# Multimodal analysis of arrhythmic risk in the spontaneously hypertensive rat

## Supporting Inter-variable Relationships Data

Prashanna Khwaounjoo,[1] Gregory B. Sands,[1] Ian J. LeGrice,[1,2] Girish Ramulgun,[1,3] Jesse A. Ashton,[1] Johanna M. Montgomery,[2] Anne M. Gillis,[4] Bruce H. Smaill,[1] Mark L. Trew[1]

[1]Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand; [2]Department of Physiology, University of Auckland, Auckland, New Zealand; [3]IHU-Liryc, University of Bordeaux, Bordeaux, France; [4]Libin Cardiovascular Institute of Alberta, University of Calgary, Calgary, AB, Canada.

### Introduction

Two approaches were used to investigate inter-variable relationships. First Pearson's linear correlation coefficient and corresponding p-values from a Student's t distribution were determined. Then an exploratory factor analysis[1] with 2 and 3 factors was conducted to assess the natural groupings of variables that best described the variable data in reduced dimensions.

### Variable Data

Ten variables were collected from each animal. Table 1 shows the 10-variable data from 15 animals.

**Table 1.** Ten variables collected from most animals. The CV anisotropy ratio is for pacing greater than 5 Hz and the APD is at the lowest pacing frequency.

| Age | VT/VF risk | AR >5 Hz | $CV_T$ >5 Hz (m/s) | Cell scale fibrosis % | APD 2.33 Hz (ms) | LV scale fibrosis % | AFA | Number of adj cells | APD dispersion >5 Hz |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 2 | 2.39 | 0.40 | 2.80 | 97.75 | 0.02 | 0.00 | 8.20 | 0.17 |
| 6 | 1 | 1.53 | 0.40 | 8.04 | 125.35 | 0.19 | 0.03 | 7.59 | 0.47 |
| 6 | 3 | 1.80 | 0.39 | 7.32 | 107.80 | 0.29 | 0.02 | 6.42 | 0.28 |
| 6 | 2 | 1.66 | 0.35 | 9.97 | 127.66 | 0.46 | 0.13 | 6.58 | 0.34 |
| 6 | 3 | 1.77 | 0.55 | | 101.99 | 0.22 | 0.05 | | 0.38 |
| 12 | 6 | 1.76 | 0.32 | 5.30 | 117.21 | 1.47 | 0.07 | 9.60 | 0.40 |
| 12 | 11 | 2.23 | 0.31 | 12.20 | 93.82 | 4.16 | 0.73 | 7.00 | 0.55 |
| 12 | 1 | 2.38 | 0.28 | | 118.45 | 0.83 | 0.17 | | 0.51 |
| 12 | 12 | 2.46 | 0.27 | 24.99 | 125.41 | 3.68 | 0.82 | 3.00 | 0.68 |
| 12 | 1 | 2.42 | 0.21 | | 138.51 | 0.68 | 0.11 | | 0.34 |
| 18 | 12 | 2.50 | 0.27 | 14.05 | 140.72 | 2.14 | 0.44 | 6.00 | 0.62 |
| 18 | 9 | 2.37 | 0.22 | 12.78 | 139.48 | 4.72 | 0.41 | 6.82 | 0.48 |
| 18 | 1 | 2.06 | 0.25 | 12.57 | 103.66 | 1.14 | 0.14 | 5.65 | 0.43 |
| 18 | 12 | 2.72 | 0.22 | 20.40 | 139.42 | 4.78 | 0.89 | 2.33 | 0.71 |
| 18 | 8 | 1.93 | 0.28 | 20.22 | 121.86 | 3.75 | 0.52 | 0.00 | 0.45 |

Shaded rows with missing data were not used for correlation or factor analysis. In all three cases no complete cells could be segmented from the 3D confocal image blocks. The data in Table 1 were normalized to have mean of 0 and standard deviation of 1 for each variable (column) so that they could be comparatively analyzed.

