Semen quality across populations

S.M. Schrader[1], M. Ayers-Cumbow[1], T.W. Turner[1], S.D. Simon[1], J. Ratcliffe[2], L. Welch[3], B. Grajewski[1], and T. Weyandt[4]

[1]National Institute For Occupational Safety and Health, Centers for Disease Control. Cincinnati, Ohio; [2] University of North Carolina, Chapel Hill, North Carolina [3]George Washington Medical School, Washington, D.C.
[4]The Pennsylvania State University, University Park, PA, USA

Send correspondence to: Dr. Steven Schrader, NIOSH, MS: C23, 4676 Columbia Parkway, Cincinnati, Ohio 45226

Abstract

A NIOSH andrology team conducted a longitudinal study of semen quality of workers with no evidence of toxicant exposure to establish baseline data in semen analysis and to provide information on the variability of semen characteristics both within and between men. The same andrology team has also conducted six cross-sectional occupational field studies assessing toxicant exposures. Each of these studies contained a concurrent comparison group of men with no known toxicant exposure. This report quantifies the relative similarities between each of the six control groups and the longitudinal group. This quantification is important for two reasons:

a) it gives an appreciation of what normally might be expected in a future control group of a cross-sectional study. This gives a means for independent assessment of the quality of that control group.

b) equally important, it helps assess whether the longitudinal data are representative of control populations measured at different times and locations.

Sperm count per ml and count per ejaculate showed large discrepancies between the six controls and the longitudinal study, both in the means (with control groups consistently larger than the longitudinal study by up to 112%) and variability (consistently smaller by up to 62%). Percent normal morphology also showed large differences in variability (consistently smaller by up to 74%) though not in the means (consistently larger, but only by 1% to 12%). There tended to be greater agreement in the means and variability for the two measures of sperm viability, for semen pH, and for semen volume. Perfect agreement is not expected. Those variables where similarities do exist reinforce the utility of using the longitudinal data in planning future population studies.

Introduction

The attempt to establish a "normal" sperm count value for humans dates back to at least 1929 when Macomber/Sanders reported the normal sperm count to be about 100 million sperm per milliliter of semen. Since that time numerous studies have reported a "normal" sperm count (Falk/Kaufman, 1950; Macleod 1951; Rehan/Sobrero, 1975; Sobrero/Rehan, 1975; Schwartz et al, 1979; David et al, 1979; Fariss et al, 1981; Schwartz et al, 1983; Wang et al, 1985; Schrader et al, 1988; Mallidis et al, 1991). Several others have debated whether a decline in sperm numbers has taken place in this century (Nelson/Bunge, 1974; James 1980; Bostofte et al, 1983; Osser et al, 1984; Menkveld et al 1986; Carlsen et al, 1992). One of the confusing points is what constitutes normal.

Population groups considered "normal" have included "proven fertile", wife currently pregnant and having fathered two other pregnancies, pre-vasectomy patients, volunteers, and workers. To further complicate these values, they were measured in different laboratories using different techniques and equipment. There is a very large range of normal sperm count values. Other semen measurements (motility, morphology, viability etc) have fewer reports but with the same confusion.

In 1988 the National Institute for Occupational Safety and Health (NIOSH) published a National Strategy for the Prevention of Disorders of Reproduction. This strategy not only called for the study of worker populations whose reproductive potential may be at risk, but also the establishment of baseline data in semen analysis. As a part of the NIOSH strategy, a longitudinal study of semen quality of workers with no evidence of toxicant exposure was conducted (Schrader et al, 1988a; Schrader et al, 1990; Schrader et al, 1991; Evenson et al, 1991; Schrader et al, 1992). These data provide information on the variability of semen characteristics both within and between men. The same andrology team has also conducted six cross-sectional occupational field studies assessing toxicant exposures. Each of these studies contained a concurrent comparison group of men with no known toxicant exposure.

This report quantifies the relative similarities between each of the six control groups and the longitudinal group. Similarities both in the mean and in the variance were assessed. This quantification is important for two reasons.

a) it gives an appreciation of what normally might be expected in a future control group of a cross-sectional study. This gives an independent assessment of the quality of that control group.

b) equally important, it helps assess whether the longitudinal data is representative of control populations measured at different times and locations.

While perfect agreement is not expected, similarities between the control data and longitudinal data reinforce the utility of using the longitudinal data in power calculations for planning future studies.

Materials/Methods

1. Populations.

The longitudinal study (LS) consisted of 45 men (25-35 years of age), each of whom provided monthly (June through February 1986-1987) semen samples for nine months in Ohio (Schrader et al 1988a).

The Colorado (CO) samples were a comparison group for a study of pesticide workers in Colorado exposed to ethylene dibromide (Schrader et al, 1988b). This group consisted of 19 men (20-35 years of age). The CO study was conducted in May 1983.

The Hawaii (HI) samples were the comparison group for a study of pesticide workers in Hawaii exposed to the ethylene dibromide (Ratcliffe et al, 1987). This group consisted of 44 men (20-61 years of age) providing semen samples in December 1983.

The Oregon (OR) samples were from the comparison group for a study of foundry workers in Oregon exposed to 2-ethoxyethanol (Ratcliffe et al 1989). This group consisted of 37 men (21-50 years of age) providing semen samples in June 1984.

The Connecticut (CT) samples were from the comparison group for a study of ship painters in Connecticut exposed to 2-ethoxyethanol (Welch et al, 1988). This group consisted of 41 men (ages 19-62 years) providing semen samples in December 1984.

The Maryland (MD) samples were from the comparison group for a study of dielectric heat sealer operators in Maryland (Grajewski et al, in preparation). This group consisted of 34 men (24-61 years of age) providing samples in May-June 1988.

The Texas (TX) samples were from the comparison group of soldiers in Texas for a study of artillerymen (Weyandt et al, submitted). This group consisted of 31 men (20-48 years of age) providing semen samples in July 1990.

2. Semen Analysis Methods.

Each man was provided information either by direct interview or video on when and how to collect and transport the semen sample. All men were asked to be at least two days abstinent, and to deliver the sample to the laboratory within one h of ejaculation. At the time of collecting the semen sample, each subject recorded his abstinence, collection time, and spillage on the jar label.

