

The Use of Partial Least Square (PLS) to Explore the Importance of Sperm Characteristics in the Prediction of Bull Fertility

Leticia Zoccolaro Oliveira^{1,2*}, Pablo Daniel Reeb^{3,4}, Fabio Morato Monteiro⁵, Janaina Torres Carreira¹ and Rubens Paes de Arruda²

¹Department of Animal Reproduction, School of Veterinary Medicine, UNIRP-Centro Universitário de Rio Preto, Unidade Universitária II (Hospital Veterinário), Rodovia Br 193 (Transbrasiliana), Km 69, 15093-450, São José do Rio Preto, SP, Brazil

²School of Veterinary Medicine and Animal Science, University of São Paulo, Pirassununga, SP, 13630-000, Brazil

³Department of Statistics, College of Agricultural Sciences, National University of Comahue, Cinco Saltos, RN, Argentina

⁴Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, 48824, USA

⁵Instituto de Zootecnia, Sertãozinho, SP, Brazil

Abstract

This work aimed to provide further information that may help researchers to use the Partial Least Square (PLS) regression analysis for selecting a set of laboratory assays that are better correlated to our response variable (field fertility), explaining how the PLS was used to explore the importance of *in vitro* sperm characteristics in the prediction of bull fertility. In the field experiment, multiparous Nelore cows (n=97) from a commercial farm underwent the same fixed-time artificial insemination protocol. Three batches of frozen semen from three Angus bulls were used. The same semen thawing protocol was performed in the laboratory to mimic field conditions. The following *in vitro* semen characteristics were assessed: Computer Assisted Semen Analysis, Thermal Resistance Test, Hyposmotic Swelling Test (HOST), assessment of plasma and acrosomal membrane integrity, assessment of sperm plasma membrane stability and of lipid peroxidation by flow cytometry, sperm morphometry and chromatin structure by Toluidine Blue staining, sperm morphology and sperm concentration. For statistical analyses, PLS regression was used to explore the importance of various sperm variables in the prediction of the response (conception rate). Twenty two (out of forty seven) *in vitro* sperm variables were determined to be important predictors of conception rate. The PLS regression was considered an interesting statistical method to identify a group of important variables in the prediction of a response for semen fertility studies.

Introduction

It has been established that individual bulls differ in their ability to fertilize oocytes and/or to develop to embryo stages following *in vitro* fertilization [1-4]. In addition, a marked variability in field fertility among bulls has been reported [5-8], the so called "bull effect" [9].

Attempting to predict bull fertility, many classical and modern laboratory tests have been developed for the assessment of semen quality [10-14]. However, the results of such *in vitro* assays do not always correlate with the real fertility of a semen sample [4,15].

Even though a practical *in vitro* semen assay for determining bull fertility would be of great benefit to reproductive programs [6], it is unlikely that the evaluation of a single sperm characteristic may reflect the real sperm fertilization capacity of a semen sample. Hence, finding a set of sperm characteristics considered important for predicting conception rate would be a useful method of monitoring and/or predicting the sire fertility [16].

According to Sudano et al. [4], the competence of a specific laboratory test in predicting the bull fertility can be directly related to the statistical analysis performed. In a previous work of Oliveira et al. [16], we used the Partial Least Square (PLS) analysis for the assessment of *in vitro* sperm characteristics and their importance in the prediction of conception rate in a bovine timed-AI program. Such statistical analysis was chosen because it allows to simultaneously explain not only the relationship between response (Y) and predictor (X) variables but also to explore the relationships between the predictor variables Esposito Vinzi and Russolillo [17]. Furthermore, PLS can be used with a relatively small set of data; a situation that is frequently observed in biological experiments. For instance, PLS methods can be used to analyze data from multiple modalities collected on the same observations Krishnan et al. [18]. Hence, with PLS regression analysis, it was possible investigating the relationships between laboratory sperm characteristics and field fertility even with few repetitions for each bull

and reduced number of bulls Oliveira et al. [16]. PLS methods have been applied in a diverse range of fields such as chemistry Wold et al. [19], genomics [20], neurobiology [21] and marketing.

In this article, we aim to provide further information that may help researchers to use PLS regression analysis for selecting a set of laboratory assays that is better correlated to field fertility, explaining how the PLS was used to explore the importance of *in vitro* sperm characteristics in the prediction of bull fertility.

Material and Methods

Assessment of field fertility

For this experiment, it was utilized data from a timed-AI program of suckled multiparous Nelore cows (n=97) located in a commercial beef farm in the state of Mato Grosso, Brazil. Cows presented body condition score (BCS) between 1.75 and 3.25 in a 1-5 scale (1=emaciated, 5=obese).

All cows receive the same timed-AI protocol beginning 30 to 40 d postpartum for first service. Cows were inseminated by two

***Corresponding author:** Leticia Zoccolaro Oliveira, Department of Animal Reproduction, UNIRP-Centro University of Rio Preto, Unidade II (Veterinary Hospital), Br Highway 153 (Transbrasiliana), Km 69, 15093-450, Saint Joseph of Rio Preto, Academic of Black River SP, Brazil, Tel: 55-17-3201-3360; Fax: 55 17-3226-7866; E-mail: leticiazoccolaro@yahoo.com.br

Received February 06, 2013; **Accepted** March 06, 2013; **Published** March 07, 2013

Citation: Oliveira LZ, Reeb PD, Monteiro FM, Carreira JT, Arruda RP (2013) The Use of Partial Least Square (PLS) to Explore the Importance of Sperm Characteristics in the Prediction of Bull Fertility. J Veterinar Sci Technol S11: 005. doi:10.4172/2157-7579.S11-005

Copyright: © 2013 Oliveira LZ, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

experienced AI technicians that inseminated the same number of cows.

Frozen semen doses from three batches of three Angus bulls were used. The semen thawing and handling protocols were performed according to the routine procedures of the farm where the experiment was conducted.

Semen from each of the 3 bulls was equally distributed in the breeding groups and AI technician in order to guarantee a randomized experimental design and a balanced number of animals per field variable. All cows were examined for pregnancy by transrectal ultrasonography 40 d after AI.