## Pearson's linear correlations

To assess inter-variable correlations, Pearson's linear correlation coefficients and corresponding p-values were found for the (normalized) data in Table 1 using the MATLAB® `corr`* function. The correlation coefficients (shaded boxes are significant ($p < 0.05$)) are given in Table 2 and the approximate equivalent $R^2$ values (the squared value of Pearson's correlation coefficient; shaded boxes are significant ($p < 0.05$)) are given in Table 3. The p values (f statistic) testing the hypothesis that there is no correlation against a non-zero correlation are given in Table 4.

**Table 2.** Pearson's correlation coefficients ($\rho \in [-1,1]$) between pairs of the 10 variables.

| | **Age** | **VT/VF risk** | **AR >5 Hz** | **$CV_T$ >5Hz (m/s)** | **Cell scale fibrosis %** | **APD 2.33Hz (ms)** | **LV scale fibrosis %** | **AFA** | **Number of adj cells** | **APD dispersion >5 Hz** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age** | | 0.60 | 0.56 | -0.94 | 0.60 | 0.41 | 0.72 | 0.57 | -0.51 | 0.62 |
| **VT/VF risk** | 0.60 | | 0.72 | -0.65 | 0.70 | 0.43 | 0.85 | 0.89 | -0.45 | 0.81 |
| **AR>5Hz** | 0.56 | 0.72 | | -0.61 | 0.49 | 0.23 | 0.63 | 0.69 | -0.34 | 0.54 |
| **$CV_T$>5Hz (m/s)** | -0.94 | -0.65 | -0.61 | | -0.70 | -0.49 | -0.80 | -0.68 | 0.51 | -0.70 |
| **Cell scale fibrosis %** | 0.60 | 0.70 | 0.49 | -0.70 | | 0.45 | 0.74 | 0.86 | -0.87 | 0.80 |
| **APD 2.33 Hz (ms)** | 0.41 | 0.43 | 0.23 | -0.49 | 0.45 | | 0.37 | 0.32 | -0.32 | 0.55 |
| **LV scale fibrosis %** | 0.72 | 0.85 | 0.63 | -0.80 | 0.73 | 0.37 | | 0.90 | -0.56 | 0.73 |
| **AFA** | 0.57 | 0.89 | 0.69 | -0.68 | 0.86 | 0.32 | 0.90 | | -0.66 | 0.85 |
| **Number of adj cells** | -0.51 | -0.45 | -0.34 | 0.51 | -0.87 | -0.32 | -0.56 | -0.66 | | -0.50 |
| **APD dispersion >5 Hz** | 0.62 | 0.81 | 0.54 | -0.70 | 0.80 | 0.55 | 0.73 | 0.85 | -0.50 | |

* mathworks.com/help/images/ref/corr.html

**Table 3.** Approximate $R^2$ values.

| | **Age** | **VT/VF risk** | **AR >5 Hz** | **$CV_T$ >5Hz (m/s)** | **Cell scale fibrosis %** | **APD 2.33Hz (ms)** | **LV scale fibrosis %** | **AFA** | **Number of adj cells** | **APD dispersion >5 Hz** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age** | | 0.36 | 0.31 | **0.88** | 0.36 | 0.17 | *0.52* | 0.33 | 0.26 | 0.38 |
| **VT/VF risk** | 0.36 | | *0.51* | 0.43 | 0.48 | 0.19 | **0.73** | **0.80** | 0.20 | **0.65** |
| **AR>5Hz** | 0.31 | *0.51* | | 0.37 | 0.24 | 0.06 | 0.40 | 0.48 | 0.12 | 0.29 |
| **$CV_T$>5Hz (m/s)** | **0.88** | 0.43 | 0.37 | | 0.48 | 0.24 | **0.64** | 0.47 | 0.26 | 0.49 |
| **Cell scale fibrosis %** | 0.36 | 0.48 | 0.24 | 0.48 | | 0.21 | *0.55* | **0.74** | **0.76** | **0.65** |
| **APD 2.33 Hz (ms)** | 0.17 | 0.19 | 0.06 | 0.24 | 0.21 | | 0.14 | 0.10 | 0.10 | 0.30 |
| **LV scale fibrosis %** | *0.52* | **0.73** | 0.40 | **0.64** | *0.55* | 0.14 | | **0.80** | 0.31 | *0.54* |
| **AFA** | 0.33 | **0.80** | 0.48 | 0.47 | **0.74** | 0.10 | **0.80** | | 0.43 | **0.73** |
| **Number of adj cells** | 0.26 | 0.20 | 0.12 | 0.26 | **0.76** | 0.10 | 0.31 | 0.43 | | 0.25 |
| **APD dispersion >5 Hz** | 0.38 | **0.65** | 0.29 | 0.49 | **0.65** | 0.30 | *0.54* | **0.73** | 0.25 | |

