January 2014

Variations In Semen Sample Parameters Among Men In A Fertility Clinic: Implications For Reproducibility In Epidemiologic Studies

Regina Esenam Edifor
Yale University, redifor1@gmail.com

Title: Variations in semen sample parameters among men in a fertility clinic: Implications for reproducibility in epidemiologic studies

By
Regina Edifor

Submitted as part of the Master of Public Health thesis requirements to Yale School of Public Health from the Department of Epidemiology of Microbial Diseases

Thesis Committee:
First Reader: Adrienne S. Ettinger, ScD, MPH
Second Reader: Jorge E. Chavarro, M.D., ScD

Yale School of Public Health
April 22nd, 2014

ABSTRACT

BACKGROUND: In population studies, one semen sample is usually collected per individual but in clinical settings it is recommended that multiple semen samples are collected per individual for analysis. The goal of this study is to estimate the size of within-person variability in semen quality parameters with the ultimate goal of figuring out how many repeat samples are needed in a semen quality study to represent this. We will also investigate how accurately one can predict semen parameter values for an individual using the long-term average as a standard.

HYPOTHESIS: We hypothesize that a maximum of 2 semen samples has enough reliability to allow us to characterize an individual as fertile or infertile in a clinical or research setting.

METHOD: This study consists of 287 men who provided a total of 654 semen samples, (range 1 to 9). Semen samples were collected over a period of about 2 years. Within-person and between-person variability was analyzed using semen parameters: sperm concentration, total sperm count, ejaculate volume, sperm morphology (% normal) and motility (% motile).

PRELIMINARY RESULTS: There were no significant differences in demographics or reproductive history according to the number of samples collected. Semen sample variation between individuals is substantial but variation within individuals ranged from 14% to 28%. Intraclass correlation values ranged from 0.72 to 0.86 signifying high reproducibilty of semen parameter values. Correlation did not diminish with time. First samples given by each individual was highly similar to their long-term within-person average.

CONCLUSIONS: Based on the results of this study, there is high reproducibility of semen parameter values and so 1 sample can provide a true representation of an individual's long-term average.

TABLE OF CONTENTS

Introduction

Infertility is a key component of reproductive health that affects men and women globally (Mascarenhas, Flaxman, Boerma, Vanderpoel, & Stevens, 2012). WHO reports depict infertility as a public health problem that affects society as a whole (WHO, 1991). According to recent studies, some form of infertility affects approximately 10 to 15% of couples in their reproductive lifetime (Singh & Jaiswal, 2011). The proportion of infertility that is attributed to male reproductive factor goes from about 20% to 40% among infertile couples ("Diagnostic evaluation of the infertile male: a committee opinion," 2012). In recent years, there has been a decrease in semen quality and an increase in male reproductive disorders and there is evidence of environmental exposures being associated with male reproductive disorders (Gaspari et al., 2011). There is the need for more semen quality studies to look at how these recent issues in semen quality can be tackled.

Semen quality parameters can be used as a surrogate measure of male infertility in clinical and research settings (Cooper et al., 2010). A few studies have looked at variations in semen quality parameters within and between individuals and some of these studies call for the measurement of multiple semen samples before classifying an individual as fertile or infertile (Keel, 2006). The problem with this suggestion is that individuals may not be willing to provide as many semen samples as are required by researchers or clinicians. Participation rates for most epidemiological studies have decreased in the last few decades (Galea & Tracy, 2007) and for longitudinal studies looking at semen sample analysis, there is evidence provided by several studies to show very poor participation rates. This could lead to selection bias, which potentially compromises the validity of these studies (Tielemans et al., 2002).

One semen sample is usually collected per individual for analysis in population studies while multiple semen samples are collected per individual for analysis in clinical settings (Stokes-Riner et al., 2007). Considering that participation rates for most epidemiological studies have declined in the last few decades and are likely to decline further in the coming years (Galea & Tracy, 2007), it is important to have a system in place where one knows how many repeat samples are needed to represent a specified within-individual variation. In resource limited settings where the cost of analyzing additional semen samples provided by patients is an issue, it would be useful to consider whether collecting one or two semen samples per individual for semen quality analysis will provide reliable data for investigating infertility.

The goal of this study is to estimate the size of within- and between-person variability in semen quality parameters. We aim to find out how many semen samples are needed in a semen quality study to give an accurate representation of the true semen parameter value for each individual. This study seeks to answer the question: how badly does one semen sample misrepresent the true semen quality parameters of an individual? This question will be addressed by checking for reproducibility within samples and comparing semen parameter values between the first semen sample collected and the long-term average of samples collected from each individual.

Materials and Methods

Study Population – Men in this study were male partners in subfertile couples who were recruited when they presented for semen analysis at the Massachusetts General Hospital (MGH) Fertility Center (Hauser, Meeker, Duty, Silva, & Calafat, 2006). Men were eligible for the study if they met the following criteria: (i) between the ages of 18 and 55 years; (ii) did not have a vasectomy; and (iii) who were partners in couples using their own gametes for intrauterine insemination or assisted reproductive technologies (Hauser et al., 2006). This study consists of 287 men who provided a total of 654 semen samples, (range 1 to 9). At enrollment, a general health questionnaire was administered to collect information on demographics, medical reproductive history and lifestyle factors. The Harvard School of Public Health and the MGH Human Subjects Committees approved this study. An informed consent was signed by all participants before joining the study.