Semen analyses were conducted in two phases. The initial evaluation of the sample was conducted when the sample arrived at the "on-site" laboratory, and consisted of recording the temperature, turbidity, color, liquefaction time, volume, and pH of the semen. Video recordings, viability assessments, sperm counts, preparation of slides and preservation of seminal plasma were also conducted at this time. Sperm viability was determined by two methods, eosin y stain exclusion (Eliasson/Treichl, 1971) and hypoosmotic swelling (HOS assay)

TABLE 1
Length of Abstinence in Days

STUDY	Percent of Population per Days of Abstinence						
	1	2	3	4	5	6	≥7
LS (June)	0.0	48	26	6.5	13.0	2.2	4.3
LS (July)	6.7	39	27	15.6	8.9	2.2	2.2
LS (Aug)	4.4	47	31	11.1	6.7	0.0	0.0
LS (Sept)	6.7	40	31	8.9	8.9	2.2	2.2
LS (Oct)	6.8	36	25	16.0	16.0	0.0	0.0
LS (Nov)	0.0	36	36	16.0	9.1	0.0	2.3
LS (Dec)	9.1	32	34	11.4	6.8	4.5	2.3
LS (Jan)	9.1	30	43	11.4	2.3	2.3	2.3
LS (Feb)	7.0	42	35	11.6	0.0	2.3	2.3
CO	5.9	53	12	11.8	5.9	0.0	11.8
HI	2.3	41	27	9.1	9.1	0.0	11.4
OR	2.7	51	22	8.1	5.4	2.7	8.1
CT	0.0	22	24	12.2	4.9	7.3	29.3
MD	0.0	50	29	11.8	2.9	2.9	2.9
TX	9.7	45	23	12.9	0.0	3.2	6.5

* 6 Missing values are not included in this table
Some Observations were rounded up from half days

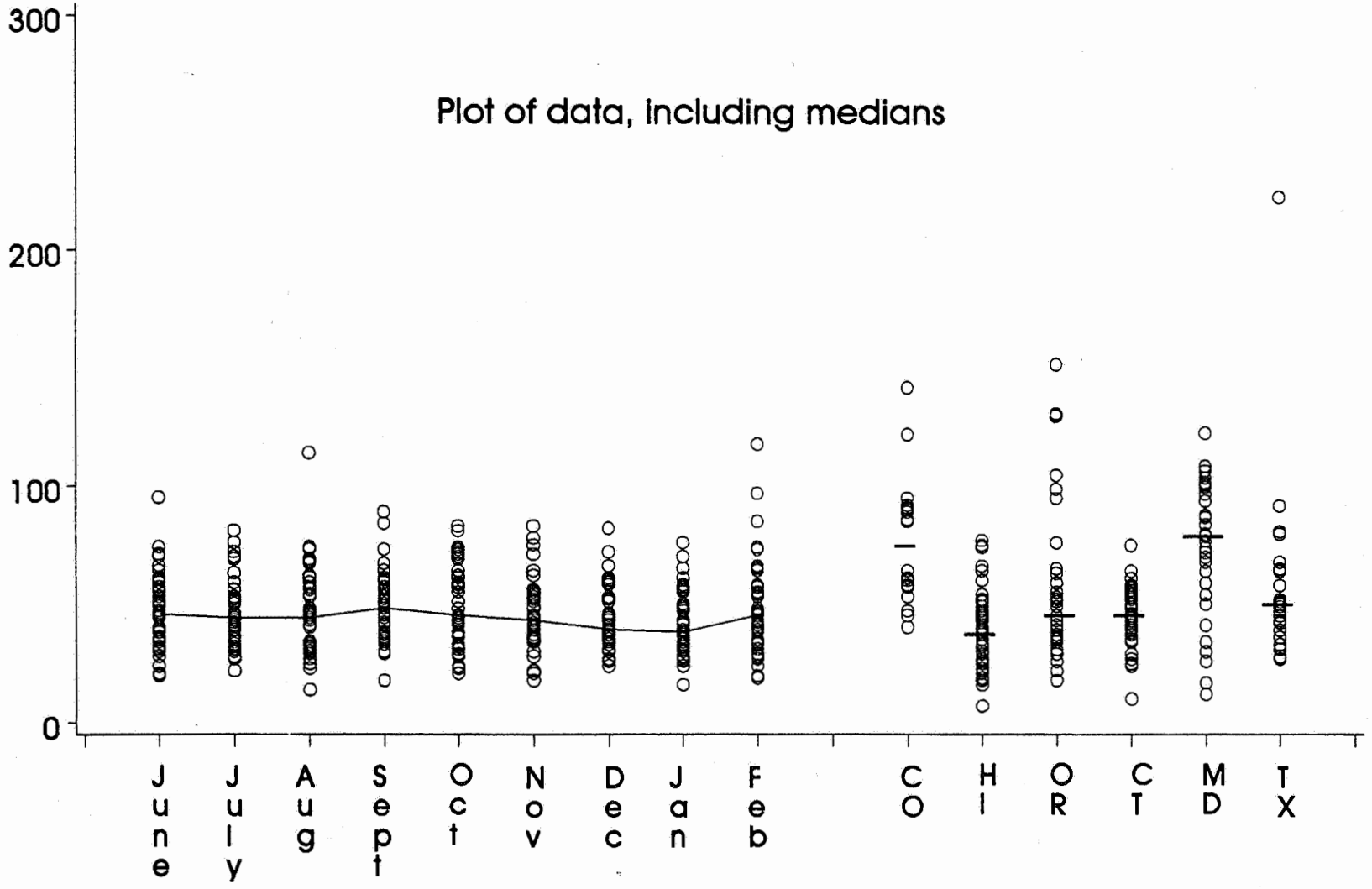


Figure 1. Sample Age
The age of the semen sample in minutes from the time of ejaculation to video recording for all semen samples distributed by study. June-Feb represent the Longitudinal Study. CO, HI, OR, CT, MD, and TX are field studies.

