Notes on Exact Power Calculations for \( t \) Tests and Analysis of Covariance

Yongqiang Tang

Tesaro, Waltham, MA

**ABSTRACT**

Tang derived the exact power formulas for \( t \) tests and analysis of covariance (ANCOVA) in superiority, noninferiority (NI), and equivalence trials. The power calculation in equivalence trials can be simplified by using Owen's \( Q \) function, which is available in standard statistical software. We extend the exact power determination method for ANCOVA to unstratified and stratified multi-arm randomized trials. The method is applied to the design of multi-arm trials and gold standard NI trials. Supplementary materials for this article are available online.

**1. Introduction**

Tang (2018a, 2018b) obtained the exact power formulas for some commonly used \( t \) tests in superiority, noninferiority (NI), and equivalence trials. The power determination for the analysis of covariance (ANCOVA) and \( t \) test with unequal variances in equivalence trials involves two-dimensional numerical integration. We show that the calculation can be simplified by using Owen's \( Q \) function, which is available in standard statistical software packages (e.g., SAS and R PowerTOST). We extend the method for ANCOVA to unstratified and stratified multi-arm randomized trials, and apply it to the power determination for multi-arm trials and gold standard NI trials (Pigeot et al. 2003).

We use the same notations as in Tang (2018a, 2018b). Let \( t(f, \lambda) \) denote the \( t \) distribution with \( f \) degrees of freedom and noncentrality parameter \( \lambda = t f, \rho \) the \( \rho \)th percentile of the central \( t \) distribution, \( \Phi(\cdot) \) the cumulative distribution function (CDF) of \( N(0, 1) \), \( F\{ t(f, \lambda) \} \) the CDF of a central \( F\{ t(f, f_2) \} \) distribution, and \( Q(\delta, \delta; 0, R) \) the \( Q \) function. We show that the calculation can be simplified by using Owen's \( Q \) function, which is available in standard statistical software packages (e.g., SAS and R PowerTOST). We extend the method for ANCOVA to unstratified and stratified multi-arm randomized trials, and apply it to the power determination for multi-arm trials and gold standard NI trials (Pigeot et al. 2003).

We use the same notations as in Tang (2018a, 2018b). Let \( t(f, \lambda) \) denote the \( t \) distribution with \( f \) degrees of freedom and noncentrality parameter \( \lambda \), \( t f, \rho \) the \( \rho \)th percentile of the central \( t \) distribution, \( \Phi(\cdot) \) the cumulative distribution function (CDF) of \( N(0, 1) \), \( F\{ t(f, f_2) \} \) the CDF of a central \( F\{ t(f, f_2) \} \) distribution, and \( Q(\delta, \delta; 0, R) \) the \( Q \) function. Let \( n g \) be the number of subjects in group \( g \), \( n \) the total size, \( M_0 \) the superiority (\( M_0 = 0 \)) or NI margin, and \( (M_1, M_a) \) the lower and upper equivalence margins. Without loss of generality, we assume high scores indicate better health.

**2. Two-Sample \( t \) Tests**

Let \( (\hat{\tau}, n^{-1} V) \) be the estimated effect and variance with true values \( (\tau_1, n^{-1} V) \) in a test based on the \( t \) distribution. Suppose \( \sqrt{\frac{\hat{\tau} - \tau_1}{n^{-1} V}} \sim N(0, 1) \) is independent of \( \xi = \sqrt{\frac{V}{\hat{\tau}}} \sim \sqrt{\frac{V}{\hat{\tau}}} \). In superiority and NI trials, we reject the null hypothesis when \( t = \sqrt{\frac{\hat{\tau} - M_0}{n^{-1} V}} > C = \frac{t f, 1 - \alpha}{2} \). If \( \hat{\tau} \) and \( V \) are known, the exact power is \( \Pr\left[t\left(\frac{\hat{\tau} - M_0}{n^{-1} V}\right) > C\right], \) or 1 minus the CDF of \( t \sim t\left(\frac{\hat{\tau} - M_0}{n^{-1} V}\right) \) evaluated at \( C \).