Semen Analysis - All semen samples were obtained on-site by masturbation into a sterile plastic cup. Sexual abstinence was requested for 48 hours before the semen sample was produced and the men were instructed to report the time period of abstinence. Semen samples which had no record of abstinence time were assigned to the most common abstinence category of 2-3 days. The semen samples were allowed to liquefy at 37°C for 20 minutes before analysis was carried out. Computer aided semen analysis (CASA) (Hamilton-Thorn Version 10HTM-IVOS) was used to obtain sperm concentration and motility at setting parameters previously described (Duty et al., 2003). The strict criteria proposed by Kruger *et al* (1988) were used to determine sperm morphology (Kruger et al., 1988). Two slides were prepared for each specimen with a minimum

of 200 spermatozoa on each slide and the percentage of sperm scored as morphologically normal was recorded (Duty et al., 2003).

Statistical Analysis – In this study, we are interested in the following semen quality parameters; sperm concentration (millions/ml), progressive motility (% motile), sperm morphology (% normal), ejaculate volume (ml), total sperm count (concentration x volume), total normal count (concentration x volume x % normal) and total motile count (concentration x volume x % motile). Sperm concentration, total sperm count, total normal count and total motile count were log-transformed to stabilize variances and to approximate a normal distribution. Ejaculate volume, motility and morphology were analyzed without transformation because they were fairly normally distributed.

To address whether participant characteristics differed according to number of samples, men who gave only 1 sample were compared with men who gave only 2 samples, only 3 samples and only 4 plus samples with respect to demographics (age, race, body mass index (BMI), and semen parameters), lifestyle factors (current smoking status) and self-reported reproductive history (primary infertility diagnosis, male factor diagnosis, previous infertility exam, varicocele, prostatitis and epididymitis). Male factor diagnosis is defined as primary or secondary cause of infertility based on Society of Assisted Reproductive Technologies (SART) diagnoses. For participants who gave more than 1 sample, the time between the first sample collection and the last sample collection was calculated for each comparison group and for all samples. A Kruskal-Wallis test was used to compare continuous variables while a chi-square test or Fisher's exact test was used to compare categorical variables.

To visualize within-person variation and between-person variation over time, we created a scatter plot with straight lines for 17 men (selected with a random number generator) for each of the semen quality parameters (sperm concentration (ml), progressive motility (%), morphology (% normal), ejaculate volume (ml), total sperm count).

To calculate within-subject variability, we used a general linear model to perform a one-way ANOVA on each individual for all semen quality parameters. The percent variation within-subject was calculated as: (model sum of squares – total sum of squares)/ Total sum of squares. The percent variation between-subject was calculated as: (error sum of squares – total sum of squares)/ total sum of squares. The Intraclass Correlation Coefficient (ICC) was calculated as: (percent variation between-subject) ÷ (percent variation between-subject + percent variation within-subject). The percentage of variation within subjects and ICC were calculated again after accounting for abstinence hours only, age only, time between samples only and all three: abstinence hours, age and time between samples. This was done by first running a regression analysis on each independent variable (abstinence hours only, age only, time between samples only, and all three: abstinence hours, age and time between samples). Using the residuals from the above regression, a one-way ANOVA was performed on each individual. Next we analyzed the change in within-person variation over time by looking at the ICC between baseline semen parameter values and 4 different time points where subsequent samples were collected. The baseline time point is 0 months (0-3 months). The subsequent time points are: 6 months (3-9 months), 12 months (9-15 months), 18 months (15-22 months), and 24 months (>22 months). To allow for easier calculations, these time points were changed into days where 1 month was considered as 30 days. Based on our calculation: 0 months is 0-90 days, 6 months is 90-270 days, 12 months is 270-450 days, 18 months is 450-660 days, and 24 months is 660 days or more. For

each of the semen parameters, we took the average of the time periods mentioned above and compared the average of 0 time period to the average of 6, 12, 18 and 24 time periods using a spearman correlation.

To find out if we can accurately predict a person's long-term semen parameter values just based on the person's first sample collected, we used paired t test to compare means of the first sample collected for each subject to the long-term within-person average for each subject. The long-term average was the calculated average of all samples given by an individual excluding the first sample. We also calculated a spearman correlation to show the correlation between the first sample and the long-term within-person average. To find out if we misclassify a subject as infertile just based on one sample, we dichotomized individuals using the WHO 2010 semen quality reference cut-off values (Cooper et al., 2010). We then defined "infertile men" as those men with abnormal semen parameters below the WHO cut-off to enable us to calculate Sensitivity, Specificity, Predictive Value Positive and Predictive Value Negative, using the long-term average as gold standard. This analysis was repeated by testing the average of the first two samples collected from each individual against the long-term average of that individual.

To conclude our analysis, we used linear and logistic regression to investigate the relationship between the semen quality parameters and age, abstinence hours, BMI and smoking. The regression analysis also allowed us to compare the first sample and the within-person average when each outcome is treated as a continuous variable or when it is dichotomized. For the model of continuous semen parameter outcomes, linear regression was used to obtain Beta and p-values of the first sample and the within-person average. Logistic regression was used to obtain odds ratio estimates with 95% confidence intervals when the semen parameter outcomes were dichotomized into fertile or infertile using the WHO 2010 cut-off values. All predictor

values were dichotomized for the regression analysis. Age was classified as above or below the median age of the sample (35.6 years), abstinence hours was classified as above or below the median abstinence hours recorded (58.0 hours), BMI was classified as normal (18.5-25 kg/m^2) or overweight/obese ($\geq$25 kg/m^2) and smoking was classified as current smoking or past/never smoked.

