Early subclinical increase in pulmonary water content in athletes performing sustained heavy exercise at sea level: ultrasound lung comet-tail evidence

Alessandro Pingitore,1,2* Erika Garbella,3,7* Paolo Piaggi,5 Danilo Menicucci,1,2 Francesca Frassi,1,2 Vincenzo Lionetti,1,7 Andrea Piarulli,1,2 Giosuè Catapano,7 Valter Lubrano,7 Mirko Passera,1,2 Gianluca Di Bella,9 Cinzia Castagnini,8 Silvia Pellegrini,8 Maria Rosa Metelli,10 Remo Bedini,1,2 Angelo Gemignani,1,2,4 and Antonio L’ Abbate2,3

1CNR Institute of Clinical Physiology; 2Extreme Centre, 3Institute of Life Sciences, Scuola Superiore Sant’Anna; Departments of 4Physiological Sciences, 5Energy and Systems Engineering, and 6Experimental Pathology, Biotechnology, Epidemiology and Infectious Disease, Pisa University; 7Fondazione CNR/Regione Toscana Gabriele Monasterio, Pisa; 8Department of Preclinical and Clinical Pharmacology, University of Florence, Florence; 9Clinical and Experimental Department of Medicine and Pharmacology, Messina University, Messina; and 10Laboratory of Experimental Clinical-Chemical Analysis, AUOP, Pisa, Italy

Submitted 18 April 2011; accepted in final form 17 August 2011

Pingitore A, Garbella E, Piaggi P, Menicucci D, Frassi F, Lionetti V, Piarulli A, Catapano G, Lubrano V, Passera M, Di Bella G, Castagnini C, Pellegrini S, Metelli MR, Bedini R, Gemignani A, L’Abbate A. Early subclinical increase in pulmonary water content in athletes performing heavy exercise at sea level: ultrasound lung comet-tail evidence. Am J Physiol Heart Circ Physiol 301: H2161–H2167, 2011. First published August 26, 2011; doi:10.1152/ajpheart.00388.2011.—Whether prolonged strenuous exercise performed by athletes at sea level can produce interstitial pulmonary edema is under debate. Chest sonography allows to estimate extravascular lung water, creating ultrasound lung comet-tail (ULC) artifacts. The aim of the study was to determine whether pulmonary water content increases in Ironmen (n = 31) during race at sea level and its correlation with cardiopulmonary function and systemic proinflammatory and cardiac biohumoral markers. A multiple factor analysis approach was used to determine the relations between systemic modifications and ULCs by assessing correlations among variables and groups of variables showing significant pre-post changes. All athletes were asymptomatic for cough and dyspnea at rest and after the race. Immediately after the race, a score of more than five comet tail artifacts, the threshold for a significant detection, was present in 23 athletes (74%; 16.3 ± 11.2; P < 0.01 ULC after the race vs. rest) but decreased 12 h after the end of the race (15 athletes; 42%; 6.3 ± 8.0; P < 0.01 vs. soon after the race). Multiple factor analysis showed significant correlations between ULCs and cardiac-related variables and NH2-terminal pro-brain natriuretic peptide. Healthy athletes developed subclinical increase in pulmonary water content immediately after an Ironman race at sea level, as shown by the increased number of ULCs related to cardiac changes occurring during exercise. Hemodynamic changes are one of several potential factors contributing to the mechanisms of ULCs.

EVEN IN HEALTHY SUBJECTS EXTREMELY demanding endurance exercise leads to functional and structural cardiac and pulmonary changes, along with local and systemic responses, reflecting oxidative, metabolic, hormonal, and thermal stress, besides immunomodulation and inflammatory reaction (4, 30, 32, 40, 41, 46). Whether interstitial pulmonary edema occurs in athletes performing heavy sea level exercise is debated (2, 6, 12, 18, 19, 27, 28, 44, 56). Interstitial pulmonary edema has been documented in endurance athletes performing heavy sea level exercise, using imaging techniques such as chest X-ray, magnetic resonance, computed tomography, and scintigraphy (17, 36, 56). Previous authors measured noninvasively the postexercise extravascular thoracic fluid volume using thoracic impedance monitoring (16). However, this method does not allow visualizing the seat of the water content in the chest. Chest sonography can be used in the field for assessment of athletes immediately after the end of exercise (11, 34, 35). This technique effectively detects and quantifies extravascular lung water, creating ultrasound lung comet-tail (ULC) artifacts from water-thickened pulmonary interlobular septa (11). In a pig model of acute lung injury, ULCs unmasked accumulation of histologically verified extravascular lung water, even at the stage when no arterial oxygen partial pressure changes could be observed (13). ULCs are accurate in detecting cardiogenic pulmonary edema, correlating with NH2-terminal pro-brain natriuretic peptide (NT-proBNP) (14), and in predicting prognosis of patients with dyspnea and/or chest pain syndrome (10). Finally, ULCs have also been shown in healthy subjects at high altitude and in breath-hold divers (11, 22, 35).