An equivalence test is significant if both \( t_1 = \sqrt{\frac{\hat{\tau} - M_0}{n^{-1} V}} > C \) and \( t_u = \sqrt{\frac{\hat{\tau} - M_0}{n^{-1} V}} < -C \). By the change of variable \( x = \sqrt{\frac{V}{\tau}} \), the exact power Equation (26) of Tang (2018b) can be rearranged in terms of Owen's \( Q \) function as

\[
\begin{align*}
P_{\text{equi}} &= \int_0^\infty \left[ \Phi(\delta_1 - C\sqrt{\xi}) - \Phi(\delta_2 + C\sqrt{\xi}) \right] dG(\xi) \\
&= Q(r(-C, \delta_2; 0, R) - Q(r(C, \delta_1; 0, R)),
\end{align*}
\]

where \( G(\xi) \) is the CDF of \( \xi \sim \frac{X^2}{f} \), \( \delta_2 = \frac{M_0 - \tau_1}{\sqrt{n^{-1} V}} < 0 \), \( \delta_1 = \frac{M_0 - \tau_1}{\sqrt{n^{-1} V}} > 0 \), and \( R = \sqrt{f(\delta_1 - \delta_2)} \).

In the \( t \) test with unequal variances (i.e., \( y_{ij} \overset{\text{iid}}{\sim} N(\mu_j, \sigma_j^2) \), \( y_{ij} \overset{\text{iid}}{\sim} N(\mu_j, \sigma_j^2) \)), the power of the superiority and NI trial is obtained from the fact (Moser, Stevens, and Watts 1989; Tang 2018b) that \( \sqrt{\frac{\hat{\tau} - M_0}{n^{-1} V}} h^* (u) = \sqrt{\frac{\hat{\tau} - M_0}{n^{-1} V}} (\frac{n-2}{(n-1)f_2^*} / (\alpha^2 + (n_0 - 1)g^2 / \sigma_j^2)) \) follows a noncentral \( t(n-2, (\frac{\hat{\tau} - M_0}{n^{-1} V})) \) distribution given \( u \)

\[
P_{\text{sup/ni}} = \int_0^\infty \Pr\left[t\left(n-2, \frac{|\tau_1 - M_0|}{\sqrt{n^{-1} V}}\right) > h(u)\right] dF_{n_1-1, n_0-1}(u).
\]

where \( \hat{\tau} = \hat{\mu}_1 - \hat{\mu}_0, s^2_g \) is the sample variance in group \( g \), \( n^{-1} V = \frac{s^2_g}{n_0} + \frac{s^2_g}{n_0} \), \( n^{-1} V = \frac{\tau_1^2}{n_1} + \frac{\tau_1^2}{n_2} \), \( u = \frac{s^2_g}{s^2_g / \sigma_j^2} \sim F(n_1 - 1, n_0 - 1) \), and

\[
h^* (u) = \sqrt{\frac{(n-2)|u\alpha^2/n_1 + \alpha^2/n_0|}{n^{-1} V (n_1-1)u + n_0 - 1}}.
\]
The exact equivalence power (Equation (A3) of Tang (2018b)) can be reexpressed as

\[
P_{\text{equi}} = \int_0^\infty \left\{ Q_{n-2} [-h(u), \delta_{21}; 0, R(u)] \right. - \left. Q_{n-2} [h(u), \delta_1; 0, R(u)] \right\} dF_{n-1|n_0-1}(u),
\]
where \( \delta_2 = \frac{M_{n-1}-\tau_1}{\sqrt{\sigma^2 V(1+q \kappa_{f_{12}})}} \), \( \delta_1 = \frac{M_{n-1}-\tau_1}{\kappa_{f_{12}} \sqrt{\sigma^2 V(1+q \kappa_{f_{12}})}} \), and \( R(u) = \frac{\sqrt{n-2}(\delta_1-\delta_2)}{2\delta_2} \).

Please see Tang (2018b) for numerical examples.

### 3. ANCOVA

Tang (2018a,b) derived the exact power formulas for ANCOVA analysis of two-arm trials. Below we present more general results for unstratified or stratified multi-arm randomized trials. Suppose subjects are randomized to \( K' = K + 1 \) treatment groups \((g = 0, \ldots, K)\) within each of \( h \) strata. In an unstratified trial, we set \( h = 1 \). Subjects in treatment group \( g \) are modeled by

\[
y_{gi} = \mu_g + z_{gi}\alpha_1 + \cdots + z_{gi}\alpha_r + x_{gi}^\top \beta + e_{gi} = \eta + \delta_g + z_{gi}\alpha_1 + \cdots + z_{gi}\alpha_r - 1 + x_{gi}^\top \beta + e_{gi},
\]