All statistical analyses were carried out using Statistical Analysis Software (SAS), version 9.3 (SAS institute Inc., Cary, NC).

Results

Semen samples from 287 men were available for analysis. The men delivered a total of 654 samples. The median age of the men was 35.6 years (interquartile range 32.7 – 38.9) at the time the first sample was taken. Majority of the men were white, non Hispanic (85%) and the median BMI of all participants was 27.0 kg/m^2 (interquartile range 24.4, 29.4) at the time the first sample was taken. Table 1 shows the results of comparisons of demographics and self-reported reproductive history between men who gave 1 sample, 2 samples, 3 samples and 4 plus samples only. There is no appreciable difference in demographics and reproductive characteristics according to the number of samples given. The frequency of male factor diagnosis was not significantly different according to the number of samples provided.

Figure 1 shows variations over time in semen parameters from 17 randomly selected subjects. There is substantial variability in semen parameters between individuals but we also see a large within-person variability, which is confirmed by a comparison to a WHO assessment of variability in semen quality over time (WHO 2010).

As expected, while variability within individuals is not trivial, the variability in semen quality observed between individuals was substantially larger. The results in Table 2a show estimates of within person variation from 14% for concentration to 28% for normal morphology. Measures of reproducibility suggest that on a population basis semen parameters are highly reproducible with ICC values from 0.72 for normal morphology to 0.86 for concentration. Within-person variation and measures of reproducibility were not substantially affected by age, abstinence hours, time between samples and age, abstinence hours and time together. We checked if this ICC diminishes over time and as is shown in Table 2b, there was not much change in the ICC when we compared baseline semen parameter values (0 months) to 6, 12, 18

and 24-month values. Further proof of correlation not changing over time is provided by Table 2c where we compare spearman correlation values over time for grouped means.

Table 3 shows a summary of semen parameters for the first sample and long-term average of each individual for men who provided two or more samples. To see how the first sample reflected the long-term average of each individual, the mean difference was calculated and a paired t-test was used to indicate significance of the difference. For all the semen parameters except ejaculate volume, there was no significant difference between the first sample and the long-term average for each individual. Mean ejaculate volume was significantly different ($p = 0.031$) between the first sample and the long-term average of participants but this difference was only 0.1ml (2.9 ml versus 2.8ml for first sample and long-term average respectively). Spearman correlation between the first sample and the long-term average showed good correlation between the two groups and values ranged from 0.527 for total sperm count to 0.705 for motility.

Individuals were classified as normal or abnormal based on WHO reference cutoffs and "infertility" was predicted in the first sample collected using the long-term average as a gold standard. The results show that based on the WHO cutoff values for abnormal semen parameters, using a first sample as a representative of long-term average is highly specific. Estimates of specificity range from 83% for motility to 95% for concentration (Table 4a). Using the average of the first two samples as representative of long-term average gave an even more accurate prediction for classifying an individual as not infertile. Estimates of specificity range from 91% for morphology to 100% for concentration and count (Table 4b).

With the exception of abstinence hours, there were no statistically significant associations between our predictor variables (Age, BMI, Smoking) and the semen quality parameters when

we treated the semen parameters as continuous or as categorical variables (Table 5a and 5b). We observed that abstinence hours is significantly associated with morphology when we look at long-term averages (Table 5a). Using WHO 2010 reference cutoffs for abnormal semen parameters, we also saw a significant association between abstinence hours and abnormal morphology for both first samples and long-term averages (Table 5b).

Discussion

In a cohort of men attending a fertility clinic, we did not find any significant difference in demographics, semen parameters and male reproductive history between individuals who gave 1 sample, 2 samples, 3 samples or 4 plus samples. Our results confirmed the presence of substantial between-person variation in semen quality parameters as well as very noticeable within-person variation. Biologic variables among humans are typically subject to a good deal of random variation (Rosner & Willett, 1988) and our analysis was able to capture that variation between and within our study subjects. However we also showed that there is some validity in using individual samples to represent a long-term average.

Few studies have looked at a comparison of one sample versus several samples of semen quality parameters among males. The Study for Future Families (SFF), a prospective study comparing men who give 1 semen sample to men who give 2 semen samples showed that semen quality parameters did not differ significantly between men who gave 1 or 2 samples. The SFF study also showed that semen parameters do not differ between the first and second sample given by men who gave 2 samples (Stokes-Riner et al., 2007).

From our study, the ICC and percentage of variation within-person showed great reproducibility of semen parameter values within each individual. After accounting for age, abstinence hours and time between samples, all factors that could influence semen parameter values, we still observed good correlation within samples. When we looked at the correlation between semen parameter values taken at a set baseline time period (0 months) and compared that to subsequent time periods (6, 12, 18 and 24 months), the correlation did not diminish considerably with time. Comparison of the first semen sample given by patients to their long-term average did not show any statistically significant difference except for ejaculate volume

(p=0.031). The correlation between the group average of all first samples and group average of all long-term averages was in the 0.5 to 0.7 range. In addition, the high specificity values obtained in our analysis, signifies that we are able to give an accurate estimate of individuals who are not infertile just by looking at the first sample or the first two samples collected from that individual. Taken together, these results show the reliability of using one semen sample as a representative of an individual's true long-term average in semen analysis studies.

