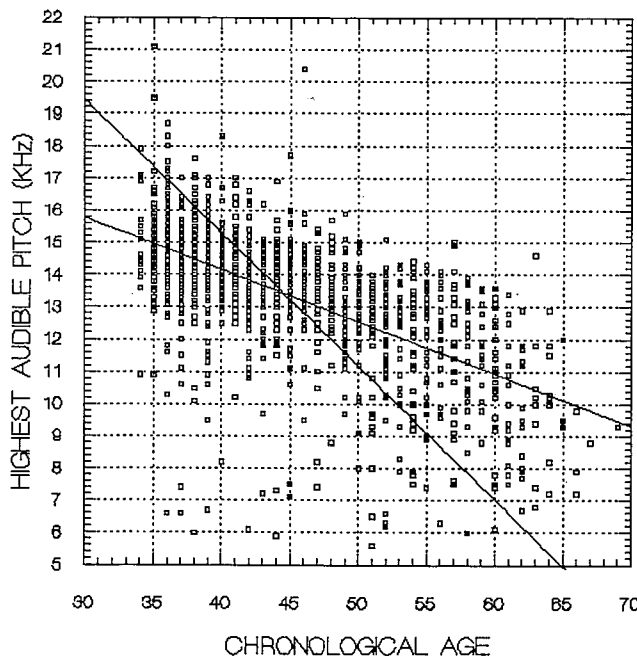

Figure 5. Highest audible pitch scores (in KHz) versus chronological age for females. Steeper line is the regression line of CA on scores, Equation (1). Less-steep line is the regression line of scores on CA, Equation (8).
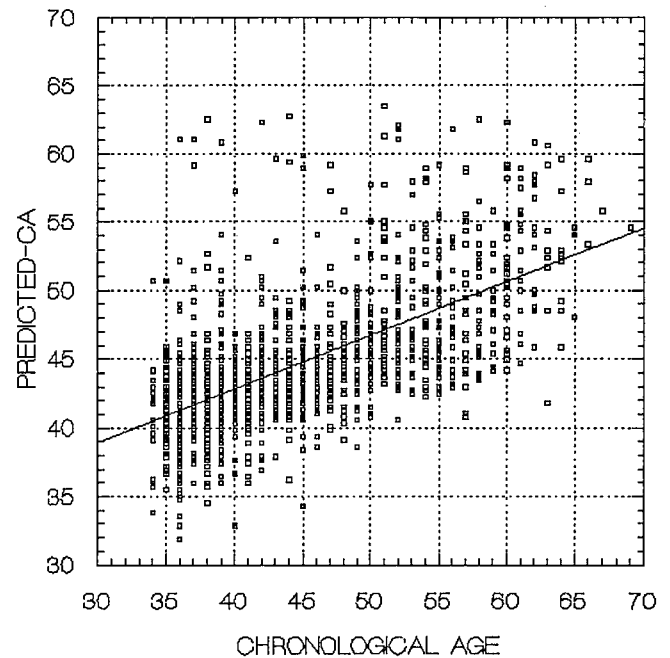

Figure 6. Predicted-CA, based on the steeper line in Figure 5, versus CA for each data point in Figure 5. By failing to pass through coordinates 30,30 and 70,70, the regression line of predicted-CA on CA illustrates regression to sample mean age.

that are within a specified number of years of CA. For example, in the distribution of 45-year-olds in Figure 4, 68.3% (±1 SD) of the norm population will obtain BA values that are within 7.2 years (0.16 × 45) of age 45. If c were assigned a value of 0.08 instead, then 68.3% of these 45-year-olds would have tested within 3.6 years of their age.

RESULTS

Illustration of Regression to Sample Mean Age

Using as an example only female data for test j = 6, highest audible pitch, Figures 5 to 8 illustrate results by the

traditional method and by the method of this article, that is, before and after reversing the regression. Figure 5 is a scatterplot of highest audible pitch scores, Y_6 (in units of kHz), vertically vs CA horizontally. Of the two lines in Figure 5, the steeper represents the traditional direction of the regression (CA on scores) as in Equation (1) with j = 6 and the summation omitted because only one test is involved. The less steep line represents the reversed regression (scores on CA) of Equation (8).

Using the steeper line to convert an obtained score to predicted-CA indicates that younger subjects tend to test too old and older subjects too young, a characteristic of the problem of regression to sample mean age. This effect is illustrated when predicted-CA is plotted vertically against actual CA horizontally, as in Figure 6. For each data point in Figure 6, predicted-CA is the X-axis coordinate of the point on the steeper line in Figure 5, whose Y-axis coordinate is the subject's highest audible pitch score. The distribution of points in Figure 6 is wider than it is high. The line shown in Figure 6 is the regression line of predicted-CA on CA (not the reverse because CA is considered as the independent variable). This line fails to pass through coordinates 30,30 and 70,70 as it would if regression to sample mean age were absent.

Results When Regression Is Reversed

In contrast, the method of this article is based on the regression of test scores on CA, Equation (8), which defines the less steep line in Figure 5. The scatterplot in Figure 7 of test age, TA_6, vertically vs CA horizontally is generated using Equation (12) and the coefficients listed for j = 6 for
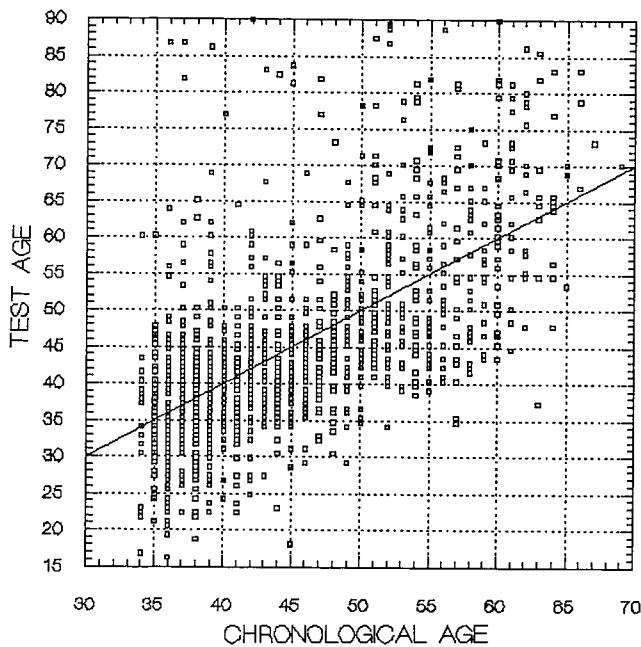
Figure 7. Test age, TA_j, based on the less steep line in Figure 5, versus CA for each data point in Figure 5. By passing through coordinates 30,30 and 70,70, the regression line of TA_j on CA indicates absence of regression to sample mean age.

Figure 8. Mean highest audible pitch for females in 15 age categories: 35, 36, 37, 38, 39, 40, 41–42, 43–44, 45–46, 47–48, 49–50, 51–52, 53–54, 55–57, and 58–61. Regression line of mean pitch on median category age almost coincides with less-steep line in Figure 5.

females in Table 4. For each data point in Figure 7, test age is the X-axis coordinate of the point on the less steep line in Figure 5, whose Y-axis coordinate is the subject's highest audible pitch score.

In comparison to Figure 6, Figure 7 illustrates that regression to sample mean age is absent when test ages are predicted by the less-steep line. (Note the contraction of the vertical scale in Figure 7. Some points lying beyond the upper and lower limits were cut off.) The line shown in Figure 7 is the regression line of TA_6 on CA. It satisfies the condition that it should coincide with the line between coordinates 30,30 and 70,70. (The reverse regression line, CA on TA_6, would regress away from sample mean age, but is not an appropriate representation of the scatterplot in Figure 4 because CA is the independent variable.)

Has Regression to the Mean Been Eliminated?

Predicting scores from CA rather than CA from scores does not do away with regression to the mean, which occurs whenever $r < 1$. In place of regression to mean age, regression to mean score occurs instead. Lower-scoring individuals, on average, have chronological ages that the regression line of scores on age predicts to be associated with somewhat higher-scoring individuals. The average predicted score of lower-scoring individuals (mostly the older portion of the sample) is too high (typical of younger individuals), and vice versa. Does this mean that results calculated for older persons still tend to make them score younger on average, and vice versa, as before?

Figure 7 illustrates that this does not happen, the explanation being that $r = 1$ (the condition for no regression to sample mean age) applies to the regression of estimated
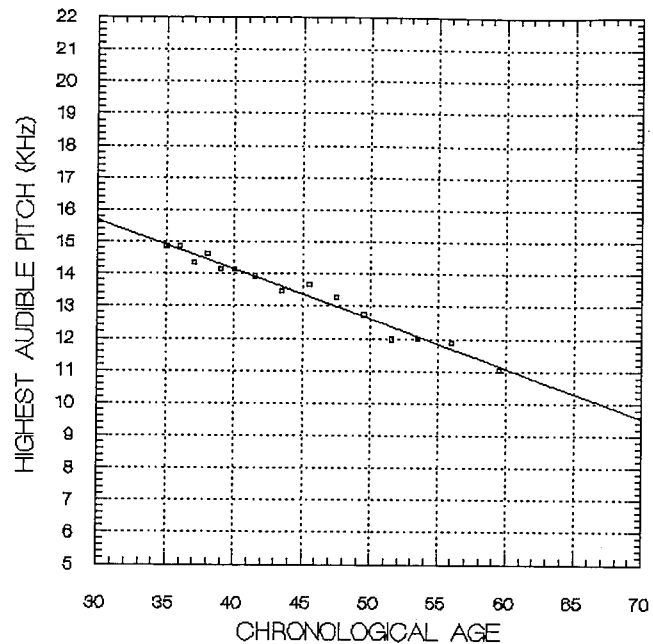
mean scores at each CA on CA. Estimated mean score at each CA is given by Equation (9), which is the equation of the less-steep line in Figure 5. Predicted-Y_6 is an estimate of sample-mean-Y_6 at each CA. This is demonstrated in Figure 8, which shows a regression line fit through 15 plotted points that are the actual sample means of Y_6 in 15 CA categories (category age range being adjusted to contain at least 80 points), namely 35, 36, 37, 38, 39, 40, 41–42, 43–44, 45–46, 47–48, 49–50, 51–52, 53–54, 55–57, and 58–61. The regression line of Figure 8 nearly coincides with the less-steep line in Figure 5.

In Figure 8, the correlation between the plotted means and CA is close to $r = 1$. For the less steep line of Figure 5, it reaches $r = 1$ (because the means are predicted) and regression to sample mean age is absent. It has been replaced by regression to sample mean score. (If the objective were, instead, to predict scores from age, it would be appropriate to trade regression to mean score for regression to mean age and use the steeper line in Figure 5. This can be accomplished by applying Equations (8) through (11) analogously, replacing age by score and score by age everywhere.)

Range of Variances of the RTA_j

For the 12 tests in Table 1, the standard deviations of the RTA_j were found to differ by more than 4-to-1 (see Table 5). The two extreme distributions for this study are illustrated in Figures 9 and 10. Both are histograms of number of cases (as "count" and "proportion" per standard unit) vertically vs RTA_j horizontally. The narrowest frequency distribution, Figure 9, with $\sigma = 0.226$, occurred for highest audible pitch (j = 6) for females. The widest distribution, Figure 10, with $\sigma = 1.075$, occurred for visual reaction time without deci-

Table 5. Standard Deviations of Relative Test Ages, RTA_j, for Test j = 1 to 12

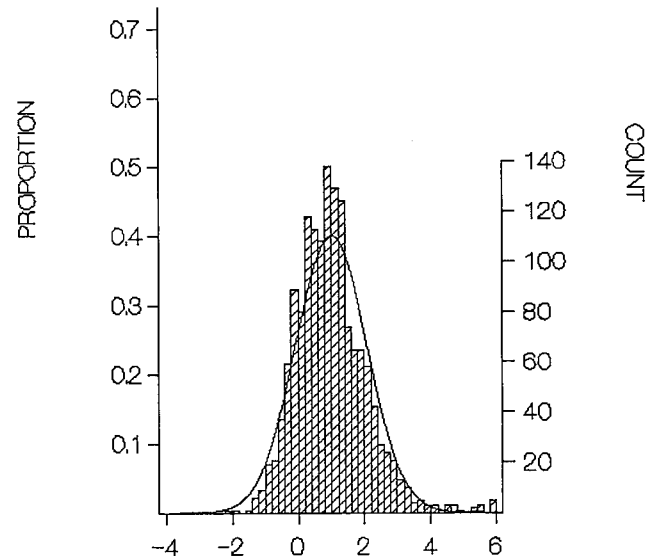| j:	1	2	3	4	5	6	7	8	9	10	11	12
Males SD	0.683	0.749	0.410	0.421	0.708	0.262	0.631	0.799	0.905	0.689	0.869	0.775
Females SD	0.636	0.890	0.374	0.366	0.660	0.226	0.660	0.809	1.075	0.706	0.788	0.609
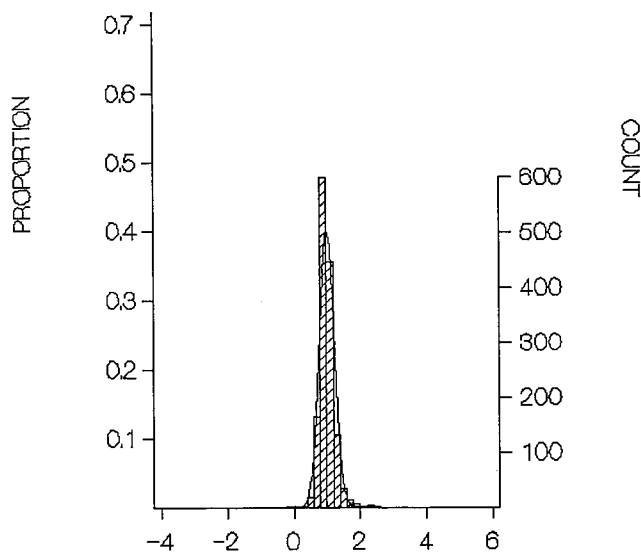
RTA FOR HIGHEST AUDIBLE PITCH

Figure 9. Frequency distribution of RTA_6, the relative test age for highest audible pitch, in females, before standardization. Mean = 1. $SD = 0.226$ (narrowest distribution in study). Normal curve with same mean and SD is superimposed.

