Human Physiology, Vol. 27, № 6, 2001, P. 725-731. Translated from Fiziologiya Cheloveka, Vol. 27, № 6, 2001, P. 88-94.

Diagnosing Aging: I. Problem of Reliability of Linear Regression Models of Biological Age

V. N. Krut'ko*, T. M. Smirnova*, V. I. Dontsov*, and S. E. Borisov**

**Institute of Systemic Analysis, Russian Academy of Sciences, Moscow, Russia*

***National Gerontological Center, Moscow, Russia*

Received April 11, 2000

Abstract—The problem of the reliability of linear regression models of biological age assessment was studied using an experimental population of patients of a geroprophylactic center. The main factors of the model quality (interpopulation difference, method of approximation of biological age, and methods of approximation of statistical significance of parameters of biological age models) were tested. New equations were derived for calculating biological age. All parameters of these equations meet the requirements of significance. It was shown that if the nonlinear character of age dynamics of biological markers of aging and the statistical significance of model parameter estimates are taken into account, the model of biological age is substantially simplified and its reliability increases.

The term biological age (BA) was introduced in age physiology research, because there was a significant difference between individual subjects of the same calendar age (CA) in the rate of age-related changes in certain functions and systems of the human body as well as in the degree of viability of the organism as a whole. Therefore, the degree of age-related "wear" of an individual subject should be characterized by a special parameter. Because the individual rate of aging is quite variable, an explicit correlation between BA and CA may be completely absent. However, in conventional gerontology, BA is regarded as a relative rather than an objective parameter of aging. This parameter was introduced to describe the degree of individual aging of a given subject relative to a normal level typical of people of the corresponding CA. Within the scope of this approach, BA is determined for an individual structural element of the human body, a group of elements, or the whole body. Biological age is determined as a parameter of the level of deterioration of structure and function of the element, group of elements, or organisms relative to the reference characteristics of deterioration averaged over the population of people of a corresponding CA. Both CA and BA are measured in the same time units. Two types of BA estimates are used to characterize the process of aging of the integral organism: 1) aging profile (set of parameters of aging of individual systems of organism), the aging profile is a vector value, and 2) the integral estimate. Usually, the aging profile is a set of parameters of aging of the most important functional systems (cardiovascular, respiratory, muscular, neuropsychical, and system of analyzers). Integral estimates are usually obtained using an explicit functional expression for the BA dependence on a certain finite set of functional and structural parameters of the human body considered as biological markers of aging ($M_1, M_2, ..., M_n$):

$$BA = f(M_1, M_2, \ldots, M_n) \quad (1)$$

Various indicator parameters of significant functions of an organism can be used as biological markers of the aging profile or integral estimates of aging. However, to be used as biological markers of aging, the functional indicator parameters should vary significantly with age (from puberty to extreme old age) and their age-related variability should be significantly larger than the individual variability. Different schools of gerontological research suggested dozens of methods for the determination of BA. The following groups of biological markers are used most frequently: morphological, psychological, neuropsychical, biochemical, immunological, and pathological parameters. Comparative analysis of the parameters of physiological functions tested in rest and under exercise load and parameters of subjective self-evaluation of the functional state are also used for this purpose. The number of biological markers recommended by various methods for the determination of BA ranges from three to several dozen [1-10]. Obviously, there should be significant between-method variability in the efficiency of the determination of BA because of the diversity in the number and character of biological markers of aging employed by these methods. However, the gerontological literature contains only a few reports on comparative trials of different methods of BA determination applied to the same population or one method applied to different populations. Usually, the quality of BA determination in each method is tested using only one

0362-1197/01/2706-0725$25.00

population (the reference population for which this method was developed).

The choice of informative biological markers appropriate for practical use is the key problem of development of the methods of the determination of BA. Significantly less attention is usually devoted to mathematical problems of the integral evaluation of BA. However, the efficiency of the evaluation significantly depends on its functional type, the selection of reference population (i.e., the population in which the parameters of the function are assessed), and the reliability of obtained estimates.

Linear regression is most frequently used for practical purposes, because it is the simplest type of presentation of BA as an explicit function of a set of biological markers:

$$BA = A + B_1M_1 + B_2M_2 + \ldots + B_nM_n. \quad (2)$$

where A, B_1, B_2, ..., B_n are constant coefficients or parameters of a linear model of BA.

