

How to get accurate sample size and power with nQuery Advisor[®]

Brian Sullivan

Statistical Solutions Ltd.

ASA Meeting, Chicago, March 2007



Sample Size

Two group t-test

χ^2 -test

Survival Analysis

2 × 2 Crossover

Odds Ratio



“What sample size do I need?”

- Formulate the study
- Specify analysis parameters
- Specify effect size for test
- Compute sample size or power
- Sensitivity analysis
- Choose sample size



“What sample size do I need?”

- Formulate the study
- Specify analysis parameters
- Specify effect size for test
- Compute sample size or power
- Sensitivity analysis
- Choose sample size



“What sample size do I need?”

- Formulate the study
- Specify analysis parameters
- Specify effect size for test
- Compute sample size or power
- Sensitivity analysis
- Choose sample size



“What sample size do I need?”

- Formulate the study
- Specify analysis parameters
- Specify effect size for test
- Compute sample size or power
- Sensitivity analysis
- Choose sample size



“What sample size do I need?”

- Formulate the study
- Specify analysis parameters
- Specify effect size for test
- Compute sample size or power
- Sensitivity analysis
- Choose sample size



“What sample size do I need?”

- Formulate the study
- Specify analysis parameters
- Specify effect size for test
- Compute sample size or power
- Sensitivity analysis
- Choose sample size



Formulate the study

- Type of analysis being undertaken
 - Clinical trial
 - Laboratory study
 - Survey research
- What is being measured?
- How is it being analyzed?



Formulate the study

- Type of analysis being undertaken
 - Clinical trial
 - Laboratory study
 - Survey research
- What is being measured?
- How is it being analyzed?



Formulate the study

- Type of analysis being undertaken
 - Clinical trial
 - Laboratory study
 - Survey research
- What is being measured?
- How is it being analyzed?



Specify analysis parameters

- What parameters are involved?
 - Significance level
 - Test type: one or two sided



Specify analysis parameters

- What parameters are involved?
 - Significance level
 - Test type: one or two sided



Specify effect size

- Use information from previous/similar studies
- What effect size would be important to detect?



Compute sample size or power

- Specify a value for sample size or power and perform the calculation solving for the other



Recalculate sample size for different scenarios about effect size

- Recalculate sample size for different levels of power
- Compare different sample sizes
- Plot sample size vs power

Choose sample size

- Based on the trade off between power attainable for the desired effect size and sample size feasibility, choose a sample size for the study



Two group t-test

Study Design

Does a new drug reduce anemia in elderly women after hip fracture?

- Two group, randomized, parallel, double-blind study
- Drug for two weeks, three times per week
- New drug vs placebo
- Equal n's



Two group t-test

Study Design

Does a new drug reduce anemia in elderly women after hip fracture?

- Two group, randomized, parallel, double-blind study
- Drug for two weeks, three times per week
- New drug vs placebo
- Equal n's



Outcome and Analysis

- Primary outcome: Mean change in hematocrit level (percentage of red blood cells in the blood) from pre-treatment to post-treatment
- Compare mean change in hematocrit level between two groups using two-group t-test
- Null Hypothesis: Mean change is the same in both groups



Outcome and Analysis

- Primary outcome: Mean change in hematocrit level (percentage of red blood cells in the blood) from pre-treatment to post-treatment
- Compare mean change in hematocrit level between two groups using two-group t-test
- Null Hypothesis: Mean change is the same in both groups



Outcome and Analysis

- Primary outcome: Mean change in hematocrit level (percentage of red blood cells in the blood) from pre-treatment to post-treatment
- Compare mean change in hematocrit level between two groups using two-group t-test
- Null Hypothesis: Mean change is the same in both groups



Calculation Parameters

- Significance level, α
- Test type: one or two sided, s
- Expected difference, $d = |\mu_1 - \mu_2|$
- Standard deviation, σ
- Effect size, $\delta = d/\sigma$
- Power, $1 - \beta$
- Sample size, n



Previous Analysis

There have been four previous studies from which data can be drawn.

Study	n	Population	Mean Hct
1	6	Elderly women with fracture	32.3%
2	32	Elderly women no fracture	33.5%
3,4		Change in placebo	0.0
3,4		Change in drug groups	2.7 - 4.7

Expected change: 0% in placebo group
2 - 2.4% in drug group



Previous Analysis

There have been four previous studies from which data can be drawn.