We hypothesized that increase in pulmonary water content may occur in healthy athletes performing strenuous exercise at sea level consisting of Ironman triathlon (3.8 km swimming, 180 km cycling, and 42 km running).

METHODS

Subjects. We enrolled healthy Ironman athletes (29 males and 2 females), with a mean age of 41 ± 13 years (range: 29–60 years) and a mean training time of 13 ± 9 years. They were all used to ultra-triathlon competitions and had normal lung, cardiac, and renal function; no known medical problems; neither history of cigarette smoking nor assumption of any drug in the 24 h before the study. The Ethics Committee at Pisa University Hospital approved the experimental protocol. After receiving the description of the procedures and potential risks, all subjects gave their written informed consent.

Experimental protocol. Measurement of arterial pressure, echocardiographic examination, chest ultrasonography, pulmonary function test, and blood samples were performed the day before the race and within 20 min after termination of the competition. They were performed in a separate room close to the competition area, using
portable noninvasive instrumental equipment. Chest ultrasonography and pulmonary function tests were repeated 12 h after the race.

**Echocardiography and chest ultrasonography.** Echocardiographic examination and chest ultrasonography were performed using a trans-thoracic two-dimensional and Doppler echocardiography (MY LAB 30; Esaote, Florence, Italy) with a broadband (2–4 MHz) phased array transducer. All images were digitized, transmitted to a personal computer, and stored on DVD for processing. All measurements were performed by a cardiologist following the recommendations of the American Society of Echocardiography (23). A dedicated software package (XStrain; Esaote) was used for quantification of longitudinal strain. This software provides angle-independent two-dimensional strain based on speckle tracking (7).

A scan of the anterior and lateral chest in the midaxillary, anterior axillary, midclavicular, and parasternal positions, from the second to the fifth (on the right side) and to the fourth (on the left side) intercostal spaces, for a total of 28 echocographic windows (34), was performed by two blinded expert echocardiographers (A. Pingitore and F. Frassi), with intraobserver and interobserver variability of the ULC score of 4% and 5.5%, respectively. The comet-tail sign was defined as an echogenic, coherent, wedge-shaped signal with a narrow origin in the near field of the image, arising from the pleural line and extending to the edge of the screen. The sum of the number of comet-tail signs in all surveyed fields yielded the overall comet-tail score (5–15 = mild, 15–30 = moderate, >30 = severe) (34). Intraobserver and interobserver variability for the other ultrasound measurements were 3% and 5%.

**Measurement of pulmonary function.** A pulmonary function test was performed on the seated subject by an automated pulmonary function unit (WinSpiroPro, Spirolab III-COL, Mir) following American Thoracic Society guidelines. The measurements included slow and forced vital capacity (VC), forced inspiratory volume in 1 s (FEV1) derived indexes [FEV1/forcedVC, maximal voluntary ventilation (MVV)] and forced expiratory flow at 25%, 50%, and 75% of VC (25%–50%–75%). Athletes practiced at least two maximal ventilatory efforts for slow VC and three for forced VC (values within 5% of each other), and the largest value was considered.

**Measurement of biohumoral variables.** Blood samples were collected and centrifuged. Both serum and plasma were then stored at −20 degrees in a portable refrigerator. Analyses were performed at the Institute of Clinical Physiology in Pisa. Plasma cytokines (cTnI) levels were determined by a microparticle enzyme immunoassay method with the TDx system (Abbott Diagnostics, Vienna, Austria), those of IL-1ra were measured by enzyme-linked immunosorbent assay (Bender Medsystems, Vienna, Austria), and those of IL-1B, IL-6, IL-5, IL-8, IL-9, IL-12, and IL-10 were measured by ELISA (module E170; Roche Diagnostics, Minneapolis, MN), and those of TNF-α were measured by enzyme-linked immunosorbent assay (Invitrogen, San Giuliano Milanese, Milan, Italy). NT-proBNP concentrations were measured by immunoradiometric assay kit (module E170; Roche Diagnostics, Mannheim, Germany). Cardiac troponin I (cTnI) levels were determined by a microparticle enzyme immunoassay (module E170; Roche Diagnostics). Serum cortisol was measured with a fluorescent polarization immunoassay method with the TDx system (Abbott Diagnostics), following the procedure indicated by the manufacturer. For the measurement of plasma catecholamines automated HPLC analyzer was used (HCL-725 CA manufactured by Tosoh, Tokyo, Japan).

