Estimating a cardiac age by means of heart rate variability

A. COLOSIMO,1 A. GIULIANI,2 A. M. MANCINI,2 G. PICCIRILLO,3 AND V. MARIGLIANO3

1Dipartimento Scienze Biochimiche, Università di Roma “La Sapienza,” 00185; and 3Istituto Prima Clinica Medica, 00161 Rome, Italy

Colosimo, A., A. Giuliani, A. M. Mancini, G. Piccirillo, and V. Marigliano. Estimating a cardiac age by means of heart rate variability. Am. J. Physiol. 273 (Heart Circ. Physiol. 42): H1841–H1847, 1997.—A data set of R-R intervals recorded for at least 15 min in 141 healthy individuals of different ages and under two different conditions (“resting” and “tilted” states) has been considered. The data have been subjected to spectral analysis by fast Fourier transform methods and considered in view of the possibility to work out a model in which the chronological and cardiac age could be compared. Understanding the results was greatly facilitated by 1) working out a number of derived variables from the original ones to highlight the presence of small but conceptually important variability factors; 2) extraction of the principal components from the original as well as from the derived variables to exclude redundancies and correlation effects; and 3) automatic clustering of the subjects in age classes, which allowed removal of individual variability within each class. The main conclusion is that, within the examined individuals, cardiac and chronological ages do not match for ages higher than ~50 years; this could reflect the presence of subtle (and difficult-to-envisage) biases in the data analysis or a real discrepancy. The latter hypothesis should be confirmed by similar observations in different systemic contexts. The use of a simple equation relating chronological and cardiac age, derived from a careful regression analysis on our data set and of general use for screening purposes, is demonstrated.

aging; power spectral analysis; statistical analysis; head-up tilt

IT HAS BEEN RECENTLY clarified that human heart rate variability (HRV) is distributed according to a broadband power spectrum that follows a hyperbolic trend called “1/f noise” (9, 11). Such behavior produces a spectral power distribution inversely correlated to the frequency (4) which is typical of the fractal dynamics or, in more general terms, of the important family of physical phenomena called “random walks,” whose best-known member is the Brownian motion. In particular, the heart rate seems to show a beat-to-beat regulation of the constrained random walk type (3) to which the sympathetic and parasympathetic modulatory influences are probably superimposed.

Even though, from a quantitative point of view, the frequency regions surely assigned to sympathetic and parasympathetic influences [i.e., high frequency (HF) and low frequency (LF)] probably play a minor role in the HRV dynamics (11), the precise identification of such a role entails a relevant diagnostic importance (10). Because of the obvious correlations existing among quantities forced to assume a fixed total value, such as the different bands in a power spectrum, the application of quite sophisticated statistical methods able to discriminate among different contributions to the HRV becomes useful. This allows us, in particular, to single out the various effects of aging on the heartbeat dynamics and, in conjunction with an appropriate database, to work out a reliable estimator of the cardiac age.

In the next sections, the approach followed in the analysis of a large data set will be described in depth with the aim to identify the variables significantly correlated with aging among those considered in our database. Subsequently, a model that describes the cardiac age and is directly derived from the aforementioned analysis will be discussed and shown to be fairly accurate, given the intrinsic high variability of clinical data.

METHODS

Data Collection Protocol

Of the 455 subjects initially enrolled for this study, 141 completed the study protocol. The more important causes of exclusion were a history or actual evidence of cardiovascular, respiratory, renal (presence of proteinuria or creatinine >132.6 μmol/l), liver, or gastrointestinal disease. Other exclusion criteria were diastolic blood pressure >90 mmHg or systolic blood pressure >150 mmHg, body mass index >26 kg/m², smoking (>5 cigarettes/day), diabetes (presence of glycosuria or fasting glycemia >6.6 mmol/l or >6.1 mmol/l at 2 h after glucose loading), plasma cholesterol >5.7 mmol/l, arrhythmias or conduction abnormalities, echocardiographic evidence of wall motion abnormalities or valvular disease, or ultrasound evidence of carotid stenosis of importance. In particular, 45 subjects were excluded because presyncopal symptoms developed during the head-up tilt test (see below).