Laboratory assessment of semen quality-laboratorial experiment

Frozen semen samples from each bull and batch (n=9) utilized in the field trial were brought to the laboratory. Each semen batch was thawed in the same thermostatically controlled thawing bath of the field experiment, in a temperature of 36°C, for 30 sec. After thawing, the following *in vitro* semen analyses were performed:

Computer-assisted semen analysis (CASA): Sperm motility was assessed by CASA (Ivos-Ultimate[®]; Hamilton Thorne Biosciences, Beverly, MA, USA). The CASA set-up was pre-adjusted for bovine sperm with the following parameters: number of frames: 30; frames per s: 60 Hz; minimum contrast: 50; minimum cell size: 6 pixels; contrast to static cells: 30; straightness: 60%; average path velocity cutoff: 30 $\mu\text{m/s}$; minimum average path velocity: 40 $\mu\text{m/s}$; straight-line velocity cutoff: 20 $\mu\text{m/s}$; cell size: 6 pixels; cell intensity: 80; static head size: 0.23 to 1.91; static head intensity: 0.56 to 1.20; static elongation: 8 to 92; magnification: X 1.89; video frequency: 60; illumination intensity: 2203; and temperature: 37°C.

For assessment of computer assisted sperm motility, 6 μL of the semen sample was placed in a standard count analysis chamber (Makler counting chamber, SEFI Medical Instruments LTD, Haifa, Israel). Six fields were randomly selected for each analysis. The following variables were analyzed by CASA: total motility (TM), progressive motility (PM), average path velocity (VAP), straight-line velocity (VSL), curvilinear velocity (VCL), amplitude of lateral head displacement (ALH), beat cross frequency (BCF), straightness (STR), linearity (LIN), and the percentage of rapidly moving cells (RAP).

Sperm thermal resistance test (TRT): The thermal-resistance test (TRT) was carried out in order to verify post-thaw sperm longevity of semen samples. Immediately after the thawing bath removal, an aliquot of 250 μL of frozen-thawed semen was put into a warmed micro centrifuge tube which remained incubated at 37°C. After 120 min of incubation (TRT 2 h), the same procedure described on section 2.2.2 for CASA was applied and the following motility parameters were assessed for TRT_2h: TM_2h, PM_2h, VAP_2h, VSL_2h, VCL_2h, ALH_2h, BCF_2h, STR_2h, LIN_2h, RAP_2h.

Hiposmotic swelling test (HOST): The Hiposmotic Swelling Test was performed by incubating 20 μL of semen with 1 mL of a 100 mOsm hiposmotic solution at 37°C for 60 min. After incubation, 20 μL of the solution was coverslipped and evaluated by contrast phase microscopy. Two hundred sperm were evaluated under magnification 1000 \times . Sperm with swollen or coiled tails were considered viable. The percentage of viable sperm in the Hiposmotic Swelling Test (HOST+cells) was calculated according to Revell and Mrode [10].

Flow cytometry analyses: These analyses were carried out using

the FACSAria[®] (Becton-Dickinson, San Jose, CA, USA) flow cytometer equipped with a 405 nm Near UV laser and 488 nm Sapphire blue laser and filters B (Band Pass 450/20), C (Long Pass 595 nm/Band Pass 610/20 nm), D (LP 556 nm/BP 575/26 nm) and E (LP 502 nm/BP 530/30 nm). The cells were simultaneously excited by an argon laser at 488 nm and by a Near UV laser at 405 nm as described by de Andrade et al. [22].

Samples for staining and flow cytometry analysis were diluted in a modified Tyrode's medium (TALPm; pH 7.4) with 114 mM NaCl, 3.2 mM KCl, 0.5 mM $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, 0.4 mM $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$, 5 mM glucose, 10 mM sodium lactate, 0.1 mM sodium pyruvate and 10 000 UI/100 mL sodium penicillin. After the addition of the dyes for each analysis, the semen samples incubated in the TALPm were analyzed in the flow cytometer (BD FACSDiva 6.0 software; Becton-Dickinson). The samples were processed through the instrument at an acquisition rate of approximately 600 to 1 000 events/s, acquiring 10,000 cells per analysis [23].

Simultaneous assessment of plasma and acrosomal membranes:

An aliquot was taken from the semen samples and added to the TALPm. The resulting samples had a concentration of 5×10^6 spermatozoa/mL in a volume of 148 μL . Then, 2 μL of Hoechst 33342 (H342; 40 $\mu\text{g/mL}$) was added. After 10 min of incubation at 37°C, 3 μL of propidium iodide (PI, 0.5 mg/mL; 28.707-5; Sigma-Aldrich, St Louis, MO, USA) and 10 μL of *Pisum sativum* agglutinin conjugated to fluorescein isothiocyanate (FITC-PSA, 100 $\mu\text{g/mL}$; L-0770; Sigma-Aldrich) were added to the samples [14]. After 10 min of incubation at 37°C, the samples were diluted with the addition of 150 μL of the TALPm to a concentration of 2.5×10^6 spermatozoa/mL to be analyzed by a flow cytometry [22].

Two-dimensional dot-plots of FITC-PSA (Filter E) vs. PI fluorescence (Filter C) from a total of 10,000 events were generated. Each quadrant represented one of the following sperm subpopulations: 1) IPIA: sperm with intact plasma and acrosomal membranes; 2) IPDA: sperm with an intact plasma membrane and a damaged acrosomal membrane; 3) DPIA: sperm with a damaged plasma membrane and an intact acrosomal membrane; and 4) DPDA: sperm with damaged plasma and acrosomal membranes [23]. The percentage of sperm cells presenting intact plasma membranes (IPM: IPIA+IPDA) and the percentage of sperm cells presenting intact acrosome (IA: IPIA+DPIA), as well as the IPIA subpopulation were considered for the regression analysis.

Assessment of sperm plasma membrane stability: Samples were incubated in the TALPm solution at a concentration of 5×10^6 spermatozoa/mL in 147 μL with 2 μL of H342 (40 $\mu\text{g/mL}$). Subsequently, 0.5 μL of fluorescent probe Yo-Pro-1 (Y3603, Molecular Probes Inc., Eugene, OR, USA) was added to the sample (7.5 μM), resulting in a final concentration of 25 nM. Subsequently, the sample was incubated for 20 min and then merocyanine 540 (M540) fluorescent probe (M 24571, Molecular Probes Inc., Eugene, OR, USA) was added (810 μM) to obtain a concentration of 2.7 μM in 150 μL ; and was incubated for 70 seconds. Samples were then diluted in 150 μL of TALPm and analysed by flow cytometry [24].