RTA FOR VISUAL REACTION TIME WITHOUT DECISION

Figure 10. Frequency distribution of RTA_9, the relative test age for visual reaction time without decision, in females, before standardization. Mean = 1. $SD = 1.075$ (widest distribution in study). Normal curve with same mean and SD is superimposed.

sion (j = 9) also for females. (Superimposed on each frequency distribution for comparison is the normal curve based on the sample's mean and standard deviation.)

Low variance does not necessarily identify a "better" biomarker of aging. An example is highest audible pitch. While it has the lowest variance for both males and females among the 12 tests of this study, highest audible pitch was inferior to 2 other tests in the case of males and to 3 other tests in the case of females in correlation with CVV (see Table 6).

Effect of Standardization

Two methods are used in this study to control the variance contributed by the individual tests to SBA, namely standardization and weighting. How standardization contributes to variance control is illustrated in Figures 11 and 12. Corresponding to Figures 9 and 10 respectively, these figures show the frequency distributions of the STA_j, the standardized transforms of the RTA_j, given by Equation (17), for the same two sets of data, j = 6 and j = 9 respectively, for females. Number of cases is plotted vertically vs STA_j horizontally, in units of SD. Both distributions have $SD = 1$ and a mean of 0. The STA_j contribute equal variance to their unweighted mean.

Weights Based on Correlation Between STA_j and CVV

As a second method of variance control, Equation (18) provides for the weighted mean of the STA_j, where the weights w_j can be based on some measure of the validity of each test as a biomarker of aging. The selected validation and weighting criterion is the Pearson correlation between STA_j and CVV. For each test, this correlation is listed as r in Table 6, along with number of subjects, N, and probability, p, for each relationship. The high values of N account for the large number of significant p values despite low r values. Whether or not the r values, and the variation of CVV explained by each test, are trivial is considered in the Discussion. The W% line was added to show more clearly the percentage weight that each test contributes to SBA.

How the weighting step affects results will be illustrated below in comparisons between Figure 13 (weighted case) and Figure 14 (unweighted case) and between columns marked "weighted" and "unweighted" in Table 7.

Explanation of Figures 13 to 31

Relationships between mean standardized biological age, SBA, and the risk factors are presented as bar graphs in Figures 13 through 31. In all of these figures, SBA is the vertical axis coordinate, while the figure captions label the
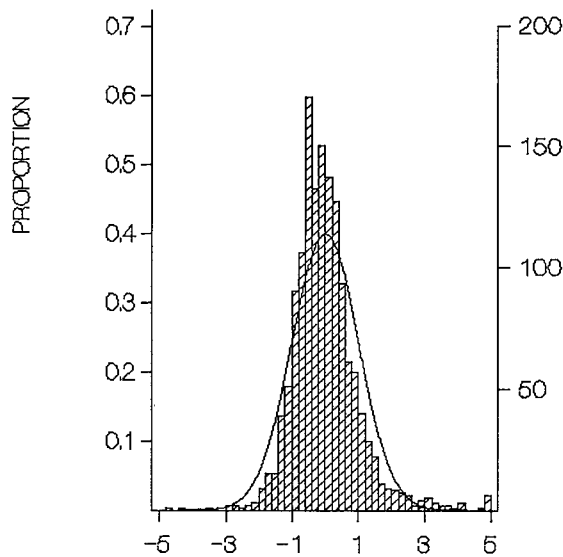
STA FOR HIGHEST AUDIBLE PITCH

Figure 11. The frequency distribution of Figure 9 after standardization. STA_6, the standardized transform of RTA_6, has mean $= 0$, $SD = 1.0$.
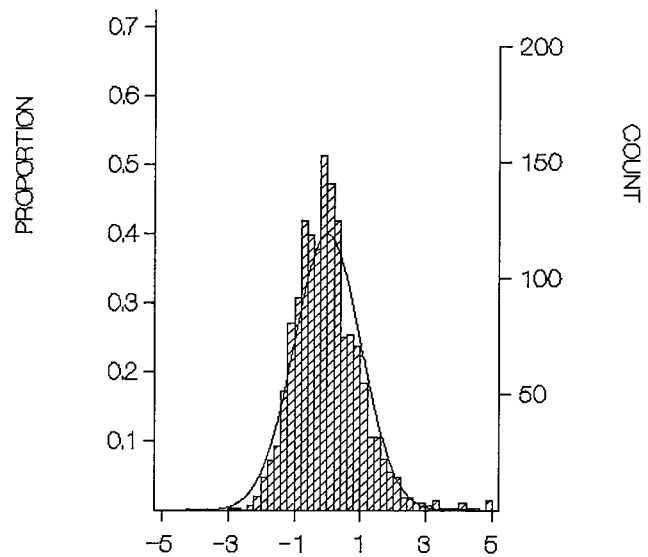
STA FOR VISUAL REACTION TIME WITHOUT DECISION

Figure 12. The frequency distribution of Figure 10 after standardization. STA_9, the standardized transform of RTA_9, has mean $= 0$, $SD = 1.0$.

Table 6. Number of Subjects, N, Correlation, r, and Probability, p, for Linear Regression of Composite Validation Variable, CVV, on Standardized Test Age, STA_j, for Each Test, j

j:	1	2	3	4	5	6	7	8	9	10	11	12
Males												
N	912	917	945	945	924	863	731	947	947	947	947	947
r	.028	.027	.220	.272	.094	.136	.018	.097	.068	.081	.037	.061
p	.405	.413	.000	.000	.004	.000	.618	.003	.037	.013	.260	.059
W%	2	2	19	24	8	12	2	9	6	7	3	5
Females												
N	1355	1279	1387	1387	1337	1338	956	1394	1394	1394	1394	1394
r	.016	.103	.216	.227	.128	.102	.091	.063	.166	.096	.093	.059
p	.550	.000	.000	.000	.000	.000	.005	.019	.000	.000	.001	.028
W%	1	8	16	17	9	8	7	5	12	7	7	4

Note. The tabulated values of r are used as weighting coefficients to calculate SBA. W% is the % of total weight contributed by each test to SBA.

horizontal axis and identify the plotted factor levels. Because the risk factor questionnaire offered from 2 to 6 possible answers per question, horizontal axes have from 2 to 6 levels per sex except for Figures 19 and 30, which are based on combinations of factors. Codes, when used to identify factor levels, are explained in the figure captions. Data for males are grouped in the left half of each chart, data for females are at right. Error bars designate standard error of the mean of SBA. The number of subjects, N, represented by each bar, is shown centered above the corresponding bar across the top of each chart (line beginning "N = ").

No questionnaire answers were missing for any participant. However, in Figures 17, 23, 27, and 28, bars for "other" replies were omitted for clarity. Accordingly, the numbers listed across the top of these figures add up to fewer than the 645 males and 881 females for whom a complete set

of all 12 physiological test scores was available (see "Missing data").

A Walk Through Figure 15

Figures 13 to 32 attempt to condense more than the usual amount of information into a small space and, for this reason, an examination of one representative figure may help to clarify what is being shown in all figures.

The data for males in Figure 15 will be used as the example. Of the 645 males for whom all 12 test scores were available, 518 were current nonsmokers. 518 is the first entry on the line labeled "N = " above the chart. It is the entry above the bar labeled "0," designating no packs smoked, on the horizontal axis. This first bar extends downward, indicating a negative (younger than average) $\overline{SBA}$ and a lower-than-average aging rate. The other bars for males

extend upward, indicating that a positive or older-than-average $\overline{SBA}$ and a higher-than-average aging rate was scored by men who smoke 0.5, 1, 1.5, and 2 packs per day on average as labeled on the horizontal axis.

The first bar extends downward to $\overline{SBA} = -0.08\ SD$. Using normal curve percentiles, this puts nonsmoking males at the 47 percentile level in terms of aging rate (which is as close to 50% as it is because 518/645 or 80% of sample males were nonsmokers). Using Equation (26) to convert to units of years and designating mean biological age, for example, at CA = 45 as $\overline{BA}_{45}$, 45-year-old nonsmoking men scored $\overline{BA}_{45} = 44.4$ years. In contrast, the 40 men who said they smoke 1 pack per day on average (third bar) scored $\overline{SBA} = +0.33\ SD$, which gives $\overline{BA}_{45} = 47.4$ years, corresponding to a difference of 0.41 SD (3.0 years at age 45) between male nonsmokers and 1-pack-a-day smokers. For 70-year-old men, this difference in mean biological age, $\overline{BA}_{70}$, corresponds to 4.6 years. Results for women are generally similar, as shown in the right half of Figure 15.

Explanation of Table Below Each Figure

A small table below each figure shows the relationship, by sex, between SBA (not $\overline{SBA}$) and the horizontal-axis factor levels, looking at the factor levels in two ways: (1) as a continuum and (2) categorically. Data in the left half of each table refer to males, data in the right half, to females.

The line labeled "REGR'N" lists the Pearson correlation, r, for the linear regression of SBA on the horizontal-axis factor (designated in the figure caption). Also shown is the F-ratio which tests the significance, listed as p, of the relationship between SBA and the horizontal-axis variable. All numerical data in these tables (as well as the bar graphs) are based on the w_j in Equation (18) being set equal to the r values in Table 6 except for data applying to the unweighted case in Figure 14 and Table 7. As a point of interest, one or two asterisks, representing $p < .05$ and $p < .01$, respectively, have been added to the p-values to indicate relationships that are significant without this added weighting step, that is for all $w_j = 1$ in Equation (18).

Regression analysis is appropriate for those questions for which the factor levels are more or less continuous (e.g., 0, 0.5, 1, 1.5, and 2 packs of cigarettes smoked per day). Regression is less appropriate, although it can be used, for risk factor questions that produce categorical factor levels (e.g., dietary preference for chicken and fish over red meat), in which case analysis of variance is the more suitable test. The table below each figure gives results for both regression and one-way ANOVA, with contrasts, for all factors regardless of their continuous or categorical nature. In most cases, both are applicable, regression being the more sensitive test.

The data on the ANOVA line were based on estimation of contrasts (Neter et al., 1985), in order to determine the probability that the SBA means were equal for 2 selected factor levels or, more commonly, for 2 selected groups of factor levels. In each figure, for each sex, the factor levels (bars) that are grouped together for the ANOVA contrast are designated by solid lines overscoring the number of subjects per bar across the top of each figure (the "N = " line).

For example, in Figure 15 the first level (nonsmokers) is contrasted with the remaining 4 levels (smokers) for each sex. The data on the ANOVA line are the corresponding F-ratio and probability, p, testing the significance of the selected contrast. As before, one or two asterisks on the p-value indicate that the equivalent contrast in the unweighted case was significant at $p < .05$ or $p < .01$ respectively. For factors having only two levels, results for regression and ANOVA are identical and are combined into one line.

Exclusion of Risk Factors From Weighting Steps

Absence of an asterisk on any p-value in the tables under Figures 15 through 31 indicates a result that is not significant for the unweighted version of SBA. This occurred for 3 of the 8 risk factors listed in Table 2 that make up CVV, namely the first 3 (Figures 15, 16, and 17). For these risk factors, it is necessary to determine whether an observed significance in the weighted case is dependent on inclusion of that risk factor in CVV. This was done by recalculating SBA substituting new weights in Equation (18), the new weights being set equal to the correlation between STA_j and a new value of CVV calculated by excluding, in Equation (5), the risk factor in question. These results are reported in the text describing Figures 15 through 22, which apply to the 8 risk factors in CVV.