Using the statistical procedure of multiple linear regression [11], the model parameters are determined to provide the closest approximation of CA in the reference population, in which the CA values and values of all biological markers are known. If the number of members of a population is k, the deviation ε_j for the j'th member of the population ($1 \le j \le k$) with the set of parameters (CA_j, M_{1j}, M_{2j}, ..., M_{nj}) is the difference between CA and the following linear function of biological markers:

$$CA_j = A + B_1M_{1j} + B_2M_{2j} + \ldots + B_nM_{nj} + \varepsilon_j. \quad (3)$$

Regression coefficients are determined from the condition of minimization of the sum of squared deviations:

$$\varepsilon^2_1 + \varepsilon^2_2 + \ldots + \varepsilon^2_{jk} \rightarrow \min.$$

The regression values of CA, i.e., the values obtained by substituting constant coefficients and values of individual biological markers into equation (2), can be regarded as the BA values:

$$BA_j = A + B_1M_{1j} + B_2M_{2j} + \ldots + B_nM_{nj}. \quad (4)$$

Let CA and BA be coordinate axes on a plane. Then, the set of individual pairs of values (CA_j, BA_j,-) produces a two-dimensional cloud (Fig. 1, A). The normal value of BA for a given population is called the proper biological age (PBA). For any given CA, the value of the PBA is taken to be equal to the BA values falling on a straight line, the cloud scattering relative to the line being minimal. The straight line is described by the following equation:

$$PBA = a + bCA. \quad (5)$$

This is the equation of simple regression of BA on CA. The deviation of an individual value of BA from the normal level is characterized by either the difference between BA and PBA (BA - PBA) or the BA/PBA ratio. A subject characterized by the inequality BA - PBA > 0 (i.e., BA > PBA or BA/PBA > 1) is considered to be biologically older than people of the same CA. Conversely, if BA < PBA, the subject is considered to be biologically younger than people of the same CA.

Because BA is an approximate value of CA and the two values are measured in the same units, the slope coefficient b of the straight line described by equation (5) is equal to the squared coefficient (r) of the correlation between the sample of individual values of CA in the reference population and the sample of corresponding values of BA. Therefore, $b = r^2$. Parameter r^2 is called the coefficient of multiple determination. The value of r^2 is the measure of accuracy of CA approximation as a function of a set of biological markers, because this value is equal to the fraction of age variation determined by the existence of a functional correlation between age and markers of biological age.

Because it is practically unfeasible to express the CA values of all members of a population with an absolute accuracy as a weighted sum of markers, the slope coefficient of the straight line PBA(CA) is less than one ($b = r^2 < 1$). At the point corresponding to the mean age of the reference population (CA_{mean}), the values of PBA and CA coincide with one another. In other words, at this point $PBA(CA_{mean}) = CA_{mean}$. Therefore, for ages younger or older than the mean age of the reference population ($CA < CA_{mean}$ or $CA > CA_{mean}$, respectively), the value of the PBA is larger or smaller than the CA, respectively. For this reason, equation (5) in the gerontological literature is often recast in an equivalent form including the multiple determination coefficients given in an explicit form:

$$PBA — CA_{mean} = r^2 (CA - CA_{mean}) \quad (6)$$

It follows from equation (6) that the PBA estimates (therefore, estimates of individual values of BA) depend both on the mean age of the reference population and on the accuracy of BA approximation achieved in a given population. Therefore, the BA estimates obtained in a group of people using the regression model are sensitive to the method of selection of the reference population from the group of people. Possible relationships between the PBA and CA at different accuracies of the approximation of BA and the mean age of a population are shown in Fig. 1, *B*. Straight lines *1* and *2* represent the plots of the PBA dependence on CA in two samples of the same group of people, in which the accuracy of BA approximation by a set of biological markers is the same (dependencies PBA(CA) in these sample are parallel to one another), whereas the mean age in one sample differs from the mean age in the other sample (different points of intersection with the straight line PBA = CA). In this case, all people whose individual values of the BA estimates fall within the band between the straight lines *1* and *2* are characterized by the BA estimates larger or smaller than the normal level of BA in this group of people, provided that population

