Sample size estimation in Phase III clinical trials

K. K. Gordon Lan (藍光國)
Janssen Pharmaceutical Companies of Johnson & Johnson

Joint2011 Symposium
Nangang, Taiwan
December 16, 2011

Collaborator: Dr. Janet Wittes

Traditionally, we consider treatment effect as fixed and estimate sample size of a trial to reach desired power.

In this presentation, assuming Phase II data are available, we try different ways to evaluate sample size for a Phase III trial under various random treatment effects.

"Sample size and probability of a successful trial"
by Chuang-Stein C (Pharmaceutical Statistics 2006)

This paper describes the distinction between the concept of statistical power and the probability of getting a successful trial. .... discusses a framework to calculate the 'average success probability' and demonstrates how uncertainty about the treatment effect could affect the average success probability for a confirmatory trial. .....Computer codes (R and SAS) to calculate the average success probability are included.

Outline:

1. Review the use of conditional power/average power in a Phase III trial.
2. Extend the same idea to Phase III trial sample size evaluation under a normal prior distribution based upon the Phase II data.
3. Consider other prior distributions.

A brief review of CP in a Phase III trial

Consider the distribution theory for a one-sample problem.

The mathematics behind a one-sample problem is very straightforward and easy to understand. Extension of the idea (not the mathematics) to the two-sample case needs only slight modifications.

Compare a new treatment T with a control treatment C.

Suppose $Y_T \sim N(\mu_T, \sigma^2)$ and $Y_C \sim N(\mu_C, \sigma^2)$, then
$X = (Y_T - Y_C)/\sigma\sqrt{2} \sim N(\Delta,1)$, where $\Delta = (\mu_T - \mu_T)/ \sigma\sqrt{2}$.

In other words, if we pair responses Y_T and Y_C, and "standardized" the difference by $X = (Y_T - Y_C)/\sigma\sqrt{2}$, then the 2-sample problem becomes an 1-sample problem.

X has mean Δ and variance 1. A positive response Δ favors the new treatment. To simplify our discussion, we assume the X's are normally distributed. The theory applies to responses different from normal if the sample size is "LARGE".

"Trend of the data" – The partial sum process

Let $X_1, X_2, \ldots, X_n, \ldots$ be iid $N(\Delta,1)$. Define $S_n = X_1 + X_2 + \ldots + X_n$.

Then $ES_n = n\Delta$ and $Var(S_n) = n$.

The expectation is a linear function of the variance.

Good news: This linear relationship gives us an easy tool to "predict" the future outcome conditional on accumulating data.

Bad news: The prediction depends on the treatment effect Δ which is unknown to us.

In clinical trials, we do not report the partial sums.

The interim test statistic is $Z(n) = S_n/\sqrt{n}$.

Example: To design a clinical trial, we test the hypothesis

$$H_0: \Delta = 0 \text{ versus } H_a: \Delta > 0.$$

If we take an one-sided $\alpha = 1.96$ and 85% power ($\beta = 0.15$), how many patients do we need to reach a 85% power?

$$Z(N) = S_N/\sqrt{N}.$$
$$EZ(N) = N\Delta/\sqrt{N} = \sqrt{N}\Delta.$$

Let us assume that the treatment effect $\Delta = \Delta_1 = 0.2$.
Solve for N from the equation:

$$EZ(N) = \sqrt{N}\Delta_1 = z_\alpha + z_\beta = 1.96+1.04 = 3, \; N = 225.$$

For a given $\Delta = \Delta_1$, the drift parameter is $\theta = E(Z) = \sqrt{N}\Delta_1$.

To evaluate sample size N, solve N from

$\theta = E(Z) = \sqrt{N}\Delta_1 = z_\alpha + z_\beta = 1.96 + 1.04 = 3.0$.(85% power)

$\Delta_1 =$	0.5	0.2	0.1	0.05	0.01	$\to 0$
N =	36	225	900	3600	90000	$\to \infty$

A fundamental equation for sample size evaluation <u>for a fixed design</u>:

$$\theta = EZ = z_\alpha + z_\beta.$$

(Change z_β to 0.84 for 80% and 1.28 for 90%.)

