Journal of Andrology, Vol. 23, No. 5, September/October 2002

Optimal Power Transformations for Analysis of Sperm Concentration and Other Semen Variables

DAVID J. HANDELSMAN

From the ANZAC Research Institute and Department of Andrology, Concord Hospital, University of Sydney, Sydney, Australia.

ABSTRACT: The nongaussian (or nonnormal) distribution of sperm concentration, and variables deriving from it, is a common practical problem in the statistical evaluation of semen data. Yet it has been little studied, and its importance to data analysis, as well as to practical remedies, is not widely appreciated. Inappropriate use of the raw scale of measurement produces inflated estimates of mean and variance, leading to false-negative (underpowered) statistical comparisons and excessive sample size estimates. This study employs the Box-Cox family of power transforms to illustrate by a simple graphical method how to identify optimal power transforms for semen data variables. Using robust statistical methods, it is shown that the nongaussian distribution is due to right skewing rather than multimodality or influential outliers. The optimal power transform, typically in the region of 0.15 to 0.35 (most easily implemented as a cube-root transformation), usually performs better than the logarithmic transformation in normalizing the data. In addition, the power transformation has an important practical advantage over the logarithmic transformation in the appropriate handling of zeros (azoospermia), a regular and important features of such data sets in practice.

Key words: Statistics, data transformation, male fertility.

J Androl 2002;23:629–634

Following the first report of semen analysis in 1929 (Macomber and Sanders, 1929), the first systematic study of large numbers of human semen samples was published in 1951 (MacLeod and Gold, 1951). In that classical study, John MacLeod involved a statistician, Ruth Gold, as coauthor. Since then, it has been well known, and readily confirmed, that sperm concentration is consistently nongaussian in distribution. Although this invalidates parametric statistical analyses in the natural scale, there has been limited consideration of the problems this creates for semen data analysis (Berman et al, 1996), and the literature demonstrates continuing widespread and uncritical usage of inappropriate statistical methods.

The distribution of sperm concentrations is typically nongaussian. This is primarily because of marked right (positive) skewing and, to a lesser extent, because sperm concentration is constrained to be nonzero. The nongaussian nature can be readily identified because, in the natural scale, the lower 95% confidence limits of the distribution (2 SDs below the mean) give biologically meaningless negative values. The arithmetic mean is characteristically vulnerable to excessive influence by high outlying values. Right skewing typically increases the arithmetic mean so that it deviates widely from the median, the value usually regarded as the best estimate of the center of the distribution. Furthermore, right skewing also inflates sample-based estimates of variance to create a falsely high standard deviation. This weakens the ability to test for real differences between group means, as these differences are compared with an unrealistically large pooled standard deviation. For the same reason, skewing also significantly distorts study power estimation, producing excessively large estimates of sample size.

Several remedies for this nongaussian distribution of semen variables are available. One traditional (Gaddum, 1945) and widely used method is the logarithmic transformation (Berman et al, 1996). This approach is often effective and is a great improvement on raw data analyses, although this hardly justifies its automatic application, as advocated by some (Keene, 1995). Log transformation has the drawback of being unable to deal properly with zeros (azoospermia), which are a frequent and important endpoint for many studies involving semen analysis data such as male infertility and contraception. In addition, the log transform makes accurate graphical representation of the data, consistent with the analysis, difficult, since a zero sperm output cannot be indicated on the log scale. Typically, an ad hoc fix is to add an arbitrary offset to each zero data point, rendering them positive and thereby remedying the inability to perform the log transform; however, altering data for analysis is both undesirable in principle and may distort the data analysis in practice, especially if the data include many zeros. In this way, the statistical analysis of a data set could be

---

Correspondence to: Prof David J. Handelsman, ANZAC Research Institute, Sydney NSW 2139, Australia (e-mail: djh@anzac.edu.au).

