Selection of a set of biomarkers and comparisons of biological age estimation models for Korean men

Haemi Jee*
Department of Sports Health Care, Namseoul University, Cheonan, Korea

Biological age (BA) represents the rate of the senescence with a set of biomarkers. The BA prediction models have not been compared to obtain an optimal BA prediction model with BA biomarkers for Korean men. The study aims to obtain a set of BA biomarkers and compare three of the reported statistical approaches for an optimal BA prediction model. The Korea National Health and Nutrition Examination Surveys data of 2009 to 2011 were used to select six BA biomarkers from 940 healthy subjects aged between 30 to 80 years. The multiple linear regression (MLR), principal component analysis (PCA), and Klemera and Doubal methods (KDM) were used to obtain three BA prediction models. Correlation coefficients \((r)\) with 95% confidence intervals (CI) and regression slopes were assessed. One of the Euro Quality of Life-5 Dimensions, mobility, was compared for feasibility test of each BA models. KDM showed greatest correlation \((r=0.88, P<0.05)\) with smallest 95% CI and regression slope (1.00). PCA also showed strong correlation \((r=0.79, P<0.05)\) with small 95% CI and regression slope (0.94). MLR \((r=0.68, P<0.05)\) showed over and underestimated BA results at the end of the age spectrum. Estimations of BA were most reliable with KDM. The PCA and MLR approaches were comparatively simple to devise for Korean men.

Keywords: Biological age, Biomarkers, Principal component analysis, Multiple linear regression

INTRODUCTION

Despite the commonly acknowledged deterioration rate for human, the actual rate is unique to an individual. Therefore, although the chronological age (CA) changes only with the function of time, various factors may influence the changes of biological age (BA). A gradual functional and structural decline occurs for increased risk for morbidity and mortality (Levine, 2013; Yin and Chen, 2005). Various internal and external factors have been known to either accelerate or deter such functional and structural deterioration rates (Dubina et al., 1983; Yin et al., 2005). Many scientists have long been working to find the most universal and optimal BA biomarkers and BA prediction model. BA is a quantitative way of expressing an aging individual’s actual life expectancy and health state (Klemera and Doubal, 2006). Therefore, BA is an exceptionally important concept in understanding longevity and living to the fullest life expectancy.

Various human subjects, biomarkers, and statistical methods have been used to select optimal biomarkers and obtain BA prediction models. Selecting the most universally representative biomarkers and applying the adequate calculative methods have been known to be some of the most critical parts of estimating correct BA (Dubina et al., 1984; Ingram et al., 2001; Jee et al., 2012; Nakamura and Miyao, 2007; Ueno et al., 2003). The rate of the aging process and the corresponding biomarkers may differ from a group to group due to different internal and external age related influences (Jee and Park, 2017; Levine, 2013). Most of the BA prediction models merge a set of biomarkers into a single variable through stepwise statistical approaches. Dubina and his colleagues were one of the first scientists to report a BA prediction model with multiple linear regression (MLR) and modified versions have been reported since the initial model (Dubina et al., 1984; Jee and Park, 2017). Other mathematical models have also been utilized to obtain more sound BA prediction models. A BA prediction

---

*Corresponding author: Haemi Jee [https://orcid.org/0000-0002-9066-1472]
Department of Sports Health Care, Namseoul University, 91 Daehak-ro, Seonghwan-eup, Sebuk-gu, Cheonan 31020, Korea
E-mail: wlgPal9@gmail.com
Received: December 12, 2018 / Accepted: January 8, 2019
model by Klemera and Doudal (2006) is one of the recent statistical approaches.

Despite previous approaches to obtain the most representative biomarkers and BA prediction models, universally optimal models have not been reported especially for the Korean male population. Therefore, the aims of this study were to utilize nationally conducted health examination results to obtain representative biomarkers and compare three of the most commonly used statistical approaches, the MLR, principal component analysis (PCA), and Klemera and Doudal methods (KDM), to provide a BA estimation model for the Korean men. The national health examination data with biomarkers referred in previous reports were utilized to reflect the aging process of the general Korean men. In addition, in order to observe the clinical applicability of the BA prediction models, one of the Euro Quality of Life-5 Dimensions (EQ-D5) factor, mobility, was used to compare the groups and BA by each model (Euroqol Group, 1990).