Standardized Biological Age vs Mortality Risk Decile

Because it comprises 8 mortality risk factor scores, CVV can be interpreted as an index of composite mortality risk. In Figures 13 and 14, CVV scores have been ranked into 10 approximately equally populated "mortality risk deciles" using Equation (7). Decile 1 represents the 10% of the sample at lowest risk. The height of the bars shows $\overline{SBA}$ per decile, negative values representing younger-than-average results and positive values, older. A significant relationship was found between $\overline{SBA}$ and mortality risk decile whether or not SBA was calculated by the routine weighted method (Figure 13) or all weights in both Equations (5) and (18) were set = 1 (Figure 14). For men, the difference in biological age between the lowest and highest mortality risk decile for the weighted case is about 1.0 SD (equivalent to 7 years at age 45). For women, it is 1.2 SD (9 years at age 45).

Comparing these figures illustrates the degree by which weighting sharpened the relationships. Referring to the tables below the figures, weighting increased the correlation between SBA and mortality risk decile for males from 0.118 to 0.260. For females, the increase was from 0.230 to 0.309. While not listed in the tables, weighting increased the correlation between $\overline{SBA}$ (the value designated by the bars) and mortality risk decile from 0.694 to 0.931 for males, and from 0.914 to 0.934 for females.

SBA vs CVV — Table 7

Some information is lost in exchange for the graphical clarity gained in ranking CVV into deciles. Table 7 gives the results of the regression of SBA (not $\overline{SBA}$) on CVV, listing number of subjects, N, correlation, r, F-ratio, and p-value, by sex, for the weighted and unweighted cases. Weighting increased the correlation between SBA and CVV from 0.119 to 0.277 for males and from 0.238 to 0.321 for females.

UNWEIGHTED

$\overline{SBA}$

Figure 14. Unweighted mortality risk decile.

	r	F-ratio	p
REGR'N	.118	8.9	.003
ANOVA		4.1	.04

	r	F-ratio	p
REGR'N	.230	46.9	.0000000
ANOVA		30.0	.0000001

$\overline{SBA}$

Figure 16. High-fat foods per day (average number of servings).

	r	F-ratio	p
REGR'N	.105	7.2	.007
ANOVA		4.2	.04

	r	F-ratio	p
REGR'N	.013	0.1	N.S.
ANOVA		0.6	N.S.*

$\overline{SBA}$

Figure 13. Mortality risk decile.

	r	F-ratio	p
REGR'N	.260	45.5	.0000000
ANOVA		25.7	.0000005

	r	F-ratio	p
REGR'N	.309	88.4	.0000000
ANOVA		54.7	.0000000

$\overline{SBA}$

Figure 15. Cigarettes smoked per day (packs).

	r	F-ratio	p
REGR'N	.154	15.6	.00009
ANOVA		15.0	.0001

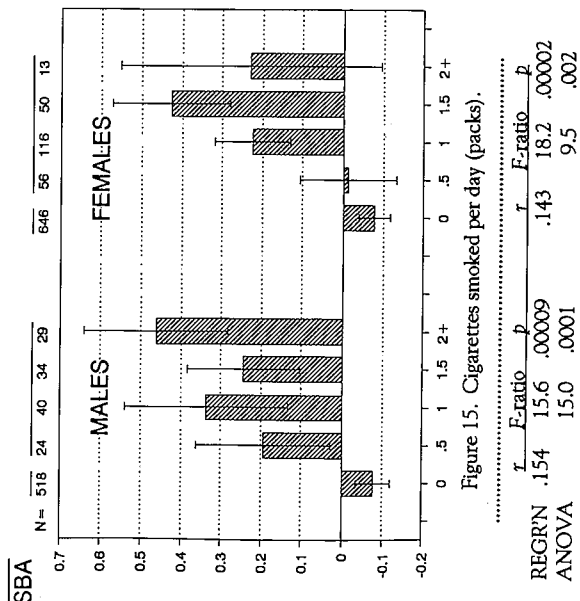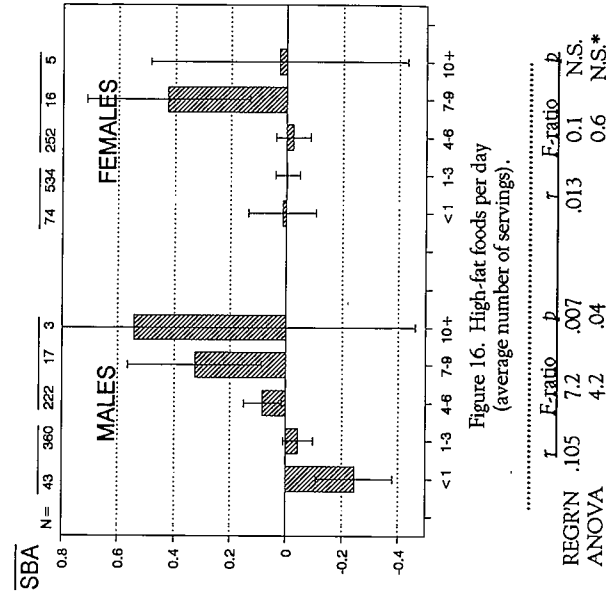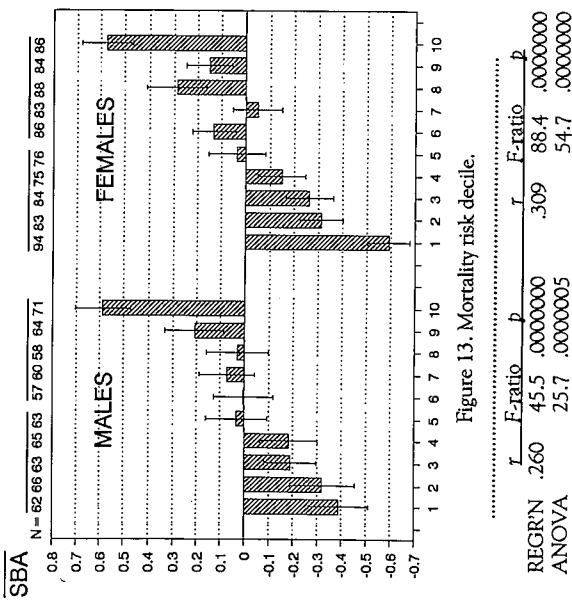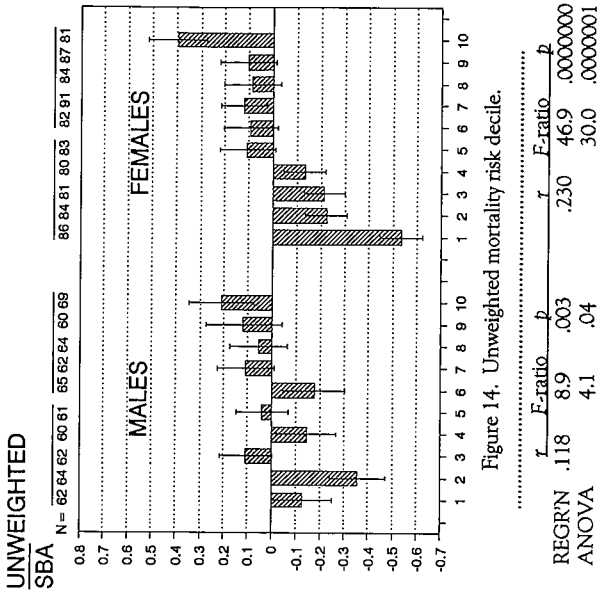|	r	F-ratio	p
REGR'N	.143	18.2	.00002
ANOVA		9.5	.002

Table 7. Number of Subjects, N, Corelation, r, F-ratio and
p-value for Linear Regression of Standardized Biological
Age, SBA, on Composite Validation Variable, CVV, for
Weighted and Unweighted Case as Described in Text

	Males		Females	
	weighted	unweighted	weighted	unweighted
N	629	629	839	839
r	.277	.119	.321	.238
F-ratio	52.2	9.0	96.1	50.1
p	.0000000	.003	.0000000	.0000000

Cigarette Smoking — Figure 15

In this and the following sections, 17 individual risk factors are examined briefly for (1) how they relate to disease and death (subsections titled *"Background"*) and (2) how they relate to calculated biological age (subsection *"Results"*).

Background. — According to the Surgeon General's annual report on smoking (U.S. Dept. of Health and Human Services, 1989), smoking is responsible for 1 out of every 6 deaths in the United States. Many studies have identified cigarette smoking as strongly associated with subsequent all-cause mortality. These include the 6,928-person Alameda County, CA, study (Breslow and Enstrom, 1980; Wingard et al., 1982), the 5,209-person Framingham Study (Kannel, 1971; Pinsky et al., 1987), and the 5,041-person Honolulu Heart Program (Benfante et al., 1985). In the Alameda County Study, current smokers had a 51% (females) to 69% (males) greater age-adjusted mortality rate than current never-smokers.

Virtually all life insurance companies have used separate premium tables for smokers and nonsmokers since shortly after 1979, when State Mutual Life announced the results of its 15 years of smoker vs nonsmoker mortality experience, showing that the average 32-year-old male who smokes takes 7.3 years off his life compared to nonsmoking males (Cowell and Hirst, 1979).

Results. — Cigarette smoking was found to be a predictor of biological age in both males and females (see Figure 15). The mean biological age of both men and women who did not smoke currently was younger by about 0.4 SD (3 years at age 45) than that of current smokers as a group. In both sexes, the smokers' biological age disadvantage tended to increase with number of cigarettes smoked per day. Asterisks missing in the table of Figure 15 indicate that results were not significant for the unweighted version of SBA. SBA, recalculated as described eliminating cigarette smoking from the weighting step, remained significantly correlated with smoking level (for males, $r = .135, p = .0006$; for females, $r = .111, p = .0009$).

High Fat Foods — Figure 16

Background. — All-cause mortality among 12,763 men under age 60 was found to be directly related to percent calories from saturated fatty acids (Keys et al., 1981). Kahn

et al. (1984) report increases in all-cause mortality with increased consumption of specific high-fat foods (but not all high-fat foods) surveyed among 27,530 Seventh-Day Adventists. Enhancement of tumorigenesis by fat is discussed by Carroll (1975), who also graphs per capita breast cancer mortality in different countries vs dietary fat intake. In a recent study of 250 Italian women with breast cancer and 499 controls, relative risks of breast cancer between the highest and lowest quartiles of consumption were a significant 2.1 for total dietary fat and 3.0 for saturated fat (Toniolo et al., 1989).

The suggested causal association of fat intake with breast cancer is supported by Phillips (1975), Miller et al. (1978), and Lubin et al. (1981), but not by a cohort study of nearly 90,000 U.S. nurses (Willett et al., 1987a). Butrum et al. (1988) offer an up-to-date review of some of the abundant evidence linking fat consumption to cancer incidence, including some studies that demonstrated no association. Willcox and Lei (1987) review the link between dietary fat and coronary heart disease.

Results. — Figure 16 reports results of answers to the following question: "Consider the foods you eat that are high in fats, such as meat, cheese, fried foods, salad dressings, french fries, creamy soups, ice cream, etc. How many servings of high-fat foods do you eat per day on average?" Number of servings of high-fat foods per day was found to be a significant predictor of biological age in men but not in women (although the ANOVA relationship reached significance for women for the unweighted version of SBA).

On average, men who generally consumed 10 or more servings of high-fat foods per day tested older by 0.8 SD (6 years at age 45) than did men who consumed less than 1 serving per day. The rather smooth dependence in Figure 16 of SBA on number of servings between these two extremes may be something of a statistical quirk considering the standard errors. Similar results were not significant for the unweighted version of SBA. However, when servings of high-fat foods were eliminated from influence on the calculation of the weighted version of SBA, the relationship between SBA and servings of high-fat foods for males remained virtually unchanged ($r = .106, p = .007$).

Preference for Red Meat vs Chicken and Fish — Figure 17

Background. — In a paper reviewing several studies of 27,529 California Seventh-Day Adventists, meat consumption was positively associated with all-cause mortality in males ($p < .001$) but not in females, with coronary heart disease in males and females, and with diabetes in males (Snowdon, 1988). Males consuming meat 6 days per week had a relative risk of death from all causes that was 1.7 times that of males consuming no meat. There is similar evidence that meat consumption is associated with fatal ischemic heart disease in men and women (Snowdon et al., 1984; Phillips et al., 1978).