where \( z_{gi} \) (\( k = 1, \ldots, r-1 \)) is the indicator variable for the pre-stratification factors, \( \mu_g \) is the effect for treatment group \( g \), \( x_{gi} \) is the \( g \times 1 \) vector of baseline covariates, \( e_{gi} \sim N(0, \sigma^2) \), \( \eta = \mu_0 \), and \( \delta_g = \mu_g - \mu_0 \). In general, \( r \) equals the number of strata \( h \). In trials with multiple stratification factors, \( r < h \) if there is no interaction between some stratification factors. By the same arguments as the proof of Equation (15) in Tang (2018a), we obtain the variance for the linear contrast with coefficients \( (l_0, \ldots, l_K)\)

\[
\text{var}\left( \sum_{g=0}^{K} l_g x_{gi} \right) = \sigma^2 V_l \left( 1 + \frac{q}{n - q - r - K + 1} \right),
\]

where \( \sum_{g=0}^{K} l_g = 0 \) and \( l_g \) is the mean of \( z_{gi} = (z_{g1i}, \ldots, z_{gri}, \ldots)\) in group \( g \), \( S_{zg} = \sum_{g=0}^{K} \sum_{i=1}^{n_g} (z_{gi} - \bar{z}_g)^2 \), \( \bar{\bar{\tau}} \) is a function of the covariate \( x_{gi}^\top \), and \( V_l = \sum_{g=0}^{K} l_g^2/n_g + \sum_{i=1}^{n_g} l_g^2 \bar{z}_g^2 / (\sum_{g=0}^{K} l_g^2 \bar{z}_g^2) \).

In a two arm trial (Tang 2018a), \( V_l = \left( \sum_{i=1}^{n_h} n_{hi} n_{gi} / n_{gi}^2 \right)^{-1} \) if there is no restriction on the stratum effect (i.e. \( r = h \)), \( n_{gi} \) is the number of subjects in stratum \( s \), treatment group \( g \). A constant treatment allocation ratio is commonly used in practice. Then \( \bar{z}_g = \cdots = \bar{z}_K \) and \( V_l = \sum_{g=0}^{K} l_g^2/n_g \). Let \( \tau_1 = \sum_{g=0}^{K} l_g \mu_g, f = n - q - r - K \), and \( f_2 = f + 1 \). When \( x_{gi}^\top \) are normally distributed, \( \bar{\bar{\tau}} \sim F(q, f_2) \) and the exact power for the superior or NI test is

\[
P_{\text{sup/ni}} = \int_0^\infty \text{Pr} \left\{ t \left( t_f, \frac{(\tau_1 - M_0)^2}{\sigma^2 V_l(1 + q \bar{\bar{\tau}}/f_2)} \right) > t_{f,1-\alpha/2} \right\} dF_{q, f_2}(\bar{\bar{\tau}}),
\]

Formula (4) also provides very accurate power estimate for nonnormal covariates (Tang 2018b). In equivalence trials, the exact power is

\[
P_{\text{equi}} = \int_0^\infty \left\{ Q_{\delta_2(\bar{\bar{\tau}})} [-t_f, 1-\alpha/2, \delta_2(\bar{\bar{\tau}}); 0, R(\bar{\bar{\tau}})] \right. - \left. Q_{\delta_2(\bar{\bar{\tau}})} [t_f, 1-\alpha/2, \delta_2(\bar{\bar{\tau}}); 0, R(\bar{\bar{\tau}})] \right\} dF_{q, f_2}(\bar{\bar{\tau}}),
\]

where \( \delta_2(\bar{\bar{\tau}}) = \frac{M_{n-1}-\tau_1}{\sqrt{\sigma^2 V_l(1+q \bar{\bar{\tau}}/f_2)}} \), \( \delta_1 = \frac{M_{n-1}-\tau_1}{\kappa_{f_{12}} \sqrt{\sigma^2 V_l(1+q \bar{\bar{\tau}}/f_2)}} \), \( \bar{\bar{\tau}} > 0 \), and \( R(\bar{\bar{\tau}}) = \sqrt{\frac{q(\delta_1-\delta_2)}{\kappa_{f_{12}}}} \). The exact power formulas (Tang 2018b, eq. (A1); Tang 2018a, eq. (30)) for two arm trials are equivalent to Equation (5) at \( K = 1 \).