For a sequential design, the drift parameter θ required for power $1-\beta$ will be slightly larger than $z_\alpha+z_\beta$. To find the value θ in the sequential setting, use software (free) created by Professor DeMets of University of Wisconsin – Madison.

http://www.biostat.wisc.edu/People/faculty/demets.htm

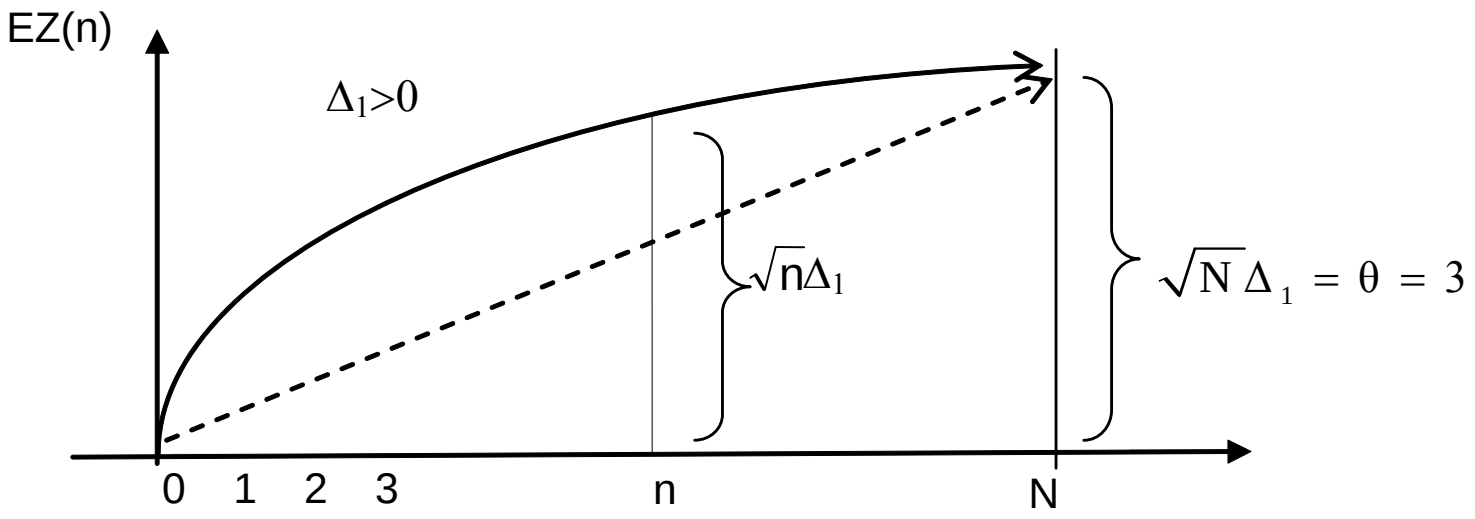
The trend of the data = Δ (Partial sums)

Interim analysis

	$X_1, X_2, \ldots, X_n,$	$X_{n+1}, \ldots X_N$
Unconditional	random	random
Conditional	fixed	random

	S_N	=	S_n	+	$(S_N - S_n)$
Unconditional	ES_N	=	$n\Delta$	+	$(N-n)\Delta$
Variance	$Var(S_N)$	=	n	+	$(N-n)$
Conditional	$E_C(S_N)$	=	S_n	+	$(N-n)\Delta$ (Δ=?)
Variance	$Var_C(S_N)$	=	$N-n$		

The trend of the data = θ (B-values)



$(n, Z(n)) \rightarrow (\tau\ ,\ Z_\tau) \rightarrow (\tau\ ,\ B_\tau)$ where $\tau = n/N$ & $B_\tau = Z_\tau \sqrt{\tau}$.

