Issues in Sample Size Calculations with Multiple Must-win Comparisons

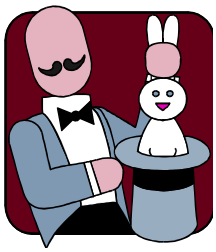
Steven A. Julious
University of Sheffield

Introduction

- Introduce the problem
- Describe some examples of multiple must-win
- Give a solution for using bioequivalence as a case study for two endpoints
- Give a solution from superiority for two or more endpoints

Sample Size Calculations

Not as simple as……



Size matters….

A study that is **too small** or **too large** poses ethical problems

Too few

You will not be able to answer the question posed

Too many

You will waste resources, and possibly give patients a treatment proven to be inferior

The Three Most Important Components of any Study Are

•Design
•Design
•Design

The sample size is just one component of the design

Why sample size calculations?

- Required by ethical committees
- Required by grant giving bodies and funding agencies
- Required by BMJ and other journals in checklist for writing up papers

Why not sample size calculations?

- Rarely enough information for precise calculations.
- Very sensitive to assumptions.
- Based on only one end-point.
- Main criteria are usually availability of patients, finance, resources and time.
  - Sample sizes based on feasibility should be disclosed
  - Still calculations that can be done – given limited resource is it still reasonable to do the study?
- More of this later…...

Why not sample size calculations?

- Rarely enough information for precise calculations.
- Very sensitive to assumptions.
- **Based on only one end-point.**
- Main criteria are usually availability of patients, finance, resources and time.
  - Sample sizes based on feasibility should be disclosed
  - Still calculations that can be done – given limited resource is it still reasonable to do the study?
- More of this later…...

The Problem for Today

The Problem

- Multiplicity in the context of the Type I error is a well known problem
- If we have multiple "or" comparisons
  - Such that a study can be significant if either is significant
- Then the significance level should be adjusted appropriately to maintain the nominal (usually) 5% level
- There is no issue with the Type II error

The Problem (cont.)

- For multiple must-win comparisons there are less issues with the Type I error
  - As all comparisons must hold for the trial to be successful
- There is an issue now with the Type II error
  - Now we have "or" comparisons as the study can fail if any comparison fails
  - There is now an issue of multiplicity for this error

When Can Multiple Must-win Endpoints Win Occur?

Example 1
Multiple Co-primary Data Sets

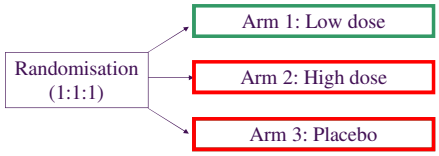
Multiple Co-primary Data Sets

- The may be instances where there is a need to show an effect in multiple patient populations
  - For non-inferiority to hold non-inferiority must be show in the ITT and PP population

Example 2
Dose Response

Trial Design

- Randomised, double-blind, parallel group trial



- Comparisons of interest:
  1. Arm 1 vs. Arm 3
  2. Arm 2 vs. Arm 3

If both must hold there is a multiplicity in Type II error

Note comparisons 1. and 2. have a correlation of 0.5

Example 3
Assessment of Superiority

Multiple Co-primary Endpoints

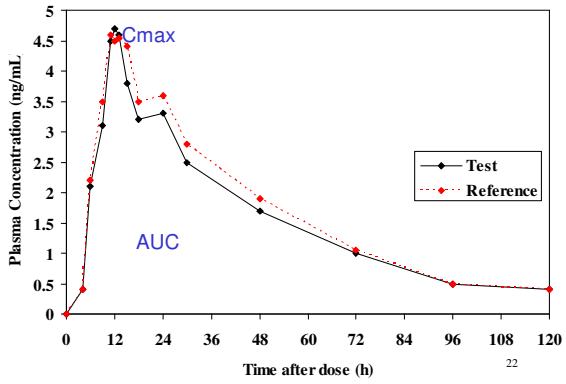
- The may be instances where there is a need to show an effect in multiple patient endpoints
  - Better reflects the multi-dimensionality of disease
- Having multiple endpoints better than the alternatives
  - Composite endpoints
  - Just using one endpoint

Example 4
Bioequivalence Study

Bioequivalence

- Bioequivalence studies are conducted to show that two formulations of a drug have similar bioavailability i.e. similar rate and extent of drug absorption
- Assumption: Equivalent bioavailability ensures equivalent therapeutic effect (both efficacy and safety)