The M540 staining showed two distinct populations of viable cells: one that was characterised by cells with low fluorescence emission (LBD) and the other that was characterised by high fluorescence (HBD) captured in the long pass-595 and band-pass-610/20 nm. The Positive Yo-Pro cells (cells with a damaged plasma membrane; YoPro+cells), LBD and HBD sperm populations were considered for relationship analysis.

Assessment of lipid peroxidation: An aliquot was taken from the

samples and TALPm was added to obtain samples with a concentration of 5×10^6 spermatozoa/mL and a final volume of 499.5 μ L. Then, the C11-BODIPY^{581/591} (1mg/mL, D-3861, Molecular Probes Inc., Eugene, OR, USA) was added. The sample was incubated for 30 min at 37°C. After this incubation period, 145 μ L of this solution was transferred to another micro tube and 2 μ L of H342 (40 μ g/mL) was added. This sample was incubated for 10 min at 37°C. After the incubation period of H342 probe, 3 μ L of PI (0.5 mg/mL) was added in the sample. Then, the sample was incubated with PI for 5 min at 37°C. Subsequently, the sample was diluted with the addition of 150 μ L of TALPm to a concentration of 2.5×10^6 spermatozoa/mL and was analyzed by flow cytometry.

Two-dimensional dot-plots of C11-BODIPY^{581/591} (Filter E) vs. PI fluorescence (Filter C) from a total of 10,000 events were generated. Each quadrant represented one of the following sperm subpopulations: 1) IPP: sperm with intact plasma membrane suffering lipid peroxidation [PI (-), C11-BODIPY^{581/591} (+)]; 2) IPNP: sperm with intact plasma membrane with no lipid peroxidation detected [PI (-), C11-BODIPY^{581/591} (-)]; 3) DPP: sperm with damaged plasma membrane and lipid peroxidation [PI (+), C11-BODIPY^{581/591} (+)]; and 4) DPNP: sperm with damaged plasma membrane and no lipid peroxidation detected [PI (+), C11-BODIPY^{581/591} (-)]. The variables IPP and IPNP were considered for relationship analysis of the present study.

Assessment of sperm morphology: An aliquot of frozen-thawed semen was diluted and fixed in pre-warmed (37°C) formaldehyde-PBS. Sperm cells (n=200) were counted under differential interference-contrast microscopy (model 80i; Nikon, Tokyo, Japan) at a magnification of 1000 \times . Morphological characteristics of sperm were classified as major defects (Maj_Def), minor defects (Min_Def) and total defects (Tot_Def) according to Blom [25].

Assessment of sperm morphometry and chromatin structure: For each sample, two smears were prepared for subsequent assessment of sperm chromatin structure and morphometry. The sperm smears were fixed with ethanol acetic acid (3:1, V/V) for 1 min and 70% ethanol for 3 min. Then, the smears were hydrolyzed for 25 min in 4 M HCl, washed in distilled water and air-dried. One droplet of 0.025% toluidine blue in McIlvaine buffer (sodium citrate-phosphate), pH 4.0, was placed over each smear and then coverslipped.

Fifty gray-level digital images of each slide were obtained randomly using a Leica DM500 microscope (Leica Microsystems Inc., Buffalo Grove, IL, USA) with a 100X objective lens (immersion) coupled to a Leica ICC50 camera (Leica Microsystems Inc.) that was connected to a PC microcomputer. Using threshold-based image segmentation [26], at least 100 sperm heads were isolated for each smear. The sperm heads were analyzed to obtain the average pixel value that made up each head. Six heads with the smallest pixel values were selected automatically and defined as standard heads. Subsequently, for each image, the difference between the standard value of the smear and the average value of each head analyzed was determined. This difference was transformed into a percentage (Dif %) of the average pixel value for the standard heads, which indicates the sperm chromatin decondensation (DIF). The coefficient of variation (CV) of the gray level intensity for each head, which indicates the sperm chromatin heterogeneity, was also calculated [13,14].

Area, Perimeter, Width, Length, Width:length ratio (WLR), Ellipticity, and Shape Factor (SF) of all sperm heads were determined using other algorithm developed in Scilab environment [13,26]. Fourier descriptors containing harmonic amplitudes from 0 to 2

(Fourier zero: Fourier_Z, Fourier one: Fourier_1 and Fourier two: Fourier_2) were also considered. Another estimated feature was the sperm head symmetry. The Side Symmetry (SS) is a measurement that identifies asymmetries along the principal (major) sperm axis. The Anterior-posterior Symmetry (SAP) is a measurement that identifies asymmetries along the horizontal (minor) sperm axis. All of these symmetries were calculated using the procedure described by Beletti and Costa [26], Kanayama and Beletti [27] which involves flipping the object along its major (or minor) axis, and then identifying the area of overlap between the original and flipped areas [13].

Assessment of sperm concentration: An aliquot of 20 μ L of frozen-thawed semen was diluted in 980 μ L of formaldehyde-PBS. The sperm cells were counted using Neubauer chamber, under optic microscopy, at a magnification of 400 \times and sperm concentration was calculated for each semen sample.

Statistical analysis

In this study, the method of Partial Least Squares (PLS) regression was the statistical procedure used to explore the importance of the *in vitro* sperm variables in the prediction of bull fertility. The PLS regression, which was developed from an initial propose of Wold [28], is a technique of multivariate data analysis used to relate one (or more) response variable (Y) (conception rate, in this case) with several predictor variables (X) (*in vitro* sperm characteristics) based on the extraction of factors (Figure 1). These factors, also called components or latent variables, simultaneously decompose the two set of variables maximizing the explanation of the variability of either Y or X, or both.

The PLS analysis was performed using proc PLS implemented in SAS v.9.3 (SAS Inst. Inc., Cary, NC, USA). The results obtained in the laboratory tests (laboratory experiment; 47 variables) were confronted with the result of field experiment (conception rate; one response variable). A matrix of scatter-plots was used to explore the patterns among variable and detect non-linear relationships.