Example:

$\Delta=0.2$, $\theta=\sqrt{N}\Delta=3 \Rightarrow N=225.$

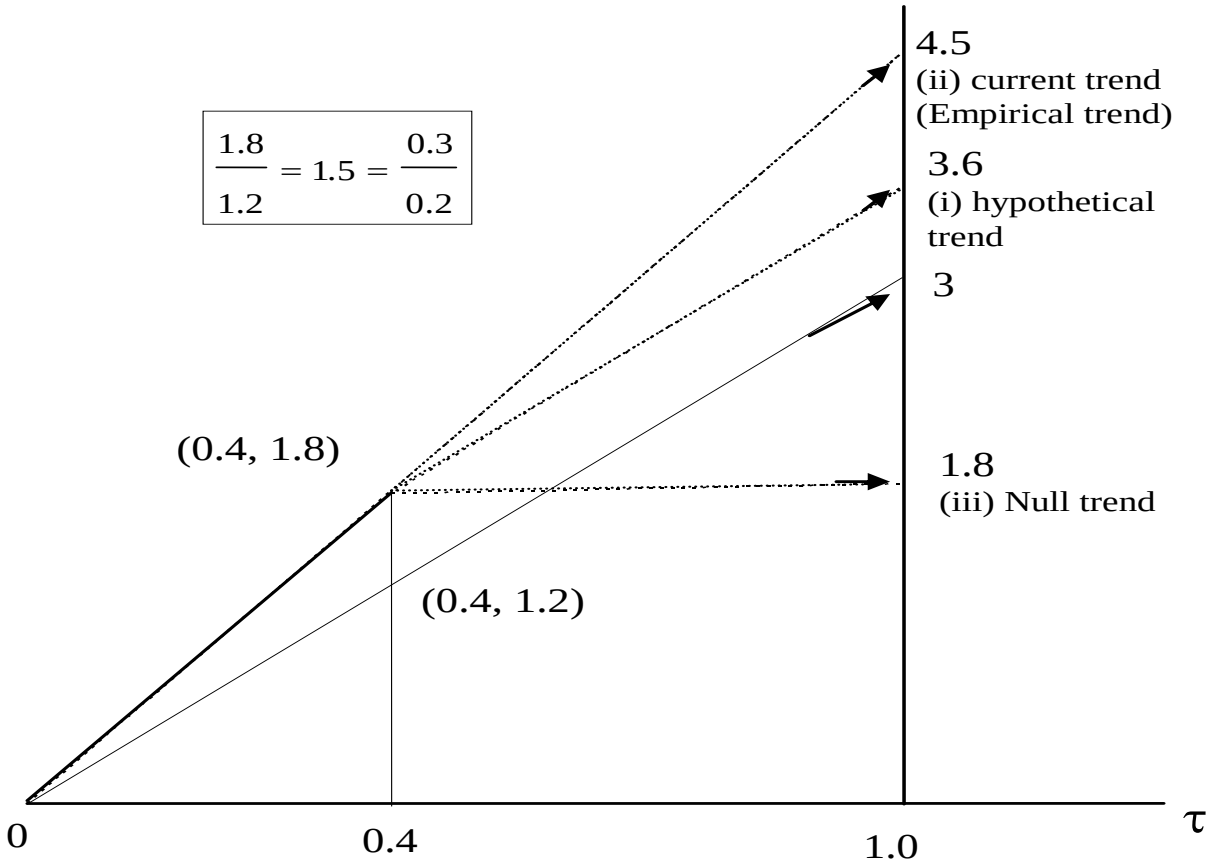
when $n=90$, $Z_{.4}=2.846$

$CP(\theta) = P(Z_1 \geq 1.96 \mid Z_{.4}=2.846, \theta) = ?$

$\tau = 90/225 = 0.4$; $B_\tau = 2.846\sqrt{\tau}=1.8.$

Note that $Z_1 = B_1$. To evaluate $CP(\theta)$:

1. Find $E_C(Z_1)$. $Var_C(Z_1) = 1-\tau$.
2. Find $P_C(Z_1 \geq 1.96)$.


$\frac{1.8}{1.2} = 1.5 = \frac{0.3}{0.2}$
4.5
(ii) current trend
(Empirical trend)
3.6
(i) hypothetical
trend
3
(0.4, 1.8)
1.8
(iii) Null trend
(0.4, 1.2)
0
0.4
1.0
τ

(i) $P[Z_1 = B_1 \geq 1.96 \big| B_{.4} = 1.8, \quad \theta = 3]$

$= P[\frac{Z_1 - 3.6}{\sqrt{.6}} \geq \frac{1.96 - 3.6}{\sqrt{.6}} \big| B_{.4} = 1.8, \ \theta = 3]$