Study	n	Population	Mean Hct
1	6	Elderly women with fracture	32.3%
2	32	Elderly women no fracture	33.5%
3,4		Change in placebo	0.0
3,4		Change in drug groups	2.7 - 4.7

Expected change: 0% in placebo group
2 - 2.4% in drug group

Previous Analysis

- From previous studies the standard deviation of hematocrit levels were between 3% and 6%.
- Here we need the standard deviation for the *change* in hematocrit level and there is less information on this.
- From previous studies it can be estimated to be between 1.5% to 2.5%.
- A reasonable estimate would be 2%.



Example

Using the following details we can proceed with the calculation:

Parameter	Value
Significance level	.05
Test type	2 sided
Expected difference	2 - 2.4
Standard deviation	2

Illustrated in nQuery Advisor file: **hematocrit.nqa**.

Sensitivity Analysis

Two group t-test of equal means (equal n's)

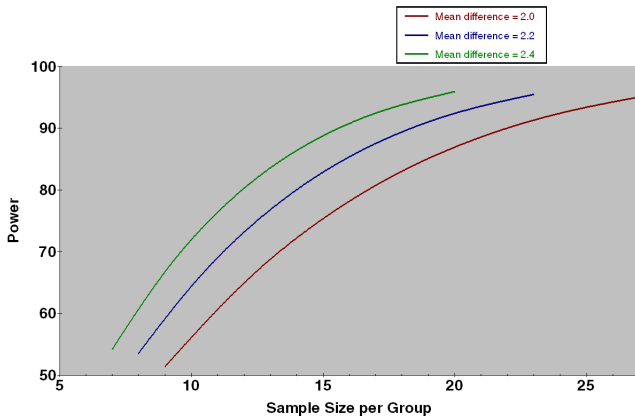


Figure: Plot of Sample Size vs Power



Choose Sample Size

A sample size of 20 in each group will have 92% power to detect a difference in means of 2.200 assuming that the common standard deviation is 2.000 using a two group t-test with a 0.050 two-sided significance level.



χ^2 -test

Study Design

Does a certain diet drug combination increase the likelihood of requiring valve surgery?

- Case-Control study
- Cases: patients who had valve surgery in defined two year period
- Controls: matched using age, gender and body-mass index



χ^2 -test

Study Design

Does a certain diet drug combination increase the likelihood of requiring valve surgery?

- Case-Control study
- Cases: patients who had valve surgery in defined two year period
- Controls: matched using age, gender and body-mass index



Outcome and Analysis

- Primary outcome: Proportion of cases and controls that have taken the diet drugs
- Compare proportions of cases and controls that have taken diet drugs using McNemar's χ^2 test for paired responses



Outcome and Analysis

- Primary outcome: Proportion of cases and controls that have taken the diet drugs
- Compare proportions of cases and controls that have taken diet drugs using McNemar's χ^2 test for paired responses



Calculation Parameters

- Significance level, α
- Test type: one or two sided, s
- Expected difference in proportions, $\delta = |\pi_1 - \pi_2|$
- Proportion of discordant pairs, $\eta = \pi_{10} + \pi_{01}$
- Power, $1 - \beta$
- Sample size, n



Previous Analysis

- It is expected that about 5% of controls would have taken the diet drugs.
- The investigators are interested in detecting an increase to 10% among the cases with 90% power.



Previous Analysis

A two by two table of proportions of cases and controls using the diet drugs is postulated as:

	Case yes	Case no	Row proportion
Control yes	0.03	0.02	0.05
Control no	0.07	0.88	0.95
Column proportion	0.10	0.90	1.00



Example

Using the following details we can proceed with the calculation:

Parameter	Value
Significance level	0.05
Test type	2 sided
Difference in proportions	0.05
Proportion of discordant pairs	0.09

Illustrated in nQuery Advisor file: **surgery.nqa**.