In each trial, a regression coefficient and a variable importance plot was obtained for each predictor (sperm variables). The regression coefficients represent the importance that each predictor (variable) has in the prediction of the response). The variable importance plot, on the other hand, represents the contribution of each predictor in fitting the PLS model for both predictors and response. It is based on the

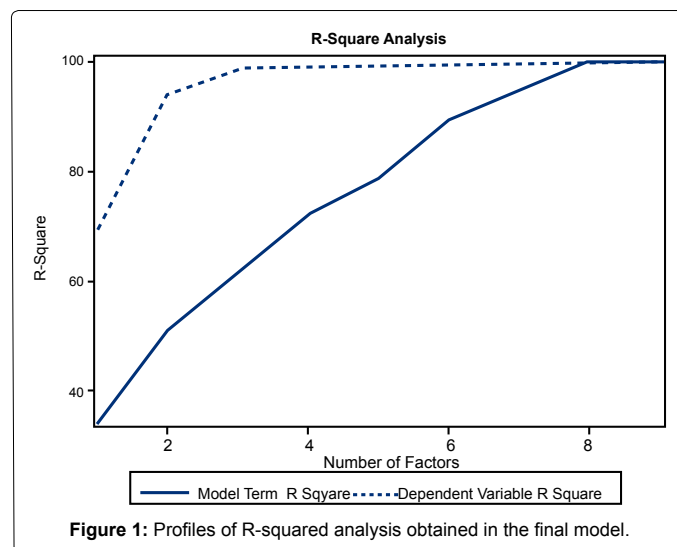


Figure 1: Profiles of R-squared analysis obtained in the final model.

Variable Importance for Projection (VIP) statistic of Wold [28] which summarizes the contribution of a variable to the model. If a predictor has a relatively small coefficient (in terms of absolute value) and a small VIP value, then it is a prime candidate for deletion. Wold's reports in Umetrics SAS [29] consider a value less than 0.8 to be "small" for the VIP.

In the present study, in order to select the laboratory sperm variables that would be important predictors of field fertility, we propose that the PLS regression was performed consecutively, until no remaining variables presented VIP lower than 0.8. Hence, the following procedure was performed: after running the PLS analysis for the first time, the predictors with VIP values lower than 0.8 were excluded. Then, the PLS was performed for the second time without the deleted variables. Again, the predictors with VIP values lower than 0.8 were excluded. Then, the PLS was performed for the third time without the deleted variables. When no other sperm variables presented VIP<0.8, the remaining variables were selected as the group of good predictors for conception rate. In each step, residual analysis, correlation loading plots, as well as X and Y scores plot for the first four factors were used to detect outliers, non-linear patterns or grouped observation.

Results

The PLS factors were assessed for each *in vitro* sperm characteristic and the variables were selected according to Wold's criterion (SAS) [29].

After the first PLS procedure, the following variables presented VIP<0.8 and were excluded: VAP, VSL, ALH, STR, LIN, ALH_2h, STR_2h, LIN_2h, HBD, minor defects, Area, Perimeter, Width, Fourier_1, Side Symmetry and DIF. Then, a second run of PLS procedure was performed with the remaining variables.

After the second PLS procedure, the following variables presented VIP<0.8 and were excluded: VCL, VSL_2h, VCL_2h, AI, LBD, Length and Shape Factor. In addition, ellipticity was excluded because it represents the same sperm parameter than WLR and YoPro was excluded because it represents the inverse of the IPM parameter. Then, a third run of PLS procedure was performed with the remaining variables.

After the third PLS procedure, no other sperm variable presented VIP<0.8. Hence, the following variables were considered important for prediction of conception rate: TM, PM, BCF, RAP, TM_2h, PM_2h, VAP_2h, BCF_2h, RAP_2h, HOST, IPIA, IPM, IPNP, IPP, Maj_Def, Tot_Def, WLR, Fourier_Z, Fourier_2, AP symmetry, CV and sperm concentration.

Table 1 demonstrates the descriptive data of conception rate and semen characteristics outcomes from sperm variables selected after the third PLS procedure.

No important curved patterns were detected in our analyses as to suggest the addition of transformed variables [19,29]. Table 2 demonstrates the Variable Importance for Projection (VIP) values, the parameters estimates for centered and scaled data, as well as the parameters for data in the original scale obtained after the last run of the Partial Least Square (PLS) procedure, performed for sperm variables in the prediction of conception rate.

Additionally, in figure 3, it is possible to observe the predictor loading profiles for the first factor extracted from the final model. Interpreting the contribution of each variable in the factors that are extracted by PLS can be also interesting because it helps to study the structure of relationships among the predictors and to relate it with the prediction of the response.

As stated above, PLS analysis extract a number of factors (components) and in the present study, it was decided to extract four factors from our analysis. Each factor is a linear combination of all the original variables. The first factor is constructed with the loadings presented in figure 3. The loadings demonstrates how the first PLS factor was constructed; i.e., it demonstrates essentially which variable, and in which sense (positive or negative), are participating in that factor (Figure 3).

Discussion and Conclusion

In the present study, sperm characteristics were assessed by classical and modern laboratory tests. Then, in order to identify a group of important variables in the prediction of conception rate, PLS analysis was performed.

	CR% (n/n)	TM (%)	PM (%)	BCF (Hz)	RAP (%)	TM2h (%)	PM2h (%)	VAP2h (%)	BCF2h (%)	RAP2h (%)	HOST (%)	IPIA (%)	IPM (%)	IPNP (%)	IPP (%)	MajDef (%)	TotDef (%)	WLR	F0	F2	SAP	CV (%)	Concent (x10 ⁶ sptz/mL)
Batch1 (Bull1)	63.64 (7/11)	46.60	41.45	22.60	43.00	30.00	25.79	62.75	24.20	26.57	41.00	43.80	44.40	33.80	0.70	19.00	23.00	0.519	3612.69	300.61	0.937	5.54	45.00
Batch2 (Bull1)	70.00 (7/10)	54.93	49.21	26.78	51.03	8.59	6.06	54.18	23.03	6.65	32.00	31.60	31.80	27.10	0.30	14.00	26.00	0.520	4120.75	329.92	0.927	12.43	35.00
Batch3 (Bull1)	60.00 (6/10)	51.59	42.18	21.80	43.94	20.85	16.05	56.84	24.77	17.35	40.00	41.90	42.20	31.90	0.60	10.00	21.00	0.516	3986.61	341.55	0.932	6.45	25.00
Batch4 (Bull2)	10.00 (1/10)	24.38	20.48	24.75	21.95	5.18	3.21	47.25	23.65	3.43	21.00	31.10	31.50	24.80	4.00	19.50	33.00	0.528	3570.51	322.39	0.932	11.03	25.00
Batch5 (Bull2)	60.00 (6/10)	42.17	34.81	22.56	36.33	10.39	7.21	54.48	22.81	8.07	54.00	42.30	42.60	34.90	1.80	26.50	35.50	0.535	3694.10	323.02	0.927	10.26	20.00
Batch6 (Bull2)	44.44 (4/9)	34.22	28.63	29.68	29.50	4.75	3.38	45.40	9.25	3.38	27.00	36.00	36.20	31.30	0.50	24.50	33.00	0.534	3628.28	341.87	0.931	10.63	50.00
Batch7 (Bull3)	60.00 (9/15)	28.13	21.56	24.52	22.44	17.64	15.27	65.33	23.66	16.18	27.00	31.10	31.20	22.10	0.20	14.00	30.00	0.515	3936.73	324.75	0.924	10.95	60.00
Batch8 (Bull3)	50.00 (4/8)	51.78	33.81	29.81	37.09	43.31	22.91	50.77	24.13	26.43	66.00	60.30	60.50	51.60	0.50	7.50	19.50	0.533	3825.65	364.78	0.929	8.52	55.00
Batch9 (Bull3)	85.71 (12/14)	58.45	40.68	19.31	46.65	16.25	9.09	50.33	26.37	9.78	45.00	50.10	50.30	48.70	0.70	15.00	36.00	0.518	3743.28	297.44	0.929	5.94	75.00

