NIH Public Access
Author Manuscript
Am Stat. Author manuscript; available in PMC 2013 September 13.

Published in final edited form as:
Am Stat. 1995 January 1; 49(1): 43–47. doi:10.1080/00031305.1995.10476111.

Computing Confidence Bounds for Power and Sample Size of the General Linear Univariate Model

Douglas J. Taylor and **Keith E. Muller**
Douglas J. Taylor is a doctoral student, and Keith E. Muller is Associate Professor, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599

Abstract

The power of a test, the probability of rejecting the null hypothesis in favor of an alternative, may be computed using estimates of one or more distributional parameters. Statisticians frequently fix mean values and calculate power or sample size using a variance estimate from an existing study. Hence computed power becomes a random variable for a fixed sample size. Likewise, the sample size necessary to achieve a fixed power varies randomly. Standard statistical practice requires reporting uncertainty associated with such point estimates. Previous authors studied an asymptotically unbiased method of obtaining confidence intervals for noncentrality and power of the general linear univariate model in this setting. We provide exact confidence intervals for noncentrality, power, and sample size. Such confidence intervals, particularly one-sided intervals, help in planning a future study and in evaluating existing studies.

Keywords

Effect size; Meta-analysis; Noncentral *F* distribution

1. Introduction

1.1 Motivation

Statisticians estimate power and sample size in many settings. These quantities are used in conducting a sensitivity analysis while planning a future study (prospective power analysis), and also help in the quantitative evaluation of an existing study (retrospective power analysis). Muller and Benignus (1992) recommended that any claim of no effect be substantiated by an appropriate retrospective power analysis. For example, the wish to demonstrate no appreciable effect arises in assessing toxicity of any compound proposed for commercial use. Following a negative result, a study's power to have detected an appreciable difference may be important. Regulatory decisions by agencies such as the U.S. Environmental Protection Agency often hinge on such negative evidence.

Researchers may propose fixed values as the true population means in order to conduct a power analysis. They often complete the power computation for a general linear univariate model (GLUM) test by using a variance estimate $\hat{\sigma}^2$ obtained from an existing study. For example, physicians describe a “clinically significant” effect as one that changes a person's health in an important way. Combining a fixed sample size (N), a fixed, clinically significant mean difference, and a variance estimate determines a power value. With estimated error variance, a power value for a fixed N must be recognized as a random variable. Similarly the sample size needed to achieve a fixed power varies randomly. Standard statistical practice requires reporting uncertainty associated with such point estimates. A lower bound for power allows stating, with specified confidence, that a study has power of at least “P” to

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

detect the effect. Similarly, an upper bound on N provides a sample size which ensures, with specified confidence, a desired target power.

Kupper and Hafner (1989) demonstrated the bias resulting from choosing sample size for a test statistic assuming known variance (a z test), when one will conduct the data analysis with an estimated variance (t test). Power calculations for both a z test and a t test start from an assumption of known variance, and diverge due to differences in the conduct of data analysis (test statistic computation). This highlights the importance of aligning power computations with actual analysis methods. Kupper and Hafner, as well as Beal (1989), provided closely related discussions in the context of choosing the N necessary to control the length of a confidence interval for a scalar parameter.

Other authors have addressed the question of confidence intervals for power. Dudewicz (1972) suggested substituting exact confidence bounds for $\hat{\sigma^2}$ into power calculations for a t test. This yields approximate confidence limits for power. He did not present any asymptotic or simulation evidence as to accuracy, nor did he note the approximate nature of the calculations. Clark (1987) presented analytic and simulation results for generalization of the method to the fixed-effects GLUM with Gaussian errors. She proved that exact confidence bounds for estimated variance lead to asymptotically unbiased confidence intervals for both noncentrality and power. She also demonstrated, via simulation, the optimistic bias in the method. We present an alternate and straightforward formulation which allows computation of exact confidence bounds for noncentrality, power, and sample size in the GLUM. These methods are applicable when testing scalar and vector hypotheses.