Sensitivity Analysis

McNemar's test (Chi-square) of equality of paired proportions

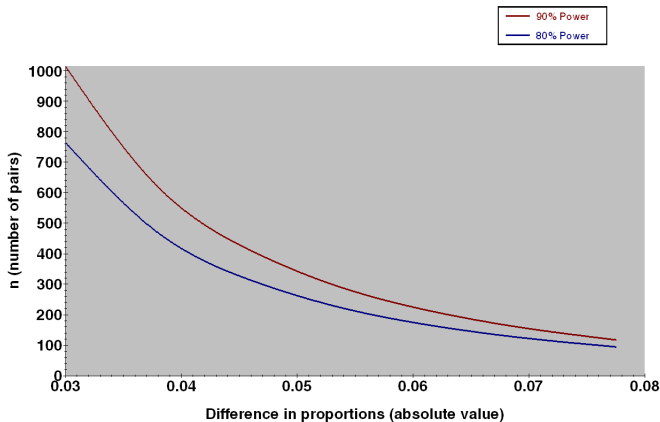


Figure: Plot of Difference in Proportions vs Sample Size



Choose Sample Size

A sample size of 342 pairs will have 90% power to detect a difference in proportions of 0.050 when the proportion of discordant pairs is expected to be 0.090 and the method of analysis is a McNemar's test of equality of paired proportions with a 0.050 two-sided significance level.



Survival Analysis

Study Design

Does TIPS surgery for bleeding esophageal varices prolong life over shunt surgery?

- Survival analysis log-rank test
- Randomization into two parallel groups
- All patients must be followed for a minimum of one year



Survival Analysis

Study Design

Does TIPS surgery for bleeding esophageal varices prolong life over shunt surgery?

- Survival analysis log-rank test
- Randomization into two parallel groups
- All patients must be followed for a minimum of one year



Outcome and Analysis

- Primary outcome: Time to death due to any cause
- Groups compared using survival analysis log-rank test
- Null Hypothesis: Time to death is the same in both groups



Outcome and Analysis

- Primary outcome: Time to death due to any cause
- Groups compared using survival analysis log-rank test
- Null Hypothesis: Time to death is the same in both groups



Outcome and Analysis

- Primary outcome: Time to death due to any cause
- Groups compared using survival analysis log-rank test
- Null Hypothesis: Time to death is the same in both groups



Calculation Parameters

- Significance level, α
- Test type: one or two sided, s
- Group proportions at time t , π_1, π_2
- Hazard ratio, $h = \frac{\ln(\pi_1)}{\ln(\pi_2)}$
- Power, $1 - \beta$
- Sample size, n



Previous Analysis

- Results from previous studies give an indication of the expected proportions.
- It is expected that 65% of patients will survive one year after shunt surgery.
- TIPS surgery would be considered to be markedly worse if only 45% of patients survived one year.



Example

Using the following details we can proceed with the calculation:

Parameter	Value
Significance level	0.05
Test type	2 sided
Group 1 proportion	0.65
Group 2 proportion	0.45

Illustrated in nQuery Advisor file: **surgery2.nqa**.

Sensitivity Analysis

Log-rank test of survival in two groups followed for fixed time, constant hazard ratio

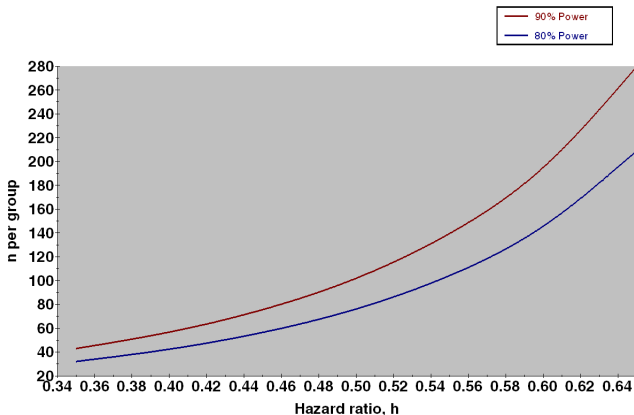


Figure: Plot of Hazard vs Sample Size



Choose Sample Size

When the sample size in each group is 131, with a Total number of events required, E , of 110, a 0.050 level two-sided log-rank test for equality of survival curves will have 90% power to detect the difference between a Group 1 proportion π_1 at time t of 0.650 and a Group 2 proportion π_2 at time t of 0.450 (a constant hazard ratio of 0.539); this assumes no dropouts before time t .



2 × 2 Crossover

Study Design

Does a new drug have similar steady state blood levels to that of a standard drug?

- Two-way crossover design
- Two period, two treatment AB, BA
- New drug vs standard drug



2 × 2 Crossover

Study Design

Does a new drug have similar steady state blood levels to that of a standard drug?