Received for publication May 22, 2000; accepted for publication March 11, 2002.

influenced by the actual value of the arbitrarily selected offset. Another alternative is the resort to nonparametric statistics. Although effective and often useful, this has the limitation of sacrificing both statistical power (especially for small sample sizes that are common in andrology) and flexibility for more complex statistical modeling, since the most sophisticated parametric statistical analyses (but those that require parametric assumptions) are not all available for nonparametric analysis. A third, more satisfactory but little-appreciated approach is the use of a power transformation. Normalizing data transformations retain the power and flexibility of parametric statistical methods for valid application of parametric modeling on a scale that renders the data effectively gaussian. Unlike the log transformation, power transformations preserve the ability to deal appropriately with zeros while simultaneously retaining access to the more powerful and flexible parametric statistical methods. Empirical applications of the logarithmic, square-root, and cube-root transformations for this purpose have been adopted for many years, but there has been no systematic study or general framework validating or comparing such empirical approaches.

This study develops a systematic framework for valid parametric statistical analysis of seminal data based on applying the Box-Cox family of power transforms (Box and Cox, 1964) to nongaussian sperm data. Applications of the Box-Cox transformations have proved useful in other areas of biology (Peltier et al, 1998; Meloun et al, 2000). The cube-root transformation is shown to be a useful rule of thumb for seminal variables. This general approach leads to the valid application of parametric statistical methods while handling azoospermia appropriately, and a simple graphical method is demonstrated to identify empirically optimal power transforms in any semen data set.

Materials and Methods

Primary Data and Variables

The main data for this study were derived from semen analysis results from 469 consecutive unselected healthy men screened as potential sperm donors in Sydney (Handelsman et al, 1984; Handelsman, 1997). Briefly, healthy men recruited without regard to marital or fertility status being screened as potential sperm donors provided semen samples prior to acceptance or rejection as donors. An additional large data set was obtained from the first semen samples of 671 men participating in 2 large World Health Organization (WHO) multicenter male contraceptive efficacy studies organized by the WHO's Human Reproduction Program (1990, 1996). These samples came from 16 centers where men in stable relationships requiring contraception were recruited to enter the study if they had no history of infertility, chronic medical illness, or reproductive pathology and had 2 normal semen analyses. Unlike the first group, this population was selected for potentially normal fertility according to their fertility history and semen analysis results.

Semen samples collected by masturbation were examined according to the standard methods described in the contemporaneous edition of the WHO manual. From each semen sample, the volume of the ejaculate (in milliliters), sperm concentration (million sperm per milliliter), and motility (percentage moving or category ''a'' + ''b'') were determined by examination of the liquefied sample in a counting chamber. Morphology of spermatozoa is determined from a stained smear and expressed as a percentage of all ejaculated spermatozoa. For this study, only the first semen sample was analyzed. From the raw data variables of semen volume, sperm concentration, motility, and morphologically abnormal forms, derived variables can be calculated to reflect the concentrations and total output of all sperm, motile sperm, and morphologically normal sperm in the ejaculate.

Analytical Approach

Three distributional criteria—the coefficients of skewness (s) and kurtosis (k) and the Shapiro-Wilks W statistic—were used. The first two were employed as standardized normal deviates calculated as centered coefficients (subtract 3 for kurtosis and 0 for skewness) divided by their respective standard errors—Z_s for skewness and Z_k for kurtosis (Sokal and Rohlf, 1995). The W statistic, being inherently standardized, was used directly. The Shapiro-Wilks W statistic (Shapiro and Wilks, 1965) is considered the most powerful omnibus test (ie, a single measure sensitive to departures both from symmetry (skewness) and spread (kurtosis) of data distributions) for normality (Royston, 1982). It estimates the deviation from a gaussian distribution by comparing the sample values (arrayed in order) with the idealized gaussian distribution for the same sample size. Using an analysis of variance (ANOVA), the W value reflects the proportion of variance explained by the idealized gaussian distribution. Data distributions close to gaussian have a higher W, approaching the maximum value of 1.0, and lower W values indicate more extreme departures from a gaussian distribution. Optima for the power transforms were estimated by linear or quadratic regression, and confidence limits were calculated by backtransformation to the natural scale. All data were analyzed using the BMDP 2D (Dixon, 1992).

The Box-Cox family of power transformations (Box and Cox, 1964) was employed as defined by $X' = (X^k - 1)/k$ for $k > 0$, and, in the limit for $k = 0$, $X' = \log(X)$, where X is the raw data, X' is the transformed data, and k is a nonzero exponent ranging from 1.0 (untransformed data) to 0. For each variable, a series of power transforms from 0.05 to 1.0 was created in increments of 0.05. For each transformation and for each variable, the 3 distributional criteria were then calculated and graphed against the exponent.