Heart rate recordings for spectral analysis in all subjects took place according to the following protocol (7). At 8:30 AM, in a quiet and comfortable environment (24°C), the subjects rested supine for at least 30 min before undergoing a 15-min electrocardiographic recording (resting state). The same subjects underwent, successively, head-upright tilt testing, a passive orthostatic maneuver obtained with a motorized tilt table. After 15 min in an upright (90°) position, the subjects underwent a second 512-beat (~5 min) electrocardiographic recording (tilted state). Transition from 0° to 90° took ~15 s. Blood pressures were recorded before tilting and at 30-s intervals during tilt. If hypotension (systolic arterial pressure drops of 20 mmHg) or symptoms indicating the onset of syncope, nausea, or gastric piroxys developed during tilt, testing was stopped and the subject was excluded from the study.

Off-Line HRV Analysis

Figure 1 reports the age distribution of the subjects analyzed in the present work. (It is instructive to compare this information with the results of clustering of the same data reported in the Results (Table 4).) The fast Fourier transform
was used to compute power spectral densities from the electrocardiographic recordings (4–6). Tracings were analyzed with software running on a Compaq computer with a 486 DX2 microprocessor. Electrocardiographic signals were digitized, stored on a hard disk, and sampled at a rate of 500 Hz, with 12 precision bits. The QRS complex (lead II) was automatically recognized as a classic derivative-threshold algorithm. Each QRS complex and each trigger point were also subsequently checked by an expert cardiologist. Adjacent complexes were used to correct ectopic beats for linear interpolation, and electrocardiographic tracings with >1% premature beats were eliminated from the analysis. The length of stationary segments of electrocardiographic recordings was 512 beats. Signals were transformed to time series of instantaneous heart rate (beats/min). The sampling rate of these series was equalized to correspond to one sample per second. Spectral power of heart rate was expressed in reciprocals squared seconds, because the HRV expressed in millisecond shows a skewed distribution.

The analyzed spectral components included total power or spectral density (TP); HF power, from 0.16 to 0.40 Hz; LF power, from 0.04 to 0.15 Hz; very low frequency (VLF) power, <0.04 Hz; and LF/HF, the ratio of LF power to HF power.

The software applications for data acquisition and storage and for the spectral analysis were designed and produced by our research group. The statistical analyses were carried out almost exclusively by means of the SAS package (6.0 release) for the IBM PC.

Data Analysis: General Strategy

Some considerations indicate that a fraction of the whole information made available by our experimental setup might be hidden at the observation scale used (see the introduction). Thus three sets of derived variables have been introduced with the aim to amplify the quantitatively minor but possibly biologically relevant components. The variables characterizing each set and their definitions in terms of the original variables are reported in Table 1.

The first set of derived variables (derived set I) is obtained by using as new variables the fraction of the total power absorbed by each band; this eliminates the major source of variability among the various bands in a power spectrum, i.e., the absolute absorbed power, and underlines the interindividual variability in terms of spectral shapes.

The second transformation of variables (derived set II) has been worked out to highlight the interindividual differences in responses to the orthostatic challenge (tilt-up) and is defined by the resting-tilted differences.

The last set of derived variables (derived set III) is generated by transforming the variables of the derived set II into their fractional counterparts. As previously discussed, such an operation has the virtue of stressing the possible interindividual differences in the shape of the resting-tilted differential spectra.

### Principal Component Analysis

By finding the principal components, a data field can be represented in terms of new variables corresponding to the directions of maximal elongation of the data cloud in the space of the original variables (1). It is worth noting that, because they are reciprocally orthogonal, i.e., uncorrelated, and have means equal to 0 and standard deviations equal to 1, the principal components of a data field allow a fair estimate of the size of the possible differences among groups of observations. Moreover, the fraction of the total variability explained by each principal component provides a good estimate of its relative importance in representing the data field.