MATERIALS AND METHODS

Subjects

Comprehensive data from the Fourth and Fifth Korea National Health and Nutrition Examination Surveys (KNHANES) were utilized for this study. Written consent was obtained for clinical investigations prior to each health examination. The assessed variables were obtained during routinely held health examinations nationwide. Use of fourth and fifth KNHANES assessment data for the years 2009 to 2011 was approved by the ethics committee of the Korea Centers for Disease Control and Prevention (approval number: 2009-01CON-03-2C, 2010-02CON-21-C, 2011-02 CON-06-C).

Inclusion and exclusion process

The inclusion and exclusion criteria were strictly followed based on previous studies (Bae et al., 2008; Dubina et al., 1983; Jee and Park, 2017; Ueno et al., 2003). First, the outliers were excluded prior to the actual application of the exclusion criteria. Routine test batteries of physiological and hematological variables were utilized to conduct correlation analysis with CA. Physiological variables were systolic blood pressure (SBP), diastolic blood pressure, body mass index (BMI), forced expiratory volume in 1 sec (FEV1), waist-to-height ratio (WHtR), body weight (g), fat mass (g), lean body mass (g), and bone mineral density. Hematological variables were insulin, fasting glucose, glutamic pyruvate transaminase, blood urea nitrogen (BUN), urine pH, Ketone, total cholesterol, high-density lipoprotein cholesterol, triglyceride, glutamic oxaloacetic transaminase, hemoglobin, hematocrit, ferritin, white blood cell, red blood cell (RBC), bilirubin, occult hematuria, urobilinogen, serum creatinine, glycosylated hemoglobin, nitrite, urine protein, Vitamin D (VitD), and urine glucose.

Since the biomarkers and the prediction models should reflect the natural aging process, the subjects with influential confounding factors should be excluded. Therefore, the subjects with medical conditions such as hypertension, diabetes mellitus, thyroid function, hypokalemia, obesity, and cancer were all excluded. Moreover, the subjects with outstanding ranges of selected variables were also excluded. The clinically normal ranges of the selected variables were as follow: SBP ($<160$ mmHg), DBP ($<90$ mmHg), waist circumference (WC; 72.5–103 cm), BMI (16-27.5 kg/m²), FEV1 (FEV1/FVC > 0.70), hemoglobin ($\leq 12.7$), BUN (6–20 mg/dL), RBC (4.04–5.82 Mil/µL), urine creatinine (Ucrea; 0.5–1.1 mg/dL), and VitD (20 ng/mL) (American College of Sports Medicine, 2014; Berk and Korenblat, 2012; Ross et al., 2011; Wang et al., 2010; WHO Expert Consultation, 2004). In order to exclude the cofounding influence of different physiology by advanced age, waist to height ratio was utilized (Corrêa et al., 2016). Out of 9 physiological and 23 hematological variables 9 variables were selected through the Pearson’s correlation analysis ($r \geq 0.15$, $P < 0.05$). Additional exclusion steps were performed by comparison for redundancy and intervariable structural relationship between the variables via the correlation analysis and loading and unloading CA in the PCA, respectively (Dubina et al., 1984; Jee and Park, 2017). Finally, 6 variables, FEV1, RBC, SBP, WHtR, Ucrea, and VitD were selected as the biomarkers from 940 healthy male subjects (Table 1).

Table 1. Correlation coefficients and first principal components of the selected variables

| Variable         | Mean ± SD | $r$   | 1st PC loadings |
|------------------|-----------|-------|-----------------|
| SBP (mmHg)       | 116.61 ± 11.70 | 0.30* | -0.38           |
| RBC (million/µL) | 4.91 ± 0.33   | -0.28* | 0.63           |
| Ucrea (m/dL)     | 161.44 ± 71.72 | -0.19* | 0.67           |
| FEV1 (L)         | 3.47 ± 0.57   | -0.60* | 0.83           |
| Vitamin D (ng/mL)| 30.31 ± 6.74 | 0.15*  | 0.04           |
| WHtR             | 0.50 ± 0.03   | 0.25*  | -0.44          |
| % Total variance  | 1.87       |       | -               |