**Table 4.** Correlation p values from the Student's t distribution.

| | **Age** | **VT/VF risk** | **AR >5 Hz** | **$CV_T$ >5Hz (m/s)** | **Cell scale fibrosis %** | **APD 2.33Hz (ms)** | **LV scale fibrosis %** | **AFA** | **Number of adj cells** | **APD dispersion >5 Hz** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age** | | 0.0385 | 0.0609 | 0.000 | 0.0385 | 0.1818 | 0.0084 | 0.0506 | 0.0875 | 0.0322 |
| **VT/VF risk** | 0.0385 | | 0.0086 | 0.0209 | 0.0120 | 0.1625 | 0.0004 | 0.0001 | 0.1428 | 0.0015 |
| **AR>5Hz** | 0.0609 | 0.0086 | | 0.0351 | 0.1027 | 0.4625 | 0.0285 | 0.0126 | 0.2807 | 0.0708 |
| **$CV_T$>5Hz (m/s)** | 0.000 | 0.0209 | 0.0351 | | 0.0120 | 0.1074 | 0.0017 | 0.0143 | 0.0903 | 0.0112 |
| **Cell scale fibrosis %** | 0.0385 | 0.012 | 0.1027 | 0.012 | | 0.1389 | 0.0061 | 0.0003 | 0.0002 | 0.0016 |
| **APD 2.33 Hz (ms)** | 0.1818 | 0.1625 | 0.4625 | 0.1074 | 0.1389 | | 0.2345 | 0.3044 | 0.3125 | 0.0643 |
| **LV scale fibrosis %** | 0.0084 | 0.0004 | 0.0285 | 0.0017 | 0.0061 | 0.2345 | | 0.0001 | 0.0585 | 0.0067 |
| **AFA** | 0.0506 | 0.0001 | 0.0126 | 0.0143 | 0.0003 | 0.3044 | 0.0001 | | 0.0197 | 0.0004 |
| **Number of adj cells** | 0.0875 | 0.1428 | 0.2807 | 0.0903 | 0.0002 | 0.3125 | 0.0585 | 0.0197 | | 0.0955 |
| **APD dispersion >5 Hz** | 0.0322 | 0.0015 | 0.0708 | 0.0112 | 0.0016 | 0.0643 | 0.0067 | 0.0004 | 0.0955 | |

## Exploratory factor analysis

The aim of exploratory factor analysis is to look for possible latent (hidden) factors (i.e. combinations of variables) in data that expose underlying relationships between the variables.[1] Factor analysis proposes a linear model of the data and determines weightings and specific variances necessary to fit the data with a pre-specified number of factors.

For each animal the 10 variables were linearly modelled as:

$$\boldsymbol{x} = \mathrm{L}\boldsymbol{f} + \boldsymbol{e}. \qquad (1)$$

Where:

$\boldsymbol{x}$ is a 10×1 vector of measurements from one animal, mapped to a standard normal distribution with mean 0 and standard deviation 1.
L is the 10×$n$ matrix of $n$ factor loadings.
$\boldsymbol{f}$ is the vector of $n$×1 common factor values (normalized).
$\boldsymbol{e}$ is a 10×1 vector of independent specific item values, i.e. what is not captured by the factors.