Figure 2

Plot of data, including medians

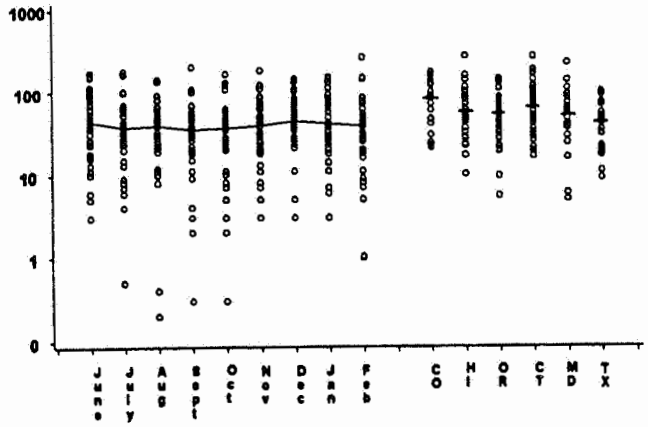


Figure 3

Confidence intervals for ratios of means

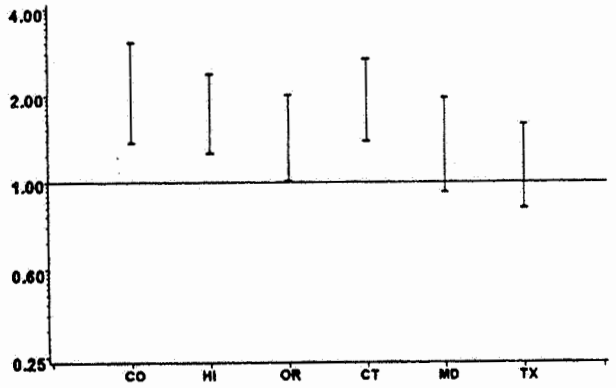
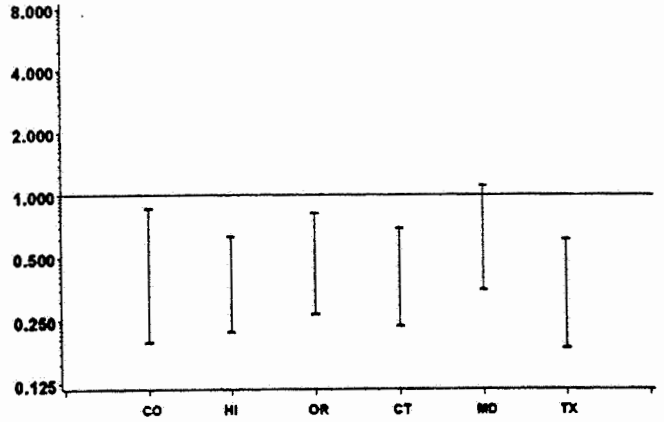


Figure 4

Confidence intervals for ratios of variances



Sperm Concentration (millions/ml)

June-Feb represent the Longitudinal Study. CO, HI, OR, CT, MD and TX are field studies.

Figure 2: The distribution of sperm counts for all semen samples (o) distributed by study and median count (-) for each study. (note: y axis is a log scale).

Figure 3: Confidence intervals for ratios of sperm count medians. The solid line at 1.00 represents the Longitudinal Study. The vertical bar represents the confidence interval for each study.

Figure 4: The confidence intervals for ratios of variances after log transformation for average count. The solid line at 1.00 represents the Longitudinal Study. The vertical bar represents the confidence interval for each study.

(Jeyendran et al, 1984). These techniques test for the structural and functional integrity of the membrane, respectively (Schrader et al, 1986).

Sperm count (sperm/ml) was measured in a Makler chamber (Makler 1980). Sperm count was calculated by averaging the count of two groups of ten Makler squares. Sperm morphology was measured after the preparation of air dried and stained smears, and sperm were classified into nine categories according to Zaneveld/Polakoski (1977). This classification system is essentially identical to that recommended by the World Health Organization (Belsey et al, 1988), and consists of normal forms (oval heads and normal tails), microcephalic heads, macrocephalic heads, tapered heads, absent heads, amorphous heads, double heads, abnormal tails, and immature forms. Two hundred sperm were read on each slide. Two slides were read for each semen sample. Motility assessments were conducted in each study, however, several different computer assisted sperm analyses systems and programs were used, making comparisons between studies impossible.

3. Statistical Analysis.

Statistical comparisons were made between each of the six control groups and the longitudinal group (e.g. CO vs LS, HI vs LS). Both the means and the variances were compared. For both the means and the variances a 95% confidence interval was computed for the ratio between the control group and the LS group. The ratio was selected because it allows variables on different scales to be compared without regard to units of measure. All ratios and their respective confidence intervals are plotted on a log scale. The use of a log scale stretches the values between 0 and 1 and compresses the values between 1 and infinity to make an appropriate symmetry between a ratio and its inverse. No adjustments were made for multiple comparisons (i.e., the six controls or the seven variables). It was necessary to calculate the degrees of freedom for the LS data since a nested analysis was used. A Satterthwaite procedure (Searle 1979) provided approximate degrees of freedom for the LS data.

Confidence intervals for a ratio of means were performed according to an adaptation of Fieller's Theorem (Casella/Berger 1990). The adaptation is included in the appendix. The confidence intervals for the ratios of means for oval morphology, pH, HOS percent, vital stain and volume were calculated under the assumption that these are normally distributed populations. This was verified using normal probability plots. Oval morphology and HOS percent showed some minor deviations from normality, so these confidence intervals should be used with caution.

The variables average count and sperm count per ejaculate are not normally distributed so a log transformation was used in the calculations of confidence intervals. To obtain a confidence interval for the ratio of means, a confidence interval for the difference of the means of the logs was computed. The endpoints were then transformed back to the original scale.

The confidence intervals for the ratios of variances were computed by the usual F-distribution method for normal populations. Since average sperm count and sperm count per ejaculate were not normally distributed, variances of the log transformed data were computed.

Figure 5

Plot of data, including medians

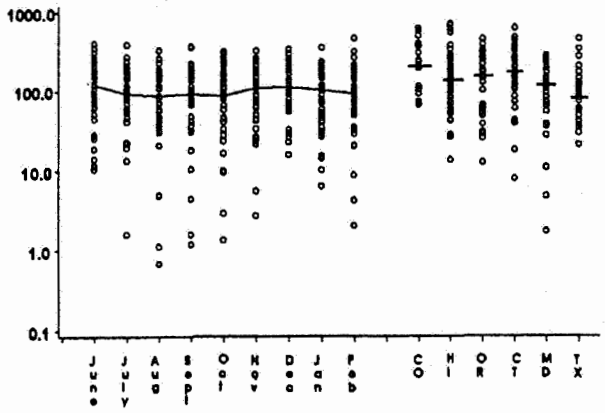


Figure 6

Confidence intervals for ratios of means

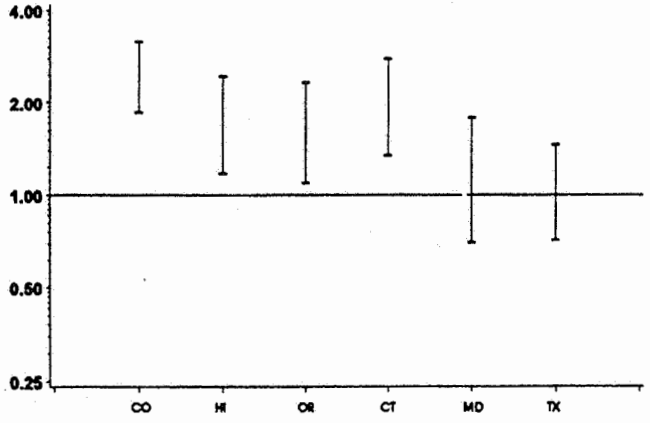
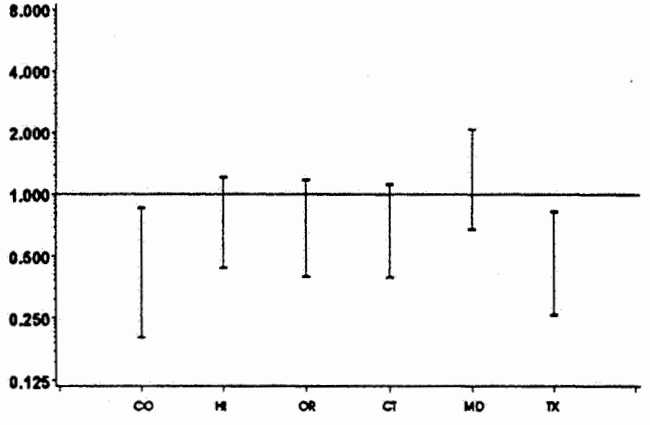