1.2 An Example

Falk, Hogan, Muller, and Jennette (1992) compared treatment of deteriorating renal function with both cyclophosphamide and corticosteroid therapies to corticosteroid therapy alone. The researchers defined a clinically significant improvement in renal function as a doubling of reciprocal serum creatinine level, corresponding to an increase of .50 dL/mg. Twelve patients were randomly assigned to each treatment in a clinical trial. A power calculation was performed with a test size of $\alpha_t = .01$ assuming true improvement due to dual therapies was .50 dL/mg. Based on a variance estimate of .068, power was approximated as .960. The authors used these calculations both to stop the trial early and to argue that the treatment likely had little value. Given these results, what interval of uncertainty surrounds the power estimate of .960?

1.3 Statement of the Model, Hypothesis, and Properties

The GLUM with N sampling units and q predictors can be stated (Searle 1971, chap. 5):

$$\underset{N\times 1}{y} = \underset{(N\times q)(q\times 1)}{\mathbf{X}\beta} + \underset{N\times 1}{\mathbf{e}}. \quad (1.1)$$

Here $\mathbf{X}$, with rank $r \leq q < N$, contains fixed, known constants and β contains fixed, unknown parameters. Also assume $e \overset{d}{=} N_N(\mathbf{0}, \sigma^2\mathbf{I})$ with $0 < \sigma^2 < \infty$. The usual hypothesis involves the secondary parameters $\theta = \mathbf{C}\beta$. Interest lies in testing the general linear hypothesis

$$H_0: \underset{\nu_1\times 1}{\theta} = \theta_0 \quad (1.2)$$

versus

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$H_A: \theta \neq \theta_0. \quad (1.3)$$

Both θ_0 and **C** contain fixed constants chosen by the data analyst. Assume full rank for **C**, with each row of **C** defining an element of θ. The usual GLUM test statistic is

$$F_{\text{obs}} = \frac{SSH(\widehat{\theta}, N)/\upsilon_1}{SSE/\upsilon_2} = \frac{SSH(\widehat{\theta}, N)/\upsilon_1}{\widehat{\sigma}^2}, \quad (1.4)$$

in which

$$SSH(\widehat{\theta}, N) = (\widehat{\theta} - \theta_0)' \left[\mathbf{C} \left(\mathbf{X}'\mathbf{X} \right)^{-} \mathbf{C}' \right]^{-1} (\widehat{\theta} - \theta_0), \quad (1.5)$$

$$SSH = y' \left[\mathbf{I} - \mathbf{X} \left(\mathbf{X}'\mathbf{X} \right) - \mathbf{X}' \right] y, \quad (1.6)$$

and $(\cdot)^{-}$ indicates a generalized inverse. The test statistic follows a noncentral F distribution with ν_1 (size and rank of **C**) numerator degrees of freedom (df), $\nu_2 = (N - r)$ denominator df, and noncentrality parameter

$$\omega = \frac{(\theta - \theta_0)' \left[\mathbf{C} \left(\mathbf{X}'\mathbf{X} \right)^{-} \mathbf{C}' \right]^{-1} (\theta - \theta_0)}{\sigma^2} = \frac{SSH(\theta, N)}{\sigma^2}. \quad (1.7)$$

We use the notation $SSH(\cdot, N)$ to emphasize that SSH is a function of sample size by way of the matrix **X**. The cumulative distribution function (CDF) of a noncentral F with ν_1 and ν_2 df and noncentrality parameter ω will be represented by $F_F[\cdot \mid \nu_1, \nu_2, \omega]$, with corresponding density function $f_F[\cdot \mid \nu_1, \nu_2, \omega]$. Using this notation, the power of the test can be written

$$P = 1 - F_F[f_{\text{crit}}(1 - \alpha_t) \mid \upsilon_1, \upsilon_2, \omega], \quad (1.8)$$

in which $f_{\text{crit}}(1 - \alpha_t)$ is the $100 \cdot (1 - \alpha_t)$ percentile from a central F with ν_1 and ν_2 df. The CDF and density of a non-central F may be expressed as Poisson-weighted infinite sums (Johnson and Kotz 1970, p. 192). In practice, CDF and density values may be precisely calculated with computer algorithms (Abramowitz and Stegun 1964, p. 947; Odeh and Fox 1991, chap. 5).

