Archives of Gerontology and Geriatrics 70 (2017) 84–91

Contents lists available at ScienceDirect

Archives of Gerontology and Geriatrics

journal homepage: www.elsevier.com/locate/archger

Full Length Article

Selection of an optimal set of biomarkers and comparative analyses of biological age estimation models in Korean females

Haemi Jee, Jaehyun Park*

Department of Information and Communication, Inha University, 100 Inha-ro, Nam-gu, Incheon, 22212, Republic of Korea

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form 4 November 2016
Accepted 9 January 2017
Available online 12 January 2017

Keywords:
Chronological age
Biological age
Multiple linear regression
Principle component analysis

ABSTRACT

To date, an optimal working model which predicts biological age (BA) with a set of working biomarkers has not been devised to represent the Korean female population. Accuracy of prediction and applicability are required of an optimal set of commonly assessed biomarkers to provide information on the health status. The goal of this study was to identify a set of biomarkers that represent the aging process and to develop and compare different BA prediction models to elucidate the most fitting and applicable model for providing information on health status in the Korean female population. Using a series of selection processes, eight clinically assessable variables were selected by analyzing relations between 31 clinical variables and chronologic age in 912 normal, healthy individuals among 3642 female participants with ages ranging from 30 to 80 years. The multiple linear regression (MLR), principal component analysis (PCA), and the Klemera-Doubal (KDM) statistical methods were applied to obtain three different sets of BA prediction models. These three models were assessed by calculating and performing the coefficient determinations (r^2), regression slopes, effect sizes, pairwise *t*-tests, and Bland-Altman plots. The BA models were further compared for the applicability by calculating the BAs of clinical risk groups. MLR showed the narrowing effects at the either ends of the age spectrum with greatest effect sizes. PCA showed the greatest degree of dispersion and deviation from the regression center. These MLR and PCA trends were also exhibited by clinically risk groups. In conclusion, the KDM BA prediction model based on the selected biomarkers was found to provide the most reliable and stable results for the practical assessment of BA.

1. Introduction

Aging is a natural phenomenon that occurs in most organisms. The age of an individual can be estimated by either calculating chronological age (CA) or biological age (BA). Normal developmental phases and rates can be estimated by measuring elapsed time since birth (Jee et al., 2012). Although CA provide a simple, clear cut method for estimating aging, CA does not provide adequate information on the rate of decline or physiological breakdown of an individual (Levine, 2013).

The concept of biological aging was proposed to provide a reliable estimation of the degree of aging process. Since the initial proposal of the age-related biological changes by Alex Comfort in Comfort in 1969, many scientists approached the concept of BA using various biomarkers and estimation algorithms (Comfort, 1969; Levine, 2013). Biological aging is an individual process, defined by a gradual functional and structural decline that increased risk of impairment, morbidity, and mortality (Levine, 2013; Yin & Chen, 2005). Multivariate factors such as the genetic composition, physical fitness, and external environmental stressors have been suggested to influence the rate of BA and various models have been proposed (Dubina, Dyundikova, & Zhuk, 1983; Hofecker, Skalicky, Kment, & Niedermuller, 1980; Ludwig & Smoke, 1980; Park, Cho, Kwon, & Lee, 2009a, 2009b).

Despite the substantial amount of work invested in elucidating the senescent process, no optimal universal method for estimating the BA of an individual has received consensus approval (Levine, 2013). However, several estimation methods based on various sets of age-dependent variables or biomarkers have been suggested to provide reliable BA prediction models (Dubina et al., 1983; Jee et al., 2012; Klemera & Doubal, 2006). Many of the prediction models involve the merging of multiple biomarkers into a single variable using a stepwise calculative process and computational algorithms, such as, multiple linear regression method (MLR), principle component analysis (PCA), or the method suggested by Klemera and Doubal (KDM) (Dubina et al., 1983; Jee et al., 2012;

* Corresponding author.
E-mail address: jhyun@inha.ac.kr (J. Park).

http://dx.doi.org/10.1016/j.archger.2017.01.005
0167-4943/

Klemera & Doubal, 2006). These methods utilize a set of representative biomarkers obtained from a group of normal, healthy individuals for comparative analysis (Jee et al., 2012; Park et al., 2009a; Ueno, Yamashita, Moritani, & Nakamura, 2003).

Requirements for assessment and diagnosis accuracy are inevitable for proper health assessment and promotion. Selecting the most correlating biomarkers from a group of individuals that represents general population, and the application of suitable calculation methods is probably one of the soundest approaches when developing BA prediction models (Jee et al., 2012; Klemera & Doubal, 2006; Levine, 2013; Park et al., 2009a). The critical issue when selecting biomarkers and developing corresponding BA estimation model is to identify the biomarkers that significantly influence the aging process in the general population and to use these to estimate the rate of biological aging with more precision. Furthermore, the most accurate computational algorithm should be applied to such representative biomarkers.

Therefore, the goals of this study were to: (a) To obtain the a set of biomarkers in a group cohort representing the general Korean female population, (b) To compare the three major BA prediction models derived using three computational algorithms, and (c) To apply the BA prediction models to a group of subjects with diagnosed clinical risks to investigate model validities.

2. Methods

2.1. Participants

In this study, comprehensive data from the Fourth and Fifth Korea National Health and Nutrition Examination Surveys (KNHANES) were utilized. The surveys included the health behavior questionnaires and details of anthropometric measurements, cardiovascular and respiratory functions, and hematological and urological indices. 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 III assessment data for the years 2009–2011 was approved by the ethics committee of the Korea Centers for Disease Control and Prevention (IRB approval no: 2009-01CON-03-2C, 2010-02CON-21-C, 2011-02CON-06-C).