Results

In the natural scale, all semen variables are severely nongaussian in distribution (Table). This is predominantly because of right skewing (nonsymmetrical), although there was also some kurtosis (flank bulge) (Sokal and Rohlf, *Location, dispersion and distributional effects of various transformations on semen analysis variables*

	Semen Volume	Sperm Density	Total Sperm	Motile Sperm Density	Total Motile Sperm	Normal Sperm Density	Total Normal Sperm
Optimal *k*	0.30	0.25	0.20	0.25	0.20	0.20	0.15
Location							
Median	3.0	74.0	208	46.8	132	41.7	125
Arithmetic mean	3.2	89.7	293	58.0	188	56.1	182
Geometric mean	2.8	69.9	198	42.2	120	39.1	112
Power mean*	2.9	74.9	216	46.3	137	42.4	122
Dispersion†							
Empirical	0.8, 7.4	6.0, 280	13.0, 1120	3.5, 176	7.7, 699	3.4, 203	8.6, 611
Natural scale	−0.1, 6.6	−34.7, 214	−275, 861	−29.3, 145	−186, 562	7.9, 104	−17, 381
Log scale	1.1, 8.7	15.8, 309	30.4, 1291	7.4, 242	15.0, 959	6.5, 236	13.7, 917
Power scale	0.8, 7.5	13.0, 251	23.7, 996	5.9, 179	10.0, 654	5.1, 187	11.0, 711
Distribution							
Skewness‡							
Natural scale	8.69	13.51	26.43	12.79	24.21	14.97	27.16
Log scale	−5.61	−5.79	−6.09	−8.59	−7.69	−4.51	−4.93
Power scale	−0.76	−0.07	0.72	−1.09	−0.32	−0.13	−0.48
Kurtosis‡							
Natural scale	5.96	13.01	61.90	11.07	52.23	18.18	71.67
Log scale	3.66	4.33	6.82	10.14	8.62	1.64	3.37
Power scale	−0.22	0.70	2.52	0.44	1.65	−0.45	0.79
W							
Natural scale	0.926	0.866	0.739	0.867	0.754	0.829	0.713
Log scale	0.965	0.962	0.969	0.952	0.958	0.967	0.971
Power scale	0.981	0.982	0.988	0.985	0.987	0.983	0.987

* Determined by backtransformation of mean of power scale distribution.

† 95% confidence limits determined from empirical or natural scale distribution or by backtransformation from log or power scales.

‡ Expressed as standardized normal deviates.

1995). Further examination of the distribution of sperm concentration showed that the asymmetrical deviation from gaussian was not due to excessively influential outliers, because none of 3 robust estimates of location (Hampel, 15% trimmed, and biweight) was any improvement over the arithmetic mean in its deviation from the median. Similarly, the empirical distribution gave no indication of multimodality as another potential explanation for the nongaussian distribution.

The 3 distributional criteria calculated for each of the 7 semen variables are illustrated (Table). For each variable, the raw data were nongaussian by all 3 criteria, and in each case, logarithmic transformation improved but did not achieve a gaussian distribution. Furthermore, for each variable, the convexity of the graph of *W* vs the exponent indicated that a power transformation always existed that was superior to logarithmic transformation.

For sperm concentration (Figure 1), as the exponent decreases from 1.00 (untransformed data) to 0.05, the rightward skew is progressively corrected to reach a gaussian distribution and then eventually to developing a left skew. Estimating the intersection of Z_s on k by linear regression ($r^2 = 0.997$) leads directly to a point estimate of 0.31 as the optimal power transform with approximately 67% confidence intervals (0.24–0.45). For kurtosis, the parabolic regression of Z_k on k ($r^2 = 0.991$) provided 2 optimal exponents ($k = 0.15$ and 0.53), giving a wide range of power transforms that rectified the relatively mild nonkurtosis. The *W* statistic displayed a smoothly parabolic function of the exponent with an optimum power transform represented by the maximum, estimated as $k = 0.37$ by quadratic regression.

Similar results were obtained for each of the other 6 variables that were rendered gaussian or nearly so by fractional power transforms (Table; Figure 2). Two other semen variables, sperm motility and morphology, both actually rescaled fractions, remained nongaussian despite power or angular transforms (data not shown).

