COMMUN. STATIST.—THEORY METH., 26(4), 839-851 (1997)

BIAS IN LINEAR MODEL POWER AND SAMPLE SIZE DUE TO ESTIMATING VARIANCE

Keith E. Muller
Dept. of Biostatistics, CB#7400
University of North Carolina
Chapel Hill, North Carolina, 27599

Virginia B. Pasour
Dept. of Biostatistics, CB#7400
University of North Carolina
Chapel Hill, North Carolina, 27599

Key Words and Phrases: *effect size; meta-analysis; confidence bounds*

ABSTRACT

Planning a study using the General Linear Univariate Model often involves sample size calculation based on a variance estimated in an earlier study. Noncentrality, power, and sample size inherit the randomness. Additional complexity arises if the estimate has been censored. Left censoring occurs when only significant tests lead to a power calculation, while right censoring occurs when only non-significant tests lead to a power calculation. We provide simple expressions for straightforward computation of the distribution function, moments, and quantiles of the censored variance estimate, estimated noncentrality, power, and sample size. We also provide convenient approximations and evaluate their accuracy. The results allow demonstrating that ignoring right censoring falsely widens confidence intervals for noncentrality and power, while ignoring left censoring falsely narrows the confidence intervals. The new results allow assessing and avoiding the potentially substantial bias that censoring may create.

1. INTRODUCTION

1.1 Motivation

Power and sample size calculations provide information important in planning and evaluating studies. Scientists often wish to estimate power and sample size with respect to a fixed, "clinically significant," treatment effect, such as a difference in means. Typically analysis of data from an earlier study provides estimates of nuisance parameters, such as the error variance. The randomness of the variance estimate causes the values of noncentrality, power, and sample size also to vary randomly. Good statistical practice requires estimating the uncertainty surrounding the point estimates.

Assume larger values of the test statistic correspond to a significant result. A data analyst may calculate power because an earlier trial yielded a significant test. *Left* censoring occurs as scientists try to replicate a promising finding. A data analyst also may calculate power because an earlier trial yielded a non-significant test. *Right* censoring occurs as scientists decide whether to increase their investment in a discouraging result. A data analyst also may calculate power only because an earlier trial yielded equivocal results, with neither an extremely small nor extremely large p-value (*double* censoring).

1.2 An Example

In comparing two treatments for deteriorating renal function, Falk, Hogan, Muller, and Jennette (1992) judged a change of 0.50 dl/mg in mean reciprocal creatinine level to have clinical significance. A T test yielded no statistical evidence of any difference between treatments. The authors estimated a power of 0.96 for a change of 0.50 dl/mg, with nominal $\alpha = 0.01$, and $\hat{\sigma}^2 = 0.068$ (dl/mg)^2 from the data. Given the concern for side effects and the high value of estimated power, the authors stopped the trial and concluded that the treatment was unlikely to have clinical significance. Assuming that the observed value provided an unbiased $\hat{\sigma}^2$, Taylor and Muller (1995) computed 95% confidence intervals for power as $[0.688, 0.999]$, two-sided, and $[0.750, 1]$, one-sided. However, the (right) censoring might have imparted bias.

1.3 Literature Review

See Muller, LaVange, Ramey, and Ramey (1992) for a general introduction to power analysis for linear models, including repeated measures and other multivariate models. O'Brien and Muller (1993) provided a tutorial on the same topic. Muller, Barton, and Benignus (1984) and Muller and Benignus (1992) reviewed basic principles in using power analysis for study planning.

Taylor and Muller (1995) described how to easily compute exact confidence bounds for noncentrality, power, and sample size for the GLUM, as a function of fixed means and estimated variance. Browne (1995) and many others (cited in Taylor and Muller, 1995) have reported approximate solutions. Taylor and Muller assumed no censoring, which implies an unbiased estimate of variance. Their results represent a special case of those in §2.

Taylor and Muller (1996) described how to compute exact confidence bounds for noncentrality, power, and sample size for the GLUM, with estimated means and estimated variance, under left or right censoring. They concluded that failing to consider censoring may badly distort sample size choice based on estimated means and estimated variance. Their results have no simple relationship to those in §2.

Asymmetric distributions and a desire for sensitivity analysis create a preference for confidence intervals over point estimates. See Taylor and Muller (1995, 1996) for further discussion.