CR: Conception rate; TM: total motility; PM: progressive motility; BCF: beat cross frequency; RAP: percentage of rapidly moving cells; TM2h: TM after 2h of thermal incubation; PM2h: PM after 2h of thermal incubation; VAP2h: Average path velocity after 2 h of thermal incubation; BCF2h: BCF after 2 h of thermal incubation; RAP2h: RAP after 2 h of thermal incubation; HOST: intact plasma membrane evaluated by Hiposmotic Swelling Test; IPIA: intact plasma and acrosomal membranes evaluated by flow cytometry; IPM: total percentage of sperm cells presenting intact plasma membranes; IPNP: intact plasma membrane with no lipid peroxidation detected; IPP: intact plasma membrane suffering lipid peroxidation; MajDef: Major Defects; TotDef: Total of defects; WLR: Width:Length ratio; F0: mathematic parameter Fourier 0; F2: mathematic parameter Fourier 2; SAP: antero-posterior symmetry; CV: chromatin heterogeneity; concent: sperm concentration.

Table 1: Descriptive data of conception rate obtained in field experiment and sperm characteristics assessed in laboratory experiment.

An ordinary multiple regression approach cannot be applied in this context due to the large number of variables compared to the number of observations, and no independency among the predictor variables (Table 3). Conversely, a principal component regression approach can be fitted, but it only maximizes the explanation of the predictor variables and does not guaranteed a relevant explanation for the response by contrast, PLS regression can find a set of linear combination of predictors (components) that best predict the response even with a relatively small sample size. Additionally, the components can be used to interpret the structure of the predictors and the relationship with the responses. Hence, with PLS statistical analysis, it was possible investigating the relationships between laboratory sperm characteristics and field fertility even with few repetitions for each bull and reduced number of bulls as described by Oliveira et al. [16].

As stated above, the PLS procedures selected the important *in vitro* variables in the prediction of field fertility. Our statistical analysis cutoff was based in Wold's criterion SAS [29-32] which considers a value less than 0.8 to be "small" for VIP in the PLS procedure. Therefore, the

variables presenting values lower than 0.8 for VIP were excluded after the sequences of PLS procedures, until no remaining variable presented a VIP<0.8.

Consequently, according to our results, twenty two (out of forty seven) *in vitro* sperm variables were considered to be important predictors of conception rate: total motility (TM), progressive motility (PM), Beat Cross Frequency (BCF), Rapidly moving cells (RAP), TM after 2 h of thermal incubation (TM_2h), PM after 2 h of thermal incubation (PM_2h), average path velocity after 2 h of thermal incubation (VAP_2h), BCF after 2 h of thermal incubation (BCF_2h), RAP after 2 h of thermal incubation (RAP_2h), intact plasma membrane evaluated by HOST, intact plasma and acrosomal membranes evaluated by flow cytometry (IPIA), total of intact plasma membrane evaluated by flow cytometry (IPM), intact plasma membrane suffering lipid peroxidation (IPP), intact plasma membrane with no lipid peroxidation (IPNP), major defects, total defects, morphometric width/length ratio (WLR), mathematic parameters Fourier 0 and Fourier 2, anterior-posterior symmetry (SAP), chromatin heterogeneity (CV) and sperm concentration.

Variables	VIP Value		
Intercept	-	0.000	1022.215
Total Motility (TM)	1.40	0.150	0.257
Progressive Motility (PM)	1.38	0.159	0.340
Beat Cross Frequency (BCF)	0.87	-0.114	-0.669
Rapidly moving cells (RAP)	1.40	0.157	0.314
TM_2h	0.88	-0.033	-0.054
PM_2h	0.84	0.016	0.039
VAP_2h	0.83	0.178	0.558
BCF_2h	0.80	-0.069	-0.286
RAP_2h	0.84	0.006	0.013
HOST +	0.83	0.098	0.143
IPIA	0.82	0.026	0.055
IPM	0.82	0.023	0.050
IPNP	0.81	0.052	0.108
IPP	1.60	-0.323	-5.596
Major Defects	0.87	0.127	0.420
Total Defects	0.95	0.136	0.445
Widht:Lenght ratio	1.00	-0.073	-180.827
Fourier_0	0.95	0.102	0.011
Fourier_2	0.85	-0.145	-0.145
Anterior-posterior simmetry	0.85	-0.188	-1011.974
Cromatin heterogeneity	0.85	-0.0414	-0.339
Sperm concentration	0.97	0.174	0.196

TM_2h: TM after 2 h of thermal incubation; PM_2h: PM after 2 h of thermal incubation; VAP_2h: Average patway velocity after 2 h of thermal incubation; BCF_2h: BCF after 2 h of thermal incubation; HOST +: intact plasma membrane evaluated by Hiposmotic Swelling Test; IPIA: intact plasma and acrosomal membranes evaluated by flow cytometry; IPM: total percentage of sperm cells presenting intact plasma membranes; IPNP: intact plasma membrane with no lipid peroxidation detected; IPP: intact plasma membrane suffering lipid peroxidation.