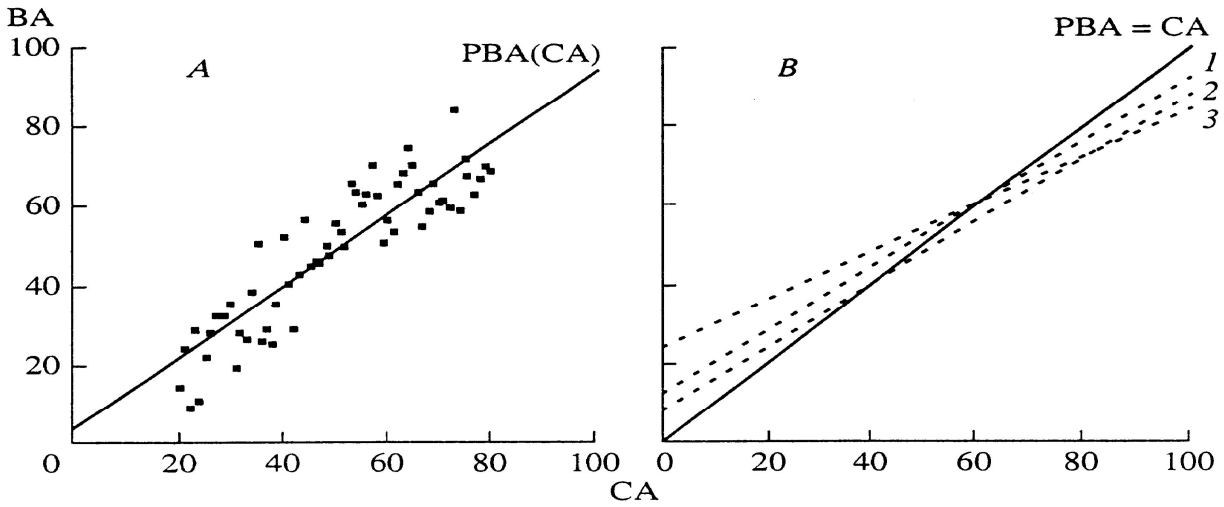

Fig. 1. Graphical presentation of relationships between calendar age (CA), biological age (BA), and proper biological age (PBA): *A,* determination of PBA by linear regression of BA on CA; *B,* different cases of the relationships between PBA and CA at different accuracies of the approximation of BA and the mean age of the population (explanation in text).

2 or *1* is chosen as the reference population, respectively. The straight line *3* corresponds to the PBA (CA) dependence if the mean age of the reference population coincides with the mean age of population *1* but the spread of individual values of BA over people of the same CA is higher than in population *1* (the coefficient of multiple determination or the slope coefficient of the straight line PBA(CA) is also correspondingly lower). In this case, the larger the difference between CA and CA_{mean}, the larger the difference between the estimates of individual values of (BA - PBA) obtained using different reference populations.

Populations may differ from each other not only in the PBA parameters but also in the coefficients of the contribution of individual markers to BA. Different rates of aging of men and women are a cause of such a difference between populations. Therefore, with rare exceptions, the parameters of BA are determined separately in men and women. Meanwhile, the range of applicability of any specific model of BA is obviously limited by a number of other factors. The correlation of the life expectancy and the character of aging with many factors is described at length in the demographic literature. These are climatic, geographical, ethnic, social, economical, and cultural factors, as well as historical changes in the process of aging.

The goals of this work were to analyze the dependence of the BA estimates on such factors as interpopulation differences, functional types of the BA model, and the strictness of requirements for statistical significance of the model parameters. Development of the BA models taking these effects into consideration was an additional goal of this work.

MATERIALS AND METHODS

A total of 195 apparently healthy subjects were tested: 159 women (from 20 to 74 years old; mean age, 43.2 years) and 36 men (from 20 to 63 years old; mean age, 42.5 years). The subjects studied in this work were patients of the National Gerontological Center and Gerophylactic Center, Medical Station № 169, Department of Biomedical and Extreme Problems, in the period 1995-1999. The specific method of examination and measurement of the physiological parameters were used in accordance with recommendations of the so-called "Kiev method" of determination of BA [3]. It should be noted that this method was adopted in 1984 by the Ministry of Health of the USSR as a type method.

This method is based on measurements of the following parameters:

(1) Systolic, diastolic, and pulse arterial pressure (AP_S, AP_d, and AP_p, respectively), mm Hg;

(2) Rate of pulse wave propagation along elastic-type blood vessels (RPW_e) at the carotid-femoral artery segment, m/s;

(3) Rate of pulse wave propagation along muscular-type blood vessels (RPW_m) at the carotid-radial artery segment, m/s;

(4) Vital capacity (VC), ml;

(5) Expiratory breath holding time (BHT), s;

(6) Lens accommodation at the distance of the closest visual point (A), diopter;

(7) Auditory threshold (AT) at 4000 Hz, dB;

(8) Static balancing (SB) on the left foot, s;

(9) Body weight (BW), kg;

(10) Self-assessment of health state (SAHS), the number of negative answers to a standard 29-point questionnaire;

(11) Vexler symbol-digital test (VT), number of correctly filled cells within 90 s.