Fig 7

Confidence intervals for ratios of variances



Sperm Count (millions/ejaculate)

June-Feb represent the Longitudinal Study. CO, HI, OR, CT, MD and TX are field studies.

Figure 5: The distribution of the sperm count per ejaculate (o) for LS and each of the field studies. The median is denoted by the solid line.

Figure 6: The confidence intervals for the sperm/ejaculate medians. The solid line at 1.00 represents the Longitudinal Study. The vertical bar represents the confidence interval for each study.

Figure 7: The variances after log transformation for sperm count per ejaculate. The solid line at 1.00 represents the Longitudinal Study. The distribution bar represents the confidence interval for each study.

Since there were two components of variance for LS, it is not immediately apparent which variance ought to be used as the basis of comparison to each of the control groups. We selected Total Variance, which is defined as the sum of Within and Between Variances. A major use of variances in a control group or in the LS study is for sample size determination. Unless a study requires more than one measurement per subject, the Total Variance would serve as input into calculations of the required sample size. In addition, data from any one month of the longitudinal study with variation equivalent to Total Variance, can be considered comparable to a control group in a future cross-sectional study. Thus, if the Total Variance in the LS is larger than the variance in a control group, that indicates that the LS would provide more conservative (i.e., larger) estimates of the required sample size.

Results

The men were asked to have two days sexual abstinence before collecting the semen samples for analysis. Because there was no method of verifying abstinence, there was no penalty for shorter or longer times, but instead the importance of truthful reporting was emphasized. Table 1 provides the distribution of sexual abstinence for all studies.

The men were instructed to deliver the semen sample to the laboratory within 1 hour of ejaculation. Again, this could not be verified and the need for recording the actual time of ejaculation was emphasized. The age of the semen sample in minutes from the time of ejaculation to video recording for all semen samples distributed by study is illustrated in Figure 1.

The confidence intervals by study for each variable were graphically displayed. The confidence intervals shown on the Figures were for the ratios of the value (either mean or variance) for the particular field study to the value for the longitudinal study. The vertical line segments on the graphs connected the endpoints for the confidence intervals. The location of a confidence interval with respect to the horizontal line at the value one indicated how the field value compares to the longitudinal study value. If the line segment was completely above the horizonal line, there is 95% confidence that the field mean or variance was greater than the longitudinal mean or variance. The reverse holds if the line segment is completely below the horizontal line. If the confidence interval line segment crosses the horizontal line at one, then the ratio was not significantly different from one, and the two means or variances were not significantly different.

The width of the confidence interval is also important. A short line segment corresponds to a narrow confidence interval. This implied that small differences between a control population and the longitudinal study could be detected statistically. If the interval was wide, only large differences could be detected. These intervals were presented on a log scale, therefore, must be compared in relation to each other rather than to a specific variable value. The confidence intervals for ratios of means are shown on the graphs using a scale from 0.25 to 4 and the confidence intervals for ratios of variances are shown using a scale from 0.125 to 8.

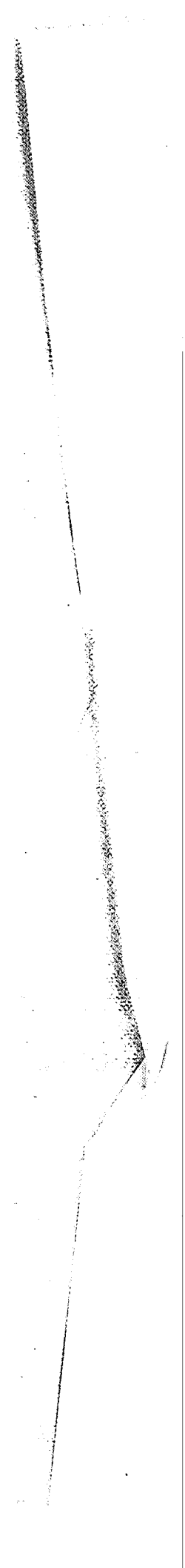


Figure 11

Plot of data, including medians

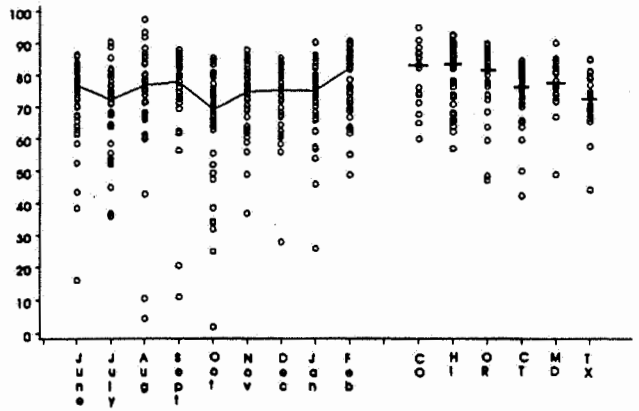


Figure 12

Confidence intervals for ratios of means

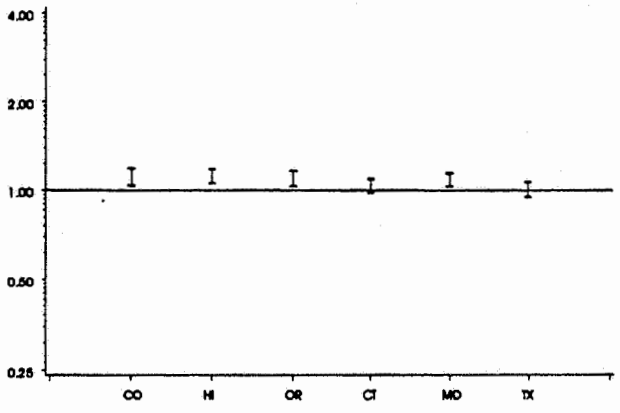
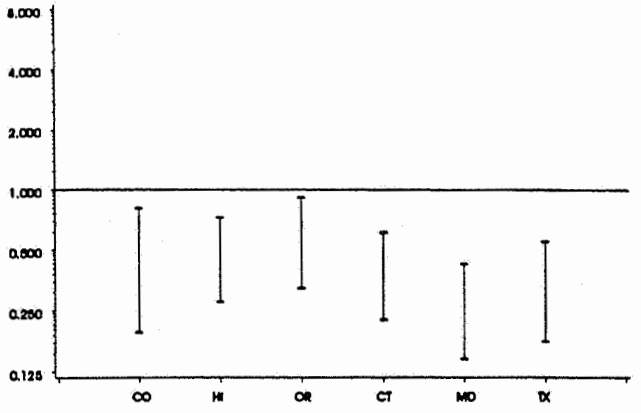