Example of Results from a Bioequivalence Study



For Bioequivalence

- The concentration time profiles for the test and reference formulations need to be super-imposable.
- This is usually done by assessing if the rate (Cmax) and extent (AUC) of absorption are the same.
- AUC and Cmax must be equivalent to declare bioequivalence

Hypotheses

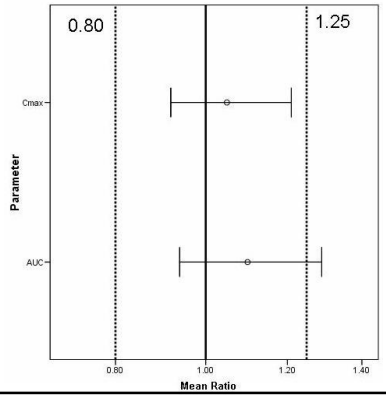
Hypothesis of interest:

$H_0 : \mu_T / \mu_R \geq 1.25$ and $\mu_T / \mu_R \leq 0.80$

$H_1 : 0.80 < \mu_T / \mu_R < 1.25$

Conclude bioequivalence if the 90% confidence interval for μ_T / μ_R is completely contained in the interval (0.80, 1.25)

Results: Test: Reference



How to Calculate the Sample Size for a Bioequivalence Study with Two Endpoints?

Solution 1: Ignore the Issue

Bioequivalence Sample Size Estimation: Normal Approximation

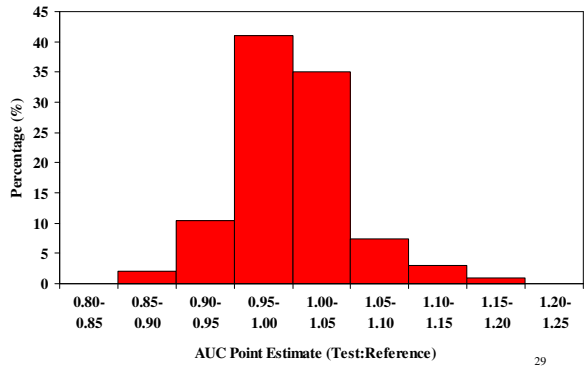
- Power (and sample size) can be calculated iteratively from:

$$1-\beta=\Phi\left(\sqrt{\frac{(\log(\mu_T/\mu_R)-\log(1.25))^2 n}{2\sigma_w^2}}-Z_{1-\alpha}\right)+\Phi\left(\sqrt{\frac{(\log(\mu_T/\mu_R)-\log(0.80)^2 n}{2\sigma_w^2}}-Z_{1-\alpha}\right)-1$$

- Where σ_w is the within subject standard deviation (on the logged scale)
- For the special case of $\mu_T / \mu_R = 1$

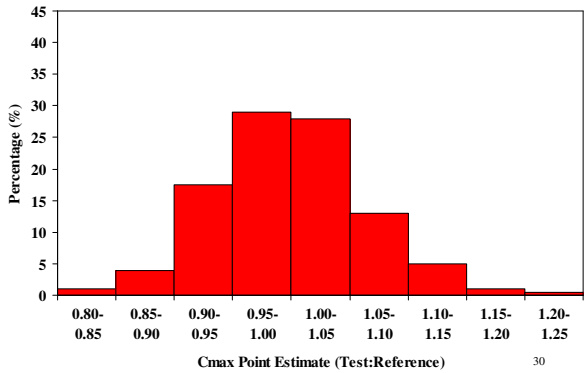
$$n=\frac{2\sigma_w^2\left(Z_{1-\beta/2}+Z_{1-\alpha}\right)^2}{(\log(1.25))^2}$$

Distribution of AUC point estimates from ANDA applications over a five year period (1996-2001)



Julious (2009) based on the work of Haider (2004)

Distribution of Cmax point estimates from ANDA applications over a five year period (1996-2001)



Julious (2009) based on the work of Haider (2004)

Solution 2: Apply a Bonferroni "Correction"

Simple Bonferroni Adjustment

- To do this, for two comparisons of interest, each with the same standardised effect size of interest, and the overall Type II error level set at 10%, we would set the Type II error per comparison to be 5.1% which comes from the following general result