$= P[N(0,1) \geq -2.12]$

$= 0.9830$

(ii) $P[Z_1 \geq 1.96 \big| B_{.4} = 1.8, \quad \theta = 4.5]$

$= P[\frac{B_1 - 4.5}{\sqrt{.6}} \geq \frac{1.96 - 4.5}{\sqrt{.6}} \big| B_{.4} = 1.8, \ \theta = 4.5]$

$= P[N(0,1) \geq -3.28]$

$= .9995$

(iii) $P[B_1 \geq 1.96 \big| B_{.4} = 1.8, \quad \theta = 0]$

$= P[N(0,1) \geq \frac{1.96 - 1.8}{\sqrt{.6}}]$

$= P[N(0,1) \geq 0.21]$

$= 0.4168$

$$\mathrm{CP}(\tau, \mathrm{B}_{\tau}, \theta_{\mathrm{E}}) = \Phi[\frac{\theta_{\mathrm{E}} - 1.96}{\sqrt{1-\tau}}\ ].$$

From conditional power (CP) to predictive power (PP)

τ=0.4, Z_τ=1.6.

-2.0 SD	0.0433
-1.5 SD	0.1353
-1.0 SD	0.3124
-0.5 SD	0.5491
Empirical	0.7690
+0.5 SD	0.9112
+1.0 SD	0.9750
+1.5 SD	0.9950
+2.0 SD	0.9993

Weighted average of CP (τ=0.4, Z_τ=1.6)

-2.0 SD	0.0433	0
-1.5 SD	0.1353	0
-1.0 SD	0.3124	0.1
-0.5 SD	0.5491	0.2
Empirical	0.7690	0.4
+0.5 SD	0.9112	0.2
+1.0 SD	0.9750	0.1
+1.5 SD	0.9950	0
+2.0 SD	0.9993	0

0.1x0.3124+0.2x0.5491+0.4x0.7690+0.2x0.9112
+0.1x0.9750 = 0.7284

Predictive power (considers θ as random)

Note that $\theta_E = B_\tau/\tau$ is a point estimate of θ. If we consider θ as random with distribution function G

$$PP = PP[\tau, B_\tau, G(\theta)] = \int CP(\tau, B_\tau, \theta)\, dG(\theta)$$
$$= \int CP(\tau, B_\tau, \theta)\, g(\theta)\, d\theta$$

(Note that we did not introduce a prior distribution and went directly to the posterior distribution of θ.)

A reasonable choice of G (for a fixed n)

Since $\overline{X}_n$ is $N(\Delta, 1/n)$,

let us consider Δ to be $N(\overline{X}_n, 1/n)$.

This is equivalent to $\theta = \Delta\sqrt{N}$ is $N(\theta_E, 1/\tau)$.

Conceptually, this is similar to calling

$[\overline{X}_n \mp 1.96\sqrt{1/n}]$ a 95% c.i. for μ.

If G* is taken to be $N(\theta_E, 1/\tau)$, then

$$PP(\tau, B_\tau, G^*) = \Phi[\frac{(\theta_E - 1.96)\sqrt{\tau}}{\sqrt{1-\tau}}].$$

Compare this expression with

$$CP(\tau, B_\tau, \theta_E) = \Phi[\frac{\theta_E - 1.96}{\sqrt{1-\tau}}].$$

Reference:

Lan KKG, Hu P, Proschan MA (2009) "A conditional power approach to the evaluation of predictive power." *Statistics in Biopharmaceutical Research*; 1: 131-136.

Two-sample comparisons, Comparison of two means

$$H_o:\mu_x = \mu_y \quad \text{vs} \quad H_a:\mu_x > \mu_y$$

$$X_1, X_2, \ldots, X_M \quad \text{iid} \quad N(\mu_X, \sigma^2)$$

$$Y_1, Y_2, \ldots, Y_M \quad \text{iid} \quad N(\mu_Y, \sigma^2)$$

$$N = M + M = 2M$$

$$Z_{(N)} = \frac{\bar{X}_M - \bar{Y}_M}{\sigma\sqrt{1/M + 1/M}} = \frac{\sum_1^M X_i - \sum_1^M Y_i}{\sigma\sqrt{M+M}} = \frac{\sum_1^M (X_i - Y_i)}{\sigma\sqrt{N}}$$