The lens accommodation index was measured in millimeters: this index was converted to diopters for the use in Kiev formulas as required by this method.

According to the method described in [3], quantitative estimates of BA can be obtained using the following empirical equations:

In men,

$$BA = 58.873 + 0.180AP_S - 0.073AP_d - 0.041\ AP_p - 0.262RPW_e + 0.646RPW_m - 0.001VC + 0.005BHT - 1.881A + 0.189AT - 0.026SB - 0.107BW + 0.320SAHS - 0.327VT. \quad (7)$$

In women,

$$BA = 16.271 + 0.280AP_S - 0.193AP_d - 0.105AP_p + 0.125RPW_e + 1.202RPW_m - 0.003VC - 0.065BHT - 0.621A + 0.277AT - 0.070SB + 0.207BW + 0.039SAHS - 0.152VT. \quad (8)$$

In addition to the basic procedure, there are two simplified versions of this method each based on four biological markers. The first version is based on four biological markers of the highest information value (RPW_e, A, AT, and VT in men and AP_S, AT, B W, and VT in women). The second version of the method is based on the biological markers simplest to measure (AP_S, BHT, SB, and SAHS in men and AP_p, BW, SB, and SAHS in women).

The results of measurements were statistically processed using standard methods of correlation and regression analysis (Student's *t*-test for comparing samples and χ^2-test for analyzing data categories).

RESULTS AND DISCUSSION

The results of estimation of the BA approximation quality in the Moscow population studied in this work using the "Kiev method" of determination of BA are given in the table. Similar parameters of the reference Kiev population studied in [3] are also given in the table for comparison.

The results of the study show that the "Kiev method" of determination of BA in the case of the Kiev population provides more adequate explanation of the correlation between age and biological markers than in the case of the Moscow population. Even the full version of the method allowed the biological marker dynamics to be brought into correlation with slightly more than 50% of the age variation parameters in women and less than 33% of the parameters in men. It is interesting to note that in the reference population any version of the method provided better age approximation in men than in women, whereas in the experimental population studied in this work the sex ratio of the age approximation accuracy was the opposite.

It follows from Fig. 2 that there is a trend toward a decrease in the BA estimates obtained using the full version of the "Kiev method." The calculated values of BA in men and women were found to be 5.9 ± 1.8 and 6.9 ± 0.7 years less than CA, respectively (the difference was significant at a confidence level $p < 0.01$ and $p < 0.0001$, respectively). If the values of BA were compared with normal values of the PBA rather than individual age values, the decrease in the BA estimates would remain statistically significant even against the background of the correction factor. The BA - PBA difference in this case was -6.9 ±1.6 years ($p < 0.001$) and -6.3 ±0.6 years ($p < 0.0001$) in men and women, respectively. Because the magnitude of the BA estimate decrease was approximately the same in men and women, further BA analysis by the "Kiev method" was performed without regard to the patient's sex. The BA decrease relative to the PBA was approximately the same in men and women, and it was less pronounced in people younger than 40 years (BA - PBA = -4.8 ± 1.0 years; the difference from zero was insignificant) than in subjects older than 40 years (BA - PBA = -7.3 ± 0.7; $p < 0.05$). The fact that the sign of the parameter of individual aging of a given subject relative to other people of the same age was negative and that the absolute value of the parameter increases with age can be explained by the fact that the experimental group contained appar-

Values of the coefficient of multiple determination of biological age (B A) as calculated according the "Kiev method" using a set of biological markers

Method of B A evaluation	Coefficient of multiple determination r^2			
	experimental Moscow population		reference Kiev population	
	men	women	men	women
Full version	0.307	0.516	0.863	0.706
The first reduced method	0.195	0.281	0.837	0.640
The second reduced method	0.171	0.235	0.629	0.581

