RELIABILITY ESTIMATES FOR A TEST OF SPERM CAPACITATION, A FUNCTIONAL COMPLEMENT TO SEMEN ANALYSIS

G. Charles Ostermeier[1], Cristina Cardona[1], T Timothy Smith[2], Melissa A. Moody[1], Alana J Simpson[1] and Alexander J Travis[3,1]

[1]Androvia LifeSciences, Mountainside, NJ; [2]ARTWorks Consulting, New York, NY; [3]Cornell University, Ithaca, NY

Abstract

Introduction: Evaluations of semen quality are subjective. This can add unwanted variability into fertility assessments. To augment traditional semen evaluation, samples were evaluated for their ability to capacitate and become fertilization competent using localization patterns of the ganglioside G_{M1}, the basis of the laboratory−developed Cap−Score™ Sperm Function Test. Here, Cap−Score precision and its variation when determined by the same and different operators were assessed.

Methods: Following liquefaction of semen samples from consenting men, sperm were washed, incubated, fixed and then evaluated via fluorescence microscopy for G_{M1} localization patterns. Student's t−Test employing unequal variance was done using Microsoft Excel (2013).

Results: Precision was evaluated by comparing the percent change about Cap−Score values (%Δ=(y2−y1)/y2) when 50, 100, 150 and 200 sperm were evaluated. Changes in values of 11, 6 and 5% were observed for each addition of 50 sperm (n≥23). This supports the view that Cap−Score precision was only modestly improved by counting more than 100 sperm. To be conservative in our studies, Cap−Score was determined by counting the G_{M1} localization patterns of at least 150 cells. To assess variation within and between readers, 20 large image files containing up to 5,000 sperm each were generated. Two different readers were trained and determined Cap−Scores by randomly resampling the images 20 times, counting 150 cells each time. When scoring the same sample, individual readers reported an average SD of 3 Cap−Score units. The difference between readers when scoring the same sample ranged from 0.00 to 1.52, with an average difference of 1 between the readers for any given sample. Applying the Bonferroni correction, no difference between readers was observed for any image file (p−values ranged from 0.02 to 0.99). These data demonstrate that the same and independent readers can replicate Cap−Score values when repeatedly evaluating the same semen donor.

Conclusions: Common measures of semen quality are subjective and can vary within and among readers, making the assessment of male fertility challenging. The Cap−Score Sperm Function test evaluates the ability of sperm to capacitate, a necessity for male fertility. The data presented here show that the Cap−Score Sperm Function Test is highly reproducible and reliable within and between readers, which are key considerations when attempting to diagnose male infertility. Funded by Androvia LifeSciences.

Introduction

There are over 73 million infertile couples globally, with >40% of infertility having a male factor. Standard semen analysis includes the descriptive parameters of sperm count, motility, and morphology, and diagnoses approximately half the cases of male infertility. The other half have defects in sperm function and are only diagnosed by repeated failed cycles of IUI. Standard semen evaluations are often subjective in nature and can add unwanted variation to the diagnosis of male fertility. Here, we evaluated the variation of a new test for sperm function/fertilizing ability, the Cap-Score, within and between readers.

Results

Table 1. Percent change in Cap-Score with increasing number of counted sperm.

# of counted sperm	Mean % Δ	STDEV	# of Obs	SEM	95% CI LL	95% CI UL
from 50 to 100	11%	9%	23	2%	7%	14%
from 100 to 150	6%	5%	26	1%	4%	8%
from 150 to 200	5%	3%	26	1%	4%	6%

%Δ=$(y_2-y_1)/y_2$ where y_2 and y_1 are the larger and smaller Cap-Scores respectively, with either the upper or lower # of counted sperm; STDEV-standard deviation; obs - observations; SEM-Standard error of the mean; CI-Confidence Interval; LL-lower limit; UL-Upper limit

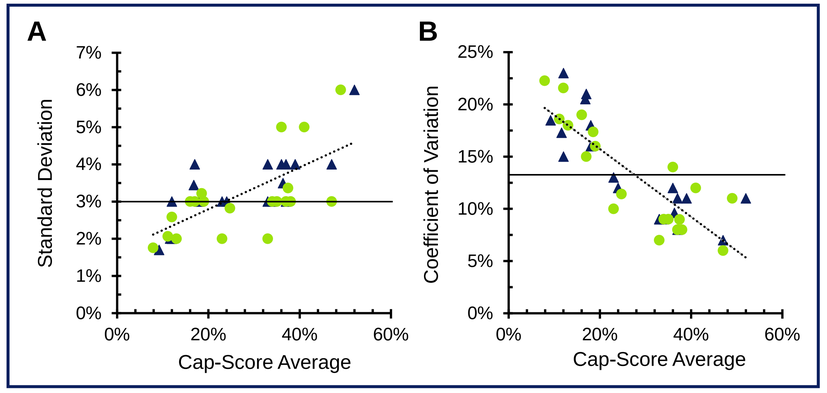

Figure 1. ***Cap-Score readings are tightly clustered about the true value.*** 20 large image files, containing up to 5,000 sperm each, were generated. Two different readers were trained (reader 1 green circles, reader 2 blue triangles) and determined Cap-Scores by randomly resampling each image 20 times, counting 150 cells each time. The average Cap-Score is shown on the x-axis and the corresponding Standard Deviation (SD; **Panel A**) and Coefficient of Variation (CoV=SD/mean; **Panel B**) are shown on the y-axes. The average SD and CoV for all images were found to be 3 and 13 and are shown by the solid horizontal lines. The dotted lines show the linear dependence of the SD (y = 0.06x + 0.02; r= 0.69; p=0.00) and CoV (y = -0.32x + 0.22; r=-0.84; p=0.00) to the Cap-Score average.

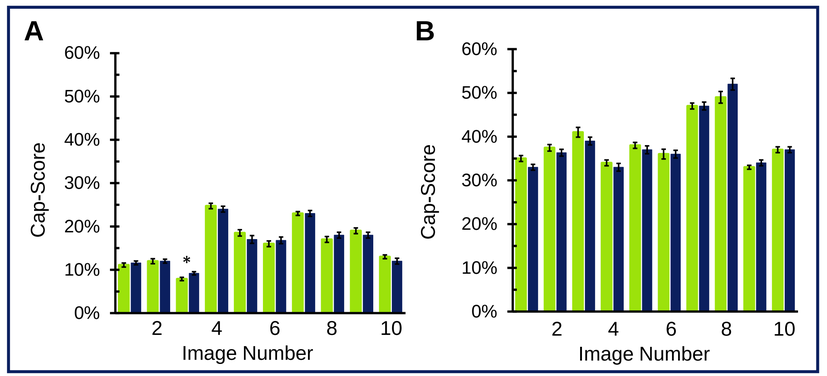

Figure 2. ***Reproducibility of mean Cap-Scores between operators.*** Ten stitched images, containing up to 5,000 sperm each, were obtained for "less than normal" (more than 1 SD below the mean Cap-Score result for a population of normal men; **Panel A**) and "presumed normal" (above 1 SD below the mean Cap-Score result for a population of normal men; **Panel B**) groups to evaluate whether reproducibility varied with higher or lower scores. Two different readers determined Cap-Scores by randomly resampling each image 20 times and counting 150 cells each time (reader 1 green bars, reader 2 blue bars). Mean Cap-Scores were not different between readers for any image file. The p-values from 2 sample T-Tests, comparing mean Cap-Score between readers, ranged from 0.02* to 0.99, with no difference being significant.

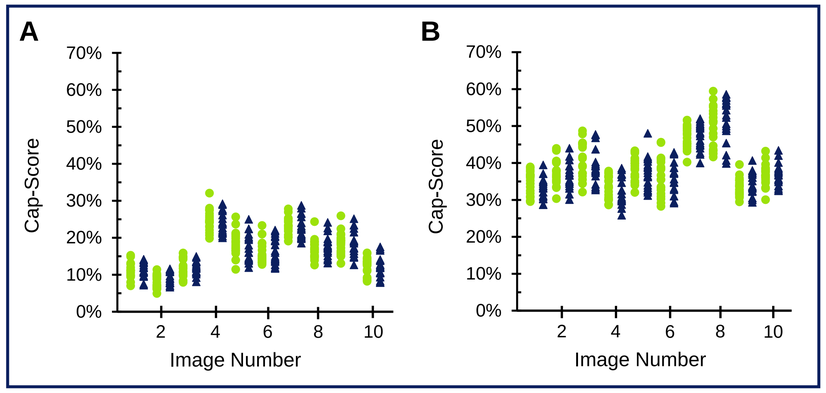

Figure 3. ***Repeatability of Cap-Score variances between operators.*** Ten stitched images, containing up to 5,000 sperm each, were obtained for "less than normal" (more than 1 SD below the mean Cap-Score result for a population of normal men; **Panel A**) and "presumed normal" (above 1 SD below the mean Cap-Score result for a population of normal men; **Panel B**) groups. Two different readers evaluated Cap-Scores by arbitrarily resampling each image 20 times, counting 150 cells each time (reader 1 green circles, reader two blue triangles). The variances in Cap-Score readings were not different between readers for any image file. The p-values for Bartlett's test of homoscedasticity ranged from 0.11 to 0.94.

Conclusions

- The Cap-Score is highly accurate within a sample; when the same population of sperm was randomly resampled by the same or different readers, Cap-Score values were tightly clustered.
- Cap-Score is highly reproducible between readers; two different readers obtained similar Cap-Scores for each of 20 populations of sperm.
- The data presented here show that the Cap−Score Sperm Function Test is highly reliable and reproducible within and between readers, which are key considerations for assays diagnosing male infertility.

Funding