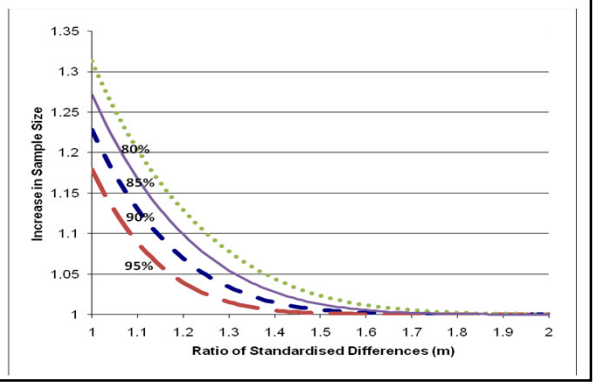
$$\beta_t = 1 - \sqrt[t]{1-\beta}$$

$$\beta_t = 1 - \sqrt[2]{1-0.1} = 0.0513$$

Adjustment for different t…

t	β_t
1	0.1000
2	0.0513
3	0.0345
4	0.0260
5	0.0209
6	0.0174
7	0.0149
8	0.0131
9	0.0116
10	0.0105

Increase in Sample Size for Two Must-win Endpoints (Correlation Assumed to be Zero)



However,

- This application is probably too conservative as there is a strong long likelihood endpoints will be correlated

Solution 3: Use the Bivariate Normal Distribution

Univariate Normal Distribution

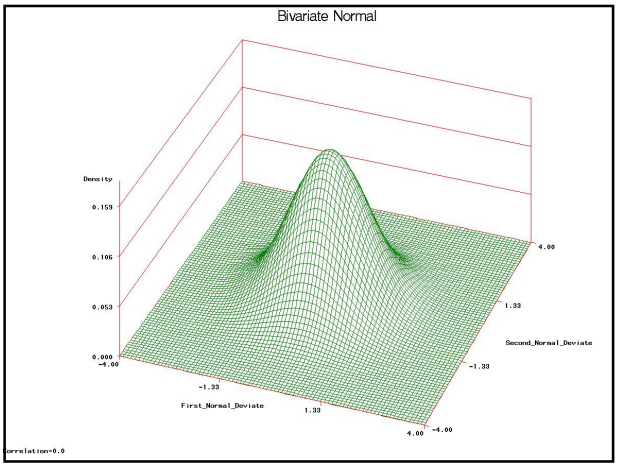
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