The WHO data set had similar properties, with the exception that the improvement in performance of the optimal power transform over the logarithmic ($k = 0$) was smaller, presumably due to the exclusion by design of lower sperm concentrations (Figure 3).

Discussion

This study confirms that the underlying distribution of all conventional semen variables is intrinsically nongaussian

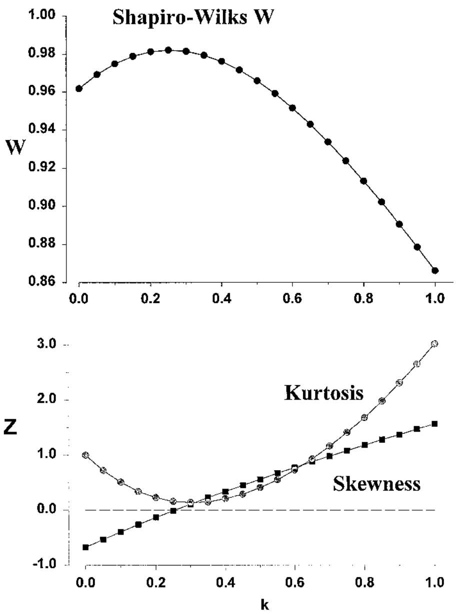

Figure 1. Plot of Shapiro-Wilks *W* statistic (upper panel) and skewness and kurtosis (lower panel) ranging from an exponent (*k*) of 1.0 (untransformed raw data) to 0 (logarithmic transformation). Note that the convexity of the *W* statistic demonstrates that a range of *k* provides more normalizing transformations than the log transform. In the upper panel, note optimum at *k* approximately equal to 0.37. In the lower panel, both skewness and kurtosis have optima near the same region as indicated for *W*.

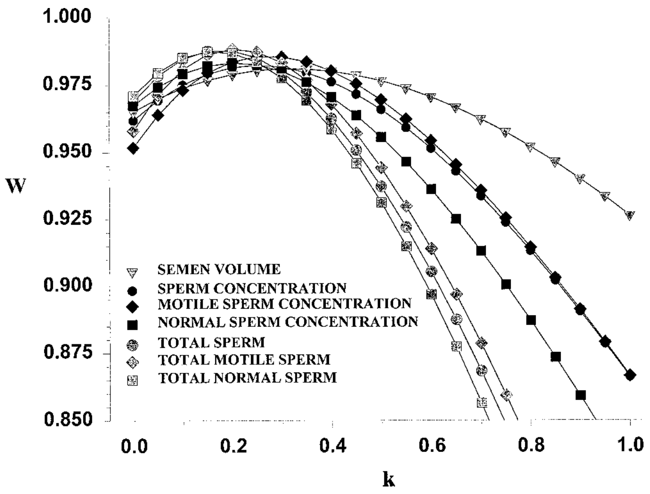

Figure 2. Plot of Shapiro-Wilks *W* statistic ranging from an exponent (*k*) of 1.0 (untransformed raw data) to 0 (logarithmic transformation). Note similarity of optimum *k* for all 7 sperm variables.

in the natural scales of measurement (MacLeod and Gold, 1951; Berman et al, 1996). Furthermore, the ubiquitous positive skewing is not attributable to multimodality or influential outliers, which is consistent with the concept that we are dealing with an intrinsic biological characteristic. Severe intrinsic skewing in the natural scale has important implications for valid statistical analysis and modeling of data derived from semen analysis. The severe, systematic distortion in estimates of location and of dispersion can invalidate frequently used parametric methods such as the *t* test, ANOVA, and linear regression. Despite this basic caveat on the use of parametric statistical analyses, misapplications are still common in the published literature.

In principle, the consistently nongaussian distribution of seminal variables could be handled by either a nonparametric approach or suitable normalizing transformations. Despite methodological advances, nonparametric methods still sacrifice power (particularly for small samples, as are most common in andrology) and flexibility (in not allowing access to the most sophisticated parametric methods generally preferred in practice). This has led to an ad hoc adoption of data transformations, most notably the logarithmic (Berman et al, 1996) but occasionally the square- and cube-root transformations (Bahamondes et al, 1979). The logarithmic transform, however, is not ideal, as it cannot directly handle zeros (azoospermia), a regular and important feature of male fertility studies. In many situations in andrology research, men may be rendered azoospermic (Handelsman, 2001), or treatment is designed to increase sperm output from azoospermia (Liu et al, 2002). Power transforms used previously have been used on a purely empirical basis without a coherent framework for their adoption or evaluation. The lack of a rational framework for valid statistical analysis of seminal data may have contributed to the limited use of both power calculations and effect size calculations in the design and analysis of clinical trials investigating various aspects of male fertility.