SD, standard deviation of each variable; $r$, Pearson correlation coefficients; PC, principal component; SBP, resting systolic blood pressure; RBC, red blood cell counts; Ucrea, uric acid level; FEV1, forced expiratory volume in 1 sec; WHtR, waist to height ratio. *$P<0.05$. 
BA prediction models

Three types of statistical approaches, the MLR, PCA, and KDM were used to acquire the BA prediction models in this study. First calculated approach was the BA prediction model based on the MLR approach (Dubina et al., 1983; Jee and Park, 2017). The MLR method is one of the first and the most popular mathematical approach for developing the BA prediction model (Dubina et al., 1983; Jee and Park, 2017).

\[
BA_{MLR} = 78.52 + 0.16 \times SBP - 6.14 \times RBC - 0.02 \times UCREA - 9.28 \times FEV_i + 0.14 \times VitD + 30.94 \times WHR
\]

The second method was the BA prediction model based on the PCA (Dubina et al., 1983; Jee et al., 2012). The PCA method combines all of the covariant biomarkers in one principal component for formation of one output of BA (Dubina et al., 1983). This method is developed based on the mathematical approaches of the MLR method. The MLR approach combined all the biomarkers in to one principal component. The PCA approach extended the MLR approach by reflecting the changes in CA and adjusting for under and overestimated BA values (Dubina et al., 1983; Jee et al., 2012; Ueno et al., 2003).

\[
BA_{PCA} = -1.08 + 0.32 \times SBP - 6.62 \times RBC - 0.04 \times UCREA - 8.26 \times FEV_i + 0.39 \times VitD + 111.77 \times WHR + 0.36 \times CA
\]

The last approach was the BA prediction model based on the KDM (Jee and Park, 2017; Klemera and Doubal, 2006). The last method by Klemera and Doubal is based on the theory that the differences in BA with equal CA correspond to individually unique aging process. That is, this approach is developed based on the concept that the actual values of the biomarkers are affected by the fluctuations of BA. Based on such hypothesis, meticulous mathematical approaches were applied (Klemera and Doubal, 2006).

\[
BA_{KDM} = \sum_{j=1}^{m} \left( x_j - q_j \right)^2 + \frac{C}{S^2} \sum_{j=1}^{m} \left( \frac{k_j}{s_j} \right)^2 + \frac{1}{S^2}
\]

Feasibility assessment

One of the items of EQ-D5, mobility, is self-reported quality of life assessment questionnaire with three levels (EuroQol Group, 1990). Level one states that ‘I have no problems in walking about (mobile group: YG1, MG1, OG1),’ and level two states that ‘I have moderate problems in walking about (less mobile group: YG2, MG2, OG2).’ Last level three was excluded from comparison analysis due to limited number of subjects in this study. The EQ-D5 mobility assessment was conducted with three aged groups of younger group (YG: CA > 45), middle-aged group (MG: 45 ≤ CA < 65), and older group (OG: 65 ≤ CA) for more detailed comparison across the age spectrum.

Statistical analyses

General data analyses were performed using IBM SPSS Statistics ver. 22.0 (IBM Co., Armonk, NY, USA). Pearson correlation analysis was conducted to compare the relationship between each potential biomarker and CA. Correlation coefficients greater than 0.15 with \( P \)-value equal to or less than 0.5 was used as the biomarker selection criteria based on previous studies. The biomarker...
results were expressed in mean and standard deviation. MLR and principal component analyses were conducted to obtain the BA prediction models of MLR and PCA. Mathematical calculations for the KDM were conducted by MATLAB 2018b of MathWorks Inc. (Natick, MA, USA). Spiro–Wilk test was used to assess whether data was normally distributed prior to conducting the comparative analysis. Comparisons between active and less-active groups were done through $t$-test. The significance level ($\alpha$) was set at 0.05.