- Two-way crossover design
- Two period, two treatment AB, BA
- New drug vs standard drug



Outcome and Analysis

- **Primary outcome: Mean steady state blood level**
- Interested to know will the mean steady state blood level for the new drug be within 15% of that of the standard
- Two group t-test (TOST) for equivalence in means
- Null Hypotheses: Difference \geq upper level and Difference \leq lower level



Outcome and Analysis

- Primary outcome: Mean steady state blood level
- Interested to know will the mean steady state blood level for the new drug be within 15% of that of the standard
- Two group t-test (TOST) for equivalence in means
- Null Hypotheses: Difference \geq upper level and Difference \leq lower level



Outcome and Analysis

- Primary outcome: Mean steady state blood level
- Interested to know will the mean steady state blood level for the new drug be within 15% of that of the standard
- Two group t-test (TOST) for equivalence in means
- Null Hypotheses: Difference \geq upper level and Difference \leq lower level



Outcome and Analysis

- Primary outcome: Mean steady state blood level
- Interested to know will the mean steady state blood level for the new drug be within 15% of that of the standard
- Two group t-test (TOST) for equivalence in means
- Null Hypotheses: Difference \geq upper level and Difference \leq lower level



Calculation Parameters

- Significance level (one sided), α
- Lower equivalence limit for $\mu_T - \mu_S$, Δ_L
- Upper equivalence limit for $\mu_T - \mu_S$, Δ_U
- Expected difference in means, $d = \mu_T - \mu_S$
- Crossover ANOVA, $\sqrt{\text{MSE}}$
- Standard deviation of differences, σ_d
- Power, $1 - \beta$
- Sample size, n



Previous Analysis

- Previous study on standard drug indicated a mean steady state blood level of 16
- Pilot study of two different doses for new drug showed standard deviation of about 2.8 for each
- The correlation between blood levels for the two periods was 0.55



Example

Using the following details we can proceed with the calculation:

Parameter	Value
Significance level	0.05
Lower limit	-2.4
Upper limit	2.4
Expected difference	0.0
Standard deviation	2.8
Correlation	0.55

Illustrated in nQuery Advisor file: **blood.nqa**.



Sensitivity Analysis

t-tests (TOST) of equivalence in means for crossover design

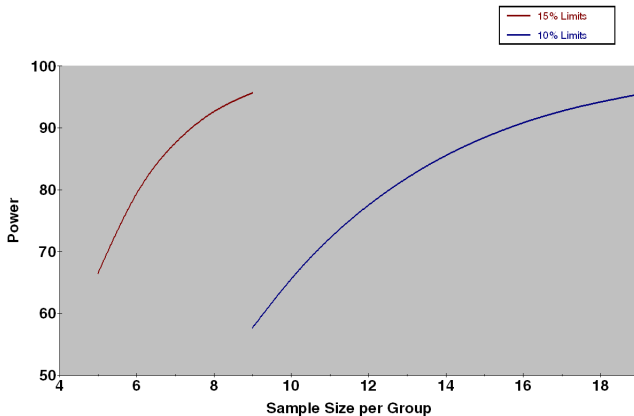


Figure: Plot of Sample Size vs Power



Choose Sample Size

When the sample size in each sequence group is 7, (a total sample size of 14) a crossover design will have 80% power to reject both the null hypothesis that the test mean minus the standard mean is below -2.400 and the null hypothesis that the test mean minus the standard mean is above 2.400 i.e., that the test and standard are not equivalent, in favor of the alternative hypothesis that the means of the two treatments are equivalent, assuming that the expected difference in means is 0.000, the Crossover ANOVA \sqrt{MSE} is 1.878 (the Standard deviation of differences, σ_d , is 2.656) and that each test is made at the 5.0% level.



Odds Ratio

Study Design

Does an educational program for expectant mothers reduce the risk of preterm birth?

- two group study, odds ratio
- cases vs controls
- cases receive education
- equal n's



Odds Ratio

Study Design

Does an educational program for expectant mothers reduce the risk of preterm birth?

- two group study, odds ratio
- cases vs controls
- cases receive education
- equal n's



Outcome and Analysis

- **Primary outcome: Proportion of preterm births**
- The educational program would be of interest if it reduced the rate of preterm births to produce an odds ratio of 0.5
- How wide is the 95% confidence interval for the odds ratio likely to be if 500 women are assigned to each group?