In the study by Toniolo et al. (1989) of 499 healthy women and 250 women with breast cancer, relative risks of breast cancer between the highest and lowest quartiles of consumption were a significant 2.9 for animal protein
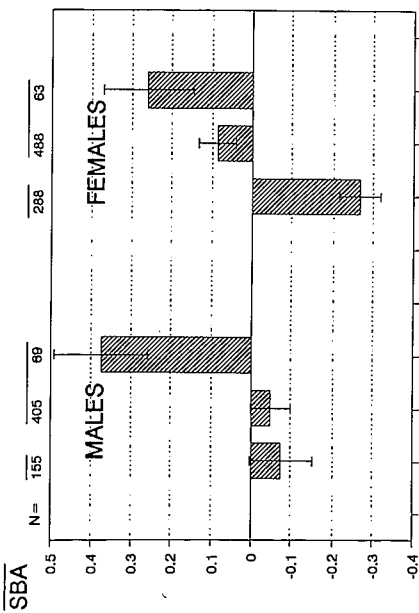
Figure 17. Which best describes your diet?
C/F: chicken and fish but little red meat. AV: average amount red
meat. >AV: greater than average amount red meat.
("Other" reply omitted).

	r	F-ratio	p
REGR'N	.100	6.4	.01
ANOVA		5.4	.02

Figure 18. Average intensity of exercise.
SE: sedentary. LI: light. MO: moderate. HI: high.
VH: very high. IN: intense.

	r	F-ratio	p	r	F-ratio	p
REGR'N	.168	18.6	.00002*	.172	26.9	.0000003**
ANOVA		16.3	.00006*		28.5	.0000001**

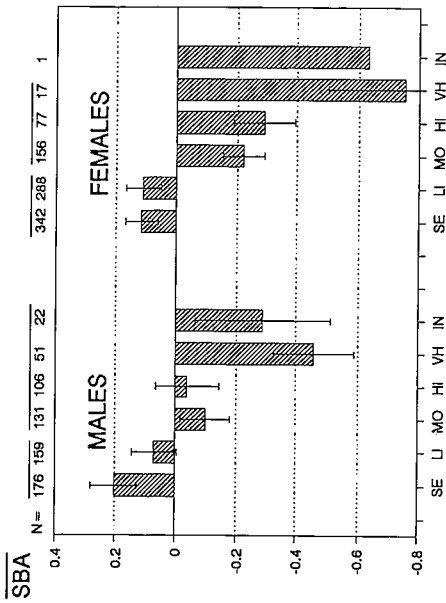

Figure 19. Exercise, total of 3 scores ÷ 2
(hours/week + intensity + years, on arbitrary scales).

	r	F-ratio	p
REGR'N	.182	22.1	.000003*
ANOVA		8.3	.004
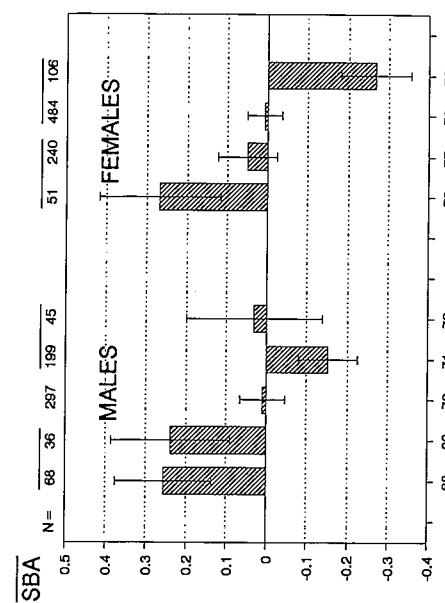
Figure 20. Life expectancy at birth in state of residence (years).

	r	F-ratio	p	r	F-ratio	p
REGR'N	.108	7.6	.006	.103	9.4	.002**
ANOVA		5.5	.02		9.5	.002**

(which included dairy products) and nonsignificant risks of 1.4 for meat, 0.9 for poultry, and 1.0 for fish.

Results. — As shown in Figure 17, the relationship between mean biological age and dietary emphasis on red meat is significant for both men and women. Participants who characterized their diet as consisting of "above average amounts of red meat" tested biologically older on average by 0.46 *SD* (3 years at age 45) for men and 0.53 *SD* (4 years at age 45) for women than persons of like sex who described their diet as "fish and chicken but little red meat." Males and females who said they consumed "average amounts of red meat" scored an intermediate mean biological age. Diets that did not fall into these three categories were omitted from the analysis. Using the unweighted version of SBA, this relationship was significant for women but not for men. When emphasis on red meat was excluded from the calculation of the weighted version of SBA, the relationships remained significant and little changed (for males, $r = .091$, $p = .02$; for females, $r = .178$, $p = .0000002$).

Physical Activity — Figures 18 and 19

Background. — Paffenbarger et al. (1986) studied 16,936 Harvard alumni aged 35–74, of whom 1,413 died during a 16-year follow-up period. Death rates of the alumni declined steadily and significantly as energy regularly expended in physical activity increased from under 500 to 3,500 kcal per week, beyond which death rates increased slightly. Rates were $\approx$ 30% lower among alumni expending 2,000 kcal or more per week than among less active men, with or without consideration of hypertension, cigarette smoking, extremes or gains in body weight or early parental death.

In the 6,928-person Alameda County Study (Breslow and Enstrom, 1980; Wingard et al., 1982), regular physical activity was associated with lower mortality from all causes. Among 16,882 British male civil servants, the incidence of coronary heart disease in those who engaged in vigorous exercise was about a third that in comparable men who did not (Morris et al., 1973).

Results. — Separate questions dealt with three measures of physical activity. Hours of physical activity per week were given as one of 6 levels: none, 1, 2, 3–4, 5–6, and >6. Aerobic intensity was selected from 5 levels: "light (like walking, golf), moderate (tennis, bicycling), high (jogging 10 minute miles), very high (running 8 min/mile), and intense (more than above)." Five levels applied to the number of years the participant had been exercising: 0–1, 1–3, 3–6, 6–10, and >10.

Figure 18 shows results for the second of these three questions, namely average aerobic intensity. Significant inverse associations were found in both men and women, with and without weighting, between mean biological age and aerobic intensity. These associations could not be confirmed at the highest intensity level, where $\overline{\text{SBA}}$ decreased nonsignificantly for men and was meaningless for women because only one woman classified herself in the "intense" category. The mean biological age difference between persons who classified themselves as "sedentary" and "very

high" was 0.66 *SD* for men and 0.87 *SD* for women (5 and 6 years at age 45).

In Figure 19, results are examined in terms of a composite exercise score, defined as the sum of the three exercise scores (hours/week + intensity + years, each on arbitrary scales from 1 to 5 or 1 to 6 corresponding to the levels listed above) divided by 2. There is a significant inverse dependence of mean biological age on increasing composite exercise score for both sexes, regardless of weighting. On average, men whose composite score was less than 3 tested biologically younger by 0.70 *SD* (5 years at age 45) than men who scored 8, the highest level. For women, the mean biological age difference between the same two categories was 0.52 *SD* (4 years at age 45).
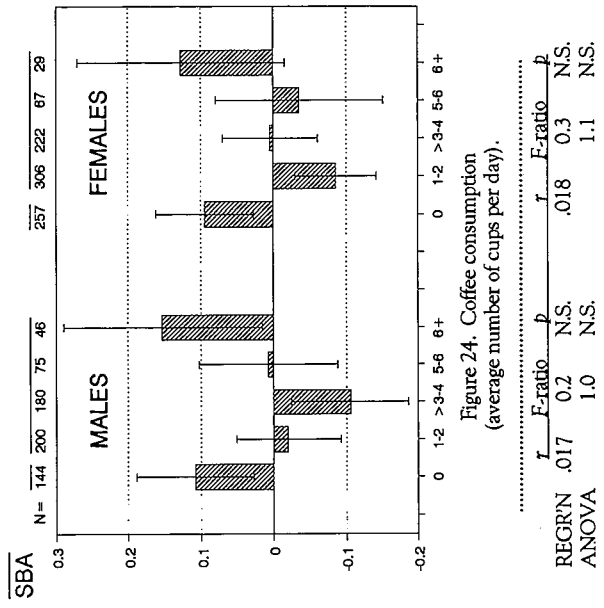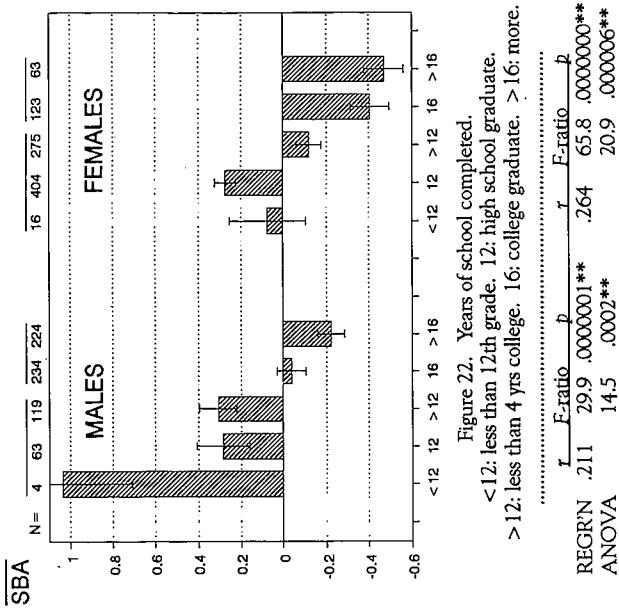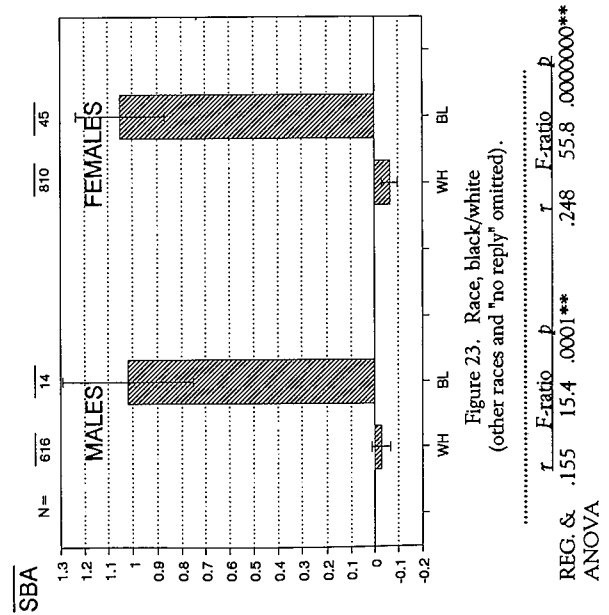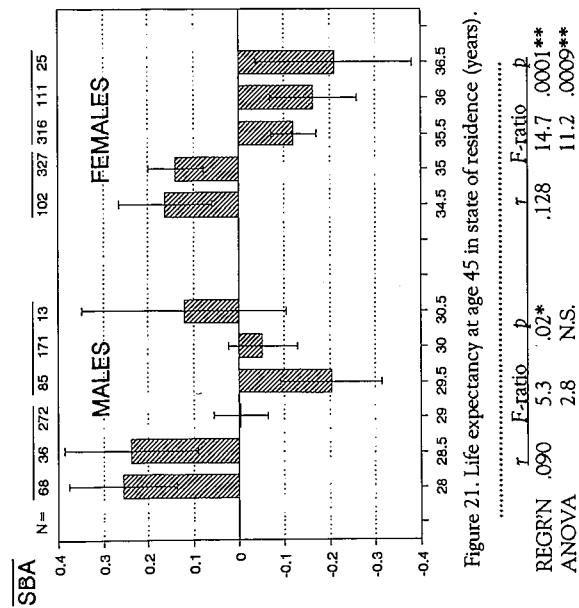
Life Expectancy in State of Residence — Figures 20 and 21

Background. — Sizable geographical variations exist in all-cause death rates in the United States. These have been examined by county (Sauer, 1980; Sauer and Parke, 1974) and by state (Metropolitan Life, 1986; National Center for Health Statistics, 1985). For men, U.S. state life expectancies at birth range from 67.6 years (Louisiana, Mississippi, South Carolina) to 74.1 years (Hawaii). For women, they range from 75.9 years (Louisiana) to 80.3 years (Hawaii). Geographical differences in life expectancy reflect a combination of variables, some of the more obvious candidates being race and ethnic origin, years of education, and regional dietary preferences. Effects associated with elevation above sea level, drinking water hardness, rural living, mining and manufacturing have also been observed (Sauer and Parke, 1974).

In organizing this study, it was of interest to examine whether persons residing in longer-lived states age less rapidly in the sense that they experience less rapid declines of the measured physiological functions, than residents of shorter-lived states. Accordingly, in inviting insurance companies to participate, an effort was made to select companies located in states representing a wide distribution of life expectancies based on 1979–81 data from the National Center for Health Statistics (1985). This goal was met only partially.

As Table 3 shows, our test states ranged in life expectancy at birth for males from 67.6 years (Mississippi) to 71.9 years (Wisconsin), 66% of the U.S. range. For females, our range was from 76.4 years (Mississippi) to 79.0 years (Kansas), 59% of the U.S. range. These ranges were, nevertheless, sufficient to show significant effects.