The power formulas (2)–(5) are of the form \( \int_0^\infty P_c(x) dF_{q, f_2}(x) \), and can be calculated as

\[
P = \int_0^\infty P_c(x) dF_{q, f_2}(x) = \int_0^1 P_c \left( F^{-1}_{q, f_2}(v) \right) dv.
\]

Below we give three hypothetical examples. Sample R code is provided in the supplementary materials. In each example, the simulated (SIM) power is evaluated based on 4,000,000 simulated datasets. There is more than 95% chance that the SIM power lies within 0.05% of the true power. In example 1, we perform the power calculation for a superiority trial. Subjects are randomized equally into \( K' = 3 \) groups (\( K = 2 \) experimental, or control treatment) stratified by gender \((z_{gi} = 1\) for male, 0 for female) and age \((z_{gi} = 1\) if old, 0 otherwise). There are 6 subjects per treatment group per stratum \((n_0 = n_1 = n_2 = 24, n = 72)\). There is no interaction between age and gender \((r = 3, h = 4)\), and the outcome is normally distributed as

\[
y_{gi} \sim N(\mu_g + 0.6 z_{gi} + 0.3 z_{gi}^2 + 0.5 x_{gi}^\top, 1),
\]

where \((\mu_0, \mu_1, \mu_2) = (0.6, 0.9, 0.9)\) and \( x_{gi} \sim N(0.2 z_{gi} + 0.4 z_{gi}^2, 1)\). We compare each experimental treatment versus control treatment at the Bonferroni-adjusted one tailed significance level of \( \alpha/2 = 0.0125 \). The exact power by formula (4) is 78.63% and 41.39%, and the SIM power is 78.62% and 41.39%, respectively, for the two tests.

Example 2 has similar setup to example 1 except that \((\mu_0, \mu_1, \mu_2) = (0, 0.05, 0.1)\) and the sample size is 30 per group per stratum \((n_0 = n_1 = n_2 = 120, n = 360)\). The aim is to establish the equivalence of each experimental treatment versus control treatment at \( \alpha/2 = 0.0125 \). The margin is \((M_0, M_0) = (-0.5, 0.5)\). The exact power by formula (5) is 79.14% and 86.72%, respectively, for the two tests, while the SIM power is 79.14% and 86.71%.

In example 3, we design a three-arm "gold standard" NI trial (Pigeot et al. 2003). It consists of placebo \((g = 0)\), an active control treatment \((g = 1)\) and an experimental treatment \((g = 2)\). The set up is similar to example 1 except that \((\mu_0, \mu_1, \mu_2) = (0, 0.1, 1.1)\), and the sample size is 10 per group per stratum \((n_0 = n_1 = n_2 = 40, n = 120)\). Two tests are conducted at the one-sided significance level of \( \alpha/2 = 0.025 \). Test 1 evaluates the superiority of treatment 1 over placebo. The power for this test (exact \( P_1 = 99.29\% \), SIM 99.28%) is very close to 1. In test 2, we assess the NI of treatment 2 to treatment
1 by demonstrating that treatment 2 preserves at least 50% of the efficacy of treatment 1 compared to placebo (i.e., \( \frac{\mu_2 - \mu_0}{\mu_1 - \mu_0} > 50\% \) or \( \mu_2 - 0.5\mu_1 - 0.5\mu_0 > 0 \)). The exact power of test 2 is \( P_2 = 86.41\% \) (SIM power 86.41%). The NI is claimed only if both tests are significant (Pigeot et al. 2003), and the overall power is at least \( P_1 + P_2 - 1 = 85.70\% \) while the simulated power is 85.80%.

**Supplementary Materials**

Sample R code is provided in the supplementary materials.

**References**

Moser, B. K., Stevens, G. R., and Watts, C. L. (1989), “The Two-Sample \( t \) Test Versus Satterthwaite’s Approximate \( F \) Test,” *Communications in Statistics—Theory and Methods*, 18, 3963–3975. [394]

Pigeot, I., Schafer, J., Rohmel, J., and Hauschke, D. (2003), “Assessing Non-Inferiority of a New Treatment in a Three-Arm Clinical Trial Including a Placebo,” *Statistics in Medicine*, 22, 883–899. [394,395,396]

Tang, Y. (2018a), “Exact and Approximate Power and Sample Size Calculations for Analysis of Covariance in Randomized Clinical Trials With or Without Stratification,” *Statistics in Biopharmaceutical Research*, 10, 274–286. [394,395]

——— (2018b), “A Noniterative Sample Size Procedure for Tests Based on \( t \) Distributions,” *Statistics in Medicine*, 37, 3197–3213. [394,395]