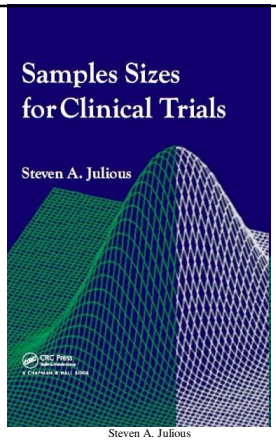
where μ = mean
σ = standard deviation

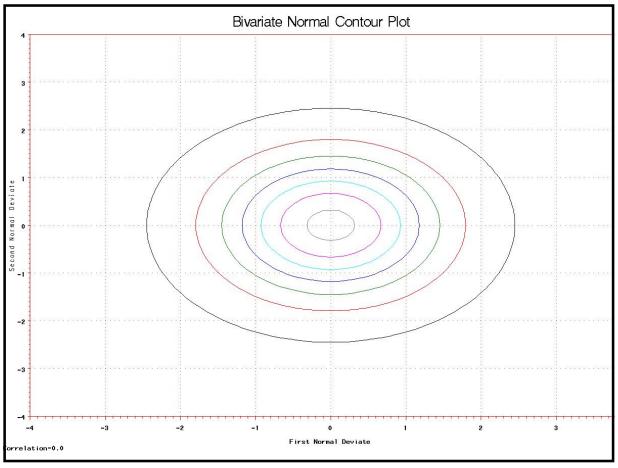
Bivariate Normal

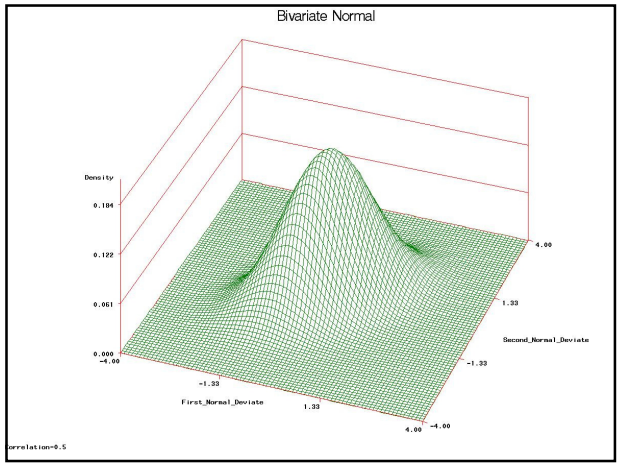
$$f(x) = \frac{1}{2\pi\sigma_{11}\sigma_{22}\sqrt{(1-\rho_{12}^2)}} e^{\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_{11}}\right)^2+\left(\frac{x_2-\mu_2}{\sigma_{22}}\right)^2-2\rho_{12}\left(\frac{x_1-\mu_1}{\sigma_{11}}\right)\left(\frac{x_2-\mu_2}{\sigma_{22}}\right)\right]}$$

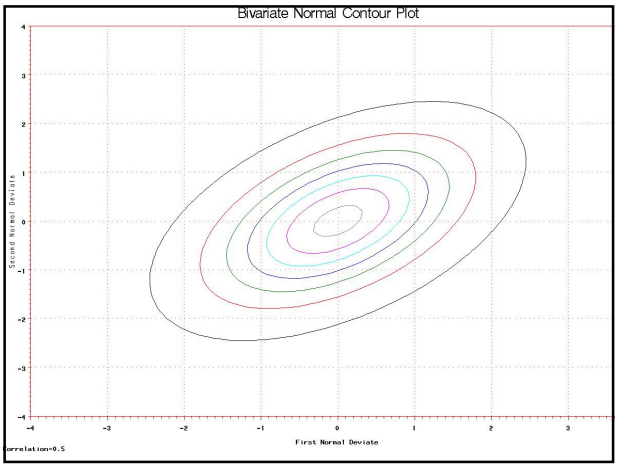
- Note when the correlation is zero this becomes two univariate Normals multiplied