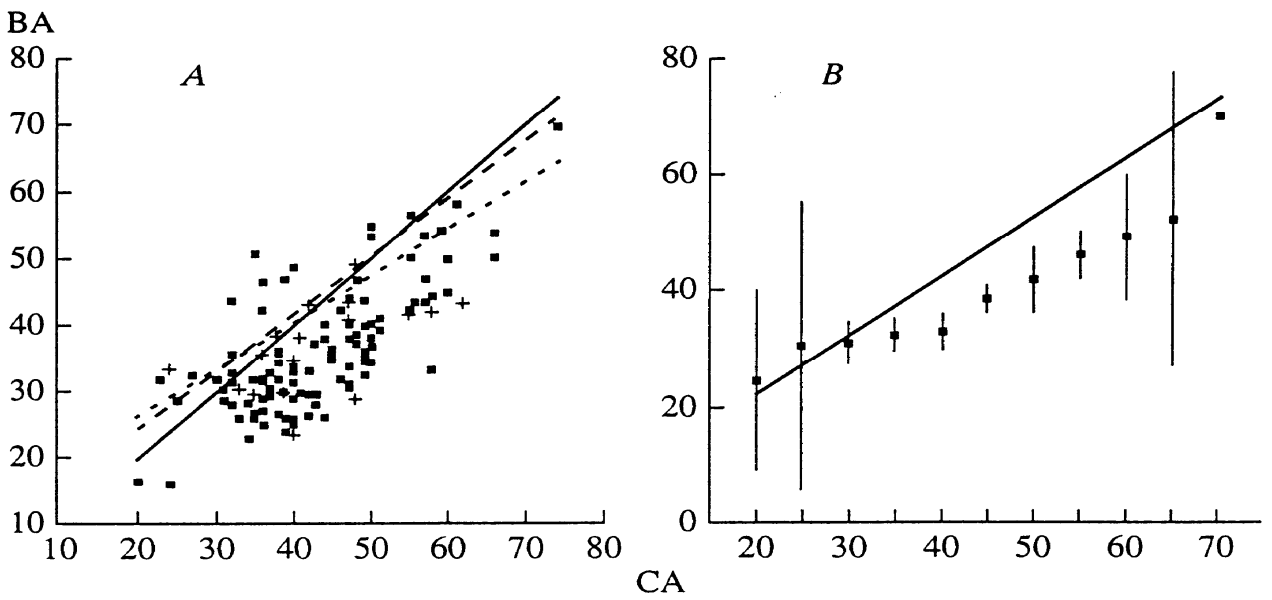

Fig. 2. Approximation of BA in accordance with the "Kiev method": *A,* individual data (crosses and dots correspond to men and women, respectively); the straight solid line is the BA = CA function. The dashed line is the PBA(CA) function for men. The dotted line is the PBA(CA) function for women; *B,* data averaged over five-year intervals (arithmetic mean together with an interval of 95% confidence). The straight solid line is the BA = CA function.

ently healthy subjects, whereas the reference group included less healthy subjects.

Averaging of the BA estimates over five-year groups of CA revealed an interesting trend (Fig. 2b): within the age range from 25 to 40 years, the mean calculated values of BA were virtually the same in all age groups (i.e., the intragroup variability was significantly higher than the age-related difference). However, a pronounced linear dependence between BA and CA was typical of the age range from 40 to 65 years. Therefore, the results of this study suggest that it is fundamentally impossible to construct a universal linear model of BA that would provide equally effective estimates within all segments of the whole age range from puberty to extreme old age.

The curves of multiple regression dependence of BA on the same set of biological markers but with coefficients different from those in equations (7) and (8) were plotted to obtain adequate estimates of BA. The coefficients of multiple determination in men and women were increased to 0.870 and 0.717, respectively. Therefore, the coefficients of multiple determination in the experimental population were close to or even higher than similar coefficients of multiple determination in the reference population tested by the full version of the "Kiev method". This result suggests that the set of biological markers used in these studies provides virtually equal efficiency of BA assessment in different populations. However, certain parameters of regression equations were insignificantly different from zero (i.e., they were determined with insufficient reliability). The following algorithm was used to obtain reliable estimates of BA. The initial BA equation included all biological markers studied except AP_p and self-assessment of health state (SAHS), because AP_p is a calculated parameter and determination of SAHS requires further refinement. After that, all parameters whose weighting coefficients were close to zero with the highest probability were sequentially excluded from consideration. The reduction in the number of biological markers continued until all weighting coefficients became significantly different from zero (p =0.05 was taken to be the critical level of significance). As a result, the following equations of multiple regression were obtained:

In men,

$$BA = 19.455 + 5.460RPW_e - 0.005VC - 0.052SB + 0.166BHT. \quad (9)$$

In women,

$$BA = 1.717 + 5.197RPW_e - 0.072SB + 0.165AT + 0.017A. \quad (10)$$

The coefficients of multiple determination for these equations of BA calculation were virtually identical in men and women (0.694 and 0.664, respectively) and were significantly higher than in the reduced versions of the "Kiev method" based on four biological markers for men and four biological markers for women. This is probably due to the biological marker selection by the criterion of reliability of their contribution to BA. The sex difference in the BA structure was found to be quite significant. The parameters of the cardiovascular and respiration systems (RPW_e, VC, and BHT) provided the dominant contribution to the regression equation in men, whereas the role of these parameters in the regression equation for women was less significant. Indeed, the absolute value of the RPW_e coefficient in equation (10) is lower than in equation (9), whereas parameters VC and BHT are not included in equation (10) at all. On the other hand, the parameters of the analyzer system in

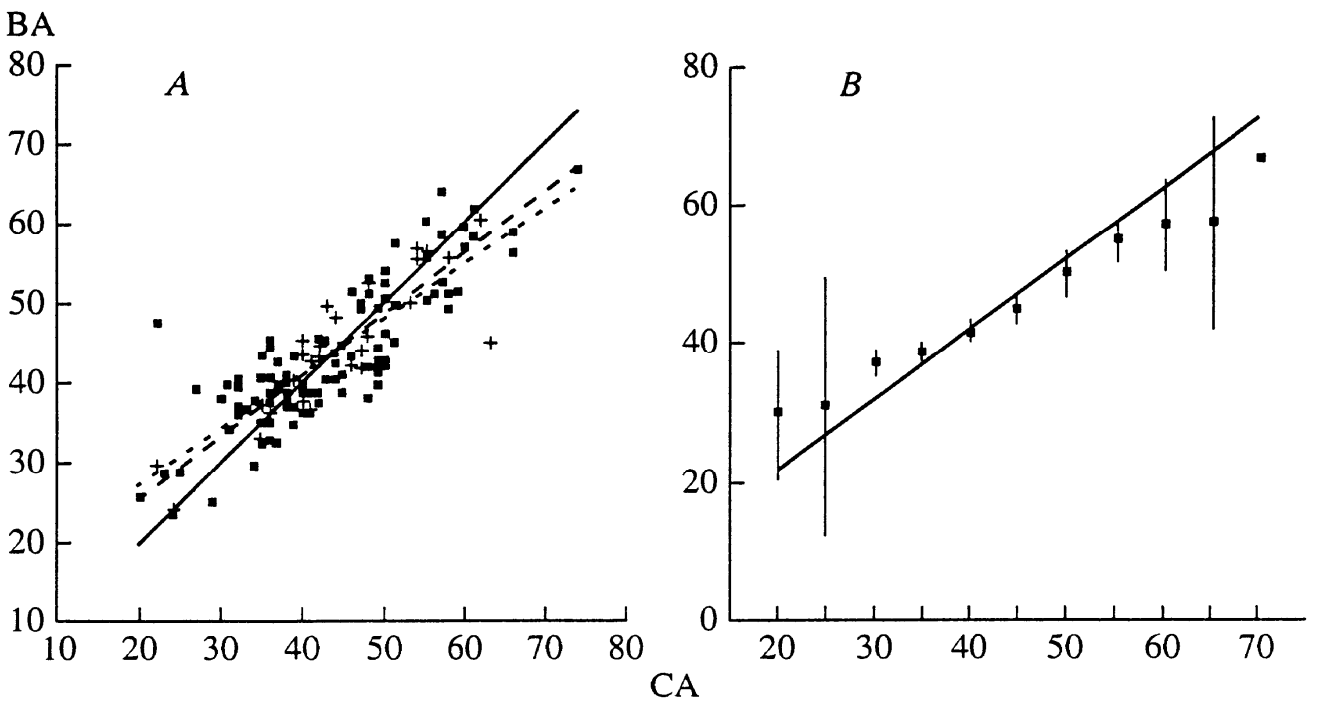

Fig. 3. Approximation of BA in accordance with the nonlinear regression model: *A,* individual data; *B,* data averaged over five-year intervals (arithmetic mean together with an interval of 95% confidence). Notation as in Fig. 2.

women play a more important role than in men: the integral equation for biological age contains parameters of vision and hearing, whereas the relative weight of parameter SB in equation (10) is higher than in equation (9).