**Data organization.** The cardiac variables were organized into five groups, and the spirometric variables were organized into three groups, according to their physiological meaning (Table 1). The acute-basal data set was composed of 31 subjects and 41 variables, which were arranged in 12 groups (5 cardiac, 3 spirometric, 3 biohumoral, and 1 for ULCs).

**Statistical analysis.** Analysis consisted of two main steps: 1) identifying features with significant changes in the acute state compared with the baseline and 2) highlighting associations between the different groups of features. In step 1, for each parameter, the difference between acute state and baseline was computed and submitted to the Student’s t-test for paired samples. In step 2, associations between feature groups were investigated by means of multiple factor analysis (MFA) (9). For each group, only significant features from step 1 were considered. MFA cope with data in which a set of individuals is described by several sets of features, called active features. At variance with the standard factor analysis, it considers the hierarchical structure of data and balances the influences of each set of features before the search of latent factors, called global axes (GAs) in the MFA jargon. We determined the number of significant GAs using a Monte Carlo simulation that enabled us to compare the structure of real data with random simulated features (see Appendix); this allowed us to identify the correct number of significant GAs to retain for the subsequent analysis (20). GAs can be described both in terms of original features and groups. The former description is based on the Pearson correlation between each GA and features; the latter is based on the concept of contribution, a measure of the influence of each group on the GA, ranging from 0 to 1 (see Appendix).

Moreover, the influences of other features or groups (called supplementary) on estimated GAs can be evaluated. In this study we performed two distinct MFAs: in the first one, we estimated GAs from the features belonging to cardiac, spirometric, and ULC groups (active groups). In the second MFA, we obtained GAs from the biohumoral groups (active groups) and assessed the relationship between GAs and cardiac, spirometric, and ULC groups (supplementary groups). Throughout the analyses, P values smaller than 0.05 were considered statistically significant. Data are presented as mean acute-basal difference ± SD.

**RESULTS**

All subjects completed the triathlon race with a mean time of 12 ± 1 h, without any clinical problems.

**Ultrasound comets.** Considering the entire group of athletes the mean ULC score was 1.6 ± 2.3 at rest; 14 athletes had an ULC score <5 and two had a ULC score of >5 (9 and 7 ULC, respectively). Immediately after the Ironman race ULC score >5 was present in 23 athletes (P < 0.01 vs. rest). The mean ULCs at this stage was 16 ± 11 (P < 0.01 vs. rest). At 12 h after the end of the race the number of athletes with ULC score >5 (13 athletes) and the number of ULCs (6.3 ± 8.0) decreased (P < 0.01; Fig. 1).

**Pre-post analysis.** All variables showed a Gaussian distribution (all P > 0.05 using Kolmogorov-Smirnov test) both in the basal and at stress, so the use of parametric t-test for the subsequent analysis was allowed. The cardiac, spirometric, ULC, and molecular variations between acute and basal values are reported in Table 1. In the former ensemble, reduction of systolic blood pressure and left ventricular end-diastolic volume and both cardiac right and left function-related variables was observed (all P < 0.01), whereas heart rate increased (P < 0.01). Of spirometric parameters, all variables (with the exception of FEV1/forcedVC) showed a reduction (P < 0.01). Regarding systemic inflammatory stress-related indexes and cardiac biohumoral markers, cTnI (P = 0.02), NTproBNP (P < 0.01), cortisol, and noradrenaline, as well as IL-6, IL-1ra, IL-10, IL-12, and IL-8 (all P < 0.01), showed an increase after the competition.

**Multivariate analysis.** MFA was applied to the difference-based data matrix regarding the cardiac, spirometric, ULC, and biohumoral groups to study their association.