$$\theta = EZ_{(N)} = \frac{\mu_x - \mu_y}{\sigma}\sqrt{\frac{N}{4}} = \frac{\mu_x - \mu_y}{\sigma}\sqrt{N}\sqrt{\frac{1}{2}\times\frac{1}{2}}$$

treatment sample two-sample

$$EZ = \Delta\sqrt{SS/4} = \Delta\sqrt{SS}\sqrt{\frac{1}{2}x\frac{1}{2}} = \Delta\ \sqrt{SS}\ \sqrt{\text{two-sample factor}}.$$

For a given $\Delta=\Delta_1$, From the equation:

$EZ = \Delta_1\ \sqrt{(SS/4)} = z_\alpha + z_\beta = 1.96 + z_\beta$.

We may solve for sample size SS (for desired power 1 - β) OR,

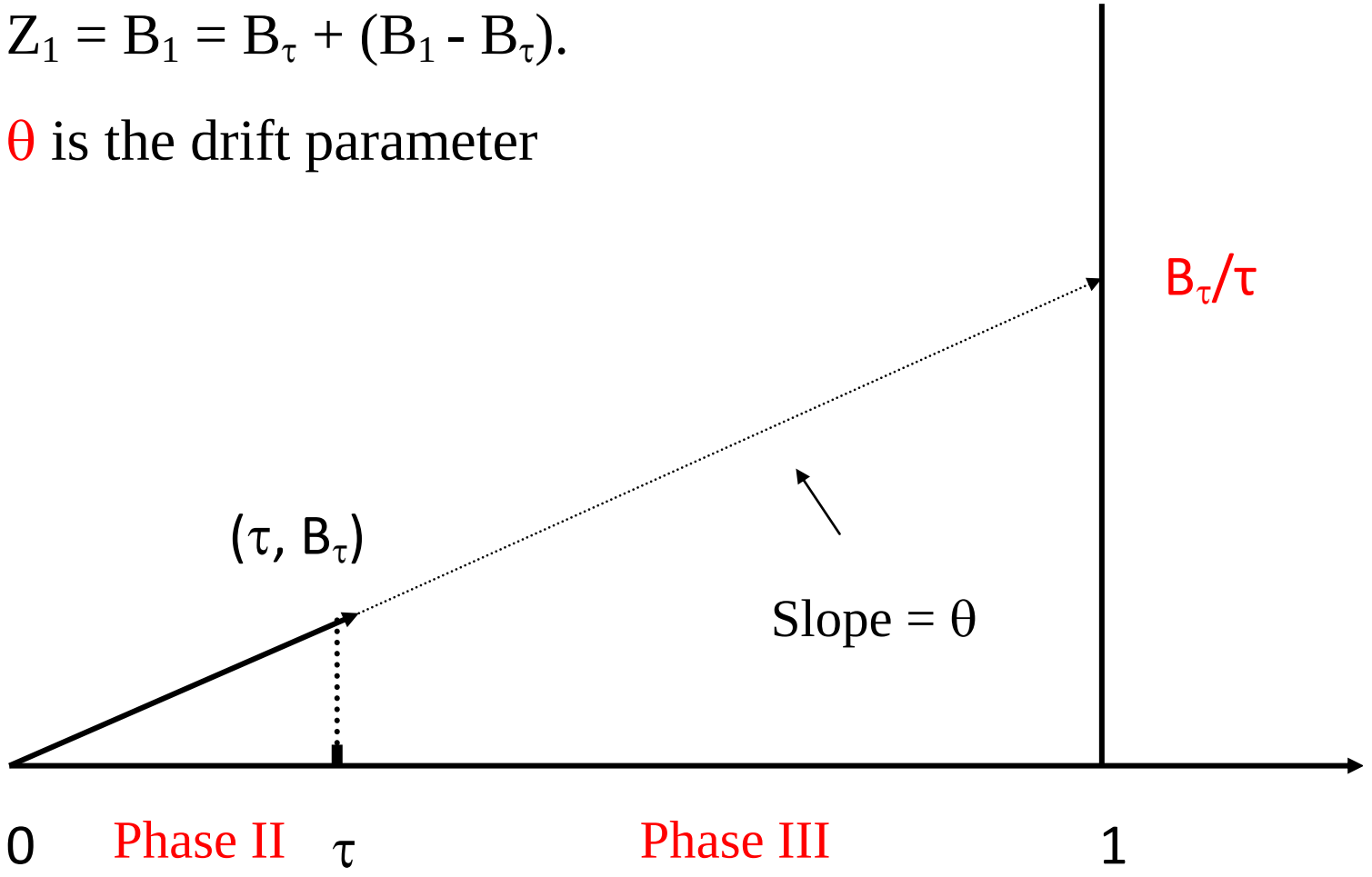
Solve for power 1 - β for given sample size SS.