Table 2: Variable importance for projection (VIP) and regression coefficients values obtained after the last run of partial least square (PLS) procedure, performed for sperm variables in the prediction of conception rate.

Number of extracted factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	33.883	33.883	69.504	69.504
2	17.201	51.084	24.421	93.925
3	11.044	62.129	4.825	98.749
4	9.582	71.710	1.056	99.806

Table 3 demonstrates the R squared obtained in the last run of Partial Least Square (PLS) procedure performed for sperm variables in the prediction of conception rate. Note that 93% of the conception rate is already explained by two factors, but only 51% of the predictor variation is explained. A model using only the first two factors can be useful for prediction purposes; however, we decided to include 4 factors in order to increase the explanation of the predictor.

Table 3: Percent variation accounted for the first 4 partial least squares factors of sperm variables in the prediction of conception rate.

Additionally, with the PLS regression procedure of the present experiment, it was possible to demonstrate, within the selected sperm characteristics, which variables presented higher importance in the prediction of conception rate. As demonstrated in figure 2, the most important predictors are the variables that presented higher absolute parameters from the centered and scaled values, and high VIP values. This information can be confronted with the contribution of each predictor in the construction of the factors (loadings).

Another interesting aspect of PLS regression is the possibility of checking the relationships among the predictors and between predictors and response by interpreting the factor loadings. For instance, in figure 3, it is possible to observe that TM, PM, RAP, TM_2 h, PM_2 h, VAP_2h, BCF_2 h, RAP_2 h, HOST, IPIA, IPM, IPNP Fourier 0, SAP and sperm concentration have positive values in the prediction of conception rate (CR) while BCF, IPP, major defects, total defects, WLR, Fourier 2 and chromatin heterogeneity have negative values.

The observation of factor 1 from the final model (Figure 3) enriches the interpretation of the results, since some sperm variables may be positively (positive values) or negatively (negative values) important in the prediction of conception rate. In this sense, it is interesting to note that some negative variables of figure 3 actually seem to be negatively correlated to conception rate according to scientific literature; and most of the positive variables actually seem to be the sperm characteristics positively correlated with semen fertility. Farrell et al. [33] demonstrated that multiple combinations of CASA variables had high correlations with bull fertility. For instance, the authors observed that the combination of Progressive Motility, ALH, BCF, and VAP presented high correlation value Farrel et al. [33] In our study, Plasma membrane functionality (HOST+), plasma membrane integrity (IPM) and also sperm cells presenting intact acrosomal and plasma

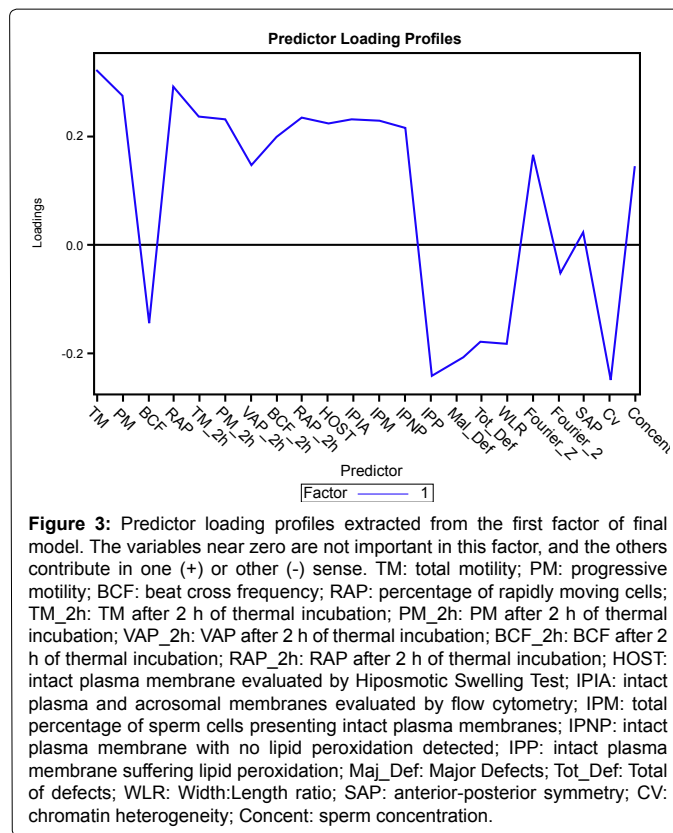


Figure 3: Predictor loading profiles extracted from the first factor of final model. The variables near zero are not important in this factor, and the others contribute in one (+) or other (-) sense. TM: total motility; PM: progressive motility; BCF: beat cross frequency; RAP: percentage of rapidly moving cells; TM_2h: TM after 2 h of thermal incubation; PM_2h: PM after 2 h of thermal incubation; VAP_2h: VAP after 2 h of thermal incubation; BCF_2h: BCF after 2 h of thermal incubation; RAP_2h: RAP after 2 h of thermal incubation; HOST: intact plasma membrane evaluated by Hiposmotic Swelling Test; IPIA: intact plasma and acrosomal membranes evaluated by flow cytometry; IPM: total percentage of sperm cells presenting intact plasma membranes; IPNP: intact plasma membrane with no lipid peroxidation detected; IPP: intact plasma membrane suffering lipid peroxidation; Maj_Def: Major Defects; Tot_Def: Total of defects; WLR: Width:Length ratio; SAP: anterior-posterior symmetry; CV: chromatin heterogeneity; Concent: sperm concentration.