This study indicates that fractional power transforms can provide a consistent framework of normalizing trans-

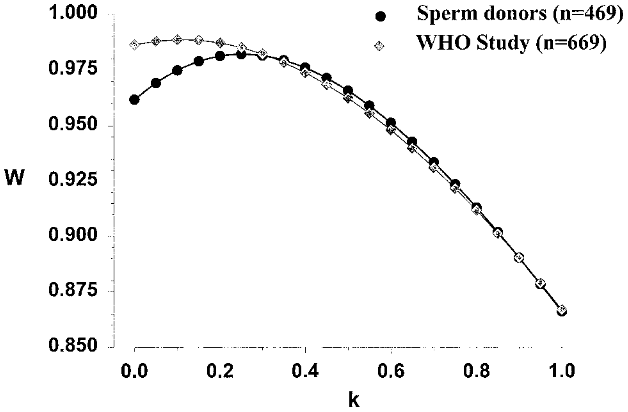

Figure 3. Plot of Shapiro-Wilks *W* statistic for a range of power transformations ranging from an exponent (*k*) of 1.0 (untransformed raw data) to 0 (logarithmic transformation). Data consist of sperm donors (filled circle) and World Health Organization (WHO) study (gray-filled diamonds). Note sharper optimum at $k = 0.37$ for sperm donors but flatter optimum over range 0–0.4 for the WHO study. In both cases, the convexity of the curve indicates that there is a range of power transforms that perform better than the log transform at transformation to gaussian distribution.

formations for semen data that unifies various ad hoc approaches, including logarithmic, square-, and cube-root transformations. The present study shows that, while the logarithmic transformation was a major improvement over the natural scale, an optimal power transform could be found that was superior to logarithmic transformation in normalizing even highly skewed nongaussian semen data. Modification of the exponent provides a very flexible range of transforms that render the raw nongaussian data suitable for parametric statistical analysis and modeling analysis. Furthermore, unlike logarithmic transform, power transforms have the advantage of appropriate handling of the structural zeros (azoospermia), which are a regular and important feature of studies of male fertility. This is desirable not only for the data analysis but also for the graphical representation of data, which ideally should be congruent with the data analysis by using the same scale. On the basis of these findings, it is suggested that the logarithmic transformation will always improve but rarely be the best transformation for semen data. In practical terms, it is proposed that the cube-root transformation be considered the first option but that, if not effective, then a comparison of the logarithmic, square-root, and cube-root transformations should provide an indication of the direction in which to search for improved estimates of *k*. Ultimately, a graphical approach similar to that demonstrated can identify the region of optimal power transform. Thus, the power transform approach has the desirable property of preserving power and efficiency while availing the investigator of the most flexible parametric methods for data. The use of power transforms is easily implemented on computer software using the exponentiation function.

Wide experience indicates that the optimal transformation identified in these 2 data sets is generally the most efficient for other studies involving semen variables. The 2 large data sets in this study were consistent in showing the superiority of a power transform in the region of 0.15 to 0.35 over logarithmic transform. The magnitude of this advantage differed between the data sets, presumably due to the different biases in their constitution, leading to systematic differences in sperm concentrations in the lower range. Such variations in recruitment biases of other data sets may yield similar differences. Whether or not this is true, the approach to identifying optimal power transforms would be generally applicable. Another limitation of power transforms is that some semen variables such as sperm motility and morphology may not be amenable to normalization. The variables failing to be normalized by power transforms were those requiring subjective evaluation and quantitated on what is effectively a fractional scale—assessment of sperm motility and morphology. These variables are characteristically those involving manual estimation and subjective judgments with the potential for rounding errors and systematic variability between observers. The failure of the traditional angular transformation (Sokal and Rohlf, 1995) to normalize the distributions of both motility and morphology (both of which are essentially fractions expressed as a percentage) may be due to the inherently subjective nature of these variables, whereby human scoring for these subjective variables may be intrinsically ordinal and hence discontinuous in nature. More objective methods of recording sperm motility and morphology using computer-based image analysis may make these variables more statistically tractable.