The principal components correspond to the independent concepts underlying a given data set, and their meanings can be rationalized using the so-called "factor loadings," i.e., the correlation coefficients with the original variables (1, 2). In

### Table 1. Original and derived variables

<table>
<thead>
<tr>
<th>Original variables</th>
<th>Resting State</th>
<th>Tilted State</th>
<th>Difference Between States</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>TP(R)</td>
<td>TP(T)</td>
<td></td>
</tr>
<tr>
<td>VLF-band power</td>
<td>VLF(R)</td>
<td>VLF(T)</td>
<td></td>
</tr>
<tr>
<td>LF-band power</td>
<td>LF(R)</td>
<td>LF(T)</td>
<td></td>
</tr>
<tr>
<td>HF-band power</td>
<td>HF(R)</td>
<td>HF(T)</td>
<td></td>
</tr>
<tr>
<td>LF/HF</td>
<td>LF/HF(R)</td>
<td>LF/HF(T)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Derived variables</th>
<th>Resting State</th>
<th>Tilted State</th>
<th>Difference Between States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set I: Fractional power of various bands</td>
<td>VLF(R)%</td>
<td>VLF(T)%</td>
<td></td>
</tr>
<tr>
<td>VLF-band % Total Power</td>
<td>LF(R)%</td>
<td>LF(T)%</td>
<td></td>
</tr>
<tr>
<td>HF-band % Total Power</td>
<td>HF(R)%</td>
<td>HF(T)%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set II: Differential power between states</th>
<th>Resting State</th>
<th>Tilted State</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP(R) – TP(T)</td>
<td></td>
<td>difTP</td>
</tr>
<tr>
<td>VLF(R) – VLF(T)</td>
<td></td>
<td>difVLF</td>
</tr>
<tr>
<td>LF(R) – LF(T)</td>
<td></td>
<td>difLF</td>
</tr>
<tr>
<td>HF(R) – HF(T)</td>
<td></td>
<td>difHF</td>
</tr>
<tr>
<td>LF(HF(R) – LF(HF(T))</td>
<td></td>
<td>dif(LF/HF)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set III: Fractional differential power between states</th>
<th>Resting State</th>
<th>Tilted State</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLF(R)% – VLF(T)%</td>
<td></td>
<td>difVLF%</td>
</tr>
<tr>
<td>LF(R)% – LF(T)%</td>
<td></td>
<td>difLF%</td>
</tr>
<tr>
<td>HF(R)% – HF(T)%</td>
<td></td>
<td>difHF%</td>
</tr>
</tbody>
</table>

For frequency ranges associated with very low frequency (VLF), low-frequency (LF), and high-frequency (HF) bands, see METHODS. TP, total power; R, resting state; T, tilted state; dif, differential.
the present work, the principal component analysis (PCA) has been carried out on the original as well as on the derived variables, thus allowing the full exploitation of the available information embodied in the HRV from different and complementary points of view.

Cluster Analysis

The correlation between the functional cardiac parameters and age was investigated by means of the Pearson's correlation coefficient (r), i.e., a model based on linear regression between continuous variables. This is in spite of the fact that aging-linked biological observations are generally classified in terms of discrete structures; in everyday language we speak of young, middle-aged, and old individuals, and in even more scientifically refined contexts, 34-, 36-, and 38-yr-old healthy individuals surely would be included into the very same age group.

The present general availability of excellent algorithms for automatic classification allowed us to make use of the discrete character of age by rearranging our database in terms of a minimum number of internally homogeneous and externally well-separated age classes. An important advantage of this strategy lies in the fact that, taking the center mass of the classes as a reference, the individual variability within classes almost disappears, with the effect of a powerful and efficient noise filtering.