Our regression analysis showed no significant association between the semen parameters and age, BMI or smoking. However we did observe an association between abstinence hours and morphology. Abstinence time is one of the factors that can affect semen quality and the WHO recommends an optimal abstinence time of 2 to 7 days however there is not much literature supporting the basis of this recommendation (WHO, 2010). Some studies have found no association between abstinence time and semen parameters while others have found strong associations between abstinence time and semen parameters. One such study found that abstinence time had no significant association with morphology (De Jonge et al., 2004) while another found that morphology was significantly associated with abstinence time (Levitas et al., 2005). Further studies will have to be carried out to investigate the true effect of length of abstinence time on each of the semen quality parameters.

One of the strengths of this study is that we had a varied number of samples represented and so this allowed us to obtain a within-person long-term average with which to compare the first sample collected from individuals. One limitation of this study is that the data was taken from male partners in subfertile couples who were visiting a fertility clinic. Thus, it is unknown if the results found here can be generalized to a population of male partners in fertile couples. In addition, majority of the study population were non-Hispanic whites and so these results are not

representative of other races not featured in the study. It is important to note that the WHO 2010 cut off limits that we used do not necessarily suggest a biological cause of infertility (WHO, 2010).

For clinical purposes, it is standard to obtain multiple samples from an individual when investigating infertility however in population studies 1 sample is usually collected from individuals for semen analysis. Based on the results of this study, for research purposes, increasing the number of participants as opposed to increasing the number of samples collected per person is likely to yield more valuable results especially in cases where high statistical power is desired or in settings where resources are limited. For clinical purposes, using 2 samples only will yield valuable results when classifying patients as fertile or infertile.

References

Cooper, T. G., Noonan, E., von Eckardstein, S., Auger, J., Baker, H. W., Behre, H. M., . . . Vogelsong, K. M. (2010). World Health Organization reference values for human semen characteristics. [Comparative Study

Meta-Analysis

Research Support, Non-U.S. Gov't

Review]. *Hum Reprod Update, 16*(3), 231-245. doi: 10.1093/humupd/dmp048

De Jonge, C., LaFromboise, M., Bosmans, E., Ombelet, W., Cox, A., & Nijs, M. (2004). Influence of the abstinence period on human sperm quality. *Fertil Steril, 82*(1), 57-65. doi: 10.1016/j.fertnstert.2004.03.014

Diagnostic evaluation of the infertile male: a committee opinion. (2012). *Fertil Steril, 98*(2), 294-301. doi: 10.1016/j.fertnstert.2012.05.033

Duty, S. M., Silva, M. J., Barr, D. B., Brock, J. W., Ryan, L., Chen, Z., . . . Hauser, R. (2003). Phthalate exposure and human semen parameters. *Epidemiology, 14*(3), 269-277.

Galea, S., & Tracy, M. (2007). Participation rates in epidemiologic studies. [Review]. *Ann Epidemiol, 17*(9), 643-653. doi: 10.1016/j.annepidem.2007.03.013

Gaspari, L., Paris, F., Jandel, C., Kalfa, N., Orsini, M., Daures, J. P., & Sultan, C. (2011). Prenatal environmental risk factors for genital malformations in a population of 1442 French male newborns: a nested case-control study. [Research Support, Non-U.S. Gov't]. *Hum Reprod, 26*(11), 3155-3162. doi: 10.1093/humrep/der283

Hauser, R., Meeker, J. D., Duty, S., Silva, M. J., & Calafat, A. M. (2006). Altered semen quality in relation to urinary concentrations of phthalate monoester and oxidative metabolites. *Epidemiology, 17*(6), 682-691. doi: 10.1097/01.ede.0000235996.89953.d7

Keel, B. A. (2006). Within- and between-subject variation in semen parameters in infertile men and normal semen donors. [Comparative Study]. *Fertil Steril, 85*(1), 128-134. doi: 10.1016/j.fertnstert.2005.06.048

Kruger, T. F., Acosta, A. A., Simmons, K. F., Swanson, R. J., Matta, J. F., & Oehninger, S. (1988). Predictive value of abnormal sperm morphology in in vitro fertilization. [Research Support, Non-U.S. Gov't]. *Fertil Steril, 49*(1), 112-117.

Levitas, E., Lunenfeld, E., Weiss, N., Friger, M., Har-Vardi, I., Koifman, A., & Potashnik, G. (2005). Relationship between the duration of sexual abstinence and semen quality: analysis of 9,489 semen samples. [Comparative Study]. *Fertil Steril, 83*(6), 1680-1686. doi: 10.1016/j.fertnstert.2004.12.045

Mascarenhas, M. N., Flaxman, S. R., Boerma, T., Vanderpoel, S., & Stevens, G. A. (2012). National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. [Research Support, Non-U.S. Gov't]. *PLoS Med, 9*(12), e1001356. doi: 10.1371/journal.pmed.1001356

Rosner, B., & Willett, W. C. (1988). Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. [Research Support, U.S. Gov't, P.H.S.]. *Am J Epidemiol, 127*(2), 377-386.

