Journal of Andrology, Vol. 27, No. 4, July/August 2006


# Two-Step Cluster Procedure After Principal Component Analysis Identifies Sperm Subpopulations in Canine Ejaculates and Its Relation to Cryoresistance

IVAN NÚÑEZ MARTÍNEZ, JOSE MARÍA MORÁN, AND FERNANDO J. PEÑA

*From the Veterinary Teaching Hospital Service of Reproduction and Obstetrics Faculty of Veterinary Medicine, University of Extremadura Avd de la Universidad s/n Cáceres, Spain.*

**ABSTRACT:** A 2-step clustering procedure, using indexes derived from principal component analysis, was used to disclose sperm subpopulations within the canine ejaculate and its relationship to sperm cryoresistance. Semen from 4 dogs was frozen-thawed by a standard protocol: before freezing, computer-assisted sperm analysis of motility and morphometry were performed; after thawing, motility analysis was performed again; and cryoresistance was estimated as the percent changes in progressive motility and sperm velocities after thawing. We used indexes derived from principal component analysis (sperm velocity index [SVI] and sperm motility index [SMI]) and the SPSS 2-step cluster method to disclose sperm subpopulations. The 2-step clustering procedure revealed the existence of 6 subpopulations. Subpopulations 4 and 6 were characterized by high values of both SVI (>200 arbitrary units) and SMI (>90 arbitrary units), subpopulations 2 and 3 were characterized by medium values (SVI 100 to 130; and SMI 30 to 40), and subpopulations 1 and 5 were characterized by low values (SVI < 100; SMI < 30). The distribution of sperm subpopulations was completely different among dogs. Four sperm subpopulations based in morphometric parameters of the sperm head and midpiece were revealed. Models including SVI and SMI indexes explained curvilinear velocity ($R^2$ = 0.997; $P < .001$), straight-line velocity ($R^2$ = 0.98; $P < .001$), and average velocity ($R^2$ = 0.99; $P < .0001$) postthaw.

Key words: Dog, computer-assisted sperm analysis.

**J Androl 2006;27:596–603**

The identification of sperm subpopulations within the mammalian ejaculate has become an issue of utmost interest for the sound evaluation of the ejaculates. The classical approach considering the whole ejaculate as a homogeneous population with a normal statistics distribution and the use of mean values to classify the ejaculates or to assess the effect of a treatment or a biotechnological procedure is nowadays considered erroneous (ESHRE Andrology Special Interest Group, 1998; Mortimer, 2000). The development of computer-assisted sperm analysis (CASA) systems has allowed a means to obtain high amounts of objective information of the kinematics and morphometric characteristics of a semen sample (Verstegen et al, 2002). Despite this obvious advantage, the high number of sperm variables and the lack of information regarding the biological meaning, and thus the practical importance of each, make using the obtained information difficult. In an attempt to improve the information obtained by CASA, a number of recent works have used clustering procedures to investigate the sperm subpopulation structure within the mammalian ejaculate (Abaigar et al, 1999; Chantler at al, 2004; Martinez-Pastor et al, 2005). These authors utilized sophisticated statistical techniques with specific software or multistep statistics. These works gave relevant new information on the biological characteristics of the mammalian ejaculate; however, the procedure implied the use of specific software or complicated multistep statistical techniques. Recently, a new, simple clustering procedure has been incorporated to the latest versions of the SPSS software. The SPSS TwoStep cluster method is a scalable cluster analysis algorithm designed to handle very large data sets, as are those generated after CASA. It requires only 1 data pass and has 2 steps: 1) precluster the cases into many small subclusters, and 2) cluster the subclusters resulting from precluster step into the final number of clusters. The precluster step uses a sequential clustering approach. It scans the data records one by one and decides if the current record should be merged with the previously formed clusters or starts a new cluster based on the distance criterion. The procedure is implemented by constructing a modified cluster feature (CF) tree. The

Supported by a grant from The Ministry of Education and Science of Spain AGL 2004-01722 (GAN). Dr Núñez Martinez holds a PhD grant from Valhondo Calaaf Foundation, Cáceres, Spain.

Correspondence to: Dr F. J. Peña, Section of Reproduction and Obstetrics Department of Herd Health and Medicine Faculty of Veterinary Medicine Avd de la Universidad s/n 10071, Cáceres, Spain (e-mail: fjuanpvega@unex.es).

Received for publication October 3, 2005; accepted for publication March 12, 2006.

DOI: 10.2164/jandrol.05153

CF tree consists of levels of nodes, and each node contains a number of entries. An entry in the leaf node represents a final subcluster. The nonleaf nodes and their entries are used to guide a new record quickly into a correct leaf node. Each entry is characterized by its CF that consists of the entry's number of records, mean and variance of each continuous variable, and counts for each category of each categorical variable. For each successive record, starting from the root node, it is recursively guided by the closest entry in the node to find the closest child node and descends along the CF tree. Upon reaching a leaf node, it finds the closest leaf entry in the leaf node. If the record is within a threshold distance of the closest leaf entry, it is absorbed into the leaf entry and the CF of that leaf entry is updated. Otherwise, it starts its own leaf entry in the leaf node. If there is no space in the leaf node to create a new leaf entry, the leaf node is split in half. The entries in the original leaf node are divided into 2 groups with the farthest pair as seeds, and the remaining entries are redistributed based on the closeness criterion. If the CF tree grows beyond its allowed maximum size, it is rebuilt based on the existing CF tree by increasing the threshold distance criterion. The rebuilt CF tree is smaller and hence has space for new input records. This process continues until a complete data pass is finished. All records falling in the same entry can be collectively represented by the entry's CF. When a new record is added to an entry, the new CF can be computed from this new record and the old CF without knowing the individual records in the entry. These properties of CF make it possible to maintain only the entry CFs rather than the sets of individual records. Hence the CF tree is much smaller and more able to be stored in main memory. An optional outlier-handling step is implemented in the algorithm in the process of building the CF tree. Outliers are considered as data records that do not fit well into any cluster. Before rebuilding the CF tree, the procedure checks for potential outliers and sets them aside. After rebuilding the CF tree, the procedure checks to see if these outliers can fit in without increasing the tree size. At the end of CF tree building, small entries that cannot fit in are outliers.

The second step (cluster step) takes subclusters (nonoutlier subclusters if outlier handling is used) resulting from the precluster step as input and then groups them into the final number of clusters. SPSS uses the agglomerative hierarchical clustering method (www.rrz.uni-hamburg.de/RRZ/Software/SPSS/Algorith.120/twostepcluster.pdf ).

This system allows the user to fix the previous maximum number of clusters or let the technique automatically choose the number of clusters with either the Bayesian information criterion or Akaike information criterion. The great advantage of this system is that all the analyses can be done in 1 step more rapidly and easily than with the techniques used so far to disclose sperm subpopulations.

Another problem derived from the analysis of the data originating from the CASA-ASMA analysis is the large variety of separated parameters that in most cases are highly correlated. When there are several correlated variables in biological systems, principal component analysis (PCA) can be used to reduce these to 1 or 2 variables that are linear functions of the original variables (Agarwal et al, 2003). Therefore, the aims of this study were to:

- Develop sperm indexes from PCA of the CASA data that can resume all the original variables
- Use a new simple clustering method to disclose sperm subpopulations
- Evaluate if the study of sperm subpopulations can be used to test the cryoresistance of a semen sample

## Material and Methods

### Animals

Four privately owned mongrel dogs (weight 15 to 20 kg; age 2 to 6 years) were used. All dogs were of known fertility.

### Experimental Design

Five ejaculates were obtained from each dog once a week for 5 consecutive weeks. After collection, an aliquot was removed for computer-assisted evaluation of motility and morphometry. After 4 weeks of storage, semen was thawed and sperm motility was evaluated again.

### Semen Collection

Semen was collected by masturbation in a prewarmed graduated test tube. After collection, sperm samples were kept at 37°C in a water bath. An aliquot was removed for sperm concentration measurement and evaluation of motility and morphology (phase contrast, microscopy). Only samples with at least 70% motility and 80% normal morphology were included in the study.

### Semen Processing

Semen was processed by the Uppsala method (Peña and Linde Forsberg, 2000) with modifications. In brief, after collection semen was diluted 1:1 in Tris-glucose extender I (200 mM Tris, glucose 70 mM, citric acid 63 mM, bovine serum albumin 3%, penicillin 1000 UI/mL, dihydroestreptomicin 1 mg/mL) and centrifuged for 8 minutes at 700 × *g*. The seminal plasma was then removed and the sperm pellet was resuspended in extender II at room temperature (200 mM Tris, 70 mM glucose, 63 mM citric acid, glycerol 3%, vol/vol egg yolk 20%, vol/vol penicillin 1000 UI/mL, dihydroestreptomicin

1 mg/mL) and cooled to 5°C over a period of 1 hour. After equilibration, an equal volume of Extender III (200 mM Tris, 70 mM glucose, 63 mM citric acid, 7% vol/vol glycerol, 20% vol/vol egg yolk, 1% vol/vol Equex STM paste [Nova Chemical Sales Inc, Sciutate, Mass] penicillin dihydroestreptomicin) was added at 5°C to a final sperm concentration of 150 to 200 $\times$ $10^6$ spermatozoa per milliliter. The sperm was then loaded in 0.5-mL straws and frozen horizontally in racks, placed 4 cm above of the surface of liquid nitrogen in a closed Styrofoam box for 10 minutes, and then plunged directly in liquid nitrogen. After 4 weeks of storage, the straws were thawed in a water bath at 70°C for 8 seconds.

### Motility Analysis

Motility was measured in fresh samples and after cryopreservation by using a CASA system (ISAS, Proiser SL, Valencia, Spain). Analysis was based on the examination of 25 consecutive, digitalized images obtained from a single field with a 10$\times$ negative-phase contrast objective. Images were taken with a time lapse of 1 second; the image capture speed was therefore 1 every 40 milliseconds. The number of objects incorrectly identified as spermatozoa was minimized on the monitor by using the playback function. With respect to the setting parameters for the program, objects with a curvilinear velocity less than 10 μm/s were considered immobile, whereas objects with a velocity greater than 15 μm/s were considered motile. Objects with velocities between 65 and 100 μm/s were understood as medium-speed objects, whereas those with a velocity greater than 100 μm/s were considered rapid objects. Spermatozoa deviating less than 10% from a straight line were designated linear motile. Sperm motion kinematics measured by CASA included the following:

| | | |
|---|---|---|
| Curvilinear velocity (VCL) | μm/s | Measures the sequential progression along the true trajectory |
| Straight-line velocity (VSL) | μm/s | Measures the straight trajectory of the spermatozoa per unit time |
| Average velocity (VAP) | μm/s | Measures the mean trajectory of the spermatozoa per unit time |
| Linearity coefficient (LIN) | % | VSL/VCL $\times$ 100 |
| Straightness coefficient (STR) | % | VSL/VAP $\times$ 100 |
| Wobble coefficient (WOB) | % | VAP/VCL $\times$ 100 |
| Average lateral head displacement (ALH) | μm | Measures the mean head displacement along the curvilinear trajectory |

### Sperm Staining for Morphometric Analysis

Before ASMA analysis, sperm samples were washed once in phosphate-buffered saline (PBS) by centrifugation at 800 $\times$ $g$ for 5 minutes and were adjusted in PBS to 100 $\times$ $10^6$ cells per milliliter. Then, 10 μL of the sperm suspension was placed on the edge of a slide and extended. Preparations were allowed to dry and were then fixed and stained for 10 minutes in an eosin solution (Panreac, Barcelona, Spain) and for 10 minutes in a methylene blue solution (Panreac). The excess of staining was removed, and the slide was allowed to dry and was permanently mounted (Eukitt, Panreac)

### Computerized Morphometric Analysis

The prepared slides were examined with a Nikon Labophot microscope equipped with a 100$\times$ bright field objective and a 3.3$\times$ photo-ocular. The video signal was acquired with a Sony CCD AVC-D7CE video camera (Sony Corporation, Tokyo, Japan) interfaced with an ISAS CASMA system (Proiser SL). The array size of the video grabber was 512 $\times$ 512 $\times$ 8 bits providing digitized images of 262.144 pixels and 256 gray levels. Resolution of images was 0.083 μm per pixel in the horizontal and vertical axes.

Sperm cells were displayed on the monitor at equivalent brightness, and all the cells that did not present any overlap with debris or other cells were considered for analysis. From each sample, heads were captured and analyzed by the program as previously described (Buendia et al, 2002). After treatment of the images, some of the cells had to be discarded because of defective binarization as observed by false correspondence between the original image and its mask. Each sperm head was measured for 9 primary parameters (head area [A], $\mu m^2$; head perimeter [P], μm; head length [L], μm; head width [W] μm; percentage of the sperm head occupied by the acrosome; midpiece width [w], μm; midpiece area [a], μm; distance [d] between the major axes of the head and midpiece, μm; angle [θ] of divergence of the midpiece from the head axis) and 4 derived parameters of head shape (FUN1 [L/W], FUN2 [$4\pi A/P^2$], FUN3 [(L − W)/(L + W)], FUN 4 [$\pi LW/4A$]).

### Statistical Analysis

The data matrix consisted of 3446 observations for fresh semen and 5773 for frozen-thawed samples for the motility analysis, whereas 2361 individual spermatozoa were evaluated for the morphometric analysis.

The main objective of the analysis was to extract sperm subpopulations by using the data obtained from each dog by means of clustering procedures. The level of significance was set at $P < .05$. The first step was to perform a PCA of the data. The purpose of the first step was to derive a small number of linear combinations (principal components [PCs]) that retain as much of the information in the original variables as possible. This allows one to summarize many variables in few, jointly uncorrelated PCs. A good result is considered if we obtain a few PCs accounting for a high proportion of the total variance. The VARIMAX method with Kaiser normalization was used as a rotation method. The second step was to perform a 2-step cluster procedure with the sperm-derived indexes obtained after the PCA. To study the distributions of observations (individual spermatozoa) within dogs and ejaculates and within subpopulations, we used the analyses of variance and $\chi^2$ tests. Linear regression analyses were used to investigate relationships among sperm indexes and sperm velocities postthaw. For comparison of mean values in fresh and frozen-thawed semen, the GLM procedure followed by

Tukey's test was used. All analyses were performed with SPSS version 12.0 for Windows software.

### *Calculation of Sperm CASA-Derived Indexes*

We used a simplified version of the technique described by Agarwal et al (2003). We first performed a PCA and then weighed variables against their eigen vectors. Using this approach, we proposed the following formulas:

Sperm velocity index (SVI) = (VCL × 0.87) + (VSL × 0.76) + (VAP × 0.90) + (ALH × 0.92)

Sperm movement index (SMI) = (VSL × 0.59) + (VAP × 0.37) + (LIN × 0.95) + (STR × 0.89) + (WOB × 0.83)

Morphometric indexes:

Index 1 = (L × −0.636)+(FUN1 × 0.961)+(FUN2 × −0.550) + (FUN3 × 0.953)

Index 2 = (L × 0.716)+(A × 0.926)+(P × 0.961)+(FUN2 × −0.387)

Index 3 = (FUN4 × 0.475) + (d(c − f) × 0.839) + (A (c − f) × 0.790)

Index 4 = (w × 0.831) + (a × 0.790)

## *Results*

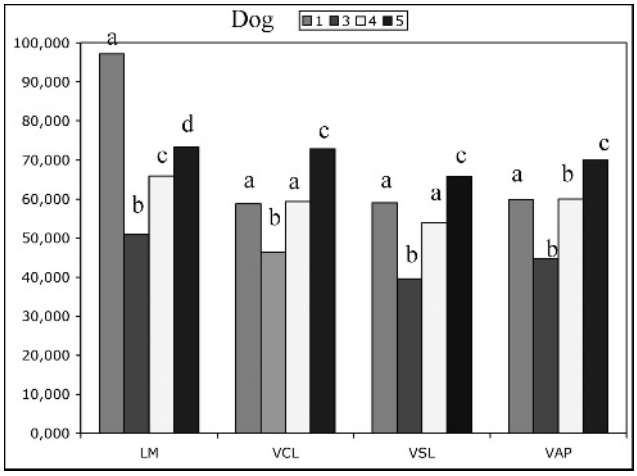
### *Freezeability of Each Dog Ejaculate*

Freezeability was defined as the percent change in the values of progressive motility, VCL, VAP, and VSL after cryopreservation (values postthaw divided by values in fresh samples multiplied by 100). Significant differences were found among dogs (Figure). Dog 1 showed the best results, whereas dog 3 had the worst results.

### *Development of Sperm Quality Indexes as Measures of Semen Quality*

PCA in CASA data revealed that 2 components accounted for more of the 91% of the cumulative variance. The first component was related to sperm velocities (SVI) and also included ALH. The second component was more related to the characteristics of the sperm movement itself (SMI). Because these scores are produced by PCA, SVI and SMI are not correlated and resume all the information obtained after the CASA. In this way, VCL, VSL, VAP, ALH, STR, WOB, and LIN are resumed in SVI and SMI.

For the morphometric values, 4 PCs accounted for more of the 70% of the variance. The first 2 PCs were related to sperm head size and shape, whereas PCs 3 and 4 were mainly related to characteristics of the midpiece. These 4 sperm indexes resumed 12 ASMA-derived sperm morphometric variables.




Freezeability of each dog. Freezeability was defined as the percent changes in each of the selected motility descriptors (values postthaw divided by values in fresh samples multiplied by 100), a-b-c-d $P < .05$. LM indicates progressive motility; VCL, curvilinear velocity; VSL, straight-line velocity; and VAP, average velocity (a-b-c-d $P < .01$).

### *Sperm Subpopulations Based on Kinematics Properties of the Sperm*

The 2-step clustering procedure revealed the existence of 6 subpopulations. Subpopulations 4 and 6 were characterized by high values of both SVI and SMI, subpopulations 2 and 3 were characterized by medium values, and subpopulations 1 and 5 were characterized by low values (Table 1). The distribution of sperm subpopulations was completely different among dogs (Table 2). Two sperm subpopulations were present in each dog: subpopulations 5 and 6 were in dog 1, subpopulations 3 and 4 were in dog 3, subpopulations 1 and 4 were in dog 4, and subpopulations 1 and 4 were in dog 5.

### *Sperm Subpopulations Based on Morphometric Properties of the Spermatozoa*

Four sperm subpopulations based on morphometric parameters of the sperm head and midpiece were revealed (Table 3). Also, the distribution of sperm subpopulations was completely different in each dog (Table 4).

### *Relationship Among Sperm-Derived Indexes and Sperm Quality Postthaw*

VSL, VAP, and VCL were used as indicators of sperm quality postthaw. Stepwise linear regression analyses using the sperm indexes as predictive variables were used to forecast these 3 sperm velocities separately. We choose these variables because they are widely considered to be good indicators of sperm quality. In addition, there are a number of studies showing a high value of

Table 1. *Results of the 2-step cluster procedure performed with sperm velocity index (SVI) and sperm motility index (SMI) as variables*

| | SVI | | SMI | |
|---|---|---|---|---|
| Cluster | Mean | SD | Mean | SD |
| 1 | 82.8 | 51.25 | 24.9 | 16.19 |
| 2 | 135.4 | 83.33 | 41.8 | 26.33 |
| 3 | 130.7 | 77.45 | 40.0 | 26.11 |
| 4 | 295.2 | 47.05 | 104.6 | 20.41 |
| 5 | 86.3 | 47.12 | 26.7 | 14.55 |
| 6 | 275.9 | 54.00 | 94.1 | 23.74 |
| Outliers cluster | 481.0 | 62.71 | 178.9 | 23.88 |

VCL in the prediction of human sperm quality (de Geyter et al, 1998; Larsen et al, 2000) and field fertility in animal models (Holt et al 1997). Thus, the maintenance of high velocities postthaw may be a good indicator of the success of cryopreservation.

The 2 models explained VCL postthaw; however, the residuals sum of squares was much lower in model 2. This model had $R^2 = 0.997$, which explained the 99% in the variation. The model included a constant, SVI, and SMI (Table 5) and was statistically significant ($P < .001$).

Two models were developed for VSL and VAP; the ones including both SVI and SMI index were elected because of the lower values of the residual sum of squares. However, this was much bigger in the case of VSL than VAP.

The model predicting values of VAP had $R^2 = 0.98$ ($P < .001$), and the model predicting values of VSL had $R^2 = 0.99$ ($P < .0001$) (Tables 6 and 7).

## Discussion

Sperm motility is measured to provide information about the likely fertility of an individual or to assess the viability of a semen sample after liquid storage or cryopreservation (Holt, 1996). In these situations, motility has to be considered in 3 ways: 1) the percentage of cells showing movement, 2) the characteristics of the movement itself, and 3) the speed of the movement. Although sperm motility evaluation is normally one of the first steps of the spermiogramme (Rodriguez Martinez, 2003), the amount of predictive information given, even using computerized systems, has been scarce and even disappointing. In spite of this, recent evidence suggests that the pitfall in the predictive value of motility relies on the use of the information given and not on the intrinsic value of the motility itself (Abaigar et al, 1999; Holt and Van Look, 2004; Martinez-Pastor et al, 2005). Growing consensus exists within the scientific community regarding the existence of sperm subpopulations within the mammalian ejaculate (Abaigar et al, 1999; Quintero-Moreno et al, 2003). Rather than a homogenous population in which all the spermatozoa show the same behavior, the mammalian ejaculate is a heterogeneous group of different sperm subpopulations showing different responses to physiological or biotechnological stimuli (Abaigar et al, 1999; Martinez-Pastor, 2005), motility patterns (Quintero-Moreno et al, 2003), and even different morphometric characteristics (Thurston et al, 2001; Peña et al, 2005).

From this evidence, it is becoming clear that the classical approach considering the ejaculate as a homogeneous population following a normal statistical distribution is no longer valid.

In this study, we used simple indexes resuming all the variables of the CASA for morphometric and kinematics data separately, and we used these data to disclose sperm subpopulations. Using this approach, we were

Table 2. *Distribution of each cluster within each dog, with absolute number of spermatozoa within each cluster and relative percentages*

| | Dog 1 | | Dog 3 | | Dog 4 | | Dog 5 | |
|---|---|---|---|---|---|---|---|---|
| Cluster | n | % | n | % | n | % | n | % |
| 1 | 0 | 0 | 0 | 0 | 375 | 61.1 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 510 | 81.3 |
| 3 | 0 | 0 | 527 | 90.1 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 55 | 9.4 | 239 | 39.8 | 113 | 18.0 |
| 5 | 242 | 45.2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 293 | 54.8 | 0 | 0 | 0 | 0 | 0 | 0 |
| Outliers | 0 | 0 | 3 | 0.5 | 0 | 0 | 4 | 0.6 |

Table 3. *Results of the 2-step cluster procedure with the morphometric indexes as variables*

| | Index 1 | | Index 2 | | Index 3 | | Index 4 | |
|---|---|---|---|---|---|---|---|---|
| Cluster | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | −2114.9 | 116.33 | 31.0 | 2.06 | 7.6 | 6.90 | 33.6 | 1.18 |
| 2 | −2225.9 | 117.27 | 32.2 | 1.95 | 8.1 | 6.41 | 33.7 | 1.71 |
| 3 | −2203.8 | 117.32 | 31.6 | 1.97 | 6.5 | 5.76 | 33.9 | 1.50 |
| 4 | −2229.4 | 114.20 | 32.0 | 1.98 | 7.7 | 5.99 | 35.3 | 1.52 |
| Outliers | −2228.3 | 260.47 | 32.1 | 5.75 | 34.1 | 26.45 | 30.1 | 1.73 |

Table 4. *Distribution of each morphometric-derived cluster within each dog, with absolute number of spermatozoa within each cluster and relative percentages*

| | Dog 1 | | Dog 3 | | Dog 4 | | Dog 5 | |
|---|---|---|---|---|---|---|---|---|
| Cluster | n | % | n | % | n | % | n | % |
| 1 | 0 | 0 | 581 | 99.3 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | | 0 | 616 | 98.2 |
| 3 | 0 | 0 | 0 | 0 | 607 | 98.9 | 0 | 0 |
| 4 | 532 | 99.4 | 0 | 0 | 0 | 0 | 0 | 0 |
| Outliers | 3 | 0.6 | 4 | 0.7 | 7 | 1.1 | 11 | 1.8 |

Table 5. *Variables in the model obtained from regression analysis showing the relation among sperm indexes and the curvilinear velocity of sperm postthaw**

| Variable | B | SE B | Beta | T | Sig T |
|---|---|---|---|---|---|
| Constant | 2.438 | 0.10 | ··· | 24.17 | 0.0001 |
| SVI | 0.906 | 0.002 | 2.03 | 412.99 | 0.0001 |
| SMI | −1.333 | 0.006 | −1.09 | −221.45 | 0.0001 |

* SVI indicates sperm velocity index; SMI, sperm motility index.

Table 7. *Variables in the model obtained from regression analysis showing the relation among sperm indexes and the straight-line velocity of sperm postthaw**

| Variable | B | SE B | Beta | T | Sig T |
|---|---|---|---|---|---|
| Constant | −2.090 | 0.117 | ··· | −17.85 | 0.0001 |
| SVI | −0.199 | 0.003 | −0.522 | −77.983 | 0.0001 |
| SMI | 1.565 | 0.007 | 1.499 | 223.95 | 0.0001 |

* SVI indicates sperm velocity index; SMI, sperm motility index.

able to identify 6 sperm subpopulations by kinematics data and 4 sperm subpopulations by morphometric data. Given the new information regarding the quality of individual ejaculates, the distribution of clusters varied significantly among dogs. Classical analysis of sperm data by using mean values did not reveal differences among dogs.

Also, the approach used in our study allowed us to predict the freezeability of a semen sample. We used as a measure of freezeability the changes in sperm velocities after freeze-thaw procedures, for it is assumed that these parameters are a manifestation of functional competence of spermatozoa and they are normally positively correlated with those of plasma membrane and sperm morphology (Peña, 2004). Also, a number of studies show that sperm velocities are sound indicators of sperm quality (de Geyter et al, 1998; Larsen et al, 2000). The predominance of a particular subpopulation in samples with better cryoresistance or fertility has been demonstrated in other studies (Quintero-Moreno et al, 2003; Martinez-Pastor et al, 2005; Peña et al, 2005).

The results from our study strongly support the hypothesis that the mammalian ejaculate is a heterogeneous cellular suspension and that the erroneous use of data generated by the CASA was attributed to frequently disappointing results when motility was used as an indicator of sperm quality (Abaigar et al, 1999; Quintero-Moreno et al, 2003).

It is clear that not all the spermatozoa in an ejaculate have the same biochemical status and thus respond in a different manner to different stimuli (Abaigar et al, 1999). It is also clear that not all the spermatozoa have the same morphometry (Peña et al, 2005), and thus the cell surface area is different (Mazur and Koshimoto, 2002),

Table 6. *Variables in the model obtained from regression analysis showing the relation among sperm indexes and the mean velocity of sperm postthaw**

| Variable | B | SE B | Beta | T | Sig T |
|---|---|---|---|---|---|
| Constant | −1.21 | 0.18 | ··· | −6.53 | 0.0001 |
| SVI | 0.35 | 0.004 | 0.918 | 87.92 | 0.0001 |
| SMI | 0.08 | 0.01 | 0.078 | 7.426 | 0.0001 |

* SVI indicates sperm velocity index; SMI, sperm motility index.

thereby affecting the exchanges of heat, water, and ions during cryopreservation in a different manner in each subpopulation (Peña et al, 2005). Clearly, the subpopulations structure approach has advantages over the use of mean values for the study of the ejaculate; the former identifies subpopulations of spermatozoa in different biochemical status, whereas the use of mean values just masks it (Abaigar et al, 1999) and also can identify morphometric subpopulations with a cell surface area more adequate to support the stress of cryopreservation (Peña et al, 2005). Also, the use of combined indexes by PCA can better reflect the energetic status of each subpopulation. In our study, SMI and SVI were able to predict sperm velocities postthaw. The statistical technique is advantageous because of its simplicity; whereas previously described techniques imply 2 or 3 steps (Abaigar et al, 1999; Martinez-Pastor et al, 2005), all the clustering procedures here are performed in 1 step. Also, the system manages outliers automatically. This statistical tool proved to be very powerful when disclosing sperm subpopulations despite the relative low number of dogs used in this study. Recent studies (Rathi et al, 2001; Quintero-Moreno et al, 2003) have suggested that studies with a relatively low number of animals can produce very interesting proposals for the study of the mammalian ejaculate if the data are adequately processed.

Interesting findings of our study are those related to the predictability of sperm cryoresistance. As stated earlier, median values of sperm kinematic parameters did not differ among dogs; however, the pattern in the distribution of sperm subpopulations varied. It is noteworthy to remark that this interpretation would not have been possible without considering sperm subpopulations. The dog with the worst ability to support cryopreservation (dog 3) had predominated sperm subpopulations with low values of SVI and SMI in fresh samples. On the other hand, the dogs showing the best cryoresistance (dogs 1 and 5) had predominated sperm subpopulations with higher values of SMI and SVI. Also, it is noteworthy that dog 1 showed higher cryoresistance. In this dog, there was only a slight decrease in the percentage of progressive motile sperm after freeze-thaw (Figure) and a clear predominance of kinematic cluster 6, characterized by the highest values

of SVI and very high values of SMI. It is plausible that these sperm indexes resume the biological status of sperm cells. It is interesting to note that different sperm velocities (VCL, VAP, VSL) are regulated in a different manner, as demonstrated by different tyrosine kinase inhibitors (Bajpai and Doncel, 2003). Therefore, these indexes may be a reliable indicator of the biochemical status of the sperm subpopulation and thus may be used to predict cryoresistance. This may be especially relevant in species showing a relatively high cryoresistance, such as dogs. In fact, from our study this theory may explain the high predictive value of this approach. By using the PCA-derived indexes, we can resume in few variables the complex structure and biochemical status of the ejaculate. Those ejaculates with a predominance of subpopulations with high SMI values will show higher sperm velocities after freezing. If we evaluate what happened in dog 5 (the dog with higher sperm velocities after freezing) kinematic clusters 2 (mid to high values of SVI and SMI) and 4 (very high values of SMI and SVI) predominated. It is very plausible to think that most of the sperm not surviving the freeze-thaw procedure belonged to cluster 2; thus, in this dog there was not a marked decrease in sperm velocities after thawing. Following this reasoning, we can also explain why dog 3 had the worst cryoresistance; in this dog, more than the 90% of spermatozoa belonged to a cluster with low values of SVI and SMI. It is very important to note that when using mean values of sperm motility we were unable to detect statistically significant differences among dogs. Cryopreservation implies much stress to the spermatozoa (Mazur, 1984). Perhaps the 2 main types of stress are osmotic stress and the formation or reshaping of intracellular ice during freezing and again during thawing. In more sensitive species such as pigs, it may be plausible that most of the cryoinjury is physical, whereas in most cryoresistant species such as dogs, the biochemical injury may be more present in the form of subtle damages of the sperm membranes, resulting in alterations of the cell physiology. Our results support this theory, because rather than a high percentage of inmotile sperm after thawing we observed mainly decreases in the percentage of progressive motile spermatozoa and decreases in all sperm velocities.

In conclusion, we have demonstrated that a simple 2-step cluster procedure can be used to disclose sperm subpopulations within a semen sample. This system is easier than the previously used statistical procedures because it allows 1-step entry of data and the automatic management of outliers. Also, we have demonstrated that the use of PCA-derived indexes can be used as variables to enter in the cluster procedure, retaining most of the information of the original variables and much more useful information regarding the characteristics of the ejaculates and its freezeability than provided by conventional semen analysis. Finally, we propose the use of a multivariate clustering procedure as standard analysis of semen samples when CASA systems are used.

## References


Abaigar T, Holt WV, Harrison RAP, del Barrio G. Sperm subpopulations in boar (*Sus scrofa*) and gazelle (*Gazella dama mhorr*) semen as revealed by pattern analysis of computer assisted motility assessments. *Biol Reprod*. 1999;60:32–41.

Agarwal A, Sharma RK, Nelson DR. New semen quality scores developed by principal component analysis of semen characteristics. *J Androl*. 2003;24:343–352.

Bajpai M, Doncel GF. Involvement of tyrosine kinase and camp dependent kinase cross-talk in the regulation of human sperm motility. *Reproduction*. 2003;126:183–195.

Buendia P, Soler C, Paolicchi F, Gago G, Urquieta B, Pérez Sánchez F, Bustos Obregón E. Morphometric characterization and classification of alpaca sperm heads using the sperm class analyzer system. *Theriogenology*. 2002;57:1207–1218.

de Geyter Ch, de Geiter M, Koppers B, Nieschlag. Diagnostic accuracy of computer assisted sperm motion analysis. *Hum Reprod*. 1998;13:2512–2520.

ESHRE Andrology Special Interest Group. Guidelines on the application of CASA technology in the analysis of spermatozoa. *Hum Reprod*. 1998;13:142–145.

Holt C, Holt WV, Moore HDM, Reed HCB, Curnock RM. Objectively measured boar sperm motility parameters correlate with the outcomes of on farm inseminations: results of two fertility trials. *J Androl*. 1997;18:312–323.

Holt W. Can we predict fertility rates? Making sense of sperm motility. *Reprod Domest Anim*. 1996;31:17–24.

Holt WV, Van Look KJW. Concepts in sperm heterogeneity, sperm selection and sperm competition as biological foundations for laboratory tests of semen quality. *Reproduction*. 2004;127:527–535.

Larsen L, Scheike T, Kold Jensen L, Peter Bonde J, Ernst E, Henrik Hjollund N, Zhou Y, Erik Skakkebaek N, Giwercman A. Danish first pregnancy planner study team. Computer assisted semen analysis parameters as predictors for fertility of men from the general population. *Hum Reprod*. 2000;15:1562–1567.

Martinez-Pastor F, García-Macias V, Alvarez M, Herráez P, Anel L, de Paz P. Sperm subpopulation in Iberian red deer epididymal sperm and their changes through the cryopreservation process. *Biol Reprod*. 2005;72:316–327.

Mazur P. Freezing of living cells: mechanism and implications. *Am J Physiol (Cell Physiol)*. 1984;16:C125–C142.

Mazur P, Koshimoto CH. Is intracellular ice formation the cause of death of mouse sperm frozen at high cooling rates? *Biol Reprod*. 2002;66:1485–1490.

Mortimer ST. CASA-practical aspects. *J Androl*. 2000;21:514–524.

Peña A, Linde-Forsberg CB. Effect of spermatozoal concentration and post-thaw dilution rate on survival after thawing of dog spermatozoa. *Theriogenology*. 2000;54:703–718.

Peña AI. Canine fresh and cryopreserved evaluation. *Anim Reprod Sci*. 2004;82/83:209–224.

Peña FJ, Saravia F, García Herreros M, Nuñez-Martinez I, Tapia JA, Johannisson A, Wallgren M, Rodriguez Martinez H. Identification of sperm morphometric subpopulations in two different portions of the boar ejaculate and its relation to post thaw quality. *J Androl*. 2005;26:716–723.

Quintero-Moreno A, Miró J, Rigau T, Rodríguez Gil JE. Identification of sperm subpopulations with specific motility characteristics in stallion ejaculates. *Theriogenology*. 2003;59:1973–1990.

Rathi R, Colenbrander B, Bevers BB, Gadella BM. Evaluation of in vitro capacitation of stallion spermatozoa. *Biol Reprod*. 2001;65:462–470.

Rodriguez Martinez H. Laboratory assessment and prediction of fertility: still utopia? *Reprod Domest Anim*. 2003;38:312–318.

Verstegen J, Iguer-Ouada M, Onclin K. Computer assisted semen analyzers in andrology research and veterinary practice. *Theriogenology*. 2002;57:149–179.