Figure 13

Confidence intervals for ratios of variances



Percent Normal Morphology

June-Feb represent the Longitudinal Study. CO, HI, OR, CT, MD and TX are field studies.

Figure 11: The distribution of percent normal morphology for all semen samples (o) distributed by study and the median count (-) for each study.

Figure 12: The confidence intervals for percent normal morphology means. The solid line at 1.00 represents the Longitudinal Study. The vertical bar represents the confidence interval for each study.

Figure 13: The variances normal sperm morphology. The solid line at 1.00 represents the Longitudinal Study. The distribution bar represents the confidence interval for each study.

1. Sperm Count.

A. Sperm per ml:

The mean sperm count for all samples of LS was 47.43 million sperm per ml with a standard deviation of 29.47 between men and 22.90 within men. The distribution of sperm counts for all semen samples distributed by study and the median count for each study are shown in Figure 2. The y axis is a log scale to better illustrate the distribution).

Figure 3 indicates that four of the field studies had a higher average count than the longitudinal study. The means for Maryland and Texas are not significantly different from the longitudinal study. The wide confidence intervals illustrate limited ability to detect changes between the control groups and the longitudinal study.

Variances (after log transformation) were used for average count since the data do not follow a normal distribution. Figure 4 shows that the field studies except for Maryland tend to be less variable than the longitudinal study for average count. These intervals are also fairly wide.

B. Count per ejaculate:

The distribution of the sperm count per ejaculate for the longitudinal study and each of the field study populations are illustrated in Figure 5. The median for each month of the LS and for each field study is denoted by the solid line. The confidence intervals for the means again were above one for Colorado, Hawaii, Oregon, and Connecticut (Figure 6). Means for Maryland and Texas were not significantly different from the longitudinal study mean.

The confidence intervals for the ratios of variances are ratios of the variances computed on the log scale, since the count variables do not follow a normal distribution (Figure 7). Variances for the field studies were not distinguishable from the longitudinal study except for the Colorado and Texas studies, which had less variance than the longitudinal study. These intervals are also fairly wide.

2. Semen Volume.

The mean volume for LS was 2.78 ml with a standard deviation of 0.91 between men and 0.74 within men. The distribution and median semen volume for each study are plotted in Figure 8. None of the field studies were significantly different from the longitudinal study (Figure 9). The intervals were narrower than the intervals for the count variables, indicating greater ability to detect small differences between the control studies and the longitudinal study. Variances for the field studies do not differ significantly from the variance for the longitudinal study (Figure 10). These intervals are fairly wide.

Figure 14

Plot of data, including medians

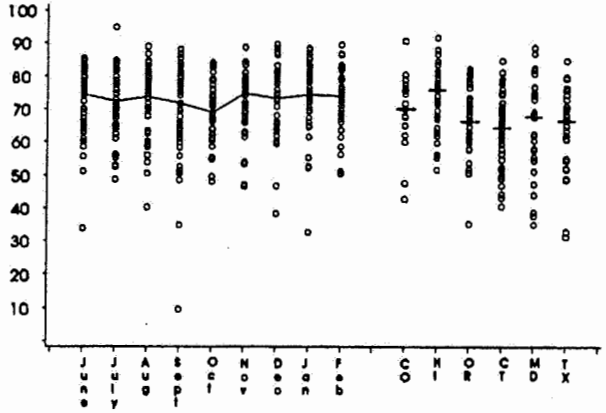


Figure 15

Confidence intervals for ratios of means

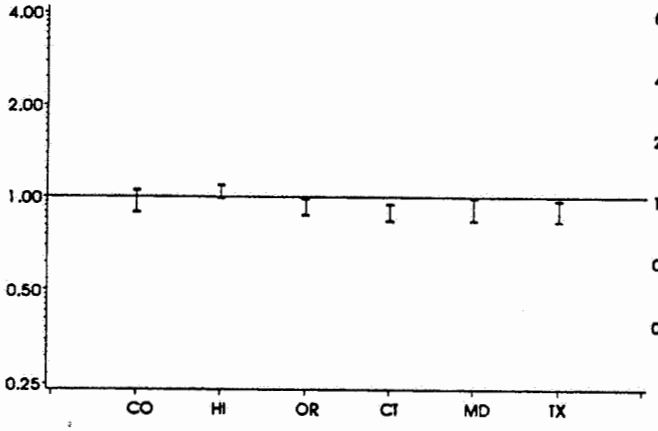
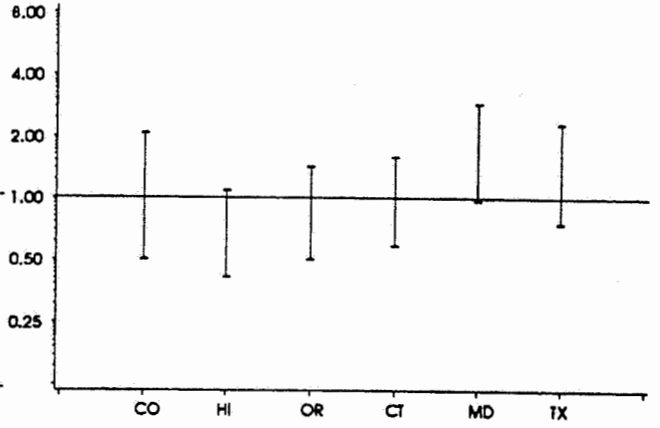


Figure 16

Confidence intervals for ratios of variances



Percent Viable - Hypoosmotic swelling

June-Feb represent the Longitudinal Study. CO, HI, OR, CT, MD and TX are field studies.

Figure 14: The distribution of percent alive (HOS) for all semen samples (o) distributed by study and the median count (-) for each study.