**Cardiospirometric-based analysis.** The global representation of features on the first two GAs is shown in Fig. 1. The graph shows the significant correlations between cardiac, spi-
rometric, and ULC features with the first (x-axis) and second (y-axis) GA obtained from MFA. The first GA (25% of the total data variance) included all spirometric features (correlation coefficients from 0.75 to 0.94; all \( P < 0.01 \)) and systolic blood pressure (inverse correlation, \( r = -0.49; \ P = 0.01 \)). The second GA (17%) mainly included variables for cardiac features, namely left ventricular (LV) end-diastolic (\( r = 0.80; \ P < 0.01 \)) and stroke volumes (\( r = 0.47; \ P = 0.02 \)), and ULC score (\( r = 0.56; \ P < 0.01 \)) along with right heart function features (\( r = 0.50; \ P = 0.01 \)), tissue-Doppler S wave (inverse correlation, \( r = -0.49; \ P = 0.01 \)), and TAPSE (inverse correlation, \( r = -0.58; \ P < 0.01 \)). The third GA (16%) was related to cardiac LV ejection fraction (\( r = 0.81; \ P < 0.01 \)), SV (\( r = 0.80; \ P < 0.01 \)), and ULC score (inverse correlation, \( r = -0.54; \ P = 0.01 \)). Analyzing the correlations between these three parameters, LV ejection fraction was inversely correlated with ULC (\( r = -0.48; \ P = 0.02 \); Fig. 2). Regarding the contribution of active groups on GAs (Table 2), the three spirometric groups explained 84% of first GA. The second GA was mostly described by the morphological and right cardiac component (66%) and ULC (16%), whereas the third GA was dominated by left cardiac component (48%) and ULC (17%) groups.

### Biochemical-based analysis

The Monte Carlo procedure indicated the significance of the first two GAs, which explained 46% of the global data variability. The global representation of active features on these two GAs is shown in Fig. 2. The graph shows the significant correlations between cardiac damage, hormones, and inflammatory markers features with the first (x-axis) and second (y-axis) GA obtained from MFA. The first GA (27% of the total data variance) included all inflammatory markers (correlation coefficients from 0.44 to 0.84; all \( P \leq 0.05 \)).
0.03) and the cardiac damage feature NTproBNP ($r = 0.82; P < 0.01$). The second GA (19%) mainly included noradrenaline ($r = 0.58; P < 0.01$) along with the cardiac damaged feature cTnI ($r = 0.62; P < 0.01$) and the cortisol (inverse correlation, $r = -0.47; P = 0.02$). Regarding the supplementary features, ULC ($r = 0.47; P = 0.02$) and MVV ($r = 0.38; P = 0.05$) were correlated with the first GA. With the analysis of the correlations between ULC, inflammatory markers, and NTproBNP, the latter was correlated with both IL-6 ($r = 0.90; P < 0.01$) and ULC ($r = 0.86; P < 0.01$), whereas correlations between MVV and inflammatory markers highlighted that MVV was related to IL-10 ($r = 0.48; P = 0.02$) and IL-1ra ($r = 0.41; P = 0.05$) (Fig. 3). There was no significant correlation between ULC and IL-6 when controlling for NT-proBNP ($P = 0.05$).

DISCUSSION

We detected asymptomatic ULCs in the presence of increased plasma level of anti-inflammatory mediators in Ironmen immediately after the race. Hodges et al. (18) showed no changes in pulmonary density, which increases in the presence of extravascular lung water, following normoxic and hypoxic midintensive exercise. Conversely, Hopkins (19) demonstrated the presence of red cells, proteins, and leukotriene B4 in the bronchoalveolar lavage after brief maximal sea level exercise, suggesting the occurrence of transient increase in pulmonary blood-gas barrier permeability. As suggested by the metanalysis of Zavorsky (56), exhaustive or near-maximal exercise is able to induce pulmonary edema. However, some questions on the pathophysiology of interstitial pulmonary edema in extreme athlete are still unanswered, in part because of the limitations of diagnostic technology. Thoracic echography allowed us to assess ULCs within 15 ± 3 min from the end of the exercise. Our results showed that ULCs were almost undetectable at rest, but appeared after an intensive race in 68% of athletes and were mostly resolved within the first 12 h. At rest ULCs were present in 14 athletes, and in two of them the number was ≤5. However, the presence of a low number of comets has been shown in normal subject and confirmed by histology in healthy pigs (14, 25). Furthermore, the presence of few ULCs in Ironman athletes at rest is difficult to explain but may be related, in part, to the training performed by these particular athletes immediately before the race. However, we described by hormones (57%) with a further contribution by the cardiac damage group (33%).