Outliers and missing data were first excluded prior to conducting the selection procedure. Female participants between the ages of 30 to 80 (mean50.87 (SD 10.72)) years were selected. Subjects aged 30 years or older were selected because significant age-related declines in major organs and physiological functions (Baker & Sprott, 1988; Jee et al., 2012).The study protocol was designed, approved and conducted in accordance with the ethical standards of the Declaration of Helsinki with approval from the Institutional Review Board of the Korean Center for Disease Control and Prevention.

In order to examine the clinical applicabilities of the three models, BAs of patients with impaired glucose tolerance (IGT, n = 350) or diabetes mellitus (DM, n = 75) diabetes mellitus were compared with those of healthy subjects. The selection criteria used to define IGT and DM were fasting plasma glucose levels of ≥100 mg/dl to <126 mg/dl and ≥126 mg/dL, respectively (American Diabetes Association, 2014). All other clinical manifestations were reviewed and excluded to allow the influence of diabetes on BA to be investigated in the absence of obvious confounders.

2.2. Test items and procedures

Routinely tested physiological, hematological and urological variables composed of 31 biomarkers were assessed for correlations with CA in 3642 study subjects: SBP (systolic blood pressure), DBP (diastolic blood pressure), heart rates (bpm), BMI (body mass index), FEV1, WC (waist circumference), insulin, FG (fasting glucose), GPT(glutamic pyruvate transaminase), BUN(blood urea nitrogen), urine pH, Ketone, TC(total cholesterol), HDL cholesterol, TG(triglyceride), GOT(Glutamic Oxaloacetic Transaminase), Hg (hemoglobin), Ht (hematocrit), Frtn (ferratin), WBC (white blood cell), RBC (red blood cell), bilirubin, occult hematuria, urobilinogen, serum creatinine, HbA1c, nitrite, urine protein, and urine glucose. Out of these 31 variables, variables with correlation coefficient of ≥0.15 ($p < 0.01$) were selected as the biomarkers of BA (Jee et al., 2012).

2.3. Biomarker selection criteria and exclusion procedure

Biomarkers of BA should be measureable parameters that accurately reflect the biological aging process, provide reproducible results, change independently with time, and represent the intrinsic progression of aging (Baker & Sprott, 1988; Jee et al., 2012). In order to observe the intrinsic biological progression of aging, the participants with cofounding factors for abnormal aging should be excluded. Therefore, we excluded participants with medically diagnosed medical conditions, such as cancer, hypertension, diabetes mellitus, dyslipidemia, and thyroid dysfunction (Bae et al., 2008; Jee et al., 2012; Park et al., 2009a; Ueno et al., 2003). In addition, clinically normal ranges of the selected biomarkers were used to exclude the clinically risk ranges. The clinically normal ranges of the variables were as follow: SBP (<160 mm Hg), WC (65.5–101.2 cm), TC (<225 mg/dL), GOT (10–40 mg/dL), Frtn (12–150 mg/dL), BUN (5–20 mg/dL), Ucrea (0.4–1.1 mg/dL), and FEV_1 ($FEV_1/FVC > 0.70$). The ranges were used to exclude subjects with abnormal variable values (ACSM, 2010; Berk & Korenblat, 2011; Giboney, 2005; Hoffman, Benz, & Silberstein, 2012; Hosten, 1990; Looker, Dallman, Carroll, Gunter, & Johnson, 1997; Park et al., 2015; Wang, Ma, & Si, 2010).

After excluding the participants with abnormal values of each variable, Pearson's correlation coefficients were computed to select the biomarkers that exhibited deteriorating in function of CA. Variables with correlation coefficients of ≥0.15 ($p < 0.01$) were selected as potential biomarkers of BA. Further elimination step was performed by testing for redundancy between the variables. As a result, data of 912 participants were selected out 3642 participants by excluding the abnormal values (Jee et al., 2012). Selected variables were assessed for redundancy and tested for inter-variable structural relationships between the variables by loading and unloading CAs (Dubina, Mints, & Zhuk, 1984; Jee et al., 2012). A total of 5 hematological and 3 physiological variables of the routinely tested items were included for this analysis (Table 1). After the initial exclusion process, other influential factors, such as quality of life as determined by EQ-D5 (Euro Quality of Life – 5 Dimensions) and

Table 1
Means (standard deviation), correlation coefficients, and principle factors of selected the selected biomarkers.

Variables	Mean ± SD	cc	principle factors
SBP (mmHg)	113.93 ± 13.71	0.45**	0.667
WC (cm)	77.98 ± 6.69	0.19**	0.369
TC(mg/dL)	184.11 ± 25.42	0.28**	0.485
GOT (mg/dL)	19.59 ± 4.75	0.29**	0.522
Frtn (mg/dL)	48.45 ± 27.34	0.30**	0.434
BUN (mg/dL)	13.27 ± 3.11	0.31**	0.365
Ucrea (mg/dL)	113.91 ± 61.53	−0.25**	−0.338
FEV_1 (mL)	2.52 ± 0.43	−0.60**	−0.633

*$p < 0.05$, **$p < 0.01$; Mean: mean of each variable; SD: standard deviation of each variable; cc: Pearson's correlation coefficients; SBP: systolic blood pressure; WC: waist circumference; TC: total cholesterol; GOT: Glutamic Oxaloacetic Transaminase; Frtn: ferratin; BUN: blood urea nitrogen; Ucrea (Urine creatinine): FEV_1: forced expiratory volume in 1 s.