Singh, K., & Jaiswal, D. (2011). Human male infertility: a complex multifactorial phenotype. [Review]. *Reprod Sci, 18*(5), 418-425. doi: 10.1177/1933719111398148

Stokes-Riner, A., Thurston, S. W., Brazil, C., Guzick, D., Liu, F., Overstreet, J. W., . . . Swan, S. H. (2007). One semen sample or 2? Insights from a study of fertile men. *J Androl, 28*(5), 638-643. doi: 10.2164/jandrol.107.002741

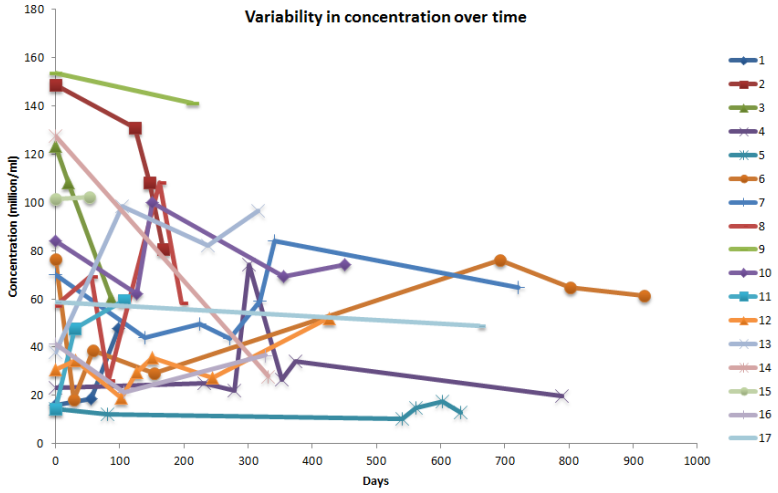
WHO. (1991). Infertility (D. o. F. H. Maternal and Child Health, Trans.) *A tabulation of available data on prevalence of primary and secondary infertility* (pp. 2). Geneva: World Health Organization.
WHO. (2010). *WHO laboratory manual for the Examination and processing of human semen* (Fifth Edition ed.). Geneva: WHO Press.

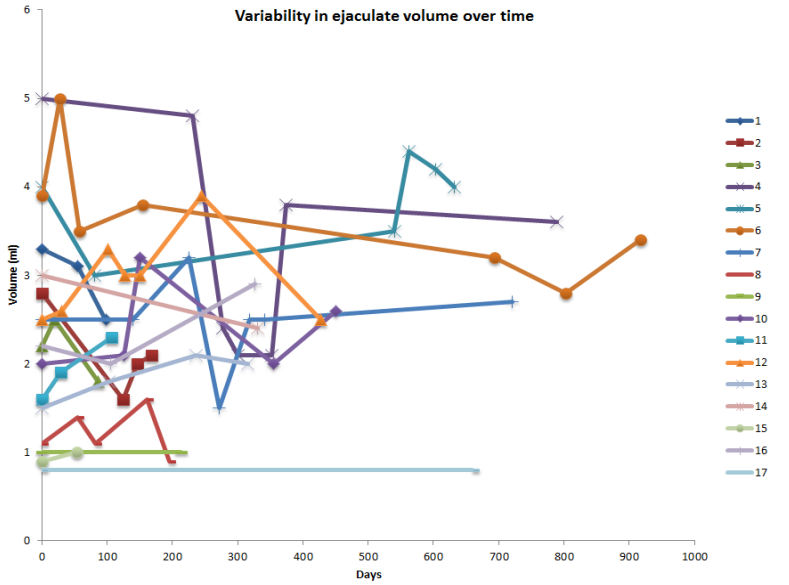
Table 1. Participants' characteristics according to number of samples: Median (IQR) or N(%)