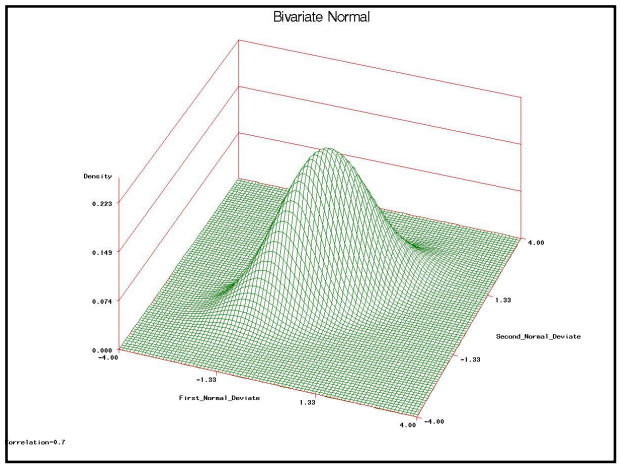


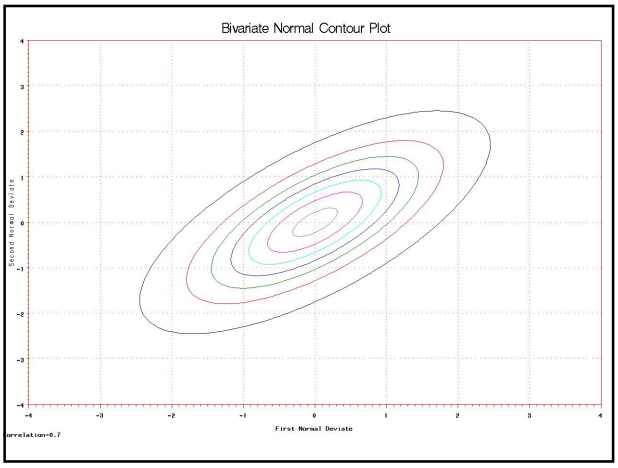


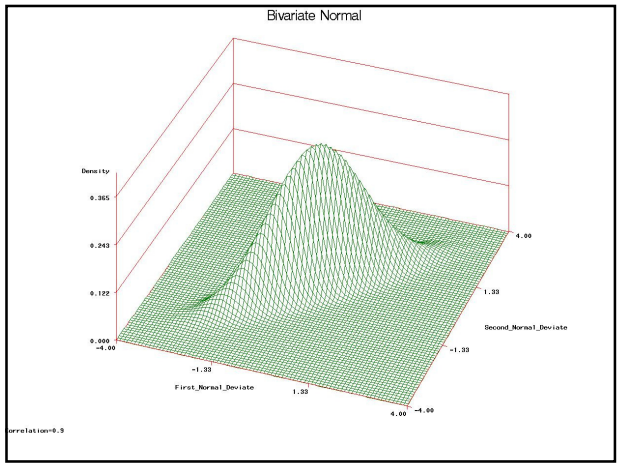


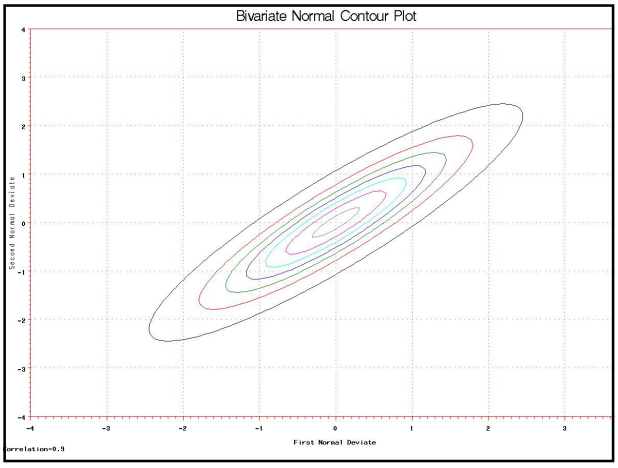


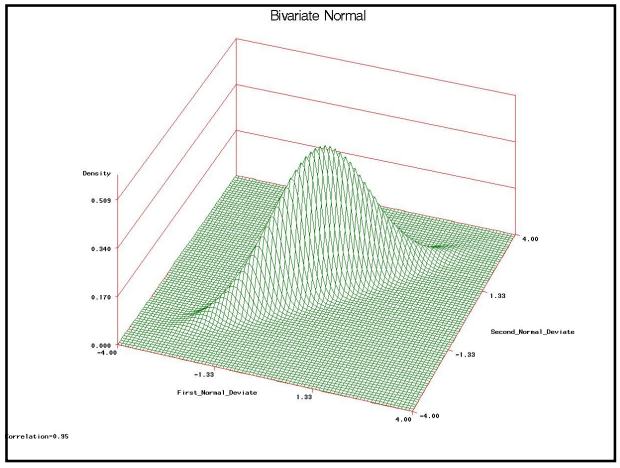

Bivariate Normal Contour Plot
Second Normal Deviate
First Normal Deviate
Correlation=0.5

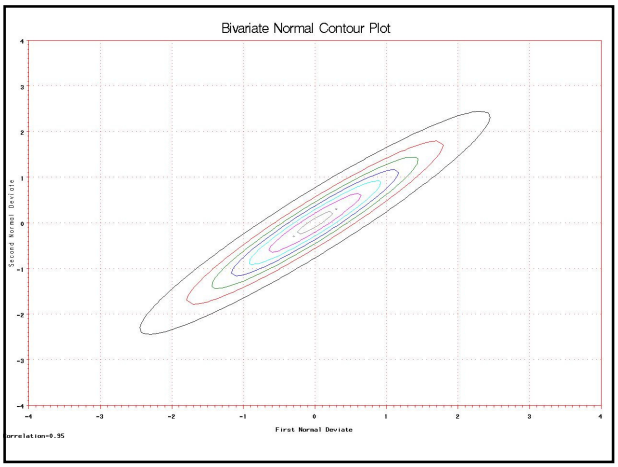

Bivariate Normal
Density
0.223
0.149
0.074
0.000
First_Normal_Deviate
Second_Normal_Deviate
Correlation=0.7