I will use notations SS = M for a Phase II study and SS = N for Phase III.

Use average power to design a Phase III trial

$Z_1 = B_1 = B_\tau + (B_1 - B_\tau)$.

θ is the drift parameter



To design a Phase II trial, we may not have a good estimate of Δ. Sample size m may depend on the budget, or, we might just pick a "reasonable" value of m.

At the end of Phase II, we observed

$$Z_{II} = \frac{\Delta_{II}}{\sqrt{4/M}} = \Delta_{II}\sqrt{M/4}\ . \quad \{\Delta_{II} = (\overline{X_T} - \overline{X_C})/\sigma\} \qquad \text{(Eq 2)}$$

(1.1) If $Z_{II} < 0$, stop the program.

(1.2) If $Z_{II} > 0$ and statistically significant, ?????????

(1.3) If $Z_{II} > 0$ and Δ_{II} looks "promising", use Δ_{II} to design Phase III.

Note that Z_{II} may not be statistically significant.

We consider case (1.3) only for determination of N in Phase III.

Two different approaches (fixed and random):

A fixed treatment effect approach

Consider $\Delta = \Delta_{II}$ (observed in Phase II) as fixed.

Let α = one-sided 0.025 and power = $1-\beta$.

Solve for sample size N from

$EZ(N) = \Delta_{II}\sqrt{(N/4)} = 1.96 + z_{\beta}$.

The fixed approach has been used extensively in Phase III sample size evaluation. Many people feel that it contributes to the fact that many Phase III studies were under-sized.

A random treatment effect approach

As an alternative to the fixed approach, one may try to consider Δ as random then consider the "average power". (Bayesian???)

Example of average power (Δ_{II}=0.3, N=400):

Δ = x =	0	.1	.2	.3	.4	.5	.6
Weight=prob.	.05	.1	.2	.3	.2	.1	.05
Fixed power	2.5%	16.8%	52.6%	85%	97.9%	99.9%	100%

Average power = 72.2% < 85%. (symmetric weights, prior belief)

In general, how do we assign distribution for Δ?

If Δ is continuous with density g, average power = $\int$ power(x)g(x)dx.

Prior distribution of Δ in Phase III

$$Z_{II} = \frac{\Delta_{II}}{\sqrt{4/M}} = \Delta_{II}\sqrt{M/4}\ . \quad \{\Delta_{II} = (\overline{X_T} - \overline{X_C})/\sigma\} \qquad \text{(Eq 2)}$$

It is common that we consider $[\Delta_{II} - 1.96\ \sqrt{(4/M)},\ \Delta_{II} + 1.96\ \sqrt{(4/M)}]$ as a 95% confidence interval for Δ.
This is similar to assuming $\Delta \sim N(\Delta_{II}, 4/M)$.

Under normality, we use $N(\Delta_{II}, 4/M)$ to evaluate "average power":

In the following slides,

"fixed power" = "power" = power of a Phase III study derived from the fixed approach;

"average power" = power from a random approach.

Default distribution: $\Delta \sim N(\Delta_{II}, 4/m)$.

Let M = sample size of a Phase II study, Z_{II} = the observed Z-value, Δ_{II} is observed treatment effect.

For the Phase III trial with one-sided $\alpha = .025$, the desired fixed power is $1-\beta = \Phi(z_\beta)$.

N = sample size is chosen to reach fixed power $1-\beta$. N satisfies $z_\beta = \Delta_{II}\sqrt{(N/4)} - 1.96$.