Finally, it is interesting to consider whether the existence of an optimal power transform for sperm data may have biological interpretation as a power law rather than merely representing an empirically convenient prelude to valid parametric statistical analysis. For example, a cube-normal distribution of sperm output might reflect output of spheroidal generators having a gaussian distribution of radii. Structurally, the testis can be viewed as comprising a limited number of long, cylindrical seminiferous tubules, each coiled and tightly packed within its own lobular compartment with both ends emptying sperm into the rete testis. The testis could then be viewed as an ellipsoid packed by a number of roughly spherical foci of sperm-producing tissue. In this fashion, a gaussian distribution of radii could then explain the apparent power law relating to sperm output. Regardless of this speculation, the observation that sperm concentration has a cube-root normal distribution has important implications for the valid and statistically efficient analysis of sperm data.

Acknowledgments

The author is grateful to the staff of the Department of Andrology, whose professionalism made this study possible. The wide-ranging contribution of an anonymous reviewer to the manuscript is acknowledged.

References

Bahamondes L, Abdelmassih R, Dachs JN. Survey of 185 sperm analyses of fertile men in an infertility service. *Int J Androl.* 1979;2:526–533.

Berman NG, Wang C, Paulsen CA. Methodological issues in the analysis of human sperm concentration. *J Androl.* 1996;17:68–73.

Box GEP, Cox DR. An analysis of transformations. *J R Stat Soc Ser B.* 1964;26:211–246.

Dixon W. *BMDP Statistical Software Manual.* Berkeley: University of California Press; 1992.

Gaddum JH. Lognormal distributions. *Nature.* 1945;156:463–466.

Handelsman DJ. Sperm output of healthy men in Australia: magnitude of bias due to self-selected volunteers. *Hum Reprod.* 1997;12:2701–2705.

Handelsman DJ. Male contraception. In: DeGroot LJ, ed. *Endocrinology.* 4th ed. Philadelphia, Pa: WB Saunders; 2001:2344–2349.

Handelsman DJ, Conway AJ, Boylan LM, Turtle JR. Testicular function in potential sperm donors: normal ranges and the effects of smoking and varicocele. *Int J Androl.* 1984;7:369–382.

Keene ON. The log transformation is special. *Stat Med.* 1995;14:811–819.

Liu PY, Turner L, Conway AJ, Handelsman DJ. Predicting pregnancy and spermatogenesis by survival analysis during gonadotropin treatment of gonadotropin deficient infertile men. *Hum Reprod.* 2002;17:625–633.

MacLeod J, Gold RZ. The male factor in fertility and infertility. II. Spermatozoan counts in 1000 men of known fertility and in 1000 cases of infertile marriage. *J Urol.* 1951;66:436–439.

Macomber D, Sanders MB. The spermatozoa count: its value in the diagnosis, prognosis and treatment of sterility. *N Engl J Med.* 1929;200:981–984.

Meloun M, Hill M, Militky J, Kupka K. Transformation in the PC-aided biochemical data analysis. *Clin Chem Lab Med.* 2000;38:553–559.

Peltier MR, Wilcox CJ, Sharp DC. Technical note: application of the Box-Cox data transformation to animal science experiments. *J Anim Sci.* 1998;76:847–849.

Royston JP. An extension of Shapiro and Wilks' W test for normality to large samples. *Appl Stat.* 1982;31:115–124.

Shapiro SS, Wilks MB. An analysis of variance test for normality (complete samples). *Biometrika.* 1965;52:591–611.

Sokal RR, Rohlf FJ. *Biometry*. 3rd ed. San Francisco, Calif: WH Freeman; 1995.

WHO Task Force on Methods for the Regulation of Male Fertility. Contraceptive efficacy of testosterone-induced azoospermia in normal men. *Lancet.* 1990;336:955–959.

WHO Task Force on Methods for the Regulation of Male Fertility. Contraceptive efficacy of testosterone-induced azoospermia and oligozoospermia in normal men. *Fertil Steril.* 1996;65:821–829.