Results. — Figure 20 shows the relationship between mean biological age and state life expectancy at birth rounded to the nearest 0.5 year. Both men and women residing in longer-lived states tested significantly younger on average than men and women living in shorter-lived states. The mean biological age of men living in states with a male life expectancy of 71 ± 0.5 years was younger by 0.40 *SD* (3 years at age 45) than that of men living in states with a male life expectancy of 68 ± 0.5 years. Women living in states with a female life expectancy of 79 ± 0.5 years were found to have a mean biological age that was 0.54 *SD* (4

Figure 21. Life expectancy at age 45 in state of residence (years).

	r	F-ratio	p		r	F-ratio	p
REGR'N	.090	5.3	.02*	.128	14.7	.0001**
ANOVA		2.8	N.S.		11.2	.0009***

Figure 22. Years of school completed.
<12: less than 12th grade. 12: high school graduate.
>12: less than 4 yrs college. 16: college graduate. >16: more.

	r	F-ratio	p		r	F-ratio	p
REGR'N	.211	29.9	.0000001**	.264	65.8	.0000000**
ANOVA		14.5	.0002**		20.9	.0000006***

Figure 23. Race, black/white
(other races and "no reply" omitted).

	r	F-ratio	p		r	F-ratio	p
REG. &	.155	15.4	.0001**	.248	55.8	.0000000**
ANOVA						

Figure 24. Coffee consumption
(average number of cups per day).

	r	F-ratio	p		r	F-ratio	p
REGR'N	.017	0.2	N.S.	.018	0.3	N.S.
ANOVA		1.0	N.S.		1.1	N.S.

years at age 45) younger than women living in states with a female life expectancy of 76 ± 0.5 years.

Similar results were found in analyzing data for state life expectancy at age 45, Figure 21. In this case, life expectancies were rounded to the nearest 0.25 year. Biological age relationship became more significant for women but less so for men. For both sexes, significance was reached regardless of weighting.

Educational Level — Figure 22

Background. — Among 312 Catholic nuns in Minnesota assessed from 1936–1987, nuns with at least a bachelor's degree lived to a median age of 89.2 years, compared to 81.4 years for nuns with less than a bachelor's degree (Snowdon et al., 1989). This study of the association of educational level with mortality rate is of particular interest because many of the usual correlates of low education in the general population, such as limited access to medical care, low income, poor housing, and smoking were not present in this population. Occupation was a factor, however. All of the more highly educated nuns were teachers, half of the less educated nuns did domestic work. The authors conclude that education is a robust marker for surviving, the causes remaining obscure.

Of 23 factors studied among 1,474 Framingham Study participants, education was one of 5 significant predictors of "survival with good function" in men, while in women, education was the only significant predictor besides age (Pinsky et al., 1987). A Metropolitan Life (1932) study of nearly 40,000 men demonstrated that college graduates lived an average 5 years longer than the population as a whole while men graduating with honors lived an average 7 years longer, an observation that has been confirmed by more recent studies (Metropolitan Life, 1975).

Among 270,000 men employed by the Bell System, coronary heart disease (incidence and death) was 30% lower among college men than noncollege men regardless of age or occupation, the difference in risk seeming to be traceable to biological differences between college and noncollege men existing at the time they were hired (Hinkle et al., 1968).

Results. — Educational level was found to be a significant predictor of biological age in both sexes regardless of weighting (see Figure 22). On average, male college graduates tested 0.33 *SD* (2 years at age 45) younger than males who stopped their formal schooling after graduating from high school. Female college graduates, on average, tested 0.67 *SD* (5 years at age 45) younger than females who stopped with a high school diploma. Additional education beyond college boosted these differences further but non-significantly.

Race — Figure 23

Background. — Racial differences in mortality in the U.S. are substantial (National Center for Health Statistics, 1985). The issue is complicated by the large number of risk factors which tend to weigh more heavily on blacks than on whites, including nutrition, preventive care, high blood pressure, smoking, and injury control (National Center for Health Statistics, 1988).

Results. — Figure 23 shows a mean biological age difference of 1.04 *SD* (8 years at age 45) between black and white males, white males testing younger. Results were significant regardless of weighting. The standard error of this result is quite large due to the fact that the analysis included only 14 black males. White females, on average, tested 1.12 *SD* (8 years at age 45) younger than black females, based on 45 black females.
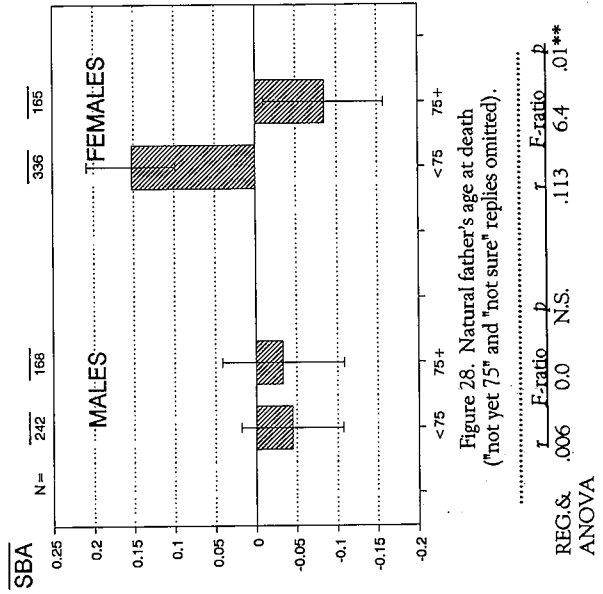
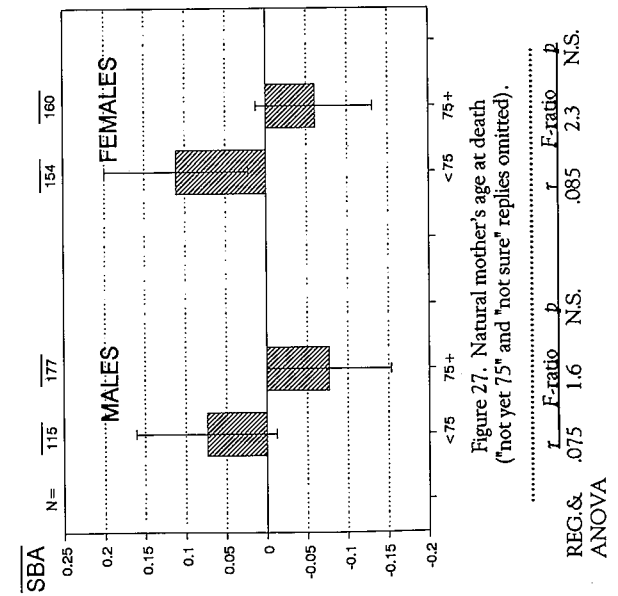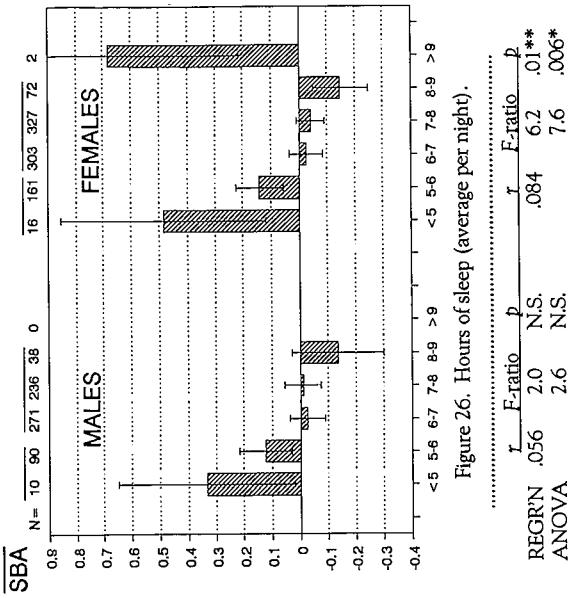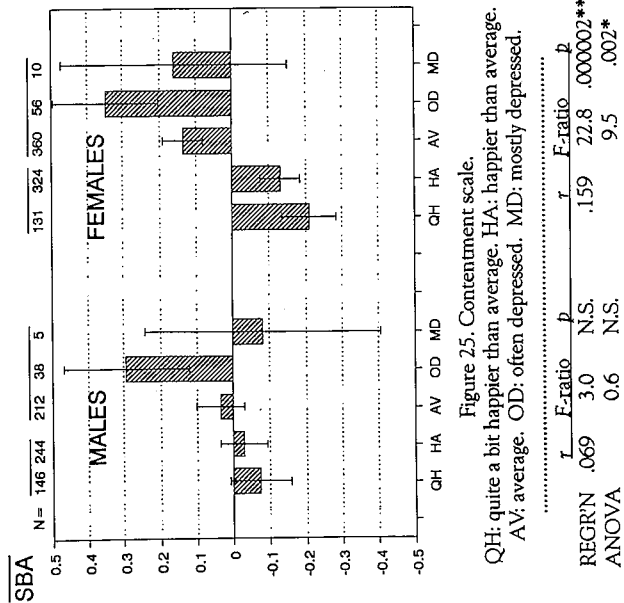*Coffee Consumption — Figure 24*

Background. — Investigations of the relationship between coffee intake and the risk of coronary heart disease have given inconsistent results. Positive associations were reported in 3 publications and no association in 7 publications cited by Yano et al. (1977), who reported that a positive association among 7,705 Japanese men living in Hawaii became insignificant when cigarette smoking was taken into account. A review by Leonard and Watson (1987) cites 2 studies that showed positive associations of intakes of 5, 6, or more cups of coffee per day with risk of myocardial infarction, one study relating coffee consumption to all-cause death, and 5 studies showing no such associations.

Results. — Figure 24 indicates that biological age was not significantly associated with average number of cups of coffee consumed per day. The possibility of a U-shaped relationship is suggested because, for both sexes, mean biological age was older at the two extremes of consuming no coffee and consuming the highest amounts (> 6 cups per day). However, standard errors were large at all levels of consumption, ruling out significance regardless of hypotheses as to shape.

Contentment Scale — Figure 25

Background. — Of 22 variables that were found to be significant predictors of longevity among 252 participants aged 60–94 at the start of a 25-year longitudinal study at Duke University, work satisfaction was one of the 3 strongest predictors in men, and health satisfaction was one of the 3 strongest predictors in women (Palmore, 1982). In the same study, a weaker but nevertheless significant predictor of longevity was happiness based on a choice of one of 6 statements such as "these are the best years of my life" and "this is the saddest time of my life."

Results. — Subjects were asked: "Where would you place yourself on the following contentment scale? 1 – Quite a bit happier than average. 2 – Happier than average. 3 – Average. 4 – More often depressed than average. 5 – Depressed most of the time." Regardless of weighting, biological age was found to be significantly associated with contentment self-rating in women but not in men (see Figure 25). Compared to women who answered "more often depressed than average," women who answered "quite a bit

Figure 25. Contentment scale.
QH: quite a bit happier than average. HA: happier than average. AV: average. OD: often depressed. MD: mostly depressed.

	r	F-ratio	p
REGR'N	.069	3.0	N.S.
ANOVA		0.6	N.S.
	.159	22.8	.000002**
		9.5	.002*

Figure 26. Hours of sleep (average per night).

	r	F-ratio	p
REGR'N	.056	2.0	N.S.
ANOVA		2.6	N.S.
	.084	6.2	.01**
		7.6	.006*

Figure 27. Natural mother's age at death ("not yet 75" and "not sure" replies omitted).

	r	F-ratio	p	r	F-ratio	p
REG.& ANOVA	.075	1.6	N.S.	.085	2.3	N.S.

Figure 28. Natural father's age at death ("not yet 75" and "not sure" replies omitted).

	r	F-ratio	p	r	F-ratio	p
REG.& ANOVA	.006	0.0	N.S.	.113	6.4	.01**

happier than average'' had a mean biological age that was younger by 0.55 *SD* (4 years at age 45).

Hours of Sleep — Figure 26

Background. — In a prospective study of 1,064,004 American men and women, sleeping <5 hours per night was associated with a large decrease in life expectancy, <6 hours per night with a small decrease (Hammond, 1964). In the Alameda County Study, the 6,928 participants were classified into 2 categories based on usual number of hours of sleep per night, namely "7 or 8 hours" and "≤6 or ≥9." Both males and females sleeping 7 to 8 hours per night had a significant ≈ 25% lower age-adjusted mortality rate then those sleeping ≤6 or ≥9 hours (Breslow and Enstrom, 1980; Wingard et al., 1982). In the Honolulu Heart Program study of 5,041 men of Japanese ancestry, hours of sleep were not associated with health during a 12-year follow-up period (Benfante et al., 1985).

Results. — In both sexes, biological age was inversely associated with average hours of sleep per night, but the association was significant only in females (regardless of weighting). As illustrated in Figure 26, women who slept 5 to 6 hours per night on average had a mean biological age that was older by 0.29 *SD* (2 years at age 45) than that of women who got 8 to 9 hours of sleep per night on average. Women who slept <5 hours per night tended to score older still. (The reversal for >9 hours of sleep is not significant because only 2 females slept that long on average.)

Parental Longevity — Figures 27 and 28

Background. — Long life runs in families, and a number of studies have explored genetic factors as contributors to longevity (Abbott et al., 1974; Pearl and Pearl, 1934; Suzuki et al., 1985; Takata et al., 1987). While mother's age at death was not a significant predictor, father's age at death was one of the weaker of 22 predictors of longevity for men (not women) identified by Palmore (1982) in the Duke 25-year follow-up study of 252 ambulatory, noninstitutionalized residents of North Carolina aged 60–94 at the start of the study. Each additional year the father lived predicted about one-third year more longevity for the men in the study.