Cohen and Sackrowitz (1996) discussed providing exact confidence intervals for a target study following a screening study. In contrast to the situation considered here, Cohen and Sackrowitz assumed that the analyst wishes to base the inference on *all* of the data.

1.4 A Comment on the Presentation

For the sake of brevity most detail has been omitted in the statement of proofs, derivations of approximations, descriptions of numerical methods, and simulation results. The focus rests on exact results and accurate approximations in accounting for censoring in sample size calculation.

2. ANALYTIC RESULTS

2.1 Notation

Let $F_U(u;\psi_1\ldots\psi_k)$ indicate the cumulative distribution function (CDF) of the random variable U, with parameters $\psi_1\ldots\psi_k$, density $f_U(u;\psi_1\ldots\psi_k)$, and pth quantile $F_U^{-1}(p;\psi_1\ldots\psi_k)$. Let $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ indicate a multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, $\gamma(\psi)$ a standard Gamma, $\chi^2(\nu,\omega)$ a noncentral χ^2 on ν

degrees of freedom with noncentrality ω, and $F(\nu_1, \nu_2, \omega)$ a noncentral F on ν_1 and ν_2 degrees of freedom with noncentrality ω (Johnson and Kotz, Chapter 17, 1970a; Chapters 26, 28, and 30, 1970b).

State the GLUM (Chapters 1-5, Searle, 1971), with N observations and q predictors, as:

$$\underset{N \times 1}{\boldsymbol{y}} = \underset{N \times q \times 1}{\boldsymbol{X\beta}} + \underset{N \times 1}{\boldsymbol{e}} \,. \tag{2.1}$$

$\boldsymbol{X}$ (fixed, known) has rank $r \leq q \ll N$, while $\boldsymbol{\beta}$ contains fixed, unknown parameters. Assuming $\boldsymbol{e} = N_N(0, \sigma^2 I)$, with $0 < \sigma^2 < \infty$, allows testing the general linear hypothesis:

$$H_0 : \underset{\nu_1 \times 1}{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 \qquad \textit{versus} \qquad H_A : \underset{\nu_1 \times 1}{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0 \,. \tag{2.2}$$

Let $\nu_2 = (N - r)$ and $\boldsymbol{M} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{C}'$, with full rank ν_1. Observe that

$$SSH(\widehat{\boldsymbol{\theta}}, N) = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{C}']^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \chi^2(\nu_1, \omega) \cdot \sigma^2, \tag{2.3}$$

$$\widehat{\sigma}^2 = \boldsymbol{y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}']\boldsymbol{y}/\nu_2 = SSE/\nu_2 = \chi^2(\nu_2) \cdot \sigma^2/\nu_2 \,. \tag{2.4}$$

The independence of $SSH(\widehat{\boldsymbol{\theta}}, N)$ and $\widehat{\sigma}^2$ allows writing the likelihood ratio test as

$$F_{obs} = \frac{SSH(\widehat{\boldsymbol{\theta}}, N)/\nu_1}{SSE/\nu_2} = \frac{SSH(\widehat{\boldsymbol{\theta}}, N)/\nu_1}{\widehat{\sigma}^2} = F(\nu_1, \nu_2, \omega) \,, \tag{2.5}$$

with noncentrality

$$\omega = \frac{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{C}']^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{\sigma^2} = \frac{SSH(\boldsymbol{\theta}, N)}{\sigma^2} \,. \tag{2.6}$$

Consider estimating variance from an initial study and conducting a power calculation under left censoring (only following a significant test), right censoring (only following a non-significant test), or both (only following non-extreme results). Power calculation may involve a hypothesis distinct from the one in the initial study.

Carefully maintain a distinction between the *screening* study (which provides the variance estimate) and the *target* study (which requires power calculation). The studies may have distinct sample sizes (N_s, N_t) error degrees of freedom (ν_{2s}, ν_{2t}), parameters tested ($\boldsymbol{\theta}_s$, $v_{1s} \times 1$, $\boldsymbol{\theta}_t$, $v_{1t} \times 1$), and test statistics ($F_s = F(v_{1s}, v_{2s}, \omega_s)$, $F_t = F(v_{1t}, v_{2t}, \omega_t)$). The target study has test size α_t, with $f_t = F_F^{-1}(1-\alpha_t;\nu_{1t},\nu_{2t})$. Let $f_L = F_F^{-1}(1-\alpha_{sL};\nu_{1s},\nu_{2s})$ indicate the value F_s must exceed (left censoring), and $f_R = F_F^{-1}(1-\alpha_{sR};\nu_{1s},\nu_{2s})$ the value F_s must not exceed (right censoring). In turn describe α_{sL} and α_{sR} as the screening levels, as distinct from α_t.