Further refinement of the method of BA determination is possible if the nonlinear character of age dynamics of individual biological markers is taken into account. Of the biological markers included in equations (9) and (10), exponential approximation of two parameters (SB in men and A in women) was found to be more effective than linear approximation. Therefore, a logarithmic plot of the inverse dependence of BA on these biological markers is more suitable. Taking into account the nonlinear effects considered above, the equations for BA calculation could be recast as follows:

In men,

$$BA = 23.400 + 5.246RPW_e - 0.004VC - 3.3711n(SB) + 0.191BHT. \quad (11)$$

In women,

$$BA = -21.337 + 4.911lRPW_e - 0.063SB + 0.173AT + 5.5121n(A). \quad (12)$$

If the value of A is expressed in diopters,

$$BA = 16.470 + 4.911RPW_e - 0.063SB + 0.173AT - 5.5121n(A). \quad (13)$$

The use of nonlinear functions allowed the aproximation quality factor to be increased to 0.762 and 0.691 in men and women, respectively. In addition, the reliability of the regression coefficients calculated from logarithmic equations is higher than in case of estimates obtained from equations (9) and (10). Indeed, the probability of the free term of the linear regression equation was found to be slightly higher than the critical level of 5%, whereas all coefficients in equations (11) -(13) differ from zero at a confidence level of $p < 0.01$. The quality of the approximation of BA by nonlinear regression equations (Fig. 3) is significantly higher than the quality of BA approximation provided by the initial version of the "Kiev method".

CONCLUSION

The results of this study show that application of the type methods of BA determination to various populations may cause a substantial shift in the resulting estimates. The quantitative estimates of the BA equation parameters are very sensitive to the procedure of selection of the reference population in which these parameters were determined. Therefore, in addition to the high information value of aging biological markers, the method of BA determination should be based on sufficiently reliable equation coefficients. Decreasing the number of biological markers and increasing the reliability of the corresponding equations can increase the efficiency of the methods of BA determination. The two goals can be achieved if the nonlinear character of age dynamics of biological markers and the statistical significance of the contribution of individual biological markers to the BA estimates are taken into account.

ACKNOWLEDGMENTS

This study was supported by the Ministry for Technologies of Russian Federation.

REFERENCES

1. Biological Aging Measurement, Dean, W., Ed., Los Angeles, 1988, 397 P.

2. Belozerova L.M., Ontogenetic Method of Determination of Human Biological Age, *Usp. Gerontol,* 1999, № 3, P. 193.

3. Voitenko V.P., Tokar' A.V., and Polyukhov A.M., Method of Determination of Human Biological Age, *Gerontologiya i geriatriya. Biologicheskii vozrast, nasledstvennost' i starenie* (Gerontology and Geriatrics: Biological Age, Heredity, and Aging), Chebotarev D.F., Ed., Kiev, 1984, P. 132.

4. Krut'ko V.N., Mamai A.V., and Slavin M.B., Classification, Analysis, and Application of Biological Age Indicators to Prognosis of Life Expectancy, *Fiziol. Chel,* 1995, № 6, P. 42.

5. Anstey K.J., Lord S.R., and Smith G.A., Measuring Human Functional Age: A Review of Experimental Findings, *Exp. Aging Res.,* 1966, Vol. 22, № 3, P. 245.

6. Bulpitt C.J., Assessing Biological Age: Practicality? (Review), *Gerontology,* 1995, Vol. 41 № 6, P. 315.

7. McClean G.E., Biomarkers of Age and Aging (Review), *Exp. Gerontol,* 1997, Vol. 32, nos. 1-2, P. 87.

8. Mooradian A.D., Biomarkers of Aging: Do We Know What to Look For?, /. *Gerontol,* 1990, Vol. 45, № 6, P. 183.

9. Practical Handbook of Human Biological Age Determination, Balin A.K., Ed., Boca Raton: CRC, 1996, 521 P.

10. Turturro A. and Hart R.W., Biomarkers of Aging: An Overview, *Biomed. Environ. Sci.,* 1991, Vol. 4, nos. 1-2, P. 130.

11. Aivazyan S.A., Enyukov I.S., and Meshalkin L.D., *Prikladnaya statistika. Issledovanie zavisimostei: spravochnoe izdanie* (Handbook of Applied Statistics: Correlation Studies), Moscow: Finansy i Statistika, 1985. 487 P.