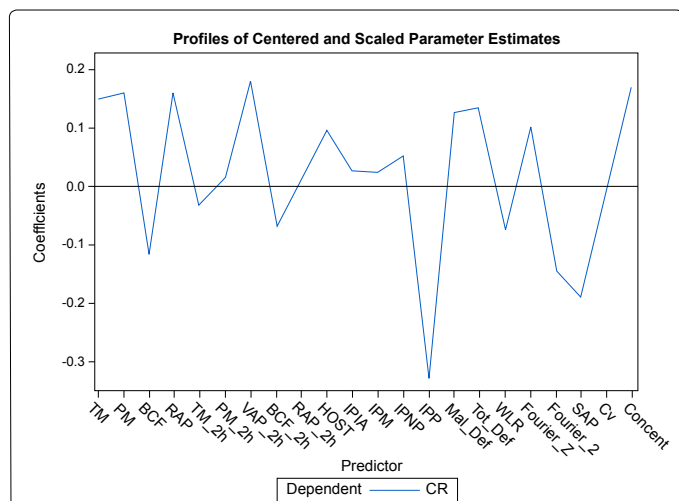


Figure 2: Profiles of centered and scaled parameters obtained after third Partial Least Square (PLS) procedure. CR: conception rate; TM: total motility; PM: progressive motility; BCF: beat cross frequency; RAP: percentage of rapidly moving cells; TM_2h: TM after 2 h of thermal incubation; PM_2h: PM after 2 h of thermal incubation; VAP_2h: VAP after 2 h of thermal incubation; BCF_2h: BCF after 2 h of thermal incubation; RAP_2h: RAP after 2 h of thermal incubation; HOST: intact plasma membrane evaluated by Hiposmotic Swelling Test; IPIA: intact plasma and acrosomal membranes evaluated by flow cytometry; IPM: total percentage of sperm cells presenting intact plasma membranes; IPNP: intact plasma membrane with no lipid peroxidation detected; IPP: intact plasma membrane suffering lipid peroxidation; Maj_Def: Major Defects; Tot_Def: Total of defects; WLR: Width:Length ratio; SAP: anterior-posterior symmetry; CV: chromatin heterogeneity; Concent: sperm concentration.

membranes (IPIA cells) presented positive loadings in the prediction of conception rate. Similarly, Januskauskas et al. [34] detected significant correlations between field fertility and plasma membrane integrity. Furthermore, Tartaglione and Ritta [35] demonstrated that HOST presented high correlation coefficient with in vitro fertility. When the sperm plasma and acrosomal membrane integrity results were included in the regression model, a higher correlation coefficient was obtained [33,34] reported that bull fertility was positively correlated to plasma membrane integrity and also to total progressive motility. The authors also demonstrated that sperm lipid peroxidation and bull fertility presented high negative correlation [34] moreover negative correlations between lipid peroxidation and IVF outcomes have already been reported in humans [35]. In the present study, sperm cells presenting lipid peroxidation (IPP) presented negative loadings in the prediction of conception rate. Sperm Major and Total Defects and Width/Length Ratio (WLR, which also represents sperm ellipticity) also presented negative loadings in the prediction of conception rate. In this sense, it has been reported that low-fertility bulls generally had high seminal content of morphologically abnormal cells Saacke [36]. Chromatin heterogeneity (CV) seems also to be an important predictor, and presented negative loadings in the prediction of conception rate. Acevedo et al. [37] reported that vulnerability of sperm DNA to acid denaturation was positively associated with potential incompetence for fertilization. Kasimanickam et al. [38] reported that sires with high sperm DNA fragmentation index presented lower sperm fertilization potential; whereas sires with lower DNA fragmentation index presented higher chance of siring calves Kasimanickam et al. [38]. However, substantial variations are commonly observed among experiments and low correlations are usually detected when these sperm characteristics are separately compared with the field fertility [4,15].

Finally, it is noteworthy that the group of components that is important to predict conception rate is a combination of positives and negatives predictor effects. This information seems to be in accordance with several researchers demonstrating that semen fertility can be better estimated from semen quality when a combination of several *in vitro* sperm analysis is performed [4,31,33]. Another way to use the provided information of PLS regression is to use the regression coefficients (and the intercept value) to build the equation for the prediction of the response variable. In our case, using the results of the present experiment, it would be possible to present a formula to calculate the prediction of conception rate for semen samples. However, for such important finding, it would be more adequate to use a large number of bulls and cows in order to present a more accurate formula.

It was concluded that PLS regression is an interesting statistical method to identify a group of important predictor variables in the prediction of a response, for *in vivo* and/or *in vitro* bull fertility studies. More observations should also be included in the analysis to provide more information about the adequacy of the model using cross validation tests. Further multivariate statistical methods, i.e., Multiple factor analysis [39,40], Generalized Procrustes Analysis [41,42], can enrich the analysis of sperm characteristic, especially if the objective is to compare experimental conditions such as batch effects, AI technician, reproductive protocols, or even experimental location.