With f_s the observed value of the screening statistic and $p_s = 1 - F_F(f_s;\nu_{1s},\nu_{2s})$ the corresponding p-value, an outcome observed under censoring has

$$0 \le f_L \le f_s < f_R \le \infty \tag{2.7}$$

and

$$0 \le \alpha_{sR} < p_s \le \alpha_{sL} \le 1\,. \tag{2.8}$$

Conducting only left censoring corresponds to $\alpha_{sR} = 0$ and $f_R = \infty$, while conducting only right censoring corresponds to $\alpha_{sL} = 1$ and $f_L = 0$. Also define $\pi_s = Pr\{f_L \le F_s < f_R\}$. While α_{sL}, α_{sR}, and α_t equal probabilities for central F's, π_s equals a probability for a noncentral F.

2.2 Distributions Under Censoring

Theorem 1. Assume the situation described in §2.1 holds, with V the random variable corresponding to the estimated variance, with particular realization z. Let $z_* = z\nu_{2s}/\sigma^2$ and

$$g(t) = F_{\chi^2}\!\left(\frac{\nu_{1s}}{\nu_{2s}} f_R t;\nu_{1s},\omega_s\right) - F_{\chi^2}\!\left(\frac{\nu_{1s}}{\nu_{2s}} f_L t;\nu_{1s},\omega_s\right). \tag{2.9}$$

Using this notation

$$F_V(z) = \pi_s^{-1}\int_0^{z_*} g(t) f_{\chi^2}(t;\nu_{2s})dt\,, \tag{2.10}$$

with corresponding density

$$f_V(z) = \frac{\nu_{2s}}{\sigma^2}\pi_s^{-1} g(z_*) f_{\chi^2}(z_*;\nu_{2s})\,. \tag{2.11}$$

Proof. Express the CDF of V as

$$\begin{aligned} F_V(z) &= Pr\{(\hat{\sigma}^2 \le z) \mid (f_L \le F_s \le f_R)\} \qquad (2.12)\\ &= \frac{Pr\{(\hat{\sigma}^2 \le z) \cap (f_L \le F_s \le f_R)\}}{Pr\{f_L \le F_s \le f_R\}} \end{aligned}$$

$$= \frac{Pr\{(\widehat{\sigma}^2 \le z) \cap (F_s \le f_R)\} - Pr\{(\widehat{\sigma}^2 \le z) \cap (F_s \le f_L)\}}{\pi_s}.$$

With $X_1 = \chi^2(\nu_{1s}, \omega_s)$, $X_2 = \chi^2(\nu_{2s})$, and $f_C \in \{f_L, f_R\}$

$$\begin{aligned} Pr\{(\widehat{\sigma}^2 \le z) \cap (F_s \le f_C)\} &= Pr\left\{\left(X_2 \frac{\sigma^2}{\nu_{2s}} \le z\right) \cap \left(\frac{X_1}{X_2}\frac{\nu_{2s}}{\nu_{1s}} \le f_C\right)\right\} \quad (2.13)\\ &= Pr\left\{X_1 \frac{\nu_{2s}}{\nu_{1s}} \le f_C X_2 \le f_C \frac{z\nu_{2s}}{\sigma^2}\right\}. \end{aligned}$$

If $f_C = f_L = 0$ then this probability equals zero. Otherwise the independence of X_1 and X_2 yields

$$Pr\left\{X_1 \frac{\nu_{2s}}{\nu_{1s} f_C} \le X_2 \le \frac{z\nu_{2s}}{\sigma^2}\right\} = \int_0^{z_*} F_{X_1}\left(\frac{\nu_{1s}}{\nu_{2s}} f_C t\right) f_{X_2}(t)dt. \quad (2.14)$$

Use this form twice, once with $f_C = f_L$ and once with $f_C = f_R$. Then substitute the results into the last line of (2.12) to complete the derivation of the CDF.