Table 2. Contributions of active groups to the first three global axes obtained from the multiple factor analysis based on cardiac, spirometric, and ULC groups.
Maximal exercise increases the risk of pulmonary capillary stress failure: increased pulmonary arterial pressure, active expiration, and increased lung inflation, respectively, cause a mechanical deformation of the pulmonary gas barrier in a dose-dependent manner (52). Alveolar cells, an element of the blood-gas barrier, also have a critical role in determining the homeostatic balance of pulmonary interstitial fluids during stress. Because alveolar cells are not elastic, Oeckler et al. (31) showed that plasma membrane-cytoskeletal adhesive interactions are important determinants of the cellular response to deforming stress. Caveolar endocytic response to transient actin remodeling following deformation of plasma membrane increases cell permeability in both alveolar (15, 49) and endothelial cells (26) in a dose-dependent manner.

It is conceivable that a combination of different biomechanisms (42, 50, 51) causes pulmonary edema under extreme conditions at sea level, leading to a greater increase in lung vascular hydrostatic pressure and vascular permeability, as well as to a downregulation of the alveolar fluid reabsorption pathways (38). Another factor modulating the magnitude of capillary pressure increase is the level of the training, which improves LV compliance and performance (45). Pulmonary arterial wedge pressure at maximal exercise was in fact lower in subjects with higher aerobic fitness, in part due to higher LV compliance (24).

We applied MFA to predict potential mechanisms involved in the extravascular lung water balance during extreme exercise. In Ironmen, early ULC detection was not related to pulmonary function parameters. Conversely, we observed acute postrace decline in vital capacity and airflow rates at mid- and low lung volumes. These results suggest functional changes at the level of small airways, which are compressed in the presence of interstitial pulmonary edema and/or local bronchial-bronchiolar mucosal inflammation (4, 48), so far causing increasing ventilation-perfusion heterogeneity (5). Moreover, we found a correlation between the intensity of the anti-inflammatory response (IL-10 and IL-1ra) and MVV at stress, suggesting a protective anti-inflammatory cascade in well-trained athletes, as hypothesized previously (30, 33).

Our data analysis suggests that hemodynamic changes are one of several potential factors contributing to the mechanisms of ULCs. In particular, the correlation between ULCs and NT-proBNP plasma levels, but not cTnI, confirms that the increased level of circulating NT-proBNP in Ironman athletes is secondary to the rise in cardiac output during sea level exercise (28) rather than to myocardial injury (14). NT-proBNP has combined natriuretic and vasodilating effects aimed at reducing myocardial wall stretch, which occurs during prolonged exercise as the effect of prolonged increase in cardiac output (40). In fact, several animal and human studies have shown that mechanical stretch elicits BNP expression (54), in a time-dependent manner (39). In addition increased circulating levels of NT-proBNP cause transient endothelial dysfunction in athletes following extreme stress (47), which might increase vascular permeability without affecting tissue compliance. Moreover, NT-proBNP circulating levels increased during recovery after prolonged stress due to a new level of water-sodium homeostasis (29). The inverse correlation between ULC and LV ejection fraction reinforces the relationship between hemodynamics and ULC development, as documented previously in patients with LV dysfunction (34). The reduction in LV ejection fraction, as well in longitudinal strain, stroke volume, and end-diastolic volumes, and the increase in cardiac troponin fall within changes that characterize cardiac fatigue following prolonged endurance (8, 53). Considering all the changes mentioned above, the reduction in ejection fraction, stroke volume, and longitudinal strain might depend on end-diastolic volume reduction, and thus preload reduction, rather than on dysfunction of myocardial contractility, also considering the absence of end-systolic enlargement and wall motion abnormalities. Surprisingly, the correlation between ULCs and increased plasma levels of IL-6 was not significant, suggesting that IL-6 expression exerts cytoprotective and growth-regulating effects, preventing cardiac remodeling (39), inhibiting TNF synthesis (3), and replacing cortisol synthesis (55) during sustained heavy sea level exercise.

Table 3. Contributions of active groups to the first two global axes obtained from the multiple factor analysis based on biohumoral groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Global Axe 1</th>
<th>Global Axe 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac damage, %</td>
<td>0.77 (42)</td>
<td>0.42 (33)</td>
</tr>
<tr>
<td>Hormones, %</td>
<td>0.26 (15)</td>
<td>0.74 (57)</td>
</tr>
<tr>
<td>Inflammatory markers, %</td>
<td>0.79 (43)</td>
<td>0.13 (10)</td>
</tr>
<tr>
<td>Total, %</td>
<td>1.81 (100)</td>
<td>1.29 (100)</td>
</tr>
<tr>
<td>Explained variance, %</td>
<td>27</td>
<td>19</td>
</tr>
</tbody>
</table>

We acknowledge that this remains speculative. After the race all athletes with ULC evidence were asymptomatic for dyspnea and cough. Our clinical findings are in line with a study showing that early detection of noncardiogenic ULCs is unconnected to clinical signs of interstitial edema and hypoxemia (13). ULCs were also detected in healthy subjects chronically exposed to extreme environments, such as hyperbarism in breath-hold divers soon after a deep dive or hypobarism in recreational climbers (11, 35).