References

- Hillery FL, Parrish JJ, First NL (1990) Bull specific effect on fertilization and embryo development *in vitro*. Theriogenology 33: 249.
- Zhang BR, Larsson B, Lundeheim N, Rodriguez-Martinez H (1997) Relationship between embryo development *in vitro* and 56-day nonreturn rates of cows inseminated with frozen-thawed semen from dairy bulls. Theriogenology 48: 221-231.
- Wei H, Fukui Y (1999) Effects of bull, sperm type and sperm pretreatment on male pronuclear formation after intracytoplasmic sperm injection in cattle. Reprod Fertil Dev 11: 59-65.
- Sudano MJ, Crespiho AM, Fernandes CB, Junior AM, Papa FO, et al. (2011) Use of Bayesian Inference to Correlate *In vitro* Embryo Production and *In vivo* Fertility in Zebu Bulls. Vet Med Int.
- Correa JR, Pace MM, Zavos PM (1997) Relationships among frozen thawed sperm characteristics assessed via the routine semen analysis, sperm functional tests and the fertility of bulls in an artificial insemination program. Theriogenology 48: 721-731.
- Ward F, Rizos D, Corridan D, Quinn K, Boland M, et al. (2001) Paternal influence on the time of first embryonic cleavage post insemination and the implications for subsequent bovine embryo development *in vitro* and fertility *in vivo*. Mol Reprod Dev 60: 47-55.
- Andersson M, Taponena J, Koskinena E, Dahlbom M (2004) Effect of insemination with doses of 2 or 15 million frozen-thawed spermatozoa and semen deposition site on pregnancy rate in dairy cows. Theriogenology 61: 1583-1588.
- Oliveira LZ, Arruda RP, De Andrade AF, Santos RM, Beletti ME, et al. (2012) Effect of sequence of insemination after simultaneous thawing of multiple semen straws on conception rate to timed AI in suckled multiparous Nelore cows. Theriogenology 78: 1800-1813.
- Oliveira LZ, Arruda RP, De Andrade AFC, Celeghini ECC, Santos RM, et al. (2012) Assessment of field fertility and several *in vitro* sperm characteristics following the use of different Angus sires in a timed-AI program with suckled Nelore cows. Livestock Science 146: 38-46.
- Revell SG, Mrode RA (1994) An osmotic resistance test for bovine semen. Anim Reprod Sci 36: 77-86.
- Verstegen J, Iguer-Ouada M, Oclin K (2002) Computer assisted semen analyzers in andrology research and veterinary practice. Theriogenology 57: 149-179.
- Brouwers JF, Gadella BM (2003) *In situ* detection and localization of lipid peroxidation in individual bovine sperm cells. Free Radic Biol Med 35: 1382-1391.
- Beletti ME, Costa LF, Guardieiro MM (2005) Morphometric features and chromatin condensation abnormalities evaluated by toluidine blue staining in bull spermatozoa. Braz J Morphol Sci 22: 85-90.
- Celeghini EC, de Arruda RP, de Andrade AFC, Nascimento J, Raphael CF (2007) Practical techniques for bovine sperm simultaneous fluorimetric assessment of plasma, acrosomal and mitochondrial membranes. Reprod Domest Anim 42: 479-488.
- Zhang BR, Larsson B, Lundeheim N, Haard MGH, Rodriguez-Martinez H (1999) Prediction of bull fertility by combined *in vitro* assessments of frozen-thawed semen from young dairy bulls entering an AI-programme. Int J Androl 22: 253-260.
- Oliveira LZ, Arruda RP, Andrade AF, Celeghini EC, Reeb PD, et al. (2013) Assessment of *in vitro* sperm characteristics and their importance in the prediction of conception rate in a bovine timed-AI program. Anim Reprod Sci 137: 145-155.
- Esposito Vinzi V, Russolillo G (2013) Partial least squares algorithms and methods. Wiley Interdisciplinary Reviews: Computational Statistics 5: 1-19.
- Krishnan A, Williams LJ, McIntosh AR, Abdi H (2011) Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. NeuroImage 56: 455-475.
- Wold S, Sjostrom M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58: 109-130.
- Boulesteix AL (2004) PLS dimension reduction for classification with microarray data. Stat Appl Genet Mol Biol 3.
- Vinzi VE, Chin WW, Henseler J, Wang H (2010) Handbook of Partial Least Squares: Concepts, Methods and Applications, Springer.
- de Andrade AF, Zaffalon FG, Celeghini EC, Nascimento J, Tarrago OF, et al. (2011) Addition of Seminal Plasma to Post-thawing Equine Semen: What is the Effect on Sperm Cell Viability? Reprod Domest Anim 46: 682-686.
- Leite TG, do Vale Filho VR, de Arruda RP, de Andrade AF, Emerick LL, et al. (2010) Effects of extender and equilibration time on post-thaw motility and membrane integrity of cryopreserved Gyr bull semen evaluated by CASA and flow cytometry. Anim Reprod Sci 120: 31-38.
- Rathi R, Colenbrander B, Bevers M, Gadella BM (2001) Evaluation of *in vitro* capacitation of stallion spermatozoa. Biol Reprod 65: 462-470.
- Blom E (1973) The ultrastructure of some characteristic sperm defects and a proposal for a new classification of the bulls permiogram. Nord Vet Med 25: 383-391.
- Beletti ME, Costa LF (2003) A systematic approach to multi-species sperm morphometrical characterization. Anal Quant Cytol Histol 25: 97-107.
- Kanayama CY, Beletti ME (2011) Computational evaluation of chromatin condensation and morphometric characteristics of rabbit (*Oryctolagus cuniculus*) sperm head. Arq Bras Med Vet Zootec 63: 94-99.
- Wold H (1966) Estimation of principal components and related models by iterative least squares. In Multivariate Analysis, Academic Press, New York 391-420.
- SAS Institute (2001) STAT® User's Guide, Release 8.2. SAS Institute Inc., Cary, NC, USA.
- Wold S (1994) PLS for multivariate linear modeling. QSAR: chemometric methods in molecular design. In: Van de Waterbeemd H, (ed.). Methods and Principles in Medicinal Chemistry, Verlag-Chemie, Weinheim, Germany.
- de Kermadec FH, Durand JF, Sabatier R (1997) Comparison between linear and nonlinear PLS methods to explain overall liking from sensory characteristics. Food Qual Prefer 8: 395-402.
- Abdi H (2010) Partial least squares regression and projection on latent structure regression (PLS Regression). Wiley Interdisciplinary Reviews: Computational Statistics 2: 97-106.
- Farrell PB, Presicce GA, Brockett CC, Foote RH (1998) Quantification of bull sperm characteristics measured by computer-assisted sperm analysis (CASA) and the relationship to fertility. Theriogenology 49: 871-879.
- Januskas A, Johannisson A, Rodriguez-Martinez H (2003) Subtle

- membrane changes in cryopreserved bull semen in relation with sperm viability, chromatin structure and field fertility. *Theriogenology* 60: 743-758.
35. Tartaglione CM, Ritta MN (2004) Prognostic value of spermatological parameters as predictors of *in vitro* fertility of frozen-thawed bull semen. *Theriogenology* 62: 1245-1252.
36. Saacke RG (2008) Sperm morphology: Its relevance to compensable and uncompensable traits in semen. *Theriogenology* 70: 473-478.
37. Acevedo NJ, Bame H, Kuehn LA, Hohenboken WD, Evenson DP, et al. (2002) Sperm chromatin structure assay (SCSA) and sperm morphology. In: Proc 19th Tech Conf Artif Insem Reprod 84-90.
38. Kasimanickam R, Kasimanickam V, Thatcher CD, Nebel RL, Cassell BG (2007) Relationships among lipid peroxidation, glutathione peroxidase, superoxide dismutase, sperm parameters, and competitive index in dairy bulls. *Theriogenology* 67: 1004-1012.
39. Zabludovsky N, Eltes F, Geva E, Berkovitz E, Amit A, et al. (1999) Relationship between human sperm lipid peroxidation, comprehensive quality parameters and IVF outcome. *Andrologia* 31: 91-98.
40. Saacke RG (2008) Sperm morphology: Its relevance to compensable and uncompensable traits in semen. *Theriogenology* 70: 473-478.
41. Escofier B, Pagès J (1994) Multiple factor analysis (AFMULT package). *Comput Stat Data Anal* 18: 121-140.
42. Gower JC (1975) Generalized procrustes analysis. *Psychometrika* 40: 33-51.