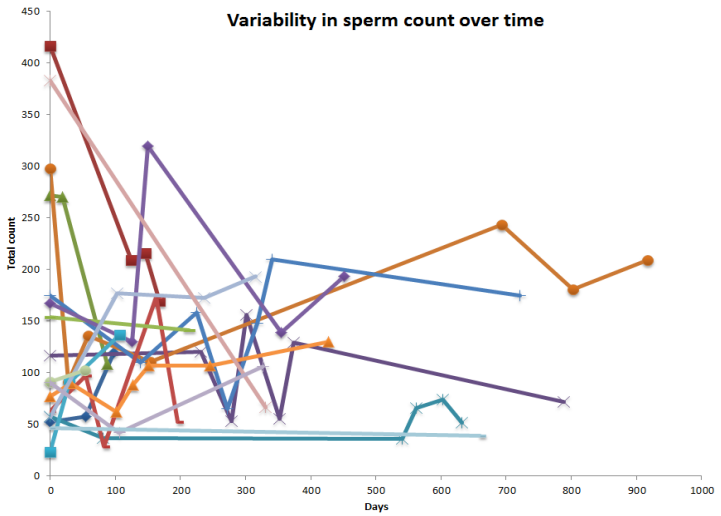
	All	1 Sample	2 Samples	3 samples	4+ samples[#]	p [*]
N	287	114	87	38	48	
Demographics						
Age, years	35.6 (32.7, 38.9)	36.1 (33.0, 39.2)	34.6 (32.3, 38.7)	35.8 (32.4, 38.4)	36.2 (33.7, 39.8)	0.304
Race/ethnicity, N (%)						0.748
White, not Hispanic	244 (85.0)	100 (87.7)	73 (83.9)	31 (81.6)	40 (83.3)	
Black	7 (2.4)	3 (2.6)	2 (2.3)	0 (0.0)	2 (4.2)	
Asian	18 (6.3)	5 (4.4)	6 (6.9)	5 (13.2)	2 (4.2)	
Hispanic or Latino	18 (6.3)	6 (5.3)	6 (6.9)	2 (5.3)	4 (8.3)	
Current smokers†	76 (26.5)	30 (26.3)	24 (27.6)	6 (15.8)	16 (33.3)	0.328
Body Mass Index, kg/m^2	27.0 (24.4, 29.4)	27.2 (24.4, 29.5)	26.9 (24.7, 28.7)	25.9 (22.7, 28.9)	27.9 (25.2, 30.7)	0.254
Sperm Concentration(10^6/ml)	64.0 (29.9, 110.0)	65.3 (28.4, 118.8)	63.7 (27.9, 109.3)	48.7 (34.7, 73.9)	64.8 (36.6, 110.8)	0.409
Semen analysis count	161.7 (83.7, 278.5)	171.0 (97.2, 261.4)	131.6 (66.2, 258.7)	120.1 (63.6, 193.1)	145.3 (69.8, 238.5)	0.220
Motility (%)	48.0 (27.0, 62.0)	48.5 (29.0, 62.0)	43.0 (23.0, 66.0)	38.5 (28.0, 55.0)	46.0 (31.5, 61.5)	0.560
Normal Morphology (%)	6.0 (4.0, 8.0)	6.0 (4.0, 9.0)	6.0 (3.0, 8.0)	7.0 (4.0, 9.0)	7.0 (5.0, 9.0)	0.183
Ejaculate Volume (mL)	2.7 (2.0, 3.8)	2.8 (2.0, 3.8)	2.6 (1.7, 3.5)	2.5 (2.0, 3.4)	2.5 (1.6, 3.5)	0.428
Self-reported reprod. history						
Primary Infertility diagnosis						0.202
Male Factor, N (%)	81 (28.2)	38 (33.3)	27 (31.0)	9 (23.7)	7 (14.6)	
Female Factor, N (%)	97 (33.8)	36 (31.6)	26 (29.9)	13 (34.2)	23 (47.9)	
Unknown, N (%)	109 (38.0)	40 (35.1)	34 (39.1)	16 (42.1)	18 (37.5)	
Male factor diagnosis ¶, N (%)	103 (35.9)	43 (37.7)	34 (39.1)	14 (36.8)	12 (25.0)	0.386
Prev. infertility exam, N (%)	223 (78.0)	89 (78.1)	70 (80.5)	32 (86.5)	32 (66.7)	0.142
Varicocele, N (%)	27 (9.4)	8 (7.0)	11 (12.6)	2 (5.3)	6 (12.5)	0.420
Prostatitis, N (%)	7 (2.4)	2 (1.8)	3 (3.5)	2 (5.3)	0 (0.0)	0.342
Epididymitis, N (%)	5 (1.7)	2 (1.8)	1 (1.2)	0 (0.0)	2 (4.2)	0.720
Abstinence hours	58.0 (47.0, 79.0)	61.0 (48.0, 77.0)	57.5 (39.0, 71.0)	59.0 (36.0, 72.0)	53.0 (38.0, 68.0)	0.280
Time (days) between first and last sample	-	-	89.0 (53.0, 150.0)	164.0 (107.0, 285.0)	334.5 (251.0, 564.5)	<0.001

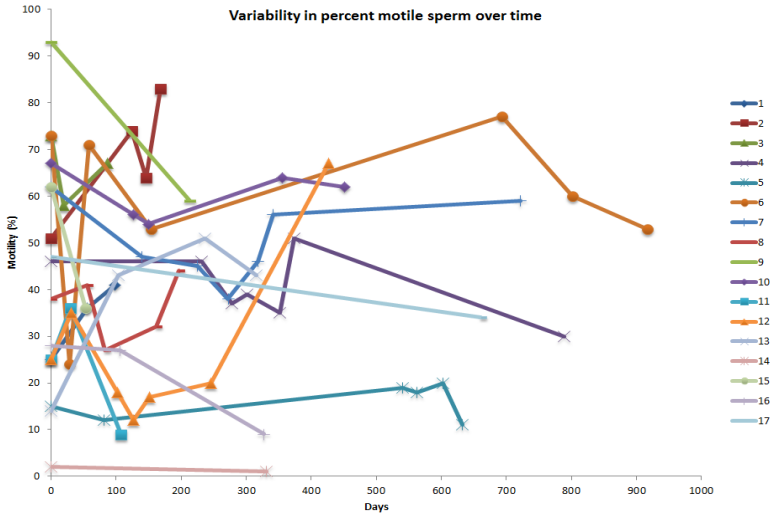
[*] [#]

Figure 1. Variation in total sperm concentration, ejaculate volume, count, motility and percent normal morphology over time in 17 randomly selected individuals






Variability in sperm count over time
Total count
Days
0
50
100
150
200
250
300
350
400
450
0
100
200
300
400
500
600
700
800
900
1000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17


Variability in percent motile sperm over time
Motility (%)
Days
0
10
20
30
40
50
60
70
80
90
100
0
100
200
300
400
500
600
700
800
900
1000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