Figure 15: The confidence intervals for percent alive (HOS) means. The solid line at 1.00 represents the Longitudinal Study. The vertical bar represents the confidence interval for each study.

Figure 16: The variances percent alive (HOS). The solid line at 1.00 represents the Longitudinal Study. The vertical bar represents the confidence interval for each study.

3. Sperm Morphology.

Figure 11 illustrates the distribution of the proportion of sperm with normal morphology for each study and the median values. Connecticut and Texas were the only states which were not significantly different from the longitudinal study, although these differences were small (Figure 12). The other four studies had slightly higher means. The variances for the field studies are lower than the variance for the longitudinal study (Figure 13). All of these intervals were rather wide.

4. Sperm Viability.

A. Hypoosmotic swelling:

Figure 14 illustrates the distribution for each study of proportion of swollen sperm using the HOS assay. Median values are also indicated. Figure 15 shows that the field study means are close to or, for Colorado and Hawaii, not significantly different from the mean for the longitudinal study. The confidence intervals for the variances were rather wide (Figure 16). However none of the studies has a variance that is significantly different from the longitudinal study variance.

B. Vital stain:

Figure 17 illustrates the distribution for each study of the proportion of unstained sperm using the vital stain assay. The confidence intervals for the study means are shown in Figure 18. Colorado, Hawaii, Oregon and Texas have slightly higher means than the longitudinal study for vital stain. Connecticut and Maryland field studies were not significantly different from the longitudinal study.

The variance estimates for Hawaii, Connecticut and Maryland were not significantly different from the longitudinal study (Figure 19). The other states were somewhat less variable than the longitudinal study. These intervals were fairly wide.

5. pH:

The Distribution of semen pH for each study is illustrated in Figure 20. The medians are illustrated by the solid line. The field study ratios of means for pH are narrow and approach unity (Figure 21). All of the confidence intervals were narrow, and the mean pH values for Colorado, Oregon, and Texas are not significantly different from the mean pH for the longitudinal study.

For the variances for pH, Oregon and Connecticut field data were less variable than the longitudinal study (Figure 22). Variances for the other states were not significantly different from the longitudinal study. The variance intervals again are fairly wide.

Figure 17

Plot of data, including medians

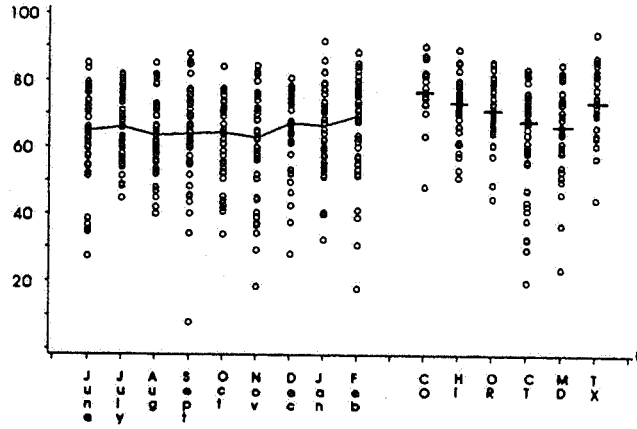


Figure 18

Confidence intervals for ratios of means

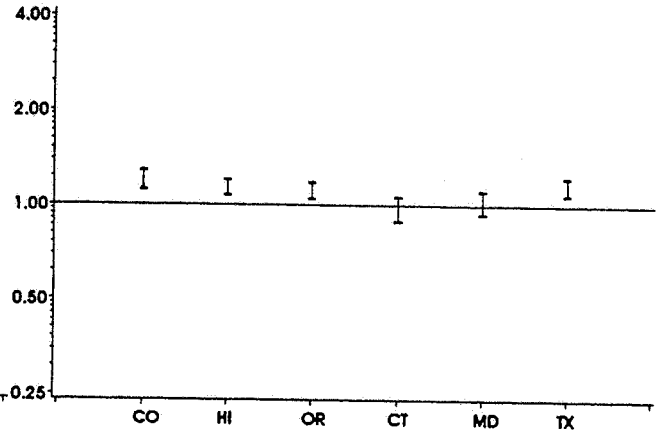
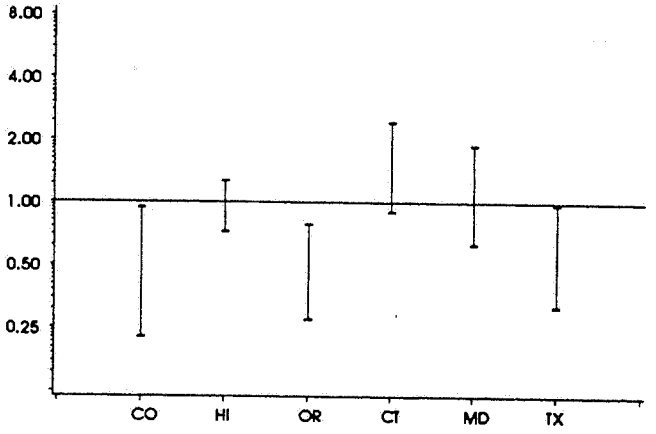


Figure 19

Confidence intervals for ratios of variances



Percent Viable - Vital Stain

June-Feb represent the Longitudinal Study. CO, HI, OR, CT, MD and TX are field studies.

Figure 17: The distribution of percent alive (vital stain) for all semen samples (o) distributed by study and the mean count (-) for each study.

Figure 18: The confidence intervals for percent alive (vital stain) means. The solid line at 1.00 represents the LongitudinalStudy. The vertical bar represents the confidence interval for each study.

Figure 19: The variances percent alive (vital stain). The solid line at 1.00 represents the Longitudinal study. The vertical bar represents the confidence interval for each study.

