Calcolo della dimensione campionaria nei clinical trials: teoria e applicazioni

Luigi Mariani
Unità di Statistica Medica, Biometria e Bioinformatica
Fondazione IRCCS Istituto Nazionale Tumori
luigi.mariani@istitutotumori.mi.it

1

Outline

Section 1
- Basic considerations
- Normally distributed data
- Binary data
- Survival data
- Equivalence/non-inferiority trials

Section 2
- Design/practical issues
- Phase II trials

Acknowledgement
John Whitehead, MPS Research Unit, Lancaster University, UK
http://www.maths.lancs.ac.uk/department/research/statistics/mps

2

SS determination
Basic considerations

Determination of the appropriate sample size is a crucial part of study design.

A study which is too small may produce inconclusive results, while a study which is too large will waste resources.

- **Precision analysis for SS computation**
  Desired precision at a fixed confidence level
- **Power analysis for SS computation**
  Achieve a desired power for detecting a scientifically meaningful difference (hypothesis test)

3

Steps for SS determination

- **Study formulation**
- **Specify parameters for planned analysis**
- **Compute sample size**
- **Sensitivity analysis**
- **Choose sample size, write statement**
- **Sample size review**

4

Information needed to perform the calculation

▪ **The magnitude of the treatment difference associated with a small chance of winning**	θ_0
▪ **The small chance of winning (e.g. 2.5%)** - significance level (Type I error)	α
▪ **The magnitude of the treatment difference associated with a high chance of winning**	θ_R
▪ **The high chance of winning (e.g. 90%)** - power	$1-\beta$
▪ **Estimates (or guesses) of parameters included in the calculation as "nuisance parameters"** e.g. variance for normally distributed data, success rates for binary data	

5

The SS calculation depends on

- **The type of treatment comparison**
  (e.g. superiority, non-inferiority, equivalence)
- **The choice of the primary response variable**
  (e.g. normally distributed, binary, survival time)
- **The measure of treatment difference**
  (e.g. difference in mean, odds ratio, hazard ratio)

6

Type of treatment comparison

- **Superiority (Inferiority)**
  Objective of showing that the response to the experimental treatment (E) is superior (inferior) to the control (C; active, placