

# Distinguishing Capacitation Status of Sperms with CASA Parameters: An Application of Decision Tree-based Methods

Yuyue Liao, Research & Pre-Clinical Statistics, Bayer AG

Workshop WG Non-Clinical Statistics, IBS-DR, November 16 & 17 2023

Capacitation is a step in the life cycle of sperms which is required for fertilization, and the capacitation rate is an indicator of the efficiency of non-hormonal contraceptives. In this study, we developed a method which uses computer-assisted sperm analysis (CASA) parameters to estimate capacitation rate in given populations. Eight CASA parameters are considered: 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 wobble (WOB).

We compared the distributions of the parameters between capacitated and incapacitated sperms based on a sample of more than 1100 sperms. Differences between the distribution of VAP, VSL, VCL, ALH, STR, and WOB are observed. These parameters and their combinations can be used to distinguish capacitated and incapacitated sperms.

After that, we applied decision tree-based methods, a diverse set of method in supervised machine learning, to distinguish capacitated and incapacitated sperms by values of their CASA parameters. The results of decision trees were used to create subsets of the sample that the presence of a sperm in the subset is highly correlated to its capacitation status.

We applied various algorithms based on decision trees to create these subsets. Methods with multiple decision trees allows the subset to include a larger proportion of capacitated sperms. On the other hand, the application of cost-sensitive decision trees allows the subset to have higher relative risk.

Of all models considered, a joint criterion of high VCL and high STR always plays a role in constructing subsets to distinguish capacitated and incapacitated sperms. With these constructed subsets, we can estimate the capacitation rate in a sample of sperms by calculating their proportions in the subsets. With these estimations, the effectiveness of non-hormonal contraception treatments can be evaluated and compared.