Corollary 1.1. Under censoring, the estimated noncentrality, $\widehat{\omega}_t$, has distribution function

$$F_{\widehat{\omega}_t}(u) = Pr\{\widehat{\omega}_t \le u\} = 1 - F_V[SSH(\boldsymbol{\theta}_t, N_t)/u] \quad (2.15)$$

and pth percentile

$$\widehat{\omega}_t(p) = SSH(\boldsymbol{\theta}_t, N_t)/F_V^{-1}(1-p). \quad (2.16)$$

Corollary 1.2. Define $w(p; f, \nu_1, \nu_2)$ by $p = F_F[f; \nu_1, \nu_2, w(p; f, \nu_1, \nu_2)]$. Under censoring, the estimated power of the target study has distribution function

$$F_{\widehat{P}_t}(q) = F_V[SSH(\boldsymbol{\theta}_t, N_t)/w(1-q; f_t, \nu_{1t}, \nu_{2t})] \quad (2.17)$$

and pth percentile

$$\widehat{P}_t(p) = 1 - F_F\left[f_t; \nu_{1t}, \nu_{2t}, SSH(\boldsymbol{\theta}_t, N_t)/F_V^{-1}(1-p)\right]. \quad (2.18)$$

Corollary 1.3. Computing the pth percentile of the estimated sample size, $\widehat{N}_t(p)$, requires fixing a target power, P_t, and solving the following equation for it:

$$P_t = 1 - F_F\{f_t; \nu_{1t}, \widehat{N}_t(p) - r, SSH[\boldsymbol{\theta}_t, \widehat{N}_t(p)]/F_V^{-1}(1-p)\}. \quad (2.19)$$

2.3 Moments Under Censoring

Lemma. For a non-negative integer m, the density of a central χ^2 random variable satisfies

$$t^m f_{\chi^2}(t;\nu) = \frac{2^m\Gamma(\nu/2+m)}{\Gamma(\nu/2)} f_{\chi^2}(t;\nu+2m) \tag{2.20}$$
$$= \prod_{i=0}^{m-1}(\nu+2i) f_{\chi^2}(t;\nu+2m)\,.$$

Theorem 2. Let $F_{sm} = F(\nu_{1s}, \nu_{2s}+2m, \omega_s)$, $f_{Lm} = f_L(\nu_{2s}+2m)/\nu_{2s}$, $f_{Rm} = f_R(\nu_{2s}+2m)/\nu_{2s}$, and $\pi_{sm} = Pr\{f_{Lm} \leq F_{sm} < f_{Rm}\}$. Then

$$\mu'_m = \mathcal{E}\left(\frac{V}{\sigma^2}\right)^m = \frac{\pi_{sm}}{\pi_s}\cdot\prod_{i=0}^{m-1}\left(\frac{\nu_{2s}+2i}{\nu_{2s}}\right). \tag{2.21}$$

Proof. Use the density in Theorem 1 to write $\mathcal{E}(V/\sigma^2)^m$ as an integral. Apply the Lemma. Use F_{sm}, f_{Lm}, f_{Rm}, and π_{sm} to define $g_m(z_*)$. Complete the proof by recognizing that the integrand equals the density of V/σ^2, based on $(\nu_{2s}+2m)$ degrees of freedom.

Corollary 2.1. The mth central moment may be expressed as

$$\mu_m = \mathcal{E}\left(\frac{V}{\sigma^2}-\mu'_1\right)^m = \sum_{j=0}^{m}\binom{m}{j}\left[\frac{\pi_{sj}}{\pi_s}\prod_{i=0}^{j-1}\left(\frac{\nu_{2s}+2i}{\nu_{2s}}\right)\right]\left(-\frac{\pi_{s1}}{\pi_s}\right)^{m-j}. \tag{2.22}$$

Theorem 3. For the model and notation of Theorem 1, assume $t < \nu_{2s}/(2\sigma^2)$. Let $f_{Lt} = f_L/(1-2t\sigma^2/\nu_{2s})$, $f_{Rt} = f_R/(1-2t\sigma^2/\nu_{2s})$, and $\pi_{st} = Pr\{f_{Lt} \leq F_s < f_{Rt}\}$. Then the moment generating function of V equals

$$M_V(t) = \mathcal{E}[\exp(tV)] = \frac{\pi_{st}}{\pi_s}\cdot\left(1-2t\sigma^2/\nu_{2s}\right)^{-\nu_{2s}/2}. \tag{2.23}$$

The proof parallels that of Theorem 2.