Some authors, notably Cohen (1987, p. 8), have discussed “effect size” as a generic concept encompassing disparate analyses (including tests of means, proportions, and categorical data). For the special cases of the GLUM considered by Cohen, effect size and noncentrality can be expressed as one-to-one functions of each other. Although distinct cases of the GLUM yield distinct formulas for effect size, they yield only a single formula for noncentrality. Hence we restrict our attention to noncentrality in order to maximize generality and avoid ambiguity.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2. Computing Confidence Bounds

2.1 Confidence Bounds for Noncentrality

For all calculations, a distinction must be carefully maintained between the sample size of the *estimation* study (which provides the variance estimate) and the sample size of the *target* study (for which power is desired). For example, the value of $f_{\text{crit}}(\cdot)$ is always calculated using the target study error df. In a retrospective analysis, the two sample sizes are generally the same. This may not be true, however, in a prospective analysis. To help facilitate the necessary distinction, we use the notation ν_{2e} and ν_{2t} to represent the error df associated with the estimation and target studies, respectively. As a further convenience, N_t represents the sample size of the target study.

For fixed means, confidence bounds for noncentrality are easily formulated using known distributional properties of the GLUM. Bounds for power and sample size follow directly from these results. Clark (1987) demonstrated that the approximate method suggested by Dudewicz (1972) yields optimistically small intervals in small samples, with bias decreasing as ν_{2e} increases. *Exact* bounds for ω can be calculated as follows. Assign lower (α_{cL}) and upper (α_{cU}) tail probabilities to define the confidence coefficient ($1 - \alpha_{cL} - \alpha_{cU}$), and let $c_{\text{crit}}(p \mid \nu)$ be the $100 \cdot (p)$ percentile from a central χ^2 with ν df. From standard Gaussian theory, recognize that $\nu_{2e} \cdot \widehat{\sigma}^2/\sigma^2 \overset{d}{=} \chi^2(\nu_{2e})$. Then observe that

$$\Pr\left\{C_{\text{crit}}(\alpha_{cL}|\nu_{2e}) < \frac{SSE}{\sigma^2} < C_{\text{crit}}(1-\alpha_{cU}|\nu_{2e})\right\} = 1 - \alpha_{cL} - \alpha_{cU}. \quad (2.1)$$

Multiplying all terms in the inequality by $SSH(\theta, N_t)/SSE$ yields

$$\Pr\left\{\frac{C_{\text{crit}}\left(\alpha_{cL}|\nu_{2e}\right)}{SSE} \cdot SSH(\theta, N_t) < \omega < \frac{C_{\text{crit}}(1-\alpha_{cU}|\nu_{2e})}{SSE} \cdot SSH(\theta, N_t)\right\} = 1 - \alpha_{cL} - \alpha_{cU}, \quad (2.2)$$

with $SSH(\theta, N_t)$ fixed under the assumption of known (nonzero) θ and sample *size* N_t. Consequently,

$$\widehat{\omega}_L = \frac{C_{\text{crit}}\left(\alpha_{cL}|\nu_{2e}\right)}{SSE} \cdot SSH(\theta, N_t) \quad (2.3)$$

and

$$\widehat{\omega}_U = \frac{C_{\text{crit}}\left(1-\alpha_{cU}|\nu_{2e}\right)}{SSE} \cdot SSH(\theta, N_t) \quad (2.4)$$

provide exact confidence bounds for ω. Choosing $\alpha_{cL} = 0$ yields a lower one-sided interval, $[0, \hat{\omega}_U]$, while choosing $\alpha_{cU} = 0$ yields an upper one-sided interval, $[\hat{\omega}_L, \infty)$.

By similar argument, one can derive a minimum length interval for ω based on the shortest unbiased confidence interval for variance (Tate and Klett 1959; Juola 1993). However, since $\omega \in \Re^+$ while power lies in the bounded interval $[\alpha_t, 1)$, minimum length intervals for noncentrality may not lead to minimum length intervals for power.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2.2 Confidence Bounds for Power

The strict monotone dependence of the noncentral F distribution function on noncentrality (Johnson and Kotz 1970, p. 193) ensures that an exact confidence interval for power follows from an exact interval for ω. The lower ($\hat{P_L}$) and upper ($\hat{P_U}$) bounds for power must satisfy the two equations

$$\widehat{P}_L = 1 - F_F[f_{crit}(1-\alpha_t)\,|\,\nu_1, \nu_{2t}, \widehat{\omega}_L] \quad (2.5)$$

and

$$\widehat{P}_U = 1 - F_F[f_{crit}(1-\alpha_t)\,|\,\nu_1, \nu_{2t}, \widehat{\omega}_U]. \quad (2.6)$$