Results. — Participants were asked separately about the longevity of their natural mother and father. Those who answered "Died under age 75" were categorized as "<75" in Figure 27 (mother) and Figure 28 (father). Subjects who answered "living and 75 or older" or "died at age 75 or older" were categorized as "75 + ." In both figures, subjects who answered "living; not yet age 75" or "not sure" were excluded. In consequence, 55% of the male subjects and 64% of the female subjects were excluded from the data in Figure 27. Likewise 36% of the male subjects and 43% of the female subjects were excluded from Figure 28.

Of the four relationships shown in these two figures, only one reached significance. On average, women whose father lived to age 75 or older scored younger biologically by 0.23 *SD* (2 years at age 45) than women whose father died at a younger age (significant regardless of weighting). No such trend was detectable in males. Mother's age at death produced trends in both males and females, but these were not significant.

Vitamin Supplements — Figures 29 and 30

Background. — A number of studies have shown that a decrease in risk of some types of cancer is associated with dietary intake of vitamin A (Kvale et al., 1983; Gregor et al., 1980; Middleton et al., 1986), beta-carotene (Pisani et al., 1986; Mettlin et al., 1979; Menkes et al., 1986; Ziegler et al., 1986), and vitamin E (Bjelke, 1978). Results for vitamin C are inconsistent, protective effects against cancer being found in some studies (Kallistratos and Fasske, 1980) but not in others (National Research Council, 1982; Reddy et al., 1982). A review of the literature and the current position of the National Cancer Institute on the subject of vitamins and beta-carotene are given by Butrum et al. (1988).

Results. — Compared to persons of like sex who answered "no," mean biological age was found to be significantly younger for women but not men who answered "yes" to the question "do you take any vitamin or mineral supplements (tablets, capsules, etc.)?". As illustrated in Figure 29, the difference for women came to 0.16 *SD* (1 year at age 45), which was significant regardless of weighting.
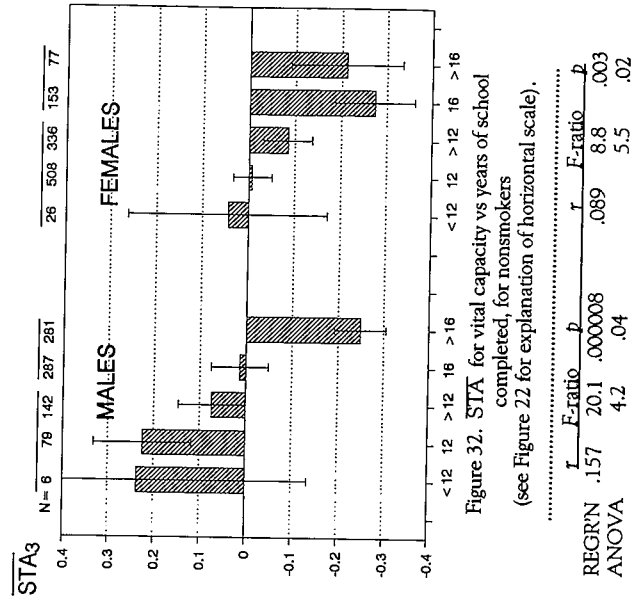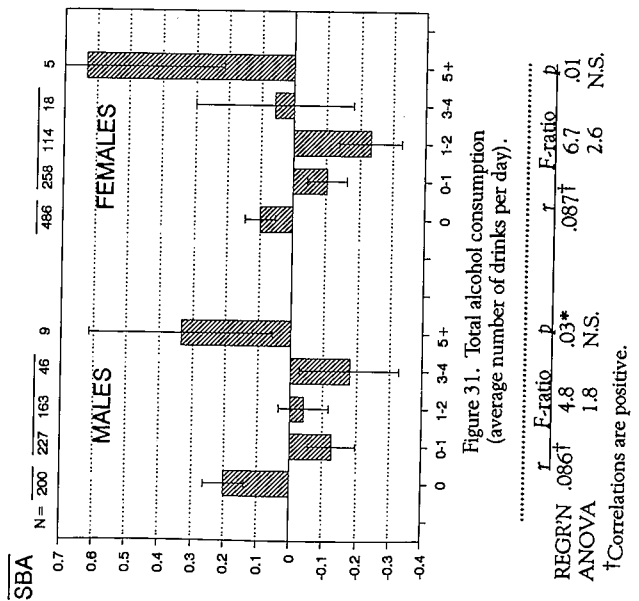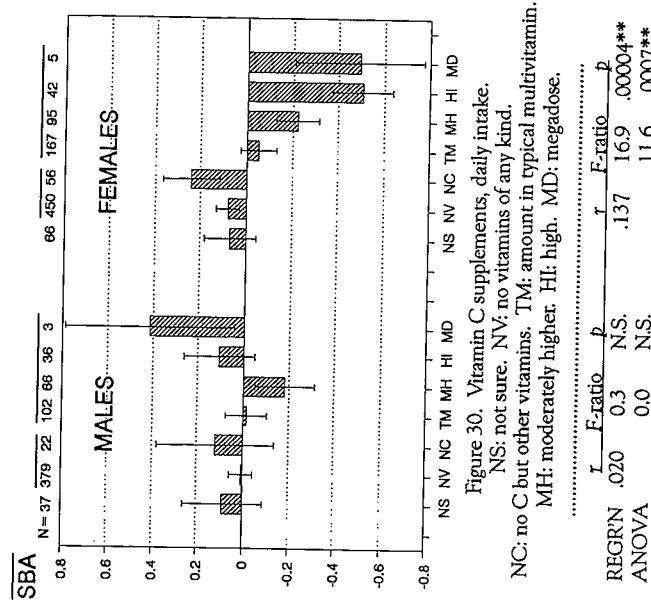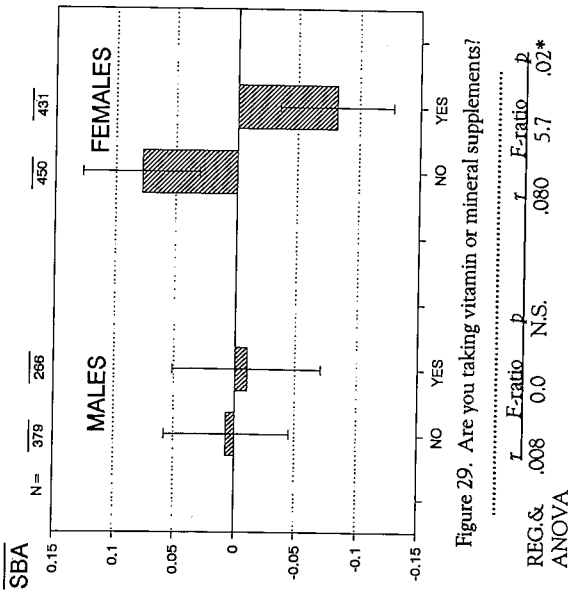
There was also a significant dose-dependent relationship (regardless of weighting) between mean biological age and intake of vitamin C in women but not men. Figure 30 shows that women who listed their vitamin C intake as "high" (but below "megadose" levels) had a mean biological age that was 0.57 *SD* (4 years at age 45) younger than women who said they took no vitamins.

Alcohol Consumption — Figure 31

Background. — Of 17 cohort and case-control studies, almost all found a positive association between alcohol consumption and risk of breast cancer (Graham, 1987). In a recent large cohort study of 89,538 nurses, Willett et al. (1987b) found that consuming 15 g of alcohol (about 1 drink) or more per day increased the risk of breast cancer by ≈ 60%. In a more representative sample of 7,200 women, Schatzkin et al. (1987) found a similar ≈ 60% increase which was dose-dependent. Butrum et al. (1988) review the association between alcohol consumption and various types of cancer and present the National Cancer Institute guidelines.

In the Honolulu Heart Program, alcohol consumption was positively associated with the risk of cancer, stroke, and asymptomatic hypertension and negatively associated with coronary heart disease (incidence and death) among 7,705 (Yano et al., 1977) and 5,041 (Benfante et al., 1985) men of Japanese ancestry. In the Framingham Study, alcohol intake predicted loss of functional ability in men but not in women (Pinsky et al., 1987).

Results. — Separate questions dealt with beer, wine, and hard liquor consumption. Average daily number of bottles or

Figure 29. Are you taking vitamin or mineral supplements?

REG.& ANOVA	r	F-ratio	p	r	F-ratio	p
	.008	0.0	N.S.	.080	5.7	.02*

Figure 30. Vitamin C supplements, daily intake.
NS: not sure. NV: no vitamins of any kind.
NC: no C but other vitamins. TM: amount in typical multivitamin.
MH: moderately higher. HI: high. MD: megadose.

	r	F-ratio	p	r	F-ratio	p
REGRN	.020	0.3	N.S.	.137	16.9	.0004***
ANOVA		0.0	N.S.		11.6	.0007**

Figure 31. Total alcohol consumption (average number of drinks per day).

	r	F-ratio	p	r	F-ratio	p
REGRN	.086†	4.8	.03*	.087†	6.7	.01
ANOVA		1.8	N.S.		2.6	N.S.

†Correlations are positive.

Figure 32. STA for vital capacity vs years of school completed, for nonsmokers (see Figure 22 for explanation of horizontal scale).

	r	F-ratio	p	r	F-ratio	p
REGRN	.157	20.1	.000008	.089	8.8	.003
ANOVA		4.2	.04		5.5	.02

cans of beer consumed were given as one of 6 levels: none, <1, 1, 2 or 3, 4 or 5, and 6 or more. Average daily wine consumption was selected from: none, <1 glass, 1 glass, 2 to 3 glasses, half bottle, full bottle. Average number of cocktails or shots of hard liquor per day was selected from: none, <1, 1, 2, 3, 4 or more. Estimates (Willett et al., 1987b) put the alcohol content of a bottle or can of beer at 13.2 g, a glass of wine at 10.8 g, and a shot of liquor at 15.1 g. Because all three are within 15% of 13 g and poured drink sizes tend to vary by more, 13 g was taken here as the unit "drink" size for each beverage, which allowed cans of beer, glasses of wine, and shots of liquor to be added to obtain average number of drinks per day. This is the horizontal-axis coordinate of Figure 31.

A positive linear regression relationship was found between biological age and alcohol consumption which was barely significant in men and only slightly more significant in women. Results are difficult to interpret because both relationships appear U-shaped. On average, persons consuming 5 or more drinks per day scored oldest biologically, while persons who drank no alcohol scored next oldest. Moderate drinkers tended to score younger biologically than the sample average. Relatively few persons said they consumed 3-to-4 or 5-and-above drinks per day, which suggests that if all categories had been more nearly equally populated, the significance of the linear regression for both men and women would have disappeared.

Relationship Between Education and Lung Function —
Figure 32

Additional insights can be gained by examining interventions in terms of their relationships to the individual physiological tests under study. In this study, 12 tests × 17 risk factors = 204 additional relationships that might be examined. A 204-item correlation matrix is not too meaningful unless supported by other analyses. In its place, just one of these 204 relationships will be examined in this article to demonstrate the potential value of this approach.

The selected relationship is not one of the most highly correlated relationships but it is one of the more interesting, namely forced vital capacity, FVC, vs years of school completed among nonsmokers. There is no obvious reason why level of education and lung function should be linked for nonsmokers except that both predict mortality rate (see section describing Figure 22 above and Introduction under "Correlation with mortality as a validating criterion").

Background. — Possible causes of the relationship between education and mortality rate are poorly understood. The cited study of 312 Catholic nuns (Snowdon et al., 1989) removed as variables some of the common correlates of education (socioeconomic status, housing, smoking, access to health care) but left a number of other correlates of education (occupation, intelligence, degree of life control, stress and stress hormones, diet).

Results. — Figure 32 illustrates the relationship between vital capacity and level of education. Cigarette smokers have been excluded. This figure differs from the other figures in that the vertical-axis coordinate is $\overline{STA_3}$, the mean value of standardized test age for vital capacity (j = 3) at each listed educational level. $STA_3 = 0$ designates subjects whose vital capacity is average for their age and sex. Positive values of STA_3 represent poorer vital capacities (an "older lung age") in units of standard deviations, and vice versa. The horizontal axis coordinate of Figure 32 is the same as that of Figure 22.

The data graphed and tabulated in Figure 32 indicate that, when cigarette smoking is excluded, level of education predicts lung age. Among nonsmokers, more highly educated men and women had significantly better average lung function for their age, height, and sex than did less educated persons. On average, men with a college education but no additional formal schooling had a "lung age" that was 0.22 *SD* (2 years at age 45) younger than men who quit school after receiving a high school degree. For women, the same comparison corresponded to a difference of 0.27 *SD* (2 years at age 45).

The observed association suggests a link between physical and mental development. Nutrition may be a factor, as may genetic influences.