2.4 Approximations

Various strategies were examined for computing approximate probabilities and quantiles. A Taylor's series expansion of $g(t)$ allowed creating approximations. Numerical derivatives (Abramowitz and Stegun, 1964, p 914) allowed approximating the noncentral χ^2 density and its derivatives by using a χ^2 CDF algorithm. Alternately express the density in terms of Bessel functions (equation (5) in Johnson and Kotz, 1970b, has $\sqrt{\omega t}$ printed incorrectly as $t\sqrt{\omega}$). Use equation 1. in §0.433 in Gradshteyn and Ryzhik (p 25, 1994), and equations 9.6.29 and 9.6.26 in Abramowitz and Stegun (1964, p376).

Expressions for the moments lead to another method of approximation. Consider approximating V by $V_* = \sigma_*^2\chi^2(\nu_*)/\nu_*$. Let M_1 and M_2 equal $\mathcal{E}V$ and $\mathcal{E}V^2$ computed with the result of Theorem 2. Solving two simultaneous equations, $M_1 = \mathcal{E}V_* = \sigma_*^2$ and $M_2 = \mathcal{E}V_*^2 = \sigma_*^4\nu_*(\nu_* + 2)$, yields

$$\sigma_*^2 = M_1 = \sigma^2 \frac{\pi_{s1}}{\pi_s} \tag{2.24}$$

and

$$\nu_* = \frac{2M_1^2}{M_2 - M_1^2} = \frac{2\pi_{s1}^2}{[\pi_s\pi_{s2} \cdot (\nu_{2s} + 2)/\nu_{2s} - \pi_{s1}^2]} \,. \tag{2.25}$$

In turn

$$F_V(z) \approx F_{\chi^2}\left(z\nu_*/\sigma_*^2; \nu_*\right) \tag{2.26}$$

and, with p the probability corresponding to the quantile of interest,

$$z_p \approx F_{\chi^2}^{-1}(p; \nu_*) \cdot \sigma_*^2/\nu_* \,. \tag{2.27}$$

Note that the approximation resolves to the exact result under no censoring and hence also asymptotically. The Satterthwaite approximation, used in various ANOVA settings, also involves matching the first two moments of a generalization of a scaled χ^2 to the moments of a scaled χ^2.

3. NUMERICAL EVALUATIONS

3.1 Methods

All probability, quantile, and moment values were calculated with SAS IML©. Numerical integrations were computed by refinement of Simpson's rule (Thisted, 1988, p271). Applying a quantile transformation to equation (2.10) creates a finite region of integration. If $p = F_{\chi^2}(t; \nu_{2s})$, then $t = F_{\chi^2}^{-1}(p; \nu_{2s})$, and $dp = f_{\chi^2}(t; \nu_{2s})dt$. Let $p_0 = F_{\chi^2}(z\nu_{2s}/\sigma^2; \nu_{2s})$. Then

$$F_V(z) = \pi_s^{-1}\int_0^{p_0} g_R\left[F_{\chi^2}^{-1}(p; \nu_{2s})\right]dp - \pi_s^{-1}\int_0^1 g_L\left[F_{\chi^2}^{-1}(p; \nu_{2s})\right]dp \,. \tag{3.1}$$

The monotone integrands insure a convergent sequence of known accuracy (Thisted, 1988, §5.1.1-5.1.3 and related exercises). A bisection algorithm (Thisted, 1988, p169) was used to invert the exact CDF, with initial values from equation (2.27). A recursion formula (Thisted, 1988, p322) refined the initial value before iteration.

Simulations were conducted in the SAS© DATA step. The FNONCT, NORMAL, and RANGAM functions were used to create $X_1/\sigma^2 = \chi^2(\nu_{1s}, \omega) = \mathrm{N}(\sqrt{\omega}, 1) + \quad 2 \cdot \gamma(\nu_{2s}/2)$ and $X_2/\sigma^2 = \chi^2(\nu_{2s}) = 2 \cdot \gamma(\nu_{2s}/2)$, $F_s = (X_1/\nu_{1s})/(X_2/\nu_{2s})$ and $V = (X_2/\nu_{2s})$. Values of V were sampled until 1000 were found with $f_L \leq F_s < f_R$. The simulated experiment used a T test design with $\alpha_t = .05$ and $\sigma^2 = 100$.

