Evaluation of Male Reproductive Safety During Clinical Development

Malle Jurima-Romet, PhD, Senior Director, Drug Development Services

How is risk to the male reproductive system identified?

Novel drug candidates may present potential toxicity risks to the male reproductive system. A difficult issue faced by pharmaceutical sponsors developing new drugs is how to proceed safely with clinical studies in male participants and patients when animal data have indicated that there is a risk of testicular toxicity. Histopathological assessment of the testes in animal studies is the most sensitive method for detecting male reproductive toxicants. Testicular histopathological abnormalities may be observed in one or more animal species, however, there is no 'best' species for predicting human risk. Rodent fertility studies may provide additional understanding of the consequences of histopathological changes in the testes; although, testicular abnormalities in any species will be viewed by toxicologists and regulatory agencies as a concern which will need to be addressed to allow clinical development to proceed safely.

Which testicular functions can be adversely affected and what tools are available to evaluate safety?

Testicular functions that may be adversely affected by a new drug candidate include spermatogenesis and Leydig cell function. The latter can be monitored by serum testosterone and luteinizing hormone (LH), which can be readily incorporated into a clinical study. However, testosterone and LH are poorly correlated with spermatogenesis. Follicle stimulating hormone (FSH), produced in the pituitary, is responsible for stimulating spermatogenesis. Serum FSH is elevated when spermatogenesis is markedly impaired, although, levels are variable and it is not a sensitive biomarker. Inhibin B is produced only in the testes, predominantly by the Sertoli cells, and controls FSH secretion by a negative feedback loop. Serum inhibin B levels have been shown to reflect the overall functional integrity of spermatogenesis in the testes, and inhibin B is considered by some investigators to be a promising biomarker of testicular toxicity (Stewart and Turner 2005). Serum inhibin B is relatively insensitive and

has not been adequately validated in clinical studies. Substantial decreased serum inhibin B levels are found only in association with severe testicular damage where there is depletion of both spermatids and spermatocytes, such as following chemotherapy or testicular irradiation. In general, spermatogenesis is more sensitive to toxic effects of drugs or chemicals than the reproductive endocrine system, and endocrine effects are rarely observed. Physical examination of the testes is of very limited value in safety evaluation and any changes in testicular size would indicate that severe damage has already occurred. Testicular biopsy, due to its invasiveness, is not an acceptable form of monitoring the status of spermatogenesis in the clinical setting. Semen analysis remains the best measure of spermatogenesis despite several challenges involved in incorporating this type of testing in clinical studies.

Semen Analysis in Clinical Studies – Challenges and Designing an Appropriate Study

Study Duration

In humans, the spermatogenesis cycle in the testes is 72 days and it takes another 10-15 days for sperm to reach the ejaculate. Therefore, a minimum of 90 days post dosing is necessary to compare baseline and post-treatment semen analysis data to determine if there was an adverse drug effect. Semen analysis conducted before 60 days does not provide any meaningful testicular safety data. Due to this requirement for a relatively long duration of safety monitoring, the study should be adequately powered to account for dropouts.

Reducing Variability

The main challenge with semen analysis is the large variability in the data: Within-participant coefficients of variation (CVs) in sperm count can be as high as 50%; between-participant CVs are much higher (Keel, 2006). Therefore, the goal in designing a testicular safety study is to reduce variability as much as possible. While in

theory, semen analysis may seem like a simple test to perform, in practice, it requires a great deal of technical expertise and care, and semen evaluation methodology is difficult to standardize between labs. To obtain reliable data from multi-center studies, comparable study populations (i.e. age range, health status) should be used and strict quality control protocols applied for the semen analysis procedures. Standardization of techniques with rigorous training of technicians at a central location has been shown to decrease variation in semen parameters (Brazil et al. 2004). It has been Celerion's experience that semen analysis conducted on site in the clinic by the same team of trained technicians, even for multi-center studies, will produce less variable semen analysis data than using more than one semen analysis lab and/or team of technicians.

Intra-individual CV in sperm concentration has been shown to decrease with increasing number of semen samples collected from the same participant. For example, in one study, the CV was 41% when two samples were analyzed, and decreased to 34% when three samples were analyzed (Carlsen et al. 2004). One semen sample per evaluation timepoint is not sufficient to obtain reliable semen analysis data. Generally, three semen samples are considered adequate. Ejaculatory frequency also contributes to variability in semen analysis data. Sperm concentration and semen volume increase with increasing duration of abstinence whereas sperm motility may decrease; sperm morphology is unaffected. Participants enrolled into a testicular safety evaluation study should be instructed to abstain for a minimum of 48 hours and maximally five days prior to baseline semen collection. Ideally, triplicate semen samples collected at each evaluation timepoint should be collected 48 hours apart with no ejaculations in between sample collection. Confinement of participants to the clinic throughout the period of semen collection can help to reduce non-compliance. Seasonal variation and geographical differences in semen data have been reported. A study comparing semen analysis data across different areas of the United States revealed that sperm concentration and motility was reduced in agricultural and semi-rural areas compared to urban areas (Swan et al. 2003). Exposure to agricultural chemicals may be a factor in reduced semen quality in agricultural and semi-rural areas. Therefore, when planning multi-site studies across different geographies, these sources of variation should be considered and their contribution minimized as much as possible.

Standardization of laboratory techniques is critical for reducing variability in semen analysis data. Although the World Health Organization (WHO) standards are recognized worldwide as the gold standard for semen testing, in practice, many laboratories are not familiar with these recommendations or may disregard them (Keel, 2004). Proficiency testing is a process of external, interlaboratory quality control in which identical semen samples are tested by participating laboratories, and the test results are compared with the collective performance of all participating labs. Sponsors planning to conduct a testicular safety evaluation study should ensure that the andrology lab(s) performing the semen analysis participates in a proficiency testing program, with an up-to-date accreditation by the College of American Pathologists (CAPA), or other accreditation-granting organization which is recognized in the site country's jurisdiction.