Figure 20

Plot of data, including medians

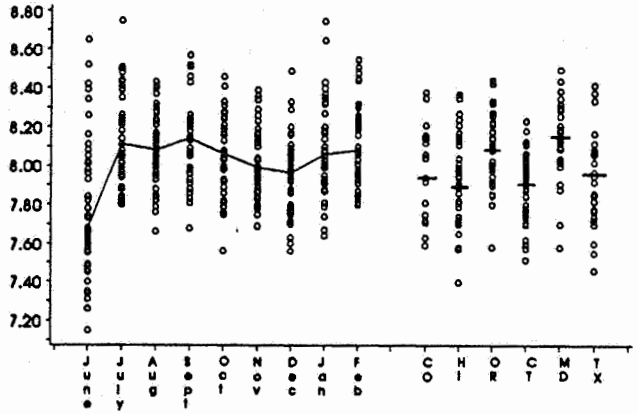


Figure 21

Confidence intervals for ratios of means

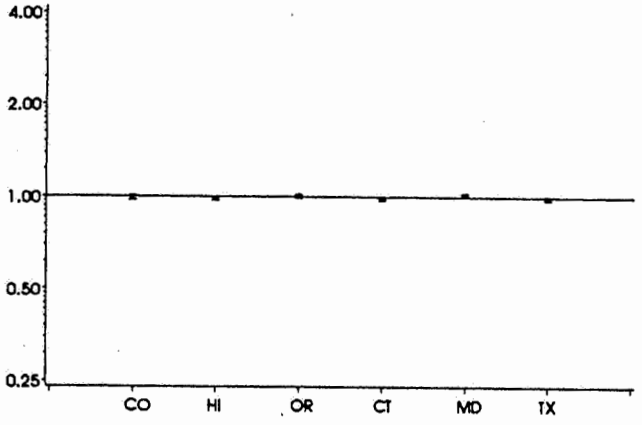
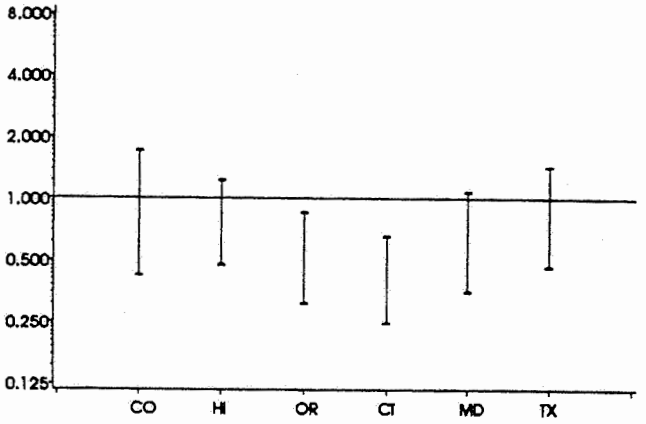


Figure 22

Confidence intervals for ratios of variances



Semen pH

June-Feb represent the Longitudinal Study. CO, HI, OR, CT, MD and TX are field studies.

Figure 20: The distribution of pH for all semen samples (o) distributed by study and the median count (-) for each study.

Figure 21: The confidence intervals pH means. The solid line at 1.00 represents the Longitudinal Study. The vertical bar represents the confidence interval for each study.

Figure 22: The variances for pH. The solid line at 1.00 represents the Longitudinal Study. The vertical bar represents the confidence interval for each study.

Discussion

Examination of control populations is very important in planning future studies. Estimates of variability in the control population are needed to determine the power and/or sample size requirements of a new study. Estimates of variability may also assist in selecting the most useful indicators of male reproductive function. Furthermore, the mean level in control populations also provides an informal benchmark for future controls. If the mean level of a new control group differs sharply from previous controls then an examination of that new control group for unsuspected exposures might be warranted.

The average sperm count and average count per ejaculate was substantially lower for LS than for all other control populations. With the large variability inherent in these two measurements, however, the MD and TX control groups are not statistically different from LS. The variance for sperm count and count per ejaculate are generally larger for LS, although again a few groups do not differ in a statistical sense. This implies that estimates of power using the LS data might be low, leading to a slightly larger than necessary sample size.

There is good consistency between the LS and the field controls in the remaining variables. Although the mean levels do sometimes differ significantly, there is no consistent deviation among these variables. This reinforces the utility of LS for representing control populations in general. For all variables, the variances for each of the control populations tend to be smaller than those of LS, although this difference is not always statistically significant. The LS data gives a more conservative estimate of sample size because it accounts for additional variation.

The longitudinal study of unexposed Cincinnati men is very useful in calculating power/sample size and serves as a benchmark for new controls. The comparison of the other control groups to this group reinforces its quality and utility in these analyses.

References

Belsey MA, Moghissi KS, Eliasson A, Paulsen CA, Gallegos AJ, Prasal MRN (1988) Laboratory manual for the examination of human semen and semen cervical mucus interaction. Press Concern (Singapore)

Bostofte E, Serup J, Rebbe H (1983) Has the Fertility of Danish Men Declined through the Years in Terms of Semen Quality? A Comparison of Semen Qualities Between 1952 and 1972. *Inter J Fert* **28**:91-95

Carlsen E, Giwercman A, Keiding N, Skakkebaek N E (1992) Evidence for Decreasing Quality of Semen During Past 50 Years. *Brit Med J* **305**:609-613

Casella G, Berger RL (1990) Statistical Inference, Wadsworth, Inc., Belmont, CA.

David G, Jouannet P, Martin-Boyce A, Spira A, Schwartz D (1979) Sperm Counts in Fertile and Infertile Men. *Fertil Steril* **31**:453-455

Eliasson R, Treichl L (1971) Supravital Staining of Human Spermatozoa. *Fertil Steril* **22**:134-137

Evenson DP, Jost LK, Baer RK, Turner TW, Schrader SM (1991) Individuality of DNA Denaturation Patterns in Human Sperm as Measured by the Sperm Chromatin Structure Assay. *Reprod Toxic* **5**:115-125

Falk HC, Kaufman SA (1950) What Constitutes Normal Semen? *Fertil Steril* **1**:489-503.

Fariss BL, Fenner DK, Plymate SR, Brannen GE, Jacob WH, Thomsaon AM (1981) Seminal Characteristics in the Presence of a Varicocele as Compared with Those of Expectant Fathers and Prevasectomy Men. *Fertil Steril* **35**:325-327

Grajewski BC, Cox C, Schrader SM, Murray WE, Edwards RM, Turner TW, Smith JM, Shekar SS, Evenson DP, Simon SD, Conover DL (in preparation) Semen Quality of Radiofrequency Heat Sealer Operators.

James WH (1980) Secular Trend in Reported Sperm Count. *Andrologia* **12**:381-388

Jeyendran RS, Van den Ven HH, Perez-Palaez M, Crabo BG, Zaneveld LJD (1984) Development of an Assay to Assess the Functional Integrity of the Human Sperm Membrane and Its Relationship to Other Semen Characteristics. *J Reprod Fert* **70**:219-228

Macleod J (1951) Semen Quality in One Thousand Men of Known Fertility and in Eight Hundred Cases of Infertile Marriage. *Fertil Steril* **2**:115-139.

Makler A (1980) The improved ten-microliter chamber for rapid sperm count and motility evaluation. *Fertil Stertil* **33**:337-338.

Mallidis C, Howard EJ, Baker HWG (1991) Variation of Semen Quality in Normal Men. *InternationalJournal of Andrology* **14**:99-107

National Institute for Occupational Safety and Health. (1988) National Strategies for the Prevention of Leading Work-Related Diseases and Injuries: Disorders of Reproduction. *DHHS(NIOSH) Publication* No. 89-133.