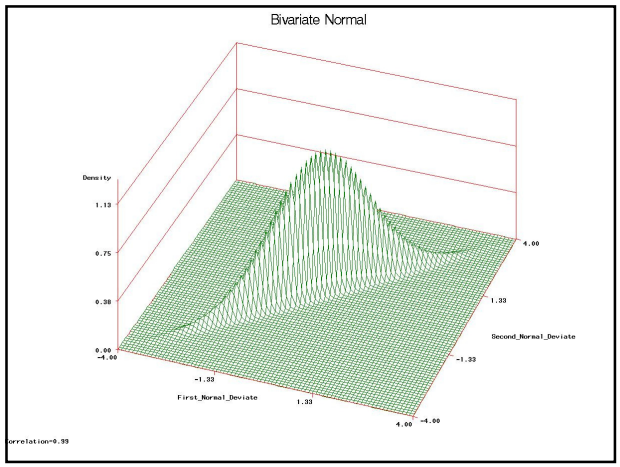

Bivariate Normal Contour Plot
Second Normal Deviate
First Normal Deviate
Correlation=0.7

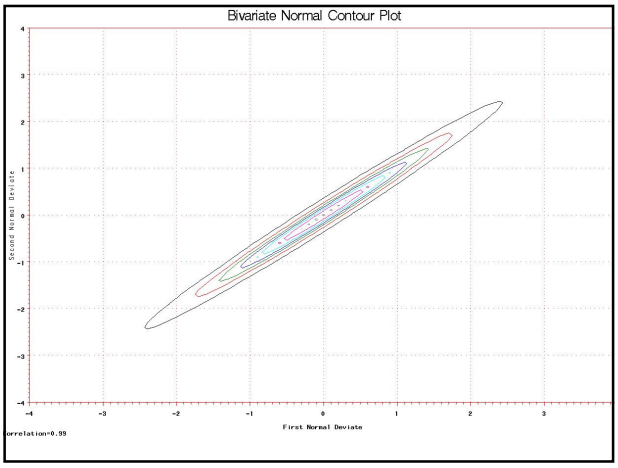

Bivariate Normal
Density
0.365
0.243
0.122
0.000
First_Normal_Deviate
Second_Normal_Deviate
Correlation=0.9

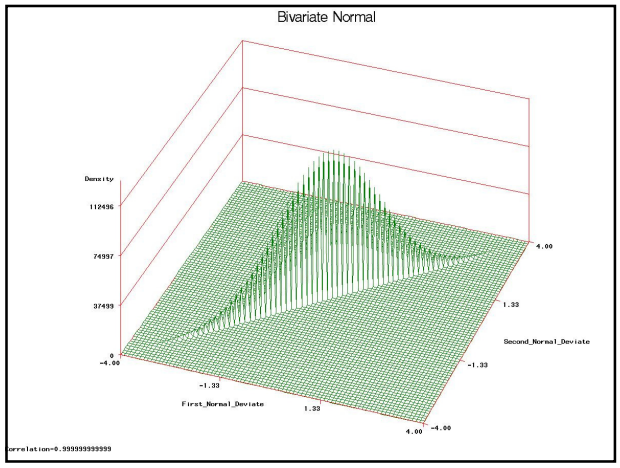

Bivariate Normal Contour Plot
Second Normal Deviate
First Normal Deviate
Correlation=0.9

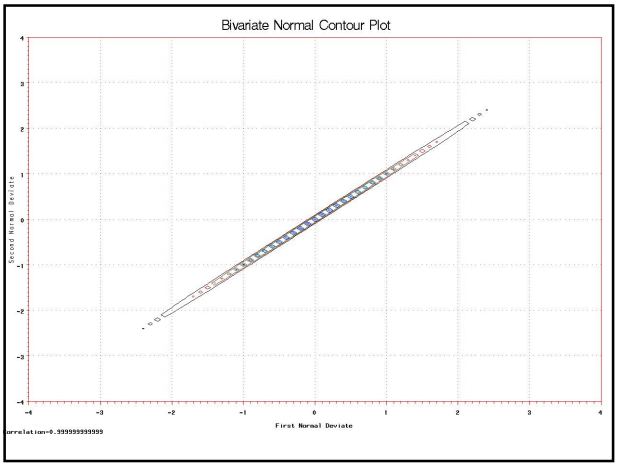

Bivariate Normal
Density
0.509
0.340
0.170
0.000
First_Normal_Deviate
Second_Normal_Deviate
Correlation=0.95











