Journal of Andrology, Vol. 24, No. 3, May/June 2003




# *Androlog* Summary

## Responses to Semen Analysis CART Report

CRAIG NIEDERBERGER

*From the Department of Urology, University of Illinois at Chicago, Chicago, Illinois.*

Correspondence to: Dr Craig Niederberger, Department of Urology, University of Illinois at Chicago, 840 South Wood Street, Chicago, IL 60612 (e-mail: craign{at}uic.edu).

Received for publication January 7, 2002; accepted for publication January 8, 2003.

*Note:* Postings to *Androlog* have been edited before publication.

Despite its limitations, the "bulk" semen analysis remains the primary screening tools for those seeking to identify male reproductive dysfunction. In an attempt to improve its clinical utility, Guzick and coauthors published a report in the New England Journal of Medicine using the modern computational classification and regression tree (CART) algorithm applied to the semen analysis. Dr Donna Vogel, one of the coauthors, asked Androlog members for their opinions regarding this tool:

> Last fall, the National Multicenter Reproductive Medicine Network (NMRMN) published an article by Guzick et al (2001). Media coverage tended to emphasize the conclusion that no single parameter could accurately predict fertility status. Two important elements were largely overlooked: First, for the first time, the ranges for "fertile" and "subfertile" were not chosen arbitrarily, but rather, the data were allowed to speak for themselves in an unbiased mathematical model. Second, different combinations of fertile or subfertile results for the 3 variables yielded strikingly different odds ratios for predicting infertility. Compared with 1.0 defined as all 3 parameters in the fertile range, odds ratios for infertility ranged from 2.2 when only parameter concentration was subfertile, to 15.8 when all three (morphology, motility, and concentration) were in the subfertile range. The work was funded in part by cooperative agreements through the National Institute of Child Health and Human Development.
>
> My question to the group is, how did you perceive this paper, and have you found the results to be of practical use in your practice or research?

David Handelsman questioned the utility of any analysis based on the fundamental limitations of the semen analysis, and the notion of discriminating fertile from infertile populations in general:

Responding to Dr Vogel's enquiry about the importance of the paper by Guzick et al (2001), my impression is that it has limited value and is unlikely to have much effect on well-informed clinical practitioners or researchers. Although Dr Vogel believes it is a virtue that the data speak for themselves (isn't all data analysis designed to achieve that?), the problem is that the data speak only for (and to) what it represents.

In developed countries, ~5% of men are infertile and ~5% are responsible for a conception in any single year; the remaining 90% have undefined fertility. The approach of the study in comparing the fertile with infertile men simply overlooks the vast majority of men with undefined fertility. How reliable are such extrapolations to the general male population? The study relied on recruiting "fertile" controls from prenatal classes but it is not clear how many such men were approached, what proportion agreed to participate and why. Do they represent all fertile men, let alone the 90% with undefined fertility? The large overlap of all raw semen variable data between fertile and infertile alone indicates that categorical classification schemes are pretty well doomed—at best, they generate parameters that are sensitive to the population from which they are derived and may not be general.

Other problems limit the study's practical usefulness even if the conclusion that semen variables provide useful information for diagnosing male infertility can be considered nontrivial. Providing odds ratios is of little use to any individual patient/couple—couples usually want to know their own expectation of fertility, not how they compare with an imaginary ideal couple. Even if this approach was a meaningful in some settings, the odds ratios do not pass the rule of thumb about retrospective case-control studies of this size: wake me up when the odds ratios reach double digits. Further, the introduction (from data analysis) of a category of indeterminate range is artificial and confusing to end users. Although it would not overcome the conceptual problems, a more flexible approach would have been to use continuous function parameters rather than creating cut points (which vary from one test sample to another). The paper also missed the opportunity to use resampling or sensitivity analysis to evaluate the robustness of the cut points.

Most of these issues have been debated for decades and exhuming the approach of comparing fertile with infertile populations, generated by convenient means for this study, just illustrates the limited creative and novel thinking that has been applied to this field.