Fig. 3. Global representation of variables and groups on the plane defined by GA1 and GA2 of MFA based on the biohumoral groups. NT-proBNP, NH2-terminal pro-brain natriuretic peptide; cTnI, cardiac troponin I; Cort, cortisol; NAd, noradrenaline.
The ULC technique is based on employing an ultrasound probe that is constantly moving, leaning and rotating until ULCS are visualized. They appear as roughly vertical, narrow-based, parallel ultrasound artifacts that show as laser-rays extending to the edge of the screen (up to at least 20 cm) from the pleural line and moving edgeways consensually with respiration, whereas a normal lung gives rise to roughly horizontal, parallel lines at regular intervals, arising from reverberation of the skin-pleural line interface.

The main limitation of the ULC technique is its dependence on operator skill. However, reproducibility was high in this study, and an intra-subject pre-post comparison was performed (1, 11). The lack of a precise topographic assessment of the lung is a further limitation. In fact, chest sonography evaluates only the subpleural cortical area, extending for 3 to 4 cm. Finally, chest sonography, like chest radiography or computed tomography, is unable to distinguish water- versus blood-thickened interlobular septa, as well as interstitial fibrosis. In any case, we excluded lung disease in our athletes by medical history and the physical examination. On the other hand, the simplicity and rapidity (less than 3 min) of performing and interpreting ULCS, with a learning curve of less than 10 examinations, and the possibility of using a basic ultrasound machine, as well as its independence from the cardiac acoustic windows, and the absence of ionizing radiation allow for open field, as in our case, and its repetition.

All variables were obtained by different methodologies and corrupted with different measurements of error. We used a complex multivariate statistical analysis (MFA), which allowed balancing the large number of blocks (each containing a different number of variables) ensuring the representation of all data. MFA is an effective exploratory data technique, capable of simultaneously discovering correlations among variables and groups of variables also in small study groups (9).

In conclusion, the study documented the occurrence of transient and asymptomatic exercise-induced interstitial pulmonary edema in Ironman athletes through the detection of ULCS at the end of maximal sea level exercise, which were related to exercise-induced hemodynamic changes rather than inflammatory response.

**APPENDIX**

MFA. Given the standardized data matrix $X$ of dimension $I \times J$ (namely, $I$ subjects and $J$ features) whose columns are arranged in $K$ groups (each of dimension $J_k$), the weighted cross-product matrix $W$ of dimension $I \times I$ is calculated by

$$W = XQX^T = \sum_k -X_kQ_kX_k^T = \sum_k -W_k$$

where $L$ is a diagonal matrix of dimension $I \times I$ whose diagonal elements are equal to $I^{-1}$ and $Q_k$ is the $k^{th}$ group weighting matrix of dimension $J_k \times J_k$ whose diagonal elements are equal to the inverse of the first eigenvalue obtained from the singular value decomposition (SVD) of $X_k^T X_k$.

The SVD of weighted cross-product matrix $W$ leads to $W = USU^{-1}$ where $S$ is a diagonal matrix that includes the squared singular values of $W$ in descending order of magnitude and $U$ being the orthonormal eigenvector matrix.

The GAs matrix (i.e., MFA scores) is obtained as $\sqrt{U} \sqrt{S}$, whereas the contributions $z_k$ of the $k^{th}$ group on GAs are calculated by $z_k = U^T W_k U$ which are expressed as percentage, after dividing it for the corresponding squared eigenvalue.

Features/groups that directly contribute to GAs (i.e., columns/groups of $X$) are called active, whereas features/groups that are only projected on obtained GAs are called supplementary, since they add ancillary information to MFA results.

**Number of significant latent factors.** For estimation of the correct number of GAs, the magnitude of MFA eigenvalues was compared with those obtained from several randomly generated data sets of the same size (i.e., same number of cases, features, and groups); only eigenvalues exceeding the corresponding 97.5th percentile derived from the random data sets were retained for further investigation.

**REFERENCES**


SELECTED REFERENCES