An interesting result:

If $\tau = M/(M+N)$, then average power = $\Phi(z_\beta\sqrt{\tau})$.

A simple mathematical expression with serious problems in application!

Example: In Phase II, M=100, $\Delta_{II} = .125$ and $Z_{II} = 0.625$.
To reach 85% fixed power, N=2304 and $\tau = 100/2404 \approx 0.04$.
(power = 85% → $\beta = 0.15$ and $z_\beta = 1.04$.)
Average power = $\Phi(z_\beta\sqrt{\tau}) = \Phi(1.04\sqrt{.04}) = \Phi(.208) \approx 58\%$.

If sample size N is increased to 5000, then
Fixed power ≈ 99.3% ,$\tau \approx .02$ and average power ≈ 64%.

If sample size N is increased to 20000, then
Fixed power = $\Phi(6.88) \approx 1$, $\tau \approx .005$ and average power ≈ 69%.
Average power will not approach 1 as $N \to \infty$.

Some interesting results:

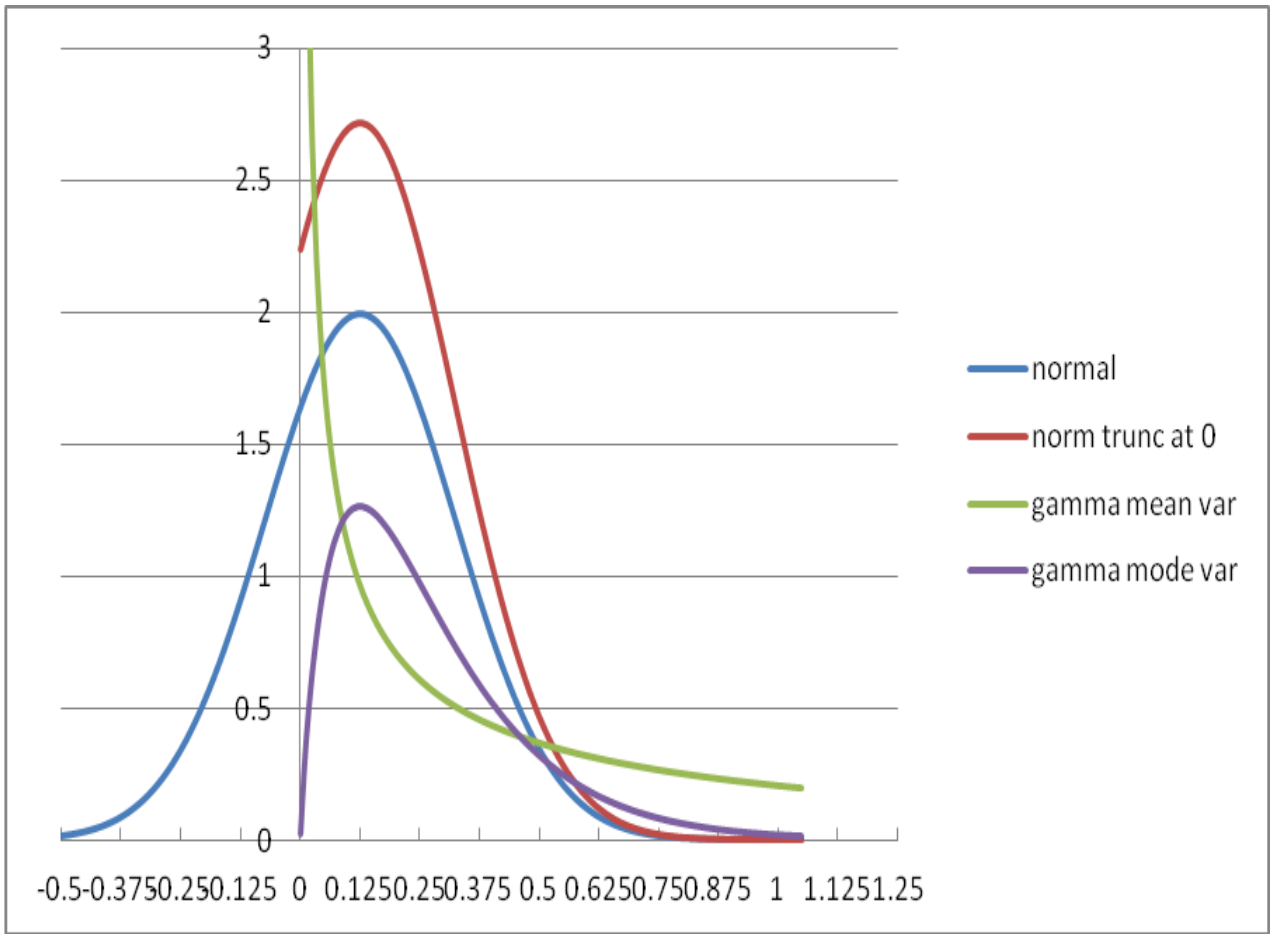
Pick any sample size N. When $\Delta = \Delta_{II}$ is considered fixed,
$\beta = 1 - \Phi[\Delta_{II}\sqrt{(N/4)} - 1.96]$ and $\tau = M/(M+N)$ are known.
It can be shown that average power $= \Phi(z_\beta\sqrt{\tau}) \leq \Phi(z_{II})$.