3.2 Results

Simulated means and variances, as well as empirical quantiles, agreed extremely well with exact calculations. The amount of bias due to censoring ranged from modest to severe, with substantial reduction as N_t increased from 10 to 50. Censoring always substantially increased the spread of the distribution, with $N_t = 50$ still having approximately double the spread under censoring.

Table I contains values of exact quantiles under censoring divided by exact quantiles under no censoring. Censoring may substantially change quantiles, especially in small samples. Reciprocals of the variance ratios in Table I provide a first approximation to the impact on sample size. In every condition studied left censoring deflated variance quantiles, while right censoring inflated variance quantiles (for fixed sample size, screening level, and power). The inverse relationship of variance to noncentrality and power implies that ignoring right censoring falsely widens confidence intervals for noncentrality and power. In contrast, ignoring left censoring falsely narrows confidence intervals for noncentrality and power.

Both the two moment and Taylor series approaches were compared to exact calculation. The two methods provided approximately the same accuracy, and nearly always gave two digits of accuracy (in probability values) for $.05 \leq p \leq .95$ and the range of conditions in Table I. Additional terms increased accuracy slowly. Some erratic values were noted for long series, likely reflecting loss of precision in summing large intermediate results of alternating sign.

3.3 The Example Revisited

Accounting for the right censoring which occurred yields a two-sided 95% interval of [.856, .999+] and a one-sided interval of [.893, 1]. The more accurate interval provides much stronger support for the author's claim of no effect of clinical importance. Figure 1 displays estimated power (the dashed line), a two-

Table I
100*Ratio of Exact Variance pth Quantile Under Censoring to Exact Variance pth Quantile Under No Censoring.

N_t	α_s	Power	$p = .05$	.10	.50	.90	.95
			Left (Require Significant)				
10	0.05	0.1	61	61	62	64	63
		0.9	97	96	95	94	93
	0.50	0.1	91	91	92	93	93
		0.9	100	100	100	100	100
50	0.05	0.1	93	93	93	93	93
		0.9	99	99	99	99	99
	0.50	0.1	98	99	99	99	99
		0.9	100	100	100	100	100
			Right (Require Non-significant)				
10	0.05	0.1	111	109	105	102	102
		0.9	193	183	155	136	132
	0.50	0.1	120	118	112	109	108
		0.9	146	143	134	127	126
50	0.05	0.1	101	101	101	101	100
		0.9	109	108	107	107	106
	0.50	0.1	102	102	102	102	102
		0.9	105	105	104	104	104

sided 95% confidence region ignoring censoring (dotted lines), and a two-sided 95% confidence region accounting for censoring. See Taylor and Muller (1995) for a sketch of a proof that point-wise computations provide correct simultaneous coverage. Accounting for the right censoring narrows the confidence region. In contrast a correction for left censoring would have widened the confidence region.

4. CONCLUSIONS AND RECOMMENDATIONS

1. Confidence regions capture uncertainty surrounding power calculation.
2. Treating censoring may greatly improve the accuracy of power analysis.
3. Ignoring right censoring falsely widens confidence intervals for power, while ignoring left censoring falsely narrows them.

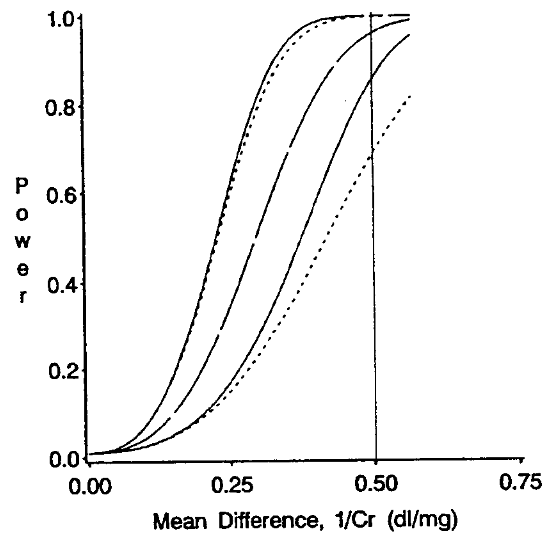

Figure 1. Estimated power for kidney study. Dashed line based on observed variance. Solid lines provide 95% confidence region accounting for censoring. Dotted lines provide 95% confidence region ignoring censoring.