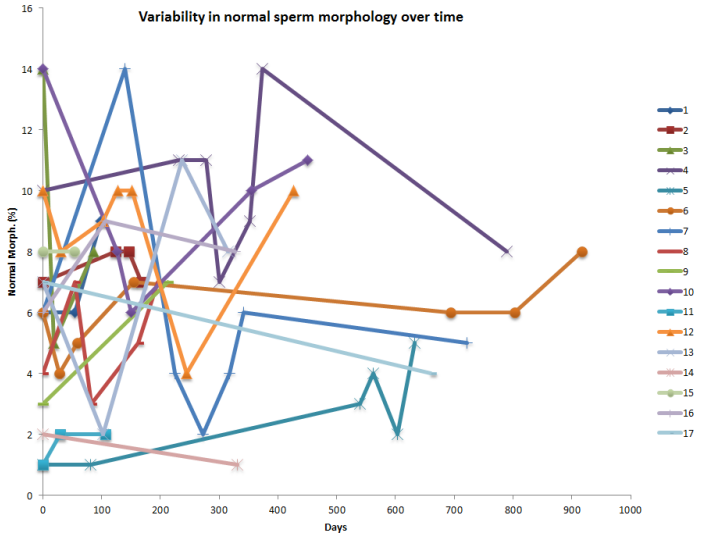

Variability in normal sperm morphology over time
Normal Morph. (%)
Days
0
2
4
6
8
10
12
14
16
0
100
200
300
400
500
600
700
800
900
1000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Table 2a. Partition of variance in semen quality (All samples 1 through 9) N= 287

Semen Analysis:	% Variation Within	ICC	Age* ICC	Abstinence hours* ICC	Time* ICC	Age + Abstinence hours* ICC	Age+Time Abstinence hours* ICC
Count ‡	22	0.78	0.78	0.79	0.78	0.79	0.79
Concentration ‡ (x 10^6 /ml)	14	0.86	0.86	0.86	0.86	0.86	0.86
Motility (%)	19	0.81	0.81	0.83	0.81	0.83	0.83
Normal morphology (%)	28	0.72	0.72	0.74	0.72	0.74	0.74
Ejaculate volume (mL)	18	0.82	0.82	0.84	0.82	0.83	0.83
Total normal count	18	0.82	0.82	0.83	0.82	0.83	0.83
Total motile count	18	0.82	0.82	0.84	0.82	0.84	0.84

‡ Count and concentration data was transformed using natural log

*ICC after adjusting for age only, abstinence hours only, time between samples only, age and abstinence hours and time between samples together

Table 2b. Partition of variance in semen quality over time (N=287)

Semen Analysis:	Overall ICC	0 months ICC	0 and 6 months ICC	0 and 12 months ICC	0 and 18 months ICC	0 and 24 months ICC
Count ‡	0.78	0.88	0.81	0.86	0.87	0.87
Concentration ‡ (x 10^6 /ml)	0.86	0.93	0.87	0.92	0.93	0.93
Motility (%)	0.81	0.90	0.83	0.88	0.89	0.89
Normal morphology (%)	0.72	0.89	0.77	0.85	0.86	0.87
Ejaculate volume (mL)	0.82	0.91	0.85	0.89	0.90	0.90

‡ Count and concentration data was transformed using natural log

Table 2c. Partition of variance in semen quality over time (N=287)

Semen Analysis:	0 and 6 months Spearman Corr*	0 and 12 months Spearman Corr	0 and 18 months Spearman Corr	0 and 24 months Spearman Corr
Sample size (N)	103	41	20	12
Count ‡	0.644	0.624	0.666	0.510
Concentration ‡ (x 10^6 /ml)	0.728	0.741	0.705	0.818
Motility (%)	0.614	0.597	0.595	0.916
Normal morphology (%)	0.630	0.770	0.572	0.690
Ejaculate volume (mL)	0.679	0.673	0.657	0.723

‡ Count and concentration data was transformed using natural log

* **Correlation between time intervals: 0months (0-3mths),** 6months (3-9mths), 12months (9-15mths), 18months (15-22mths) and 24months (>22mths)

Table 3. Comparison of semen parameter averages: For men with two or more samples (Mean (95% CI)) N=173

Semen Analysis	First sample	Within-person average	Difference (First sample-average)	p[a]	Spearman Corr*
Concentration (x10^6/ml)	73.2 (65.4, 81.0)	73.6 (64.3, 82.9)	-0.5 (-8.6, 7.7)	0.913	0.561
Ejaculate Volume (mL)	2.9 (2.7, 3.1)	2.7 (2.5, 2.9)	0.2 (0.0, 0.3)	0.031	0.667
Count	197.4 (172.9, 221.9)	180.2 (156.5, 203.9)	17.2 (-6.2, 40.6)	0.149	0.527
Motility	46.3 (42.8, 49.8)	44.8 (41.5, 48.1)	1.5 (-1.1, 4.1)	0.260	0.705
Morphology	6.1 (5.6, 6.7)	6.2 (5.8, 6.7)	-0.1 (-0.6, 0.3)	0.589	0.597
Total normal count	13.2 (11.1, 15.3)	12.9 (10.4, 15.5)	0.3 (-1.9, 2.5)	0.790	0.578
Total motile count	111.2 (92.6, 129.8)	97.3 (81.1, 113.6)	13.9 (-1.8, 29.5)	0.083	0.603