Nelson CKM, Bunge RG (1974) Semen Analysis: Evidence for Changing Parameters of Male Fertility Potential. *Fertil Steril* **26**:503-507

Osser O, Liedholm P, Ranstam J Depressed Semen Quality: A Study Over Two Decades. *Arch Andrology* **12**:113-116

Ratcliffe JM, Schrader SM, Steenland K, Clapp DE, Turner TW, Hornung RW (1987) Semen quality in papaya workers with long term exposure to ethylene dibromide. *Brit J Ind Med* **44**:317-326

Ratcliffe JM, Schrader SM, Clapp DE, Halperin WE, Turner TW, Hornung RW. (1989) Semen Quality in Workers Exposed to 2-Ethoxyethanol. *Brit J Ind Med* **46**:399-406.

Rehan NE, Sobrero AJ, Fertig JW (1975) The Semen of Fertile Men: Statistical Analysis of 1300 Men. *Fertil Steril* **26**:492-502

Schrader SM, Breitenstein M, Turner TW, Simon SD. (1992) Longitudinal Study of Semen Quality of Unexposed Workers: Seminal Plasma Characteristics. In: Comparative Spermatology, 20 Years After. B Baccetti (ed). Serono Symposia Publications. Volume 75:905-908. Raven Press, New York

Schrader SM, Platek SF, Zaneveld LJD, Perez-Pelaez M, Jeyendran RS (1986) Sperm Viability: A Comparison of Analytical Methods. *Andrologia* **18**:530-538

Schrader SM, Turner TW, Baeiterstein MJ, Simon SD (1988a) Longitudinal Study of Semen Quality of Unexposed Workers: *I Study Overview Reprod Toxicol* **2**:183-190.

Schrader SM, Turner TW, Ratcliffe JM (1988b) The Effects of Ethylene Dibromide on Semen Quality: A Comparison of Short Term and Chronic Exposure. *Reprod Toxic* **2**:191-198.

Schrader SM, Turner TW, Simon SD (1990) Longitudinal Study of Semen Quality of Unexposed Workers - Sperm Head Morphometry. *J Androl* **4** 11:32-39.

Schrader SM, Turner TW, Simon SD (1991) Longitudinal Study of Semen Quality of Unexposed Workers: Sperm Motility Characteristics. *J Androl* **12**:126-131.

Schwartz D, Laplanche A, Jouannet P, David G (1979) Within Subject Variability of Human Semen in Regard to Sperm Count, Volume, Total Number of Spermatozoa and Length of Abstinence. *J Reprod Fert* **57**:391-395

Schwartz D, Mayauz MJ, Spira A, Moscato ML, Jouannet P, Czgilk F, David G (1983) Semen Characteristics as a Function of Age in 833 Fertile Men. *Fertil Steril* **39**:530-535.

Searle SR (1979) Linear Models, Wiley and Sons, Inc., New York, NY.

Sobrero AJ, Rehan NE (1975) The Semen of Fertile Men. II. Semen Characteristics of 100 Fertile Men. *Fertil Steril* **26**:1048-1056

Wang C, Chan SYW, Leung A, Ng RP, Ng M, Tang LCH, Ma HK, Tsoi WL, Kwan M (1985) Cross-sectional Study of Semen Parameters in a Large Group of Normal Chinese Men. *Int J Androl* **8**:257-274

Welch LS, Schrader SM, Turner TW, Cullen MR (1988) Effects of exposure to ethylene glycol ethers on shipyard painters: I. Male Reproduction. *Am J Ind Med* **14**:509-526.

Weyandt TB, Schrader SM, Turner TW, Simon SD (submitted) Semen Analysis of Military Personnel Associated with Military Duty Assignments. *Reprod Toxicol*

Zaneveld LJD, Polakoski KL (1977) Collection and physical examination of the ejaculate. In: *Techniques of Human Andrology*. Hafez ESE (ed) 147-172, Elsevier, NY

Appendix

Let **A** and **B** be two independent random variables which are normally distributed with means, μ_A and μ_B, and with variances, σ^2_A and σ^2_B, respectively. In our application **A** represents the mean of CO, HI, OR, CT, MD, or TX and **B** represents the mean of LS. We wish to create a confidence interval for the ratio of the two population means, μ_A/μ_B. The material below very closely parallels the proof of Fieller's theorem, as described in [37], pages 459-460.

Define

$$\theta = \mu_A/\mu_B, \text{ and}$$

$$\mathbf{C} = \mathbf{A} - \theta\,\mathbf{B}.$$

Note that **C** is normally distributed with mean zero and variance

$$\sigma^2_C = \sigma^2_A + \theta^2\,\sigma^2_B.$$

The set

$$\{\theta;\ C^2/\sigma^2_C \le z^2_{\alpha/2}\}$$

is a $(1-\alpha/2)$ confidence interval for θ, where $z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution. This set is equivalent to the following set based on a quadratic inequality in θ

$$\{\theta;\ (B^2 - z^2_{\alpha/2}\,\sigma^2_B)\,\theta^2 - (2\,A\,B)\,\theta + (A^2 - z^2_{\alpha/2}\,\sigma^2_A) \le 0\}.$$

There are several conditions where a quadratic inequality could yield an interval of infinite length, the union of two intervals of infinite length, or the empty set. If we assume, however, that $(A^2 - z^2_{\alpha/2}\,\sigma^2_A)$ and $(B^2 - z^2_{\alpha/2}\,\sigma^2_B)$ are both positive, then the solution to the above quadratic inequality is a nice finite interval. Many of the infinite length intervals or other pathological cases occur when μ_B is close to zero, relative to σ^2_B, which implies, intuitively, that the confidence interval is unstable when the denominator is small relative to its variability.

For all of the confidence intervals computed in this paper, the above conditions are met, insuring a finite length interval.

The above interval requires one more modification. In practice, σ^2_A and σ^2_B are often unknown and must be estimated using S^2_A and S^2_B, respectively. If the sample sizes are large enough, then one still may use a standard normal percentile, since there is not sufficient variation in S^2_A and S^2_B to justify using a t-distribution. Alternately, one can apply the Satterthwaite approximation, as described in [37] pages 287-289, to the quantity S^2_C, which is a linear combination of two Chi-Square distributions (i.e., $S^2_A + \theta^2\,S^2_B$). This is further complicated by the fact that S^2_B is itself a linear combination of two Chi-Square distributions, since it reflects both between and within subject variability from the longitudinal study. The computations for the Satterthwaite approximation are tedious but not difficult. They yield an approximate degrees of freedom; the **Z** percentile in the quadratic inequality above is then replaced with the appropriate **t** percentile.