DISCUSSION

Biomarkers of aging are needed to evaluate proposed treatments to retard aging rates. While there are many age-sensitive physiological tests, age-sensitivity is not a sufficient condition to validate a biomarker of aging. A fundamental problem in assembling an experimentally useful battery of biomarkers has been the lack of a practical method for their validation.

This study sought to investigate the criterion that a valid biomarker of aging should be responsive to interventions that influence the natural life span. Twelve physiological tests (Table 1) were investigated as potential biomarkers of aging, being administered to 2,462 study subjects who were also surveyed for 17 health and mortality risk factor "interventions." A mathematical approach was introduced for calculating a composite index of aging, standardized biological age, SBA. The approach addressed a variety of problems which, for reasons given in the Introduction, appear to invalidate traditional methods of calculating biological age that used multiple regression. Relationships between the 12 prospective biomarkers, 17 risk factors, SBA, and a composite validation variable, CVV, were examined in various aspects.

Total Variance of CVV Explained by the 12 Tested Functions

The correlation between STA_j, Equation (17), and CVV, Equation (6), was used as the relative measure of the validity of test j as a biomarker of aging and to weight the STA_j in combining them into SBA, Equations (18) and (19). These correlations are given in the "r" lines of Table 6 by test number and sex.

For each sex, only 2 of 12 such correlations were found to exceed $r = .2$. The rest ranged from .027 to .136 for males and .016 to .166 for females. Squaring and summing these r's show that the total variation of CVV explained by the 12 tested functions is 18% for males and 20% for females. Of

these amounts, the amount attributable to the two best-correlated functions, FVC and FEV-1 (j = 3 and 4), taken together is 12% for males and 10% for females. The remaining 10 functions explain 6% for males and 10% for females of the variability of CVV. These values err on the high side because the tested functions are interrelated. When the correlations for the weighted case given in Table 7 are squared, SBA is found to account for 8% of CVV variance for males, 10% for females.

Are some of the physiological tests included in this study experimentally useful biomarkers of aging, and if so, at what r^2 are we to draw the line? To answer this question, it is useful to estimate the highest r^2 that might be expected in terms of correlations of biomarker scores, transformed to STA_j, with CVV.

In this connection, an important issue is how well CVV predicts mortality. CVV comprises 8 of a much larger number of possible mortality risk factors that were not surveyed. The 8 factors are, to some extent, interrelated. Moreover, they represent crude estimates of risk because questions had a limit of 6 possible answers and, in most cases, provided only a rough assessment of the factor in question. There is also the unknown element of how accurately the questions were answered.

In summary, (a) CVV is an imperfect measure of the risk factors which it incorporates and (b) CVV is largely insensitive to many other influences on mortality risk, including known and unknown risk factors, genetic influences, chance events (illnesses, exposure to damaging agents), and so on. Therefore, CVV, as constituted in this study, cannot realistically be expected to explain more than perhaps a quarter or a third of the variance of mortality rate, nor can a hypothetical battery of ideal biomarkers of aging be expected to explain more than that proportion of the variance of CVV. While this helps to put in perspective the finding that the variance shared between SBA and CVV in this study was only about 10%, it suggests considerable room for improvement in (a) the selection of the battery of biomarkers composing SBA and (b) the selection and scoring of risk factors (or other interventions) composing CVV.

Trial Acceptance Criterion For a Valid Biomarker of Aging

The results provide some guidelines for establishing a tentative quantitative acceptance criterion for a useful biomarker of aging based on CVV composed as in this study. If a value of $r^2 = .20$ were set as the lowest acceptance limit for SBA as an index of aging, then 12 unrelated tests constituting SBA would make an average individual contribution of $r^2 = .017$ which corresponds to $r = .129$. This number is low as an individual-biomarker-acceptance limit because actual biomarkers are likely to be interrelated. But it is high considering that it represents the average and that it may be useful to accept tests that are as much as 50% below average, particularly if they meet other criteria such as the quality of life criterion (Introduction). In round numbers, $r = .1$ may be a reasonable trial acceptance limit for a biomarker of aging. Of the 12 physiological tests included in this study, only 3 qualify as biomarkers of aging under this criterion for males, while 6 qualify for females.

Certainly the 12 selected tests do not make up an ideal test battery for determining SBA. A few tests seem to be next-to-useless for the purpose. Likewise, the risk-factor composition of CVV might be considerably improved. Despite these limitations, the study demonstrated the application of a quantitative approach for validating biomarkers of aging. This general approach might be used to evaluate other candidate biomarkers of aging in the future, leading to improved biomarker batteries.

Risk Factor Behaviors Predict Rates of Functional Declines In Healthy Persons

Secondarily to its main objective, the study asked whether risk factors that predict mortality also predict rates of functional declines with age. In looking at the results from this perspective, biomarker validation is not an issue. Rather than as putative biomarkers of aging, the 12 tests are viewed as what they are, namely tests of function. Likewise, SBA is seen not as a composite index of aging but as a composite index of function.

SBA was found to be responsive to a number of health risk behaviors of healthy persons. On average, individuals who scored more favorably with respect to various risk factors were found to have better function at any age than individuals who scored less favorably. Results suggest that, on average, individuals who modify their diet and life style in the direction of lower risk of major diseases are also delaying or retarding the onset of important functional declines in healthy persons.

Figures 15 to 31 show that many risk factors are significantly associated with SBA for both sexes, providing a split-sample validation of results. These risk factors were cigarette smoking, consumption of red meat, alcohol intake, intensity of exercise, a composite exercise score, life expectancy at birth and age 45 in the state of residence, years of school completed, and race. In males only, SBA was significantly associated with consumption of high-fat foods, while in females only, significant associations were found with contentment score, hours of sleep, father living to age 75, intake of vitamin/mineral supplements and intake of vitamin C supplements. SBA was not significantly related in either sex to mother living to age 75 or coffee drinking.

Causation

Clearly, none of the reported relationships can be considered as causative. For example, persons emphasizing fish and chicken instead of red meat might be taking better care of themselves in other ways, some of which, in turn, could have influenced test scores. Some persons who engage in little physical activity might be disinclined to exercise for reasons of health or heredity, the reasons rather than the lack of exercise accounting for poorer functional scores. (Do runners have better lung function because they run or do they become runners because they have better lung function?) Nor is it clear that alcohol consumption in any quantity has beneficial effects on the aging process. Distinctions represented by alcohol may be cultural, socioeconomic, educational, genetic, or based on personality, all of which have been linked to life expectancy.

Can a Handful of Physiological Tests Be Useful In Evaluating Treatments To Retard Aging?

Assuming that it becomes possible, as seems consistent with the results of this study, to identify a dozen or more biomarkers of aging that are valid by the proposed criterion or by a stricter one, and to base SBA on these, can such an index be expected to be experimentally useful in evaluating treatments to retard aging rates? It is reasonable to argue that the aging process is too complex and too limitless in its diversity of expression to be defined by a battery of biomarker tests, whatever its size. Ludwig (1989) responds to this argument with the suggestion that "in science, not only pure reason governs, but also common sense (Max Planck), and common sense tells us that if the rate of progress of a variety of unrelated markers that correlate with chronological age is demonstrably decelerated by a given experimental variable (e.g., caloric restriction), the role of aging of the entire organism . . . has been modified too."

Ludwig suggests that a relatively small number of biomarkers can be representative of many and that it is unnecessary to look at the full spectrum of age changes to know what is going on. The results of this study support that view. While the selection of the 12 physiological tests investigated in this study was far from ideal, results show that even an inferior index of aging is responsive to interventions that predict life expectancy in healthy persons, suggesting that the same index may also be responsive to other than the tested interventions. As additional tests are qualified as biomarkers of aging, it should be possible to develop increasingly sensitive tools for evaluating treatments to retard aging rates.

It is interesting that physiological correlates of risk factor behaviors can be detected by a handful of tests that were selected as generalized indicators of aging. The sensitivity of SBA to the risk factors suggests that SBA is measuring an aspect of aging beyond the obvious decline in the measured functions.

REFERENCES

Abbott, M. H.; Murphy, E. A.; Bolling, D. R. et al. The familial component in longevity: a study of offspring of nonagenarians II. preliminary analysis of the complete study. Johns Hopkins Med. J. 134:1–16; 1974.

Anderson, H. R. Respiratory abnormalities, smoking habits and ventilatory capacity in a highland community in Papua New Guinea: prevalence and effect on mortality. Int. J. Epidemiol. 8:127–135; 1979.

Ashley, F.; Kannel, W. B.; Sorlie, P. D.; Masson, R. Pulmonary function: relation to aging, cigarette habit, and mortality. Ann. Intern. Med. 82:739–745; 1975.

Baker, G. T.; Sprott, R. L. Biomarkers of aging. Exp. Gerontol. 23:223–239; 1988.

Beaty, T. H; Cohen, B. H.; Newill, C. A. et al. Impaired pulmonary function as a risk for mortality. Am. J. Epidemiol. 116:102–113; 1982.

Beaty, T. H.; Newill, C. A.; Cohen, B. H. et al. Effects of pulmonary function on mortality. J. Chronic Dis. 38:703–710; 1985.

Benfante, R.; Reed, D.; Brody, J. Biological and social predictors of health in an aging cohort. J. Chronic Dis. 38:385–395; 1985.

Bjelke, E. Dietary factors and the epidemiology of cancer of the stomach and large bowel. Aktuel. Ernaehrungsmed. Klin. Prax. 2(suppl):10–17; 1978.

Borkan, G. A.; Norris, A. H. Assessment of biological age using a profile of physical parameters. J. Gerontol. 35:177–184; 1980.

Bowden, D. M.; Short, R.; Williams, D. D. Constructing an instrument to measure the rate of aging in female pigtailed macaques (*macaca nemestrina*). J. Gerontol. Biol. Sci. 45:B59–B66; 1990.

Breslow, L.; Enstrom, J. E. Persistence of health habits and their relationship to mortality. Prev. Med. 9:469–483; 1980.

Butrum, R. R.; Clifford, C. K.; Lanza, E. NCI dietary guidelines: rationale. Am. J. Clin. Nutr. 48:888–895; 1988.

Carroll, K. K. Experimental evidence of dietary factors and hormone-dependent cancers. Cancer Res. 35:3374–3383; 1975.

Chodzko-Zajko. W. J.; Ringel, R. L. Physiological fitness measures and sensory and motor performance in aging. Exp. Gerontol. 22:317–328; 1987.

Comfort, A. Test-battery to measure ageing-rate in man. Lancet, Dec. 27, 1969:1411–1415; 1969.

Costa, P. T.; McCrae, R. R. Functional age: A conceptual and empirical critique. In Haynes, S. G.; Feinleib, M., eds. Epidemiology of aging. NIH Pub. No. 80-969. Washington, DC: U.S. Government Printing Office; 1980:23–46.

Cowell, M. J.; Hirst, B. L. Mortality differences between smokers and non-smokers. Report, Oct. 22, 1979. Worcester, MA: State Mutual Life Assurance Co. of America; 1979.

Dirken, J. M., ed. Functional age of industrial workers.. Groningen, Netherlands: Wolters-Noordhoff; 1972:173–174 and 201–205.

Dubina, T. L.; Mints, A.Ya.; Zhuk, E. V. Biological age and its estimation. III. Introduction of a correction to the multiple regression model of biological age in cross-sectional and longitudinal studies. Exp. Gerontol. 19:133–143; 1984.

Duffy, P. H. Personal communication; 1990.

Duffy, P. H.; Feuers, R. J.; Leakey, J. A.; Nakamura, K. D.; Turturro, A.; Hart, R. W. Effect of chronic caloric restriction on physiological variables related to energy metabolism in the male Fischer 344 rat. Mech. Ageing Dev. 48:117–133; 1989.

Furukawa, T.; Inuoe, M.; Kajija, F. et al. Assessment of biological age by multiple regression analysis. J. Gerontol. 30:422–434; 1975.

Graham, S. Alcohol and breast cancer (editorial). N. Engl. J. Med. 316:1211–1212; 1987.

Gregor, A.; Lee, P. M.; Roe, F. C.; Wislon, M. J.; Melton, A. Comparison of dietary histories in lung cancer cases and controls with special reference to vitamin A. Nutr. Cancer 2:93–97; 1980.

Hammond, E. C. Some preliminary findings on physical complaints from a prospective study of 1,064,004 men and women. Am. J. Public Health 54:11–23; 1964.

Harrison, D. E. Experience with developing assays of physiological age. In: Reff, M. E.; Schneider, E. L., eds. Biological markers of aging. Washington, DC: National Institutes of Health Publication No. 82-2221; 1982:237–240.

Harrison, D. E. Jackson Laboratory. Bar Harbor, ME. Personal communication; 1988.

Harrison, D. E.; Archer, J. R. Physiological assays for biological age in mice: relationship of collagen, renal function, and longevity. Exp. Aging Res. 9:245–251; 1983.

Harrison, D. E.; Archer, J. R. Genetic differences in effects of food restriction on aging in mice. J. Nutr. 117:376–382; 1987.

Harrison, D. E.; Archer, J. R.; Astle, C. M. The effect of hypophysectomy on thymic aging in mice. J. Immunol. 129:2673–2677; 1982.