In the example above, $Z_{II} = .125\sqrt{(100/4)} = .625$.
Average power $\leq \Phi(.625) \approx$ 73%.
When $N \to \infty$, average power $\to \Phi(z_{II})$.

Note that in our example, $\Delta \sim N(.125, .040)$ & $P[\Delta \leq 0] = 27\%$.
When $\Delta = x = 0$, power = 2.5%. Power for $x<0$ is ≈ 0.

Prior distribution for Δ Δ_{Ph2}=0.25; $\nu = 0.04 = 4/100$, N=576)

	Average power
Fixed μ=0.25	0.85
N(μ, ν)	.655
Gamma with mean μ and variance ν	.603
Gamma with mode μ and variance ν	.828
N(μ, ν) truncated to (0, ∞)	.732


3
2.5
2
1.5
1
0.5
0
-0.5-0.3750.250.125 0 0.1250.250.375 0.5 0.6250.750.875 1 1.1251.25
normal
norm trunc at 0
gamma mean var
gamma mode var

Sample size required under three different approaches to sample size calculation ($\Delta_{\text{Phase 2}}$=0.22, the sample size M in Phase 2=96, $Z_{\text{Phase 2}}$=1.08 and one-sided α=0.05)

Power	Sample size		
	Fixed effect	Average power calculated from a truncated normal prior (TAP)	[redacted]
0.8	512	826	
0.85	596	1332	
0.90	710	2640	
0.95	896	8890	

A new approach (compromise)

Find N so that traditional (fixed effect) power = 1-β.

Power (fixed) = $\Phi(z_\beta)=\Phi(z_\beta\sqrt{1})$

Average power (normal prior) = $\Phi(z_\beta\sqrt{w})$, w=M/(M+N)

Find a prior so that
Average power (new prior) = $\Phi(z_\beta\sqrt{(w+1)/2})$

Find a new N* so that average power (new prior) = $\Phi(z_\beta\sqrt{(w+1)/2})$.

Sample size required under three different approaches to sample size calculation ($\Delta_{\text{Phase 2}}$=0.22, the sample size M in Phase 2=96, $Z_{\text{Phase 2}}$=1.08 and one-sided α=0.05)

Power	Sample size		
	Fixed effect	Average power calculated from a truncated normal prior (TAP)	Average power calculated from a new prior
0.8	512	826	662
0.85	596	1332	800
0.90	710	2640	990
0.95	896	8890	1304

Final Comments

Solicit prior distribution from clinicians if possible. Use their choice as prior to pick sample size N.

We do not start a Phase III trial unless Δ_{II} is "promising". Therefore, there is a hidden bias (over-estimate) of Δ in Phase III studies.

References:

1. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Wiley: Chichester, 2004; pp. 193-194.
2. Chuang-Stein C. Sample size and the probability of a successful trial. *Pharm Stat* 2006; 5:305-309.3.
3. Lan K, Hu P, Proschan MA. Conditional power approach to the evaluation of predictive power. *Statistics in Biopharmaceutical Research* 2009; 1:131-136.

THANK YOU!