The chosen clustering algorithm was the "k-means" (1), a non-hierarchical procedure to identify classes without using any a priori knowledge. Clusters come from the structural characteristics of the data set, maximizing the interclass variance and minimizing the intraclass variance (1). In the case of n units described by m variables, this is obtained by the following steps: 1) a nontrivial number of classes, k, is defined, with k < n < 1; 2) k aggregation points in an m-dimensional space are arbitrarily chosen; 3) each of the n units is assigned to the nearest aggregation point; 4) a new set of aggregation points are reckoned as barycenters of the classes of data; and 5) steps are repeated from step 3 until no class change occurs for any unit.

Evaluation of a Cardiac Age from HRV Data

To estimate the cardiac age of a single subject, i.e., the best guess of the real age from a cardiovascular perspective, a multiple regression analysis was used in which the dependent variable was the chronological age and the independent variables (regressors) were the principal components that were extracted from the HRV database (original and derived sets) and were highly correlated with age; the regressions were computed over the 141 individuals included in our data set. Because of the properties of the principal components (see Principal Component Analysis), this allows us to exclude any intrinsic or randomly arising, spurious correlations between redundant variables.

An extra advantage of the latter choice lies in the fact that the principal components, once their physiological meanings are identified, represent single and independent contributions to the observed complex phenomena. It should be remembered, however, that the chosen definition of cardiac age is only endowed with an operational meaning and does not reflect any deep physiological thinking.

RESULTS

Original Variables and PCA Analysis

Table 2 reports the factor loadings between the first four principal components extracted from the original variables and the variables as such (rows). [The ordering of principal components (first, second, etc.) reflects their significance in emerging from the background noise, as also indicated by the decreasing fractions of the total variability explained.]

The main contribution to the first component (PC1orig), by far the most important in terms of the fraction of total variability explained (50%), is provided by the TP for both the resting and the tilted states; moreover, all the original variables are positively correlated with PC1orig. This points to the identification of PC1orig with a size-type component (2), which means that 1) the absolute size of the power spectra (TP) represents the most conspicuous source of interindividual variability and 2) PC1orig includes the interindividual variability embodied in the different absolute sizes of the power spectra independently of their specific shapes.

The second principal component (PC2orig), different from the first one, is characterized by the opposite sign of its correlation with the resting and tilted variables. This means that PC2orig is linked to the shape of the relationships between the variables relative to the two states (2), as confirmed by the high and negative correlation with the first component of the derived variables for set III (see Set III), describing the resting-tilted difference (Pearson's r = −0.70, P < 0.001).

According to the above-mentioned criteria, both PC3orig and PC4orig are also shape dependent; in particular, PC3orig is linked to LF/HF in the resting state, whereas PC4orig is related to the variability in the response to the tilting, as shown by the opposite correlation with LF/HF in the resting and tilted states.

Only PC1orig shows a significant negative correlation with age (last row of Table 1); this confirms the side-by-side presence of a clear effect of aging on the
decreasing variability of heart rate as well as of other aging-independent factors influencing its regulation.

Derived Variables and PCA Analysis

The results obtained by extracting the principal components from the three sets of derived variables described in the METHODS are summarized in Table 3. For the sake of clarity, these results will be commented on with reference to each single set.

Set I. (These variables represent the fraction of TP included in each spectral band.) None of the principal components extracted from this data set can be exclusively of the size type, because the fractional power of each band is calculated for the resting and tilted states separately, with respect to a TP specific for each individual. In fact, because the fractional power variables implied a fixed value (100%) for each individual, the possibility of finding all correlations to be positive is ruled out. Thus the first principal component for set I (PC1%) reflects the most important source of interindividual variability (40%) based on the spectral shapes. As in the case of the original variables, however, here too the first principal component refers to some common feature of the resting and tilted states, which is the relative importance of the VLF bands compared with the others. As a matter of fact, the VLF bands, because of the 1/f character of the power spectra, represent their most relevant fractions. In synthesis, PC1% measures the relative importance of nonoscillatory (VLF) and true oscillatory (HF, LF) (3, 8, 11) regulation of the heart rate. It is worth noting the significant correlation between PC1% and age, which indicates a more marked decrease of the oscillatory regulation with age compared with the nonoscillatory component (3, 11, 12).