The fact that power is a smooth, strictly increasing function of ω ensures that

$$\Pr\{\widehat{\omega}_L \le \omega \le \widehat{\omega}_U\} = \Pr\left\{\widehat{P}_L \le P \le \widehat{P}_U\right\}. \quad (2.7)$$

Alternately, consider using a distinct, and therefore approximate method to produce, say, $\tilde{\omega_L}$ and $\tilde{\omega_U}$. Then an approximate confidence interval for P results by observing

$$\Pr\{\tilde{\omega}_L \le \omega \le \tilde{\omega}_U\} = \Pr\left\{\tilde{P}_L \le \omega \le \tilde{P}_U\right\}. \quad (2.8)$$

2.3 Confidence Bounds for Sample Size

An upper confidence bound for sample size (N_U) provides the smallest N that ensures, with confidence ($1 - \alpha_{cL}$), a desired target power. To achieve the result, one must solve the equation

$$\widehat{\omega}_{N_U} = \frac{C_{\text{crit}}\left(\alpha_{cL}\,|\,\nu_{2e}\right)}{SSE} \cdot SSH(\theta, N_U) \quad (2.9)$$

for N_U such that

$$P = 1 - F_F[f_{crit}(1-\alpha_t)\,|\,\nu_1, \nu_{N_U}, \widehat{\omega}_{N_U}]. \quad (2.10)$$

Here ν_{N_U} is the error df for the target study, and is a function of N_U. The value of N_U is most easily obtained by iterative solution of Equations (2.9) and (2.10). The strict monotonicity of the power function makes this task straightforward. In practice, the use of discrete N introduces some inexactness in coverage.

2.4 Simultaneous Bounds for Power and SampleSize

A confidence region for an entire power curve may be desired. For example, consider power as a function of the difference between two means, δ. One may seek a confidence region for power over an interval of $\delta > 0$. For a particular value of δ, say δ_0,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\widehat{P}_L(\delta_0)=1-F_F[f_{\text{crit}}(1-\alpha_t)|\nu_1,\nu_{2t},\widehat{\omega}_L(\delta_0)] \quad (2.11)$$

and

$$\widehat{P}_U(\delta_0)=1-F_F[f_{\text{crit}}(1-\alpha_t)|\nu_1,\nu_{2t},\widehat{\omega}_U(\delta_0)] \quad (2.12)$$

provide exact confidence bounds. Considering a continuous set of δ_0 generates two curves, which in turn define a confidence region. The boundaries are defined by the locus of all points generated by computing (2.11) and (2.12) for all $\delta_0 > 0$ in the set. The strict monotonicity of the function ensures that

$$\Pr\left\{\widehat{P}_L(\delta) \le P(\delta) \le \widehat{P}_U(\delta)|\delta>0\right\}=1-\alpha_{cL}-\alpha_{cU}. \quad (2.13)$$

Hence the pointwise bounds provide simultaneous confidence intervals. See Figure 1, discussed in Section 4, for an example.

Similar logic applies to sample size and other power related functions. Consider the triplet (P, δ, N). Fixing any one of the three, and computing a second as a function of the third, leads to a potentially informative plot. Given the results of Sections 2.1–2.3, these applications lead to exact coverage (ignoring the use of discrete N).

3. Simulations For Coverage Accuracy

3.1 Simulation Methods

In order to assess both the Dudewicz/Clark and exact methods, simulations were conducted using SAS IML©. Bounds for the power of a one-way ANOVA were calculated, and results were tabulated for each method under the following conditions: significance level of the test ($\alpha_t \in \{.01, .05\}$), target power ($P \in \{.2, .5, .9\}$), number of groups (G $\in \{2, 4\}$), cell size ($N_G \in \{4, 8,16\}$), and target coverage (C $\in \{.80, .95\}$, with $\alpha_{cL} = \alpha_{cU}$). Without loss of generality, $\sigma^2 = 1$ for all simulations. A vector of cell means was chosen as $\beta = \gamma \cdot [0\ 1 \ldots\ G-1]'$, in which γ was calculated by iterative use of a power program (Muller, LaVange, Ramey, and Ramey 1992, app. A), to ensure the desired target power for each combination of α_t, P, G, N_G, and C. A total of 1,500 random samples of Gaussian data were generated for each combination and method. The statistic $\hat{\sigma}^2$ was calculated, as were lower and upper bounds for power. The proportion of events for which confidence bounds captured true power was determined for consecutive sets of 150 replications. These were used to test the significance of α_t, P, G, and N_G and their interactions as predictors of coverage using factorial ANOVA's with 35 model and 324 error degrees of freedom. The 150 cases for each binomial observation, the balanced design, and the small predicted differences in population proportions led us to expect accurate inference with the use of ANOVA's.