RESULTS

The mean values of the selected biomarkers, SBP, RBC, Ucrea, FEV$_1$, VitD, and WHtR, the Pearson correlation coefficients with CA, and the first principal components are shown in Table 1. Estimated BAs with corresponding CAs based on the BA estimation models were drawn and shown in Fig. 1. The mean regression slopes for the MLR, PCA, and KDM were 0.46, 0.96, and 1.00, respectively. The mean ages and correlation coefficients of CA, KDM, PCA, and MLR are shown in Table 2. Table 2 also contains the confidence interval of each BA model. Same calculations were conducted for the different age groups of YG, MG, and OG for the EQ-D5 mobility comparisons. Additional calculations on the differences between CA and BA were also conducted for each age group (Table 2). Greatest differences in BA with CA in both the mobile and less mobile groups were shown in MLR.

DISCUSSION

The quest for elucidating the factors that influence the aging process and the deterioration rate is an ongoing process that has been challenged by the scientists for centuries. Several statistical approaches have been observed and utilized to predict biological aging or the rate of senescence. This study aimed to find an optimal set biomarkers of BA and utilized the most commonly used and recently derived mathematical approaches, MLR, PCA, and KDM, to develop and compare between these three BA prediction models. Nationally conducted health examination results were utilized for more general application in Korean men.

Through a tedious exclusion and inclusion selection process, six biomarkers of SBP, RBC, Ucrea, FEV$_1$, VitD, and WHtR were obtained. The most significantly changing biomarker in function of CA was FEV$_1$. Cardiorespiratory function has been reported by several previous BA related studies (Jee et al., 2012; Jee and Park, 2017; Ueno et al., 2003). Clinical influential factors on FEV$_1$ have been associated with increased morbidity and mortality (Coffman et al., 2017). However, without the clinical influences for deterioration, normal aging has also been associated deterioration of respiratory function which may include rigidity of the chest wall, reduced recoil of the lung, and reduced spirometric performance (Vaz Fragoso et al., 2016).

Second factor that was showed to have the greater correlating relationship was SBP. Various studies also reported of changes in SBP in relationship with aging. Previous BA study on Korean women also reported of increase SBP in function of CA (Jee and Park, 2017). Cardiac stiffness and fibrosis, left ventricular hypertrophy, valvular deterioration, diastolic dysfunction, reduced physical fitness, have been reported in association with the increased SBP and the aging process (Sorriento et al., 2015).

Reduced RBC counts in association with aging indicate structural changes of hemoglobin and reduced oxygenation capacity. Organ functions such as erythropoietin production of the kidney is supported by the scientific community. The quest for elucidating the factors that influence the aging process and the deterioration rate is an ongoing process that has been challenged by the scientists for centuries. Several statistical approaches have been observed and utilized to predict biological aging or the rate of senescence. This study aimed to find an optimal set biomarkers of BA and utilized the most commonly used and recently derived mathematical approaches, MLR, PCA, and KDM, to develop and compare between these three BA prediction models. Nationally conducted health examination results were utilized for more general application in Korean men.
and RBC production of the bone marrow decline with age (Stauder et al., 2018). Anemia in older persons is common. Interestingly, a study on anemia reported dramatic increase in the prevalence of anemia to nearly 50% in men over 80 years of age (Stauder et al., 2018).

Reports on the importance of VitD especially with the advanced age have been gaining popularity in recent years. VitD deficiency has been related to not only musculoskeletal health but a wide range of diseases such as some cancers, autoimmune disease, infectious disease, diabetes mellitus, neurological diseases, and mortality (Hossein-nezhad and Holick, 2013). Physiological factors including low VitD intake, inadequate sun exposure, low ambient ultraviolet radiation level, and medication lead to low body VitD concentration. Since all of the factors significantly reduce VitD level in the elderly population, risk for clinical complication can rise accordingly (Hossein-nezhad and Holick, 2013).

Reduced renal function is another major age related decline in function. With advanced age, significant loss of nephrons, abnormal level of electrolyte, dysfunctional metabolic and enzymatic activities of the renal tubular cells have been reported. With reduction rate of the renal blood circulation rate, deterioration of renal function with changes in creatinine level have been reported (Timiras, 2007). Ucrea was also one of the biomarkers of BA in the female population (Jee and Park, 2017).