Christopher Schrepferman, Dolores Lamb, and Larry Lipshultz contend that the limitations of the semen analysis manifest in the report by Guzick et al argue for a better and more robust approach to the task of identifying male subfertility:

In the age of molecular biology and broad advances in reproductive technology, it is surprising that the National Cooperative Reproductive Medicine Network (Guzick et al) would attempt to define male infertility using a simple semen analysis, a technique first reported in the 1920s (Macomber and Sanders, 1929). With inconsistencies that exist in sperm density, motility, and morphology in individual patients, it is almost intuitive that fertile and infertile men cannot be clearly identified using these parameters. Despite the contention that "modern methods" were used in the current study, the authors' results are remarkably similar, as they admit, to those reported in 1951 by MacLeod, who also attempted to define male infertility by semen analyses (MacLeod, 1951). A man with an abnormal semen analysis may be fertile. Conversely, a normal semen analysis does not ensure that the sperm are functionally capable of traversing the female reproductive tract

or penetrating and fertilizing an egg. It is the fertility potential of the couple that ultimately defines the ability to conceive.

Contrary to the authors' conclusions that male fertility potential can be categorized into 3 groups, we contend that this study clearly demonstrates that a simple semen analysis alone cannot be used as a reliable diagnostic tool to evaluate infertile men. Rather, the study emphasizes the need for a comprehensive evaluation of men in all infertile couples, including a careful physical examination, hormonal evaluation, functional semen testing, and genetic analysis, if indicated. Only by modernizing and expanding the evaluation of the man can more accurate diagnostic and therapeutic interventions be achieved.

Peter Schlegel expresses concern regarding the interpretation of the Guzick report by nonandrologists:

I believe the manuscript by Guzick et al reflects a landmark paper, and has been well-referenced by people who have knowledge of what's going on. However, this article and the other NMRMN papers are often misquoted (at least with the obstetricians and gynecologist with whom I have discussed it.

For example, the articles are suggested to show that male factor is not important for intrauterine insemination or in prediction of fecundity, whereas the opposite is correct—as expected, male factor plays an important role. Even more data from NMRMN would be more helpful, in my opinion.

Grace Centola spoke from the trenches of a clinical andrology laboratory:

Of note is that Dr Guzick presented a talk at the 2002 annual meeting of the American Association of Bioanalysts. I would also be interested in comments from those who were in attendance. As I saw it, there really was no strong predictor of fertility. He emphasized some data on morphology; unfortunately, the data have not changed our so-called normal ranges on semen analysis reports in my laboratories, nor the interpretation of our results.

Rune Eliasson critically delineated fundamental flaws in the bulk semen analysis as currently used:

I fully agree with the responses by David Handelsman and Christopher Schrepferman. The article by Guzick et al represents old paradigms, which for decades, have prevented a scientific development of andrology in general and semen analysis in particular. Semen analyses are essential in the study of the testes and the male accessory genital glands providing that the analyses can reflect functional aspects and are performed according to rules that are valid for other clinical laboratory analyses.

Sperm concentration is still given a lot of attention even though it is a noninformative variable. The reasons are that sperm concentration does not reflect any specific organ function and has very little relation to a man's fertility. The concentration depends on, among other things:

1. The testicular volume. Some men have one testis with a volume of 12 mL, others have two testicles, each with a volume of 30 mL.
2. The functional capacity of the seminiferous tubules in the testis.
3. The size and secretory function of the prostate, seminal vesicles, and epididymides.

4. The time between two ejaculations.

In addition, the fertilizing capacity of the spermatozoa depends more on factors other than the number of sperm per milliliter of seminal plasma.

In a monogamic society we should refrain from classifying men as fertile or infertile on the basis of some semen characteristics because it is both unscientific and unfair to the couple. If one partner is classified as fertile, who is then to be blamed for the infertility?

I fully agree that the World Health Organization manuals for semen analysis do not meet rigorous technical and statistical standards. But it was not in recognition of these limitations that the nomenclature was changed from "normal" to "reference" values. The praxis has changed over the last 2 decades and today there are clear rules for the establishment of reference intervals for clinical laboratory analyses. Hopefully, those who work with semen analysis will also recognize and apply these rules.