3.2 Simulation Results

Simulation results are summarized in Table 1. For the Dudewicz/Clark method, the main effects of G and N_G were significant at the .01 level ($p \le .001$). Results agreed with Clark's proof that the method provides asymptotically unbiased confidence intervals for power. Small error degrees of freedom, however, yielded rather liberal results. For example, with C = .95 and 6 error df, the observed coverage was only .872 (std err = .0012).

With the exact method, the $\alpha_t \times P \times G \times N_G$ interaction term for C = .80 was significant at the .01 level (p = .007). We discounted this lone small p value in light of the small amount

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of variance associated with the effect and the multiplicity of tests conducted. Empirical results did not deviate substantially from predicted values, and accurate coverage was obtained for all conditions.

4. Example Revisited

Consider the study of patients with deteriorating renal function (Falk et al. 1992) described in Section 1.2. Assuming true improvement due to dual therapies was .50 dL/mg, estimated power for $\alpha_t = .01$ was .960. The exact method with $\alpha_{cL} = \alpha_{cU} = .025$ leads to a 95% confidence interval for ω of [11.01, 36.88]. The 95% interval for power follows as [.688, .999]. Figure 1 provides the simultaneous confidence bands for the power function associated with a range of reciprocal serum creatinine differences.

One can also determine the minimum sample size necessary to ensure, with confidence $(1 - \alpha_{cL})$, a specified level of power. For this example, a sample size of 17.95 per treatment group ensures, with probability .975, a power of .900. Hence the addition of six subjects per treatment group increases the lower bound on power from .688 to .900. The practical necessity of choosing an integer number of subjects introduces some inexactness.

5. Discussion

Strong consideration should be given to one-sided confidence intervals for power. This contrasts sharply with the usual practice of rarely considering one-sided intervals or tests. The reasons for this recommendation are twofold. First, interest typically lies in one-sided statements like, “Power of at least….” Second, the asymmetry of the distribution of noncentrality and the restriction of power to the interval $[\alpha_t, 1)$ frequently mean one bound from a two-sided interval provides information of little practical value. Consider again the example of Section 4. Compare the two-sided 95% confidence interval for power of [.688, .999] with a one-sided 95% confidence interval of [.750, 1]. The change from a one-sided to a two-sided confidence interval has little effect on the upper bound, but a large effect on the lower bound.

Although the discussed methods apply to fixed means, in some cases estimated means must be considered in a power calculation. For example, government regulators often desire evidence of any toxicological effect of a compound, regardless of clinical significance. The practical importance of such an effect must be interpreted in terms of impact for the entire population. Some relevant work has been done on providing confidence intervals for noncentrality parameters from χ^2 and F distributions in this situation (Venables 1975; Lam 1987; Wright and O'Brien 1988). Venables suggested applying a truncated Cornish-Fisher expansion to provide estimated confidence intervals for noncentrality. Wright and O'Brien extended results of Dwass (1955) to provide approximate intervals for noncentrality and power. Lam described algorithms, based on a Newton−Raphson approach, that can provide confidence limits for the noncentrality parameters. Hardison, Langston, and Quade (1980) and Guirguis (1990) presented computational algorithms that are very useful in this context. With proper treatment of certain boundary conditions, these algorithms allow computing unbiased confidence bounds for noncentrality and power with simple random selection of an observed value. These methods have been discussed by Taylor (1993). The potential for severe biases in the mean estimates makes this a complex issue, particularly for retrospective power calculations. Research is in progress to provide a general solution.

Clearly, uncertainty exists for power calculations with models other than the GLUM. For example, when testing the difference between two binomial proportions, the work of Dozier and Muller (1993) can be combined with the present results to provide approximate intervals

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

for power. Similar techniques could be useful in methods as disparate as survival analysis and multivariate linear models.