Changes in body composition with advanced age have been repeated reported by previous studies. Reduced muscle mass with muscular strength and increase body fat composition are some of the characteristic changes with age. One of the significant changes in both men and women with age is increased waist circumference (Stevens et al., 2010). In previous study on BA with women, waist circumference was significantly increased with age (Jee and Park, 2017). Strong relationships of WHtR or waist to height ratio to metabolic syndrome, diabetes mellitus, hypertension, and hyperlipidemia has been reported in men (Ashwell et al., 2014). The process of aging is associated with redistribution of fat tissue from subcutaneous to visceral depots from late middle age for both men and women. Fat accumulation occurs in bone marrow, muscle, liver, and other site reflecting age-dependent dysfunction (Stevens et al., 2010).

As for the comparisons of the BA prediction models by the MLR, PCA, and KDM approaches, the KDM approach showed the most linearly fitting distributions to CA. The MLR approach showed over- and underestimated BA toward the ends of the age spectrum. The PCA model which reflected CA and corrected for skewed age spectrum showed comparatively higher coefficient of determination ($r^2$). Such results were similar to previous studies (Jee and Park, 2017; Klemera and Doubal, 2006). The approach by Klemera and Doubal assumes the differences of BAs and CAs are related to unique difference of the biomarkers and equality of CA for BA in a population. Such approach of KDM provides a fitting linearity with CA. The correlating relationship between BA and CA by the age groups (YG, MG, and OG) showed greatest coefficient of determination ($r^2$) for the middle aged group (45 to 65 years) for all three models. This indicates that BA estimation for the middle aged group most fitting among all age groups.

Feasibility of the BA prediction models was also conducted by comparison EQ-D5 mobility. The age prediction was smallest for the middle aged group. The MLR estimated either greater or smaller BA for the younger group (YG: 5.90 years) and older group (OG: 9.84 years) with no mobility difficulties. Moreover, the BA prediction with the MLR approach was negative (-11.24 years) in older less mobile men (OG2). Such results indicate unadjusted skewed age spectrum of the MLR prediction model may only be reliable for the middle of the age spectrum. Therefore, BA predictions by the MLR approach should be carefully considered especially for the younger or older aged men.

There were several limitations to the study. First, this study used the KNHANES data between the years 2009 to 2011. Recent data excluded some of the physiological assessment results necessary as BA biomarkers. Although KNHANES results reflect the national health of the Korean adults, there are some of the necessary markers that are important as the BA biomarkers. Therefore, the KNHANES data offer limited information. Other vital biomarkers of BA such as physical fitness variables should also be considered in the BA prediction model. Another limitation is that the study only compared three of the published BA prediction models. Other BA prediction approaches exist and have been reported in other studies. Finally, the feasibility test was conducted with only one quality of life variable. Although the mobility factor provided adequate information on applicability of three BA prediction models, the applicability should be conducted with other clinical complications to assess universal application as BA prediction model.

In conclusion, six biomarkers of FEV1, RBC, SBP, WHtR, Ucrea, and VitD were used in the KNHANES data between the years 2009 to 2011. There were several limitations to the study. First, this study used the KNHANES data between the years 2009 to 2011. Recent data excluded some of the physiological assessment results necessary as BA biomarkers. Although KNHANES results reflect the national health of the Korean adults, there are some of the necessary markers that are important as the BA biomarkers. Therefore, the KNHANES data offer limited information. Other vital biomarkers of BA such as physical fitness variables should also be considered in the BA prediction model. Another limitation is that the study only compared three of the published BA prediction models. Other BA prediction approaches exist and have been reported in other studies. Finally, the feasibility test was conducted with only one quality of life variable. Although the mobility factor provided adequate information on applicability of three BA prediction models, the applicability should be conducted with other clinical complications to assess universal application as BA prediction model.
tage of simplicity in formulation. Various clinical groups should be compared for universal application of the BA prediction models.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

American College of Sports Medicine. ACSM’s guidelines for exercise testing and prescription. 9th ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2014.

Ashwell M, Mayhew L, Richardson J, Rickayzen B. Waist-to-height ratio is more predictive of years of life lost than body mass index. PLoS One 2014;9:e103483.