The equivalent covariance model for computation (correlation can be substituted as the data distributions have unit variance) is:

$$\mathrm{cov}(\boldsymbol{x}) = \mathrm{corr}(\boldsymbol{x}) = \mathrm{corr}(\mathrm{L}\boldsymbol{f}) + \mathrm{cov}(\boldsymbol{e}) = \mathrm{L}\,\mathrm{corr}(\boldsymbol{f})\mathrm{L}^T + \Psi.$$

Where $\Psi$ is the specific variance for each variable, i.e. how much of the variance is not described by the $n$ factors.

To determine the unknown factor loadings and specific variance for the SHR data in Table 1, maximum likelihood factor extraction was used together with promax oblique rotation of the factors, as implemented in the MATLAB® `factoran`† function.

Both 2 and 3 factor models were considered and scree plots of the factor loadings (L) were used to assess the natural groupings of variables in the factors. Figure 1 shows scree plots for a 2 factor model. Factor 1 loadings in Figure 1A suggest the variables to the left of the red line should be included in that factor and Factor 2 loadings in Figure 1B suggest that Age and Transverse CV are strong components of that factor and, in order to complete the 10 variable spread, APD could also be added to the factor.

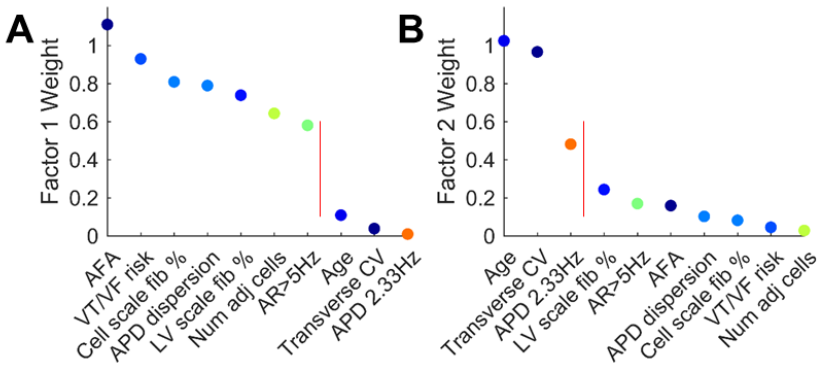



**Figure 1.** Scree plots of factor loadings for a two factor model. Variables are sorted from highest factor loading to smallest. Color is specific variance for each variable, from low (blue: most variance described by the factor model) to high (red: most variance is not described by the factor model). **A.** Factor 1. **B.** Factor 2.

† mathworks.com/help/images/ref/factoran.html

Figure 2 shows scree plots for a 3 factor model. The factors remain similar to the 2 factor model except that the cell scale measurements are shifted from Factor 1 and form their own Factor 3. Factor 2 remains the same as for the 2 factor model. For both 2 and 3 factor models (and 4-5 factor models – data not shown), the APD is least well described by the factor models of all the variables.

Both 4 and 5 factor models retained the AFA index, VT/VF risk, LV scale fibrosis % and AR in their first factor. However, the APD dispersion was then either associated with APD (4 factor model) or in a factor of its own (5 factor model). In the five-factor model, factors 4 and 5 comprised only a single variable: APD (factor 4) and APD dispersion (factor 5). The results of exploratory factor analysis indicate that the combination of AFA index, VT/VF risk, LV scale fibrosis %, AR and APD dispersion variables is a good descriptor of the experimental data collected from four independent measurement approaches.