The exact methods discussed in this paper only require algorithms for computing percentiles from central chi-square, central F, and noncentral F densities. The validity of the methods depends upon meeting the assumptions of the fixed-effects GLUM, especially homogeneity of variance. The many possible sources of heterogeneity should instill caution. Although thoroughly studied in the context of data analysis, the impact of violations on power has received much less attention.

6. Conclusions

Statisticians emphasize the need to associate a measure of uncertainty with any parameter estimate. Hence the information added to the example by considering exact confidence bounds for estimated power and sample size should not surprise any statistician. Such bounds provide valuable tools for the planning and evaluation of scientific research.

We also strongly recommend plots as in Figure 1. Graphical displays of this nature have been extremely well received by scientists with whom we have consulted. These plots convey large amounts of information, help clarify design tradeoffs, and return sample size decisions to the scientist.

Acknowledgments

Taylor's work was supported in part by NIEHS Training Grant 5-T32-ES07018, and NCI Program Project Grant P01 CA47 982-04. An earlier version of this paper was submitted by D. J. Taylor in partial fulfillment of the requirements for the M.S. degree in biostatistics. K. E. Muller's work was supported in part by NICHD Mental Retardation Research Center Grant 2-P30-HD03110-21, NCI Program Project Grant P01 CA47 982-04, and NIH Clinical Research Center Grant RR00046. The authors gratefully acknowledge the assistance of Dr. Paul W. Stewart in formulating the discussion of simultaneity in Section 2.4, and the comments of Dr. Lisa M. LaVange, Dr. Dana Quade, and Joseph A. Galanko, as well as anonymous reviewers, on earlier versions of the paper.

References

Abramowitz, M.; Stegun, IA. Handbook of Mathematical Functions. New York: Dover Publications; 1964.

Beal SL. Sample Size Determination for Confidence Intervals on the Population Mean and on the Difference Between Two Population Means. Biometrics. 1989; 45:969–977. [PubMed: 2790131]

Clark, A. MS thesis, Department of Biostatistics. University of North Carolina; Chapel Hill: 1987. Approximate Confidence Bounds for Estimated Power in Testing the General Linear Hypothesis.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. rev.. New York: Academic Press; 1977.

Dozier WG, Muller KE. Small-Sample Power of Uncorrected and Satterthwaite Corrected T Tests for Comparing Binomial Proportions. Communications in Statistics: Simulation and Computation. 1993; 22:245–264.

Dudewicz EJ. Confidence Intervals for Power with Special Reference to Medical Trials. Australian Journal of Statistics. 1972; 14:211–216.

Dwass M. A Note on Simultaneous Confidence Intervals. Annals of Mathematical Statistics. 1955; 26:146–147.

Falk RJ, Hogan SL, Muller KE, Jennette JC. Treatment of Progressive Membranous Glomerulopathy. Annals of Internal Medicine. 1992; 6:438–445. [PubMed: 1371211]

Guirguis GH. A Note on Computing the Noncentrality Parameter of the Noncentral F-Distribution. Communications in Statistics (B). 1990; 19:1497–1511.

Hardison, CD.; Langston, RD.; Quade, D. SUGI Supplemental Library User's Guide—Version 5. Cary, NC: SAS Institute, Inc.; 1980. SAS Functions to Compute Chi-Squared, F, and T

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Distributions and Their Inverses Allowing for Noncentrality and Fractional Degrees of Freedom; p. 413-416.

Johnson, NL.; Kotz, S. Continuous Univariate Distributions. Vol. 2. Boston: Houghton Mifflin; 1970.

Juola RC. More on Shortest Confidence Intervals. The American Statistician. 1993; 47:117–119.

Kupper LL, Hafner KB. How Appropriate are Popular Sample Size Formulas? The American Statistician. 1989; 43:101–105.

Lam Y. Confidence Limits for Noncentrality Parameters of Noncentral Chi-Square and F distributions. Proceedings of the Statistical Computing Section, American Statistical Association Annual Meetings. 1987:441–443.

Muller KE, Benignus VA. Increasing Scientific Power with Statistical Power. Neurotoxicology and Teratology. 1992; 14:211–219. [PubMed: 1386138]

Muller KE, LaVange LM, Ramey SL, Ramey CT. Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications. Journal of the American Statistical Association. 1992; 87:1209–1226.

Odeh, RE.; Fox, M. Sample Size Choice. New York: Marcel Dekker; 1991.