Sample Size Calculation

- The sample size can be estimated from

$$1-\beta = probbnrm\left(\frac{\sqrt{(\log(\mu_{T1}/\mu_{R1})-\log(1.25))^2 n}}{\sigma_{w1}} - t_{1-\alpha,n-2}, \frac{\sqrt{(\log(\mu_{T2}/\mu_{R2})-\log(1.25))^2 n}}{\sigma_{w2}} - t_{1-\alpha,n-2}, \rho\right) + probbnrm\left(\frac{\sqrt{(\log(\mu_{T1}/\mu_{R1})+\log(1.25))^2 n}}{\sigma_{w1}} - t_{1-\alpha,n-2}, \frac{\sqrt{(\log(\mu_{T2}/\mu_{R2})+\log(1.25))^2 n}}{\sigma_{w2}} - t_{1-\alpha,n-2}, \rho\right) - 1$$

- Where

μ_{T1} = AUC on test
μ_{R1} = AUC on reference
σ_{w1} = SD of logs for AUC
μ_{T1} = Cmax on test
μ_{R1} = Cmax on reference
σ_{w2} = SD of logs for Cmax

ρ is the correlation between AUC and Cmax

54

Sample Size Inflations for Different Correlations

CV	Ratio	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
0.30	0.85	1.000	1.112	1.149	1.176	1.196	1.211	1.223	1.233	1.243	1.248	1.253
	0.90	1.000	1.111	1.148	1.176	1.194	1.204	1.222	1.231	1.241	1.241	1.250
	0.95	1.000	1.096	1.135	1.154	1.173	1.192	1.212	1.212	1.231	1.231	1.231
	1.00	1.000	1.077	1.128	1.128	1.154	1.154	1.179	1.179	1.179	1.179	1.179
	1.05	1.000	1.120	1.160	1.180	1.200	1.200	1.220	1.240	1.240	1.240	1.240
	1.10	1.000	1.109	1.141	1.174	1.196	1.207	1.217	1.228	1.239	1.239	1.250
	1.15	1.000	1.107	1.150	1.173	1.192	1.210	1.220	1.234	1.238	1.248	1.252

How to Calculate the Sample Size for a Superiority Study with Two or More Endpoints?

Two Endpoints

Type equation here.Superiority Trial

- Assuming a bivariate Normal distribution the power for a given sample size can be estimated from

$$1-\beta = probbnrm\left(\frac{\sqrt{rn_{A_1}}d_1}{\sqrt{(r+1)\sigma_1^2}} - t_{1-\alpha/2,n_{A_1}(r+1)-2}, \frac{\sqrt{rn_{A_2}}d_2}{\sqrt{(r+1)\sigma_2^2}} - t_{1-\alpha/2,n_{A_2}(r+1)-2}, \rho\right)$$

- If we assumed same mean difference and standard deviation then this can be simplified to

$$1-\beta = probbnrm\left(\frac{\sqrt{rn_A}d}{\sqrt{(r+1)\sigma^2}} - t_{1-\alpha/2,n_A(r+1)-2}, \frac{\sqrt{rn_A}d}{\sqrt{(r+1)\sigma^2}} - t_{1-\alpha/2,n_A(r+1)-2}, \rho\right)$$

Two or more endpoints

Worked Example – A Superiority Study

The Study Design

- A study is being designed in a Osteoarthritis population to compare two treatments.
- There are three co-primary "and" endpoints
  - WOMAC – Pain
  - WOMAC – Function
  - Patient Global Assessment
- All three endpoints are Normally distributed with approximately similar variances and effect sizes of interest

Sample Size Calculations

- The Bonferroni option is not considered here
- Neither can we just ignore the fact we have multiplicity in Type II error
- We do have a solution for two endpoints

$$1-\beta = probbnrm\left(\frac{\sqrt{n}d_1}{\sqrt{2}\sigma_1}-t_{1-\alpha,n2-2},\frac{\sqrt{n}d_1}{\sqrt{2}\sigma_1}-t_{1-\alpha,2n-2},\rho\right)$$

- Could power on the two noisiest?

Sample Size Increases by Correlation for Two Endpoints

Stand Diff	Correlation										
	1.0	**0.9**	**0.8**	**0.7**	**0.6**	**0.5**	**0.4**	**0.3**	**0.2**	**0.1**	**0.0**
0.05	1.000	1.100	1.135	1.159	1.177	1.191	1.202	1.211	1.218	1.224	1.228

Extending the work of Sankoh

- For multiple "or" comparisons Sankoh gave a solution to adjust the significance level to maintain the nominal level
- This can be extended here so that we adjust the Type II error

$$\beta_t = 1-\left(\sqrt[k]{1-c\beta}\right)$$

- Where

$$k = t^{1-\rho}$$

- t is the number comparisons and ρ is the average correlation between endpoints.
- c comes from….