[a]Used Paired t test

*Correlation between first sample and within-person average

Table 4a. Proportion of men with abnormal semen parameters as determined by WHO 2010 reference values (N = 173)

Semen Analysis	First sample	Within-person Average	Sens	Spec	PVP*	PVN*
Concentration (<15 million/mL), %	8.7	7.5	53.9	95.0	46.7	96.2
Ejaculate Volume (<1.5 mL), %	14.5	15.0	53.9	92.5	56.0	91.9
Count (<39 million), %	7.5	7.5	30.8	94.4	30.8	94.4
Motility (<40% motile), %	39.9	44.5	68.8	83.3	76.8	76.9
Morphology (≤ 4% normal), %	35.3	29.5	84.3	85.3	70.5	92.9

*PVP: Predictive Value Positive; PVN: Predictive Value Negative; Sens: Sensitivity; Spec: Specificity

Table 4b. Proportion of men with abnormal semen parameters as determined by WHO 2010 reference values (n = 86)

Semen Analysis	Average of first 2 samples	Within-person Average	Sens	Spec	PVP*	PVN*
Concentration (<15 million/mL), %	4.7	9.3	50.0	100.0	100.0	95.1
Ejaculate Volume (<1.5 mL), %	9.3	12.8	63.6	98.7	87.5	94.9
Count (<39 million), %	0.0	5.8	-	100.0	-	94.2
Motility (<40% motile), %	41.9	40.7	74.3	80.4	72.2	82.0
Morphology (≤ 4% normal), %	25.6	22.1	84.2	91.0	72.7	95.3

*PVP: Predictive Value Positive; PVN: Predictive Value Negative; Sens: Sensitivity; Spec: Specificity

Table 5a. Association between predictor variables and semen parameters (Continuous, N=173)

	First sample	p	Average	p
	β (SE)		β (SE)	
Concentration (N)				
Age (82)	-0.034 (0.126)	0.786	0.137 (0.129)	0.290
Abstinence hrs (66)	-0.249 (0.130)	0.057	-0.072 (0.133)	0.587
BMI (121)	0.006 (0.137)	0.966	0.055 (0.141)	0.699
Smoking (46)	0.022 (0.143)	0.880	-0.045 (0.146)	0.760
Count				
Age (82)	-0.237 (0.132)	0.075	0.083 (0.124)	0.503
Abstinence hrs (66)	-0.169 (0.137)	0.219	0.009 (0.128)	0.944
BMI (121)	-0.084 (0.144)	0.562	0.045 (0.135)	0.741
Smoking (46)	0.039 (0.150)	0.794	-0.049 (0.141)	0.727
Motility				
Age (82)	0.257 (3.585)	0.943	1.356 (3.338)	0.685
Abstinence hrs (66)	-4.993 (3.697)	0.179	-6.698 (3.442)	0.053
BMI (121)	1.571 (3.907)	0.688	0.653 (3.638)	0.858
Smoking (46)	1.582 (4.060)	0.697	0.662 (3.780)	0.861
Morphology				
Age (82)	0.250 (0.552)	0.651	0.532 (0.466)	0.256
Abstinence hrs (66)	-0.996 (0.569)	0.082	-1.004 (0.481)	0.038
BMI (121)	-0.055 (0.602)	0.927	-0.666 (0.508)	0.191
Smoking (46)	-0.026 (0.625)	0.967	-0.111 (0.528)	0.833

*β (SE) for each predictor variable controlling for other predictor variables in each model

Table 5b. Association between predictor variables and semen parameters (OR for infertility, N=173)

	First sample	Average
	OR§ (95% CI)	**OR§ (95% CI)**
Concentration	(N=15)	(N=13)
Age	1.797 (0.607, 5.324)	0.468 (0.138, 1.587)
Abstinence hrs	2.041 (0.697, 5.971)	1.001 (0.309, 3.248)
BMI	0.812 (0.260, 2.535)	0.677 (0.208, 2.199)
Smoking	1.076 (0.320, 3.616)	0.816 (0.212, 3.150)
Count	(N=13)	(N=13)
Age	0.639 (0.197, 2.075)	0.470 (0.139, 1.592)
Abstinence hrs	1.291 (0.407, 4.098)	1.386 (0.440, 4.370)
BMI	2.636 (0.556, 12.497)	0.963 (0.280, 3.320)
Smoking	0.205 (0.026, 1.634)	0.826 (0.214, 3.184)
Motility	(N=69)	(N=77)
Age	0.940 (0.510, 1.733)	1.056 (0.577, 1.934)
Abstinence hrs	1.190 (0.635, 2.230)	1.419 (0.763, 2.641)
BMI	0.860 (0.443, 1.667)	0.897 (0.465, 1.732)
Smoking	0.973 (0.486, 1.946)	0.762 (0.382, 1.522)
Morphology	(N=61)	(N=51)
Age	0.679 (0.357, 1.292)	0.794 (0.405, 1.557)
Abstinence hrs	2.301 (1.202, 4.405)	2.346 (1.196, 4.603)
BMI	1.130 (0.559, 2.285)	1.546 (0.720, 3.320)
Smoking	1.341 (0.654, 2.749)	0.966 (0.448, 2.079)

* OR§ (95% CI) for each predictor variable controlling for other predictor variables in each model
§OR: Modelling abnormal semen parameters based on WHO 2010 cut-off values