Searle, S. Linear Models. New York: John Wiley; 1971.

Tate RF, Klett GW. Optimal Confidence Intervals for the Variance of a Normal Distribution. Journal of the American Statistical Association. 1959; 54:674–682.

Taylor, DJ. Confidence Intervals for Power in the Fixed Effects General Linear Univariate Model; paper presented at Biometrics Society/ENAR; 1993; 1993.

Venables W. Calculation of Confidence Intervals for Noncentrality Parameters. Journal of the Royal Statistical Society, Ser B. 1975; 37:406–12.

Wright S, O'Brien RG. Power Analysis in an Enhanced GLM Procedure: What it Might Look Like. SAS Users Group International Conference Proceedings. 1988:1097–1100.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

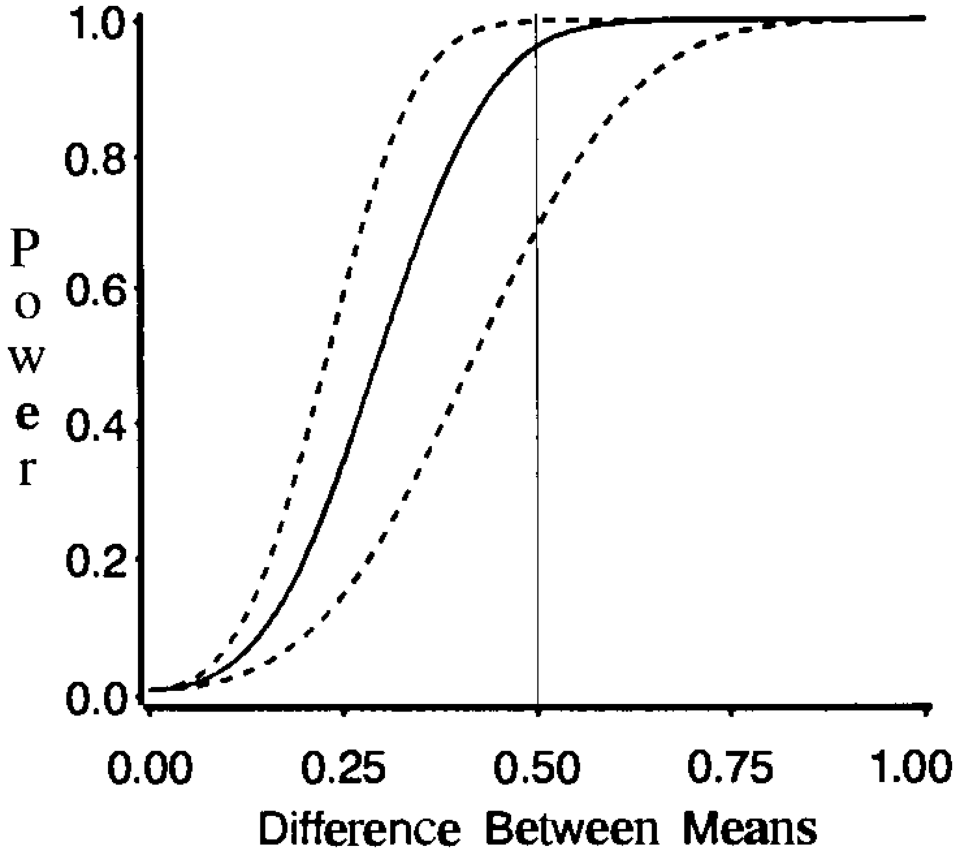

Figure 1.
Power to Detect a Difference in Mean Reciprocal Creatinine at α_t =.01. Solid line is estimated power. Dashed lines are exact 95% confidence bands on power. Vertical line corresponds to clinically significant difference of .50 dL/mg.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Overall ANOVA Tests of α_t, P, G, N_G and All Interactions as Predictors of Coverage (F Tests Involve 35 Model and 324 Error df)

	Coverage			Model		
Method	**Target**	**Observed**	**Std Err**	**F**	**R^2**	**p value**
Dudewicz/Clark	.95	.914	.0012	12.0	.581	<.001
Exact		.951	.0010	.94	.092	.566
Dudewicz/Clark	.80	.776	.0018	4.38	.321	<.001
Exact		.799	.0018	1.11	.107	.314

Am Stat. Author manuscript; available in PMC 2013 September 13.