Table of c values

	Correlation									
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	1	**0.94**	0.91	**0.87**	0.83	**0.78**	0.75	**0.71**	0.79	**0.76**
3	1	**0.89**	0.84	**0.78**	0.71	**0.63**	0.62	**0.60**	0.64	**0.68**
4	1	0.86	0.78	0.70	0.63	0.6	0.56	0.55	0.59	0.69
5	1	**0.85**	0.77	**0.68**	0.63	**0.58**	0.52	**0.46**	0.49	**0.51**
6	1	**0.85**	0.75	**0.66**	0.61	**0.55**	0.50	**0.44**	0.47	**0.49**
7	1	0.85	0.74	0.65	0.60	0.55	0.49	0.44	0.47	0.49
8	1	0.85	0.73	0.64	0.59	0.54	0.48	0.43	0.46	0.48
9	1	0.85	0.71	0.62	0.57	0.52	0.47	0.42	0.44	0.46
10	1	**0.84**	0.65	**0.56**	0.51	**0.45**	0.41	**0.37**	0.39	**0.41**

General Result

- The inflation in sample size compared to two groups can be estimated from

$$\text{Inflation Factor} = \frac{\left(Z_{1-\alpha/2} + Z_{1-\beta^{1/m}}\right)^2}{\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}$$

Inflation Factors Based on General Result

No. of Cats	Correlation									
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	1.23	1.23	1.22	1.21	1.20	1.20	1.19	1.19	1.13	1.12
3	1.36	1.36	1.35	1.34	1.33	1.34	1.31	1.28	1.23	1.17
4	1.45	1.46	1.44	1.44	1.43	1.40	1.38	1.34	1.27	1.17
5	1.52	1.52	1.50	1.49	1.47	1.45	1.43	1.42	1.35	1.28
6	1.58	1.57	1.56	1.54	1.51	1.49	1.47	1.45	1.37	1.30
7	1.62	1.62	1.60	1.58	1.55	1.52	1.49	1.47	1.38	1.31
8	1.67	1.65	1.64	1.62	1.58	1.54	1.52	1.49	1.40	1.32
9	1.70	1.69	1.68	1.65	1.61	1.57	1.54	1.50	1.42	1.34
10	1.73	1.72	1.73	1.71	1.67	1.63	1.59	1.55	1.47	1.38

Inflation Factors from Simulation

No. of Cats	Correlation									
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	1.22	1.22	1.21	1.2	1.19	1.19	1.17	1.17	1.15	1.09
3	1.35	1.35	1.31	1.29	1.29	1.28	1.26	1.23	1.21	1.17
4	1.45	1.45	1.44	1.41	1.37	1.37	1.33	1.30	1.28	1.20
5	1.53	1.53	1.48	1.47	1.46	1.43	1.40	1.38	1.29	1.21
6	1.60	1.58	1.54	1.53	1.52	1.45	1.44	1.40	1.34	1.25
7	1.66	1.62	1.57	1.57	1.53	1.50	1.46	1.43	1.37	1.26
8	1.64	1.64	1.62	1.61	1.60	1.55	1.51	1.45	1.37	1.28
9	1.71	1.67	1.65	1.64	1.63	1.58	1.52	1.49	1.42	1.30
10	1.71	1.67	1.65	1.64	1.63	1.58	1.52	1.49	1.42	1.30

Summary

- Introduced the problem of must win sample size calculations
- Described solutions for the sample size calculation for two or more endpoints

References

Julious SA and Fernandes NE. Sample sizes for trials involving multiple correlated must-win comparisons. Pharmaceutical Statistics 11(2):177-85 (DOI: 10.1002/pst.515)

References

- Haider, S (2004). Bioequivalence of Highly Variable Drugs: Regulatory Perspectives. FDA Advisory Committee for Pharmaceutical Science. Available at URL: http://www.fda.gov/ohrms/dockets/ac/04/slides/4034s2.htm
- Julious SA. (2004) Tutorial in Biostatistics: Sample sizes for clinical trials with Normal data. Statistics in Medicine 23:1921-86.
- Sankoh AJ, Huque AJ and Dubey SD (1997). Some comments on frequently used multiple endpoint adjustment methods in clinical trials. Statistics in Medicine16, 2529-2542