Harrison, D. E.; Archer, J. R.; Astle, C. M. Effects of food restriction on aging: separation of food intake and adiposity. Proc. Natl. Acad. Sci. USA 81:1835–1838; 1984.

Higgins, M. W.; Keller, J. B. Predictors of mortality in the adult population of Tecumseh. Arch. Environ. Health 21:418–424; 1970.

Hinkle, L. E.; Whitney, L. H.; Lehman, E. W. et al. Occupation, education and coronary heart disease. Science 161:238–246; 1968.

Hochschild, R. Improving the precision of biological age determinations. Part 1: a new approach to calculating biological age. Exp. Gerontol. 24:289–300; 1989a.

Hochschild, R. Improving the precision of biological age determinations. Part 2: Automatic human tests, age norms and variability. Exp. Gerontol. 24:301–316; 1989b.

Hollingsworth, J. W.; Hashizume, A.; Jablon, S. Correlations between tests of aging in Hiroshima subjects: an attempt to define "physiologic age." Yale J. Biol. Med. 38:11–36; 1965.

Ingram, D. K. Toward the behavioral assessment of biological aging in the laboratory mouse: Concepts, terminology, and objectives. Exp. Aging Res. 9:225–238; 1983.

Ingram, D. K. Biological age: a strategy for assessment. In: Chebortarev, F.; Tokar, A. V.; Voitenko, V. P., eds. Geriatrics and gerontology yearbook (U.S.S.R.). Kiev: Institute for Gerontology; 1984:30–38.

Ingram, D. K. Key questions in developing biomarkers of aging. Exp. Gerontol. 23:429–434; 1988.

Ingram, D. K; Reynolds, M. A. Assessing the predictive validity of psychomotor tests as measures of biological age in mice. Exp. Aging Res. 12:155–162; 1986.

Jensen, A. R. Bias in mental testing. New York: The Free Press; 1980: 85–86.

Kahn, H. A.; Phillips, R. L.; Snowdon, D. A.; Choi, W. Association between reported diet and all-cause mortality. 21 year follow-up on 27,530 adult Seventh-Day Adventists. Am. J. Epidemiol. 119:775–787; 1984.

Kallistratos, G.; Fasske, E. Inhibition of benzo[a]pyrene carcinogenesis in rats with vitamin C. J. Cancer Res. Clin. Oncol. 97:91–96; 1980.

Kannel, W. B. Habits and heart disease. In: Palmore, E.; Jeffers, F. C., eds. Prediction of life span. Lexington, MA: Heath; 1971.

Kannel, W. B.; Hubert, H. Vital capacity as a biomarker of aging. In: Reff, M. E.; Schneider, E. L., eds. Biological markers of aging. Bethesda, MD: National Institute of Health, Publication No. 82-2221; 1982.

Kannel, W. B.; Lew, E. A.; Hubert, H. B.; Castelli, W. P. The value of measuring vital capacity for prognostic purposes. Trans. Assoc. Life Insur. Med. Dir. Am. 64:66–83; 1980.

Keys, A.; Aravanis, C.; van Buchem, F. S. P. et al. The diet and all-causes death rate in the Seven Counties Study. Lancet ii:58–61; 1981.

Kvale, G.; Bjelke, E.; Gart, J. J. Dietary habits and lung cancer risk. Int. J. Cancer 31:397–405; 1983.

Leonard, T. K.; Watson, R. R. The cardiovascular effects of caffeine. In: Watson, R. R., ed. Nutrition and heart disease, Vol II. Boca Raton, FL: CRC Press; 1987:101–110.

Lubin, J. H.; Burns, P. E.; Blot, W. J. et al. Dietary factors and breast cancer risk. Int. J. Cancer 28:685–689; 1981.

Ludwig, F. C. University of California, Irvine. Personal communication; 1989.

Masoro, E. J. A physiological approach to the study of aging. Medical Times 117(5):63–67; 1989.

Menkes, H. A.; Cohen, B. H.; Beaty, T. H.; Newill, C. A.; Khoury, M. J. Risk factors, pulmonary function, and mortality. Prog. Clin. Biol. Res. 147:501–521; 1984.

Menkes, M. S.; Comstock, G. W.; Vuilleumier, J. P.; Helsing, K. J.; Rider, A. A.; Brookmeyer, R. Serum beta-carotene, vitamins A and E, selenium, and the risk of lung cancer. N. Engl. J. Med. 315:1250–1254; 1986.

Metropolitan Life. College honor men longer-lived. New York: Metropolitan Life. Statistical Bulletin 13:5–7; 1932.

Metropolitan Life. Socioeconomic mortality differentials. New York: Metropolitan Life. Statistical Bulletin 56:3–5; 1975.

Metropolitan Life. Longevity gains by state. New York: Metropolitan Life. Statistical Bulletin 67(4):12–17; 1986.

Mettlin, C.; Graham, S.; Swanson, M. Vitamin A and lung cancer. JNCI 62:1435–1438; 1979.

Middleton, B.; Byers, T.; Marshall, J.; Graham, S. Dietary vitamin A and cancer — a multisite case-control study. Nutr. Cancer 8:107–116; 1986.

Miller, A. B.; Kelly, A.; Choi, N. W. et al. A study of diet and breast cancer. Am. J. Epidemiol. 107:499–509; 1978.

Morris, J. N.; Chave, S. P.; Adam, C.; Sirey, C.; Epstein, L.; Sheehan; D. J. Vigorous exercise in leisure-time and the incidence of coronary heart-disease. Lancet, Feb. 17:333–339; 1973.

Nakamura, E.; Kimura, M.; Nagata, H.; Miyao, K.; Ozeki, T. Evaluation of the progress of aging based on specific biological age as estimated by various physiological functions. Nippon Eiseigaku Zasshi (Jap. J. Hygiene) 36:853–862; 1982.

National Center for Health Statistics. U.S. Decennial Life Tables 1979–81, Vol. 1 No. 1 and Vol. 2, Nos. 1–51. U.S. Dept. of Health and Human Services Pub. No. (PHS) 85-1150-1 and (PHS) 86-1151-1-51. Hyattsville, MD: DHHS; 1985.

National Center for Health Statistics. Health promotion and disease prevention, United States, 1985. Data from the National Health Survey. Vital and Health Statistics, Series 10, No. 163. U.S. Dept. of Health and Human Services Pub. No. (PHS) 88-1591. Hyattsville, MD: DHHS; 1988.

National Research Council, Assembly of Life Sciences, Committee on Diet, Nutrition and Cancer. Diet, nutrition and cancer. Washington, DC: National Academy Press; 1982.

Neter, J.; Wassermann, W.; Kutner, M. H. Applied linear statistical models, 2nd ed. Homewood, IL: R.D. Irwin; 1985;Sects. 16.2–16.9 and 17.1.

Paffenbarger, R. S.; Hyde, R. T.; Wing, A. L.; Hsieh, C. Physical activity, all-cause mortality, and longevity of college alumni. N. Engl. J. Med. 314:605–613; 1986.

Palmore, E. B. Predictors of the longevity difference: a 25-year follow up. Gerontologist 22:513–518; 1982.

Pardee, N. E.; Winterbauer, R. H.; Morgan, E. H.; Allen, J. D. Determination of mortality risk in obstructive pulmonary syndromes: physical examination compared with measurement of forced expired volume in one second. Thorax 36:928–931; 1981.

Pearl, R.; Pearl, R. de W. The ancestry of the long-lived. London: H. Milford; 1934.

Petty, T. I.; Pierson, D. J.; Dick, N. P.; Hudson, L. D.; Walker, S. H. Follow-up evaluation of a prevalence study for chronic bronchitis and chronic airway obstruction. Am. Rev. Respir. Dis. 114:881–890; 1976.

Phillips, R. L. Role of life-style and dietary habits in risk of cancer among Seventh-Day Adventists. Cancer Res. 35:3513–3522; 1975.

Phillips, R. L.; Lemon, F. R.; Beeson, L.; Kuzma, J. W. Coronary heart disease mortality among Seventh-Day Adventists with differing dietary habits: a preliminary report. Am. J. Clin. Nutr. 31:S191–S198; 1978.

Pinsky, J. L.; Leaverton, P. E.; Stokes, J. Predictors of good function: the Framingham Study. J. Chronic Dis. 40, Suppl. 1:159S–167S; 1987.

Pisani, P.; Berrino, F.; Macaluso, M. et al. Carrots, green vegetables and lung cancer: a case-control study. Int. J. Epidemiol. 15:463–468; 1986.

Reddy, B. S.; Hirota, N.; Katayama, S. Effect of dietary sodium ascorbate on 1,2-dimethylhydrazine- or methylnitrosourea-induced colon carcinogenesis in rats. Carcinogenesis 3:1097–1099; 1982.

Sauer, H. Geographic patterns in the risk of dying and associated factors, ages 35–74 years. U.S. Dept. of Health and Human Services Pub. No. 80-1402. Hyattsville, MD: DHHS; 1980.

Sauer, H.; Parke, D. W. Counties with extreme death rates and associated factors. Am. J. Epidemiol. 99:258–264; 1974.

Schatzkin, A.; Jones, D. Y.; Hoover, R. N. et al. Alcohol consumption and breast cancer in the epidemiologic follow-up study of the first National Health and Nutrition Examination survey. N. Engl. J. Med. 316:1169–1173; 1987.

Schneider, E. L.; Finch, C. E.; Weksler, M. Potential application of biological markers for assessing interventions of physiological aging. In: Reff, M. E.; Schneider, E. L., eds. Biological markers of aging. Washington, DC: National Institutes of Health Pub. No. 82-2221; 1982:237–240.

Snowdon, D. A. Animal product consumption and mortality because of all causes combined, coronary heart disease, stroke, diabetes, and cancer in Seventh-day Adventists. Am. J. Clin. Nutr. 48:739–748; 1988.

Snowdon, D. A.; Phillips, R. L.; Fraser, G. E. Meat consumption and fatal ischemic heart disease. Prev. Med. 13:490–500; 1984.

Snowdon, D. A.; Ostwald, S. K.; Kane, R. L. Education, survival and independence in elderly Catholic sisters, 1936–1987. Am. J. Epidemiol. 130:999–1012; 1989.

Suzuki, M.; Mori, H.; Asato, T. et al. Medical researches upon centenarians on Okinawa 1: case controlled study of family history as hereditary influence on longevity. Jpn. J. Geriat. 22:457–467; 1985.

Takata, H.; Suzuki, M.; Ishii, T.; Sekiguchi, S.; Iri, H. Influence of major histocompatibility complex region genes on human longevity among

Okinawan-Japanese centenarians and nonagenarians. Lancet, Oct. 10:824–826; 1987. See also editorial, ibid:835–836; 1987.

Toniolo, P.; Riboli, E., Protta, F. et al. Calorie-providing nutrients and risk of breast cancer. J. Natl. Cancer Inst. 81:278–285; 1989.

U.S. Dept. of Health and Human Services. Reducing the health consequences of smoking: 25 years of progress. A report of the Surgeon General. U.S. Dept. of Health and Human Services Pub. No. (CDC) 89-8411, Jan. 11. Rockville, MD: Public Health Service. Centers for Disease Control. Office on Smoking and Health; 1989.

Vital Statistics of the United States. Vol. II, Part A, Section 1. Washington, DC: U.S. National Center for Health Statistics; 1980.

Voitenko, V. P.; Tokar, A. V. The assessment of biological age and sex differences of human aging. Exp. Aging Res. 9:239–244; 1983.

Webster, I. A.; Logie, A. R. A relationship between functional age and health status in female subjects. J. Gerontol. 31:546–550; 1976.

Weindruch, R.; Walford, R. L. The retardation of aging and disease by dietary restriction. Springfield, IL: Charles C Thomas; 1988.

Willcox, K. K.; Lei, K. Y. Dietary fats and the aging adult. In: Watson, R. R., ed. Nutrition and heart disease, Vol I. Boca Raton, FL: CRC Press; 1987:23–25.

Willett, W. C.; Stampfer, M. J.; Colditz, G. A. et al. Dietary fat and the risk of breast cancer. N. Engl. J. Med. 316:22–28; 1987a.

Willett, W. C.; Stampfer, M. J.; Colditz, G. A. et al. Moderate alcohol consumption and the risk of breast cancer. N. Engl. J. Med. 316:1174–1180; 1987b.

Wingard, D. L.; Berkman, L. F.; Brand, R. J. A multivariate analysis of health-related practices. A nine-year mortality follow-up of the Alameda County Study. Am. J. Epidemiol. 116:765–775; 1982.

Yano, K.; Rhoads, G. G.; Kagan, A. Coffee, alcohol and risk of coronary heart disease among Japanese men living in Hawaii. N. Engl. J. Med. 297:405–409; 1977.

Ziegler, R. G.; Mason, T. J.; Stemhagen, A. et al. Carotenoid intake, vegetables and the risk of lung cancer among white men in New Jersey. Am. J. Epidemiol. 123:1080–1093; 1986.