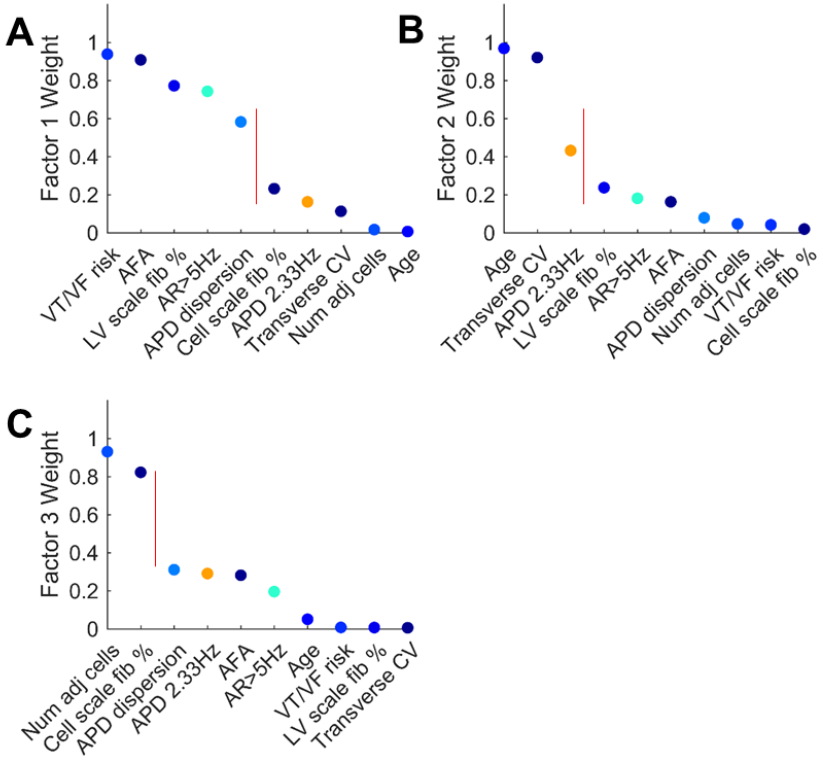



**Figure 2.** Scree plots of factor loadings for a three factor model. Variables are sorted from highest factor loading to smallest. Color is specific variance for each variable, from low (blue: most variance described by the factor model) to high (red: most variance is not described by the factor model). **A.** Factor 1. **B.** Factor 2. **C.** Factor 3.

Figure 3 plots the factor scores (original data transformed into a lower dimensional factor space) from Equation 1, for two and three-factor models.

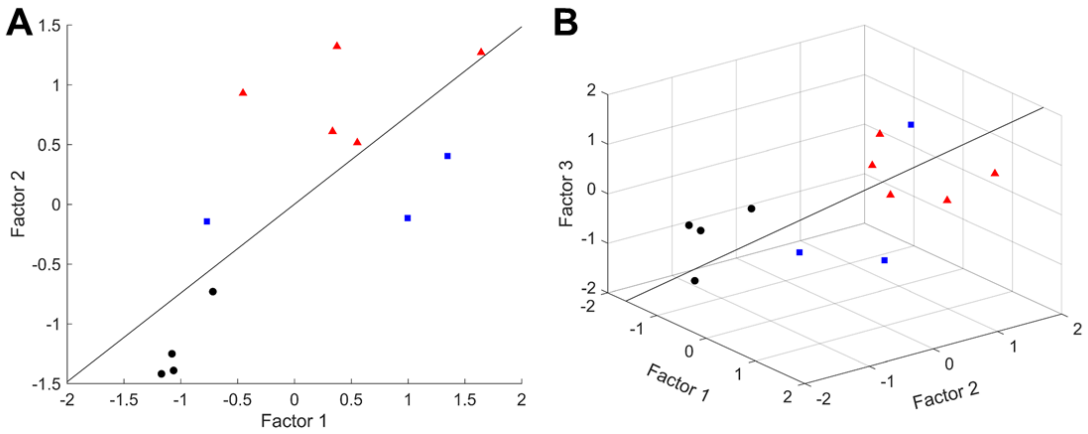



**Figure 3.** Factor score plots. **A.** Two-factor model. For best-fit linear regression line shown, $R^2 = 0.55$ and $p = 0.006$. **B.** Three-factor model. For best-fit linear regression line shown, $R^2 = 0.54$ and $p = 0.03$. Black circles are 6 month SHR, blue squares are 12 month SHR and red triangles are 18 month SHR.

## Reference

1. Tabachnick BG, Fidell LS. Using multivariate statistics. 7 ed. Boston, MA: Pearson. 2019.