Outcome and Analysis

- Primary outcome: Proportion of preterm births
- The educational program would be of interest if it reduced the rate of preterm births to produce and odds ratio of 0.5
- How wide is the 95% confidence interval for the odds ratio likely to be if 500 women are assigned to each group?



Outcome and Analysis

- Primary outcome: Proportion of preterm births
- The educational program would be of interest if it reduced the rate of preterm births to produce and odds ratio of 0.5
- How wide is the 95% confidence interval for the odds ratio likely to be if 500 women are assigned to each group?



Calculation Parameters

- Significance level, α
- Test type: one or two sided, s
- Group proportions, π_1, π_2
- Odds ratio, $\psi = \frac{\pi_2(1-\pi_1)}{\pi_1(1-\pi_2)}$
- Log odds ratio, $B = \ln(\psi)$
- Distance from B to limit, ω
- Sample size, n



Previous Analysis

- Previous studies of similar populations have found preterm birth rates between 7% and 9%.
- A reasonable estimate would be 8%.



Example

Using the following details we can proceed with the calculation:

Parameter	Value
Significance level	0.05
Test type	2 sided
Control proportion	0.08
Odds ratio	0.5
n per group	500

Illustrated in nQuery Advisor file: **education.nqa**.



Sensitivity Analysis

Confidence interval for $\ln(\text{odds ratio})$ (large equal n 's)

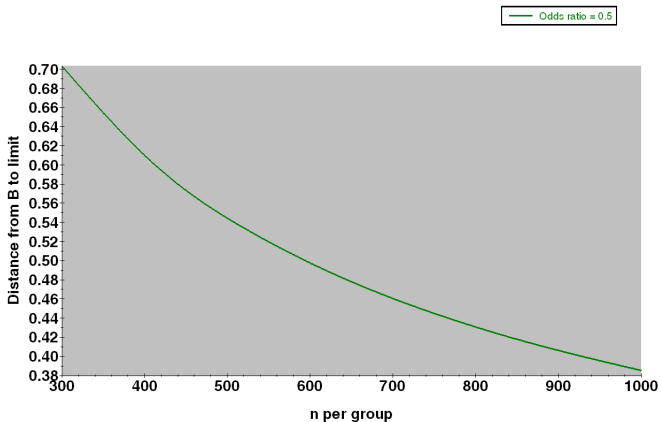


Figure: Plot of Sample Size vs Distance from B to limit



Choose Sample Size

When the sample size in each group is 500, the Control proportion, π_1 , is 0.080, and the Education proportion, π_2 , is 0.042, a two-sided 95.0% confidence interval for a $\ln(\text{odds ratio})$ expected to be -0.693 will extend 0.545 from the observed $\ln(\text{odds ratio})$ (corresponding to confidence limits of 0.290 and/or 0.862 for an odds ratio of 0.500).



Bibliography



Elashoff J.D. nQuery Advisor Version 6.01 User's Guide. (2005) Los Angeles, CA.



Dixon, W.J., Massey, F.J. Introduction to Statistical Analysis. 4th Edition McGraw-Hill (1983)



O'Brien, R.G., Muller, K.E. Applied Analysis of Variance in Behavioral Science Marcel Dekker, New York (1983) pp. 297-344



Miettinen, O.S. "On the matched-pairs design in the case of all-or-none responses" Biometrics 24(1986) pp. 339-352



Freedman, L.S. "Tables of the number of patients required in clinical trials using the logrank test" Statistics in Medicine 1(1982) pp. 121-129



Collett, D. Modelling Survival Data in Medical Research Chapman and Hall (1994) Section 9.2



Chow, S.C, Liu, J.P. Design and Analysis of Bioavailability and Bioequivalence Studies Marcel Dekker, Inc. (1992)



Schuurmann, D.J. "A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability" Journal of Pharmacokinetics and Biopharmaceutics 15(1987) pp. 657-680



Phillips, K.E. "Power of the two one-sided tests procedure in bioequivalence" Journal of Pharmacokinetics and Biopharmaceutics 18(1990) pp. 137-143



Owen, D.B. "A special case of a bivariate non-central t-distribution" Biometrika 52(1965) pp. 437-446



O'Neill, R.T. "Sample size for estimation of the odds ratio in unmatched case-control studies" American Journal of Epidemiology 120(1984) pp. 145-153