Standardization of equipment used for semen analysis is also important. Even the type of counting chamber used for sperm counts can contribute to inter-lab variability. For example, sperm concentrations determined by the MicroCell chamber have been found to be lower and with less intra- and inter-technician variability than those obtained with a hemacytometer chamber (Brazil et al. 2004).

Semen Parameters

Parameters that are evaluated in semen analysis include semen volume, sperm concentration, sperm motility and viability, and sperm morphology. Semen volume can vary considerably between and within participants, with factors such as ejaculatory frequency contributing to the variability, as discussed above. Low semen volume can have a number of causes including improper sample collection and ejaculatory dysfunction, which is not uncommon. Low testosterone levels can also result in reduced semen volume. The sperm concentration is commonly referred to as the sperm count. The total number of sperm in the ejaculate is often termed the total sperm count. Sperm movement is measured by two parameters: percent of sperm that demonstrates flagellar movement, which is termed motility and ranges from 0 to 100%, and an assessment of the speed at which sperm move in a forward direction, which is termed forward progression. In most cases of non-motility, the sperm are non-viable, however, in some cases of ultrastructural damage to spermatozoa, the non-motile sperm may still be viable. Therefore, sperm viability testing is used to determine if non-motile sperm are alive or dead. Viability testing can be performed using one of two approaches, dye exclusion or hypoosmotic sperm swelling. Morphology evaluation requires the preparation of stained cytologic smears, and is the most labor intensive part of semen analysis.

celerion
Applied Translational Medicine

There are several scoring methods in use and proper clinical interpretation requires the investigator to be familiar with the scoring system of the laboratory. Strict criteria, such as Kruger's criteria, classify sperm as having normal morphology only if the sperm shape falls within strictly defined parameters of shape (Sigman and Zini 2009).

Data Analysis

Clinical studies to evaluate testicular safety are typically designed as non-inferiority studies comparing "response rates" between drug- and placebo-treated groups, where "responder" is defined as a participant who demonstrates a 50% or greater decrease from baseline in either sperm concentration or total motile sperm count. Because sperm concentrations can be highly variable within the same participant, a decrease from baseline of at least 50% is considered clinically significant for individual participants.

Secondary endpoints may include changes from baseline in other semen analysis parameters such as semen volume, sperm motility and morphology. Since semen data are not part of routine safety analysis in Phase I studies, principal investigators may not be familiar with interpreting out-of-normal-range results for individual participants. A qualified andrology laboratory can assist with data interpretation or if study oversight is provided by an external data safety monitoring board, urology expertise should be included among the membership.

To determine if the drug distributes to the genital tract, drug exposure in seminal fluid should be evaluated if possible. This requires the development of a bioanalytical method to measure the drug and any major metabolites in semen.

Baseline responder rates may vary from study to study, depending on the study population (age, healthy participants or patients), whether the study is a single-center or multi-center study, location of sites, and whether the semen analysis is performed by a central laboratory or multiple laboratories. If background responder rates are unknown (which often is the case), it is prudent to use a conservative approach when calculating the sample size required for demonstrating a difference between drug and placebo groups. Responder rates in placebo-treated participants of 2-7% have been reported in a population of older (>45 years) healthy men or with mild erectile dysfunction (Hellstrom et al. 2008). Importantly, risk-benefit factors must be considered and the minimum number of participants should be exposed to the drug while maintaining adequate statistical power. The Informed Consent Form (ICF) should discuss the potential risks in appropriate detail and language.

Conclusion

Designing and conducting a clinical study to evaluate the potential for testicular toxicity of a new drug is challenging and requires specialized knowledge and expertise. Reducing the variation in semen analysis data is a key goal to obtaining meaningful results. Celerion has conducted several testicular safety evaluation studies in recent years and can assist the design and conduct of these studies to meet regulatory acceptance.

References

1 Stewart J, Turner KJ: Inhibin B as a potential biomarker of testicular toxicity. Cancer Biomarkers 1: 75-91 (2005).

2 Keel BA: How reliable are results from the semen analysis. Fertility and Sterility 82: 41-44 (2004).

3 Hellstrom WJG, Gittelman M, Jarow J, Steidle C, McMurray J, Talley D, Watts S, Mitchell CL, McGill JM: An evaluation of semen characteristics in men 45 years of age or older after daily dosing with tadalafil 20 mg: results of a mutlicenter, randomized, double-blind, placebo-controlled, 9-month study. European Urology 53: 1058-1065 (2008).

4 Carlsen E, Petersen JH, Andersson A-M, Skakkebaek NE: Effects of ejaculatory frequence and season on variations in semen quality. Fertility and Sterility 82: 358-366 (2004).

5 Keel BA: Within- and between-subject variation in semen parameters in fertile men and normal semen donors. Fertility and Sterility 85: 128-134 (2006).

6 Swan SH, Brazil C, Drobnis EZ, Liu F, Kruse RL, Hatch M, Redmon JB, Wang C, Overstreet JW, and The Study for Future Families Research Group: Geographic differences in semen quality of fertile U.S. males. Environmental Health Perspectives 111: 414-420 (2003).

7 Sigman M, Zini A: Semen analysis and sperm function assays: what do they mean? Seminars in Reproductive Medicine 27: 115-123 (2009).

8 Brazil C, Swan SH, Drobnis EZ, Liu F, Wang C, Redmon JB, Overstreet JW, and The Study for Future Families Research Group: Standardized methods for semen evaluation in a multicenter research study. Journal of Andrology 25: 635-644 (2004).