4. In contrast to usual practice, one-sided confidence intervals are often best.

5. Not surprisingly, censoring has most effect in small samples, such as in using pilot study estimates for power analysis.

6. Use the approximation in §2.4 if exact calculations prove inconvenient.

7. More efficient algorithms and multivariate extensions merit consideration.

APPENDIX

The power software described in Muller, LaVange, Ramey, and Ramey (Appendix A, 1992) resides at *ftp://ftp.uga.edu/pub/sas/contrib/cntb0014*. To FTP the files connect to HOST *ftp.uga.edu* as USER *anonymous* with PASSWORD *your_email_address*. Change to remote directory */pub/sas/contrib/cntb0014* and get all seven files. Alternately send E-mail with SUBJECT *cntb0014: download* to USERID *sascontrib@sasserv.uga.edu*.

ACKNOWLEDGEMENTS

Muller's work supported in part by NCI program project grant P01 CA47 982-04, NIH Clinical Research Center grant M01 RR000-46-33, and NIEHS grant N01-ES-35356. The authors gratefully acknowledge helpful comments on an earlier draft of this paper by Douglas J. Taylor. A portion of the work was described in a paper submitted by V. B. Pasour in partial fulfillment of the requirements for the M. S. degree in Biostatistics.

BIBLIOGRAPHY

Abramowitz, M., and Stegun, I. A. (1964). *Handbook of Mathematical Functions*, New York: Dover Publications.

Browne, R. H. (1995). "On the use of a pilot sample for sample size determination," *Statistics in Medicine*, **14**, 1933-1940.

Cohen, A., and Sackrowitz, H. B. (1996). "Lower confidence bounds using pilot samples with an application to auditing," *Journal of the American Statistical Association*, **91**, 338-342.

Falk, R. J., Hogan, S. L., Muller, K. E., and Jennette, J. C. (1992). "Treatment of progressive membranous glomerulopathy," *Annals of Internal Medicine*, **6**, 438-445.

Gradshteyn, I. S. and Ryzhik, I. M. (1994). *Table of Integrals, Series, and Products* (5th edition, Alan Jeffrey, Ed), New York: Academic Press.

Johnson, N. L., and Kotz, S. (1970a). *Continuous Univariate Distributions-1*, New York: Wiley.

Johnson, N. L., and Kotz, S. (1970b). *Continuous Univariate Distributions-2*, Boston: Houghton Mifflin.

Muller, K. E., Barton, C. N., and Benignus, V. A. (1984). "Recommendations for appropriate statistical practice in toxicologic experiments," *Neurotoxicology*, **5**, 113-126.

Muller, K. E., and Benignus, V. A. (1992). "Increasing scientific power with statistical power," *Neurotoxicology and Teratology*, **14**, 211-219.

Muller, K. E., LaVange, L. M., Ramey, S. L., and Ramey, C. T. (1992). "Power calculations for general linear multivariate models including repeated measures applications," *Journal of the American Statistical Association*, **87**, 1209-1226.

O'Brien, R. G., and Muller, K. E. (1993). "A unified approach to statistical power for t-tests to multivariate models," Chapter 8, *Applied Analysis of Variance* *in Behavioral Sciences*, L. K. Edwards, ed., p297-344, New York: Marcel Dekker.

Searle, S. R. (1971). *Linear Models*, New York: John Wiley.

Taylor, D. J., and Muller, K. E. (1995). "Computing confidence bounds for power and sample size of the general linear univariate model," *American Statistician*, **49**, 43-47.

Taylor, D. J., and Muller, K. E. (1996). "Bias in linear model power and sample size calculation due to estimating noncentrality," *Communications in Statistics: Theory and Methods*, **25**, 1595-1610.

Thisted, R. A. (1988). *Elements of Statistical Computing*, New York: Chapman and Hall.

Received October, 1996.