The second component of this set, PC2%, is linked to the relative importance of HF in the resting state [HF(R)%] and LF in the tilted state [LF(T)%]. Because HF(R)% and LF(T)% refer to different states, PC2% deals with the interindividual differences in the shape changes that arise in the spectra while shifting from the resting to the tilted state. PC3% is also linked to the different redistribution of the power in the two states, as indicated by the opposite pattern of signs in the factor loadings between PC3% and the original variables from the resting and tilted states. This interpretation is confirmed by the high negative correlation between PC3% and the first principal component from the third set of derived variables (Pearson's r = −0.93, P < 0.0001).
Finally, because the principal components are linearly uncorrelated among each other, the significant correlation of both PC1% and PC2% with age indicates the presence of two different and linearly irreducible factors in the overall influence of aging on the cardiovascular functions.

Set II. (These variables represent the differences between homologous bands in the resting and tilted states.) What is emerging here, again, is the size nature of the first principal component (PC1dif), which deals with the difference in the TP between the resting and tilted states (difTP) as indicated by the factor loading (0.98) with difTP. In contrast, the second principal component (PC2dif) is of the shape type and indicates the presence of specific nervous reflexes, because its most important factor loading is with the difference in LF/HF between the resting and tilted states.

Set III. (These variables represent the fractional contribution of the difference between corresponding bands to difTP.) The first principal component of this set (PC1dif%) is linked to the balance between the nonoscillatory (VLF) and the oscillatory (LF, HF) regulations and practically coincides with the fractional difference between the resting and tilted states in terms of VLF (loading = -0.985). PC2dif% has almost nothing to do with VLF, whereas it is strongly related to the different individual responses to the tilted state as expressed by both LF and HF bands.

The behavior of the second principal component has a clear interpretation in terms of sympathetic-parasympathetic balance and, in this set, is the only component to be significantly modified by aging.

Setup of a Cardiac Age Estimator

The observed significant correlations between the HRV descriptors and age allowed us to envisage a general empirical index of the cardiac age by collecting in a single score all the age-linked components. Such a score was built by letting the age act as the dependent variable and using as regressors the components significantly related to age. Application of a least-squared

<table>
<thead>
<tr>
<th>Cluster</th>
<th>n</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean ± SD</th>
<th>Chronological Age</th>
<th>Cardiac Age, mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>18</td>
<td>19</td>
<td>29</td>
<td>25.7 ± 3.5</td>
<td>38.2 ± 12.7</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>34</td>
<td>30</td>
<td>45</td>
<td>37.0 ± 5.0</td>
<td>45.9 ± 9.8</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>41</td>
<td>48</td>
<td>63</td>
<td>56.8 ± 4.4</td>
<td>56.3 ± 9.4</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>29</td>
<td>64</td>
<td>78</td>
<td>68.9 ± 3.9</td>
<td>65.6 ± 9.5</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>19</td>
<td>79</td>
<td>91</td>
<td>82.3 ± 3.3</td>
<td>65.0 ± 6.7</td>
<td></td>
</tr>
</tbody>
</table>

Means ± SD values for chronological and cardiac ages are graphically represented in Fig. 2B. Minor SD of chronological age is explained by the fact that, over this variable, age-clustering was actually built up with the declared intention to define internally homogeneous classes.

Finally, because the principal components are linearly uncorrelated among each other, the significant correlation of both PC1% and PC2% with age indicates the presence of two different and linearly irreducible factors in the overall influence of aging on the cardiovascular functions.