Bae CY, Kang YG, Kim S, Cho C, Kang HC, Yu BY, Lee SW, Cho KH, Lee DC, Lee K, Kim JS, Shin KK. Development of models for predicting biological age (BA) with physical, biochemical, and hormonal parameters. Arch Gerontol Geriatr 2008;47:253-265.

Berk P, Korenblat K. Approach to the patient with jaundice or abnormal liver tests. In: Goldman L, Schafer AI, editors. Goldman’s Cecil medicine. 24th ed. Philadelphia (PA): Elsevier/Saunders; 2012. p. 956-966.

Coffman KE, Carlson AR, Miller AD, Johnson BD, Taylor BJ. The effect of aging and cardiorespiratory fitness on the lung diffusing capacity response to exercise in healthy humans. J Appl Physiol (1985) 2017;122:1425-1434.

Corrêa MM, Thumé E, De Oliveira ER, Tomasi E. Performance of the waist-to-height ratio in identifying obesity and predicting non-communicable diseases in the elderly population: a systematic literature review. Arch Gerontol Geriatr 2016;65:174-182.

Dubina TL, Dyundikova VA, Zhuk EV. Biological age and its estimation. II. Assessment of biological age of albino rats by multiple regression analysis. Exp Gerontol 1983;18:5-18.

Dubina TL, Mints AiYa, Zhuk EV. 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 1984;19:133-143.

EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.

Hossein-nezhad A, Holick MF. Vitamin D for health: a global perspective. Mayo Clin Proc 2013;88:720-755.

Ingram DK, Nakamura E, Smucny D, Roth GS, Lane MA. Strategy for identifying biomarkers of aging in long-lived species. Exp Gerontol 2001;36:1025-1034.

Jee H, Jeon BH, Kim YH, Kim HK, Choe J, Park J, Jin Y. Development and application of biological age prediction models with physical fitness and physiological components in Korean adults. Gerontology 2012;58:344-353.

Jee H, Park J. Selection of an optimal set of biomarkers and comparative analyses of biological age estimation models in Korean females. Arch Gerontol Geriatr 2017;70:84-91.

Klemera P, Doubal S. A new approach to the concept and computation of biological age. Mech Ageing Dev 2006;127:240-248.

Levine ME. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? J Gerontol A Biol Sci Med Sci 2013;68:667-674.

Nakamura E, Miyao K. A method for identifying biomarkers of aging and constructing an index of biological age in humans. J Gerontol A Biol Sci Med Sci 2007;62:1096-1105.

Ross AC, Talyor CL, Yaktine AL, Del Valle HB, editors. Institute of Medicine (US) Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. Dietary reference intakes: calcium, vitamin D. Washington (DC): National Academies Press (US); 2011.

Serraino D, Franco A, Rusciano MR, Maione AS, Soprano M, Illario M, Iaccarino G, Ciccarelli M. Good at heart: preserving cardiac metabolism during aging. Curr Diabetes Rev 2015;11:90-99.

Stauder R, Valennt P, Theurl I. Anemia at older age: etiologies, clinical implications, and management. Blood 2018;131:505-514.

Stevens J, Katz EG, Huxley RR. Associations between gender, age and waist circumference. Eur J Clin Nutr 2010;64:6-15.

Timiras PS, editor. Physiological basis of aging and geriatrics, Boca Raton: CRC Press; 2007.

Ueno LM, Yamashita Y, Moritani T, Nakamura E. Biomarkers of aging in women and the rate of longitudinal changes. J Physiol Anthropol Appl Human Sci 2003;22:37-46.

Vaz Fragoso CA, McAvay G, Van Ness PH, Metter EJ, Ferrucci L, Yaggi HK, Concato J, Gill TM. Aging-related considerations when evaluating the forced expiratory volume in 1 second (FEV1) over time. J Gerontol A Biol Sci Med Sci 2016;71:929-934.

Wang Z, Ma J, Si D. Optimal cut-off values and population means of waist circumference in different populations. Nutr Res Rev 2010;23:191-199.

WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004;363:157-163.

Yin D, Chen K. The essential mechanisms of aging: Irreparable damage accumulation of biochemical side-reactions. Exp Gerontol 2005;40:455-465.