Classification of sperm morphology according to strict criteria has been accepted by gynecologists and many others interested in semen analysis to a degree that is entirely unfounded. It has become a new paradigm and will take a considerable amount of work and time to get rid of it.

It will be sufficient to claim that a method that considers every "defect" (small, doubtful, or large) equally important for classifying a cell as "not normal" has a fundamental error built into it. Moreover, it prevents us from learning which defects may be important or not important.

Guzick et al found the mean percentage of normal spermatozoa be 14 ± 5 in the fertile group and 11 ± 6 in the infertile group. It implies a range of 2 significant differences, from 4% to 24% in the fertile group, and from 0% to 23% in the infertile group. This information is hardly helpful to anyone working with infertile couples.

Urologist, gynecologists, and many other specialists interested in andrology are, one hopes, well aware of the specifics in their own field of expertise. Too few seem to understand that clinical laboratory medicine is also a specialty and that it takes years of education, training, and experience to be a specialist in that field. Strict criteria, sperm concentration, and other popular simplifications (including incorrect statistics) have no place in serious laboratory work related to such a complex clinical situation as human reproduction.

Handelsman and Schrepferman et al correctly emphasized the need for more creativity, novel thinking, and comprehensive evaluations of the man in andrology and human reproduction research. I want to add the need for an unbiased and functional approach that includes the possibility of subclinical genital infections as causes of infertility.

When poor manuscripts related to semen analyses can be published in journals such as the *New England Journal of Medicine* and *Lancet,* the educational work will more difficult but not less important and urgent. The Androlog forum plays an important role in this educational process. We should all be thankful for this and for the work done by the moderators.

Finally, Paul Claman notes that the report by Guzick et al brings to semen analysis the notion of an indeterminate category, that of subfertility, as opposed to the categories of fertile or nonfertile:

> Dr Eliason misrepresents the claims of Guzick et al. Because of the wide standard deviation range noted by Dr Eliason, Guzick's paper suggested that strict sperm morphology of between 9% and 12% is what they determined as indeterminate. Because was little crossover below that level, below 12% was termed *subfertile* and not *infertile* (as was suggested by Dr Eliason). Independent of the physiological limitations of using semen analysis to determine normalcy, the article in the *New England Journal of Medicine* presented compelling data to suggest that <11% normal strict morphology is a cause of subfertility. Guzick's in vivo study is also backed up by a myriad of papers showing that very low sperm morphology (<5% using strict morphology) is associated with failed fertilization during standard in vitro fertilization (IVF).
>
> Given the rigorous investigations done by Guzick et al demonstrating that the female partner had normal investigations, the data appear to be very good in suggesting that men who are in a subfertile relationship and who have strict sperm morphology of <9% are at least part of the cause for the couple's infertility. Clinically, this information (the first to my knowledge to carefully attempt to evaluate semen analysis as a diagnostic tool) is particularly important because there are now appropriate interventions to facilitate pregnancy in cases of male subfertility (ie, IVF-intracytoplasmic sperm injection). Rather than wasting these couples' time and emotional and financial resources on less effective treatment strategies (eg, superovulation and intrauterine insemination) by incorrectly diagnosing idiopathic infertility, these data allow physicians to counsel these couples regarding more effective therapeutic interventions for the problem of a male factor as a cause for a couple's infertility problem.

Clearly, in response to Donna Vogel's query, many debated the utility of applying any analytical methodology, no matter how powerful, to an assay as poor as the bulk semen analysis. Perhaps it is indeed time for a better test. Shall we relegate the bulk semen analysis to the dark ages of the 20th century, and discover a way to make the 21st century the age of male reproductive reason?

## *References*


Guzick DS, Overstreet JW, Factor-Litvak P, et al. Sperm morphology, motility and concentration in fertile and infertile men. *N Engl J Med.* 2001;345:1388 –1393.[Abstract/Free Full Text]

Macomber D, Sanders MB. The spermatozoa count: its value in the diagnosis, prognosis, and treatment of sterility. *N Engl J Med.* 1929; 200:981 .

MacLeod J. Semen quality in 1,000 men of known fertility and 800 cases of infertile marriages. *Fertil Steril.*1951; 2:115 .[Medline]