Table 4. Age clustering and estimate of cardiac age from database

Means ± SD values for chronological and cardiac ages are graphically represented in Fig. 2B. Minor SD of chronological age is explained by the fact that, over this variable, age-clustering was actually built up with the declared intention to define internally homogeneous classes.
regression generated the following equation

\[
\text{cardiac age} = 53.96 - 4.51 \text{PC1orig} - 6.19 \text{PC1}\% + 5.06 \text{PC2}\% - 4.21 \text{PC2dif}\%
\]

which is endowed with a very high statistical significance \((r = 0.71; P < 0.0001)\). This means that a highly significant agreement exists between the chronological age of the patients and the age estimated by the HRV analysis (Fig. 2A).

**Clustering of Ages**

The application of the k-means algorithm to the age distribution of the population under study produces a five-class solution (Table 4), accounting for 95% of the total observed variability. This means that, in considering the age classes, we retain the great part of the information embodied in the data with respect to the age distribution. The cardiac age estimator (see Setup of a Cardiac Age Estimator) allows the description of the age classes in terms of cardiac age by simply computing Eq. 1 for each element in a class, followed by averaging. Table 4 reports the chronological and cardiac ages relative to the classes. It is worth noting the decoupling of the chronological and cardiac ages at the level of cluster D, corresponding to 69 years. (Fig. 2B).

**DISCUSSION**

An accurate interpretation of the correlation between aging and the principal components extracted from the various sets of variables used in this paper, as depicted in Fig. 3, is a fascinating issue, surely worthy of further and fully dedicated efforts. In planning our work, however, another goal was in mind, i.e., to get, from the bulk of the available information on HRV, a systematic and statistically sound comparison between the cardiovascular and the chronological age.

In such a frame, the salient clinical finding emerging from our study is that the two patterns of chronological and neuroautonomic aging do not completely overlap. Although chronological aging and the age-related loss of autonomic cardiovascular control have a similar pattern until the sixth decade of life (12), from this age onward the cardiovascular age begins to plateau (see Fig. 2). Although a clear-cut and nonambiguous explanation of this effect seems out of reach at the moment, given an increased dispersive effect in the two groups, the older individuals may be biased in terms of greater functional reserves. If so, then our >70-yr-old population should be considered as a selected one, not wholly comparable with the younger counterpart. It is quite unusual, in fact, to find such a large fraction of people over 70 yr old with no cardiovascular disease; very probably, only a few of our subjects under the age of 60 yr will reach the age of 70 yr without any disease. In our opinion, in fact, any sensible choice between the above-mentioned alternatives is unfeasible with the kind of data presently available to us; thus, to verify in particular the second hypothesis, a longitudinal epidemiologic study designed to follow the age-related loss of cardiovascular function in the very same population is needed.

From the practical point of view, the present work demonstrates the possibility, starting from the spectral analysis of the HRV under the resting state and after head-up tilting, to reckon a cardiac age potentially useful for screening purposes. It is intuitive that, because Eq. 1 for cardiac age refers to healthy individuals, it is not applicable to unhealthy subjects; to do that, more studies are needed. From a more speculative point of view, it is worth noting that the main components of the HRV were, in order of decreasing importance, 1) the amount of variability, 2) the relative importance of the nonoscillatory and oscillatory regulations, and 3) the sympathetic-parasympathetic balance. Such a hierarchical ordering was found in all the original and the derived data sets considered in the present work and could provide the basis for a general description of heart rate regulation.

This work has been partially supported by a grant of the Italian Ministero dell’ Università e della Ricerca Scientifica e Tecnologica to A. Colosimo.

Address for reprint requests: A. Colosimo, Dip.to Sc. Biochimiche, Univ. Roma La Sapienza, P.le A. Moro, 5, 00185 Rome, Italy.

Received 19 December 1996; accepted in final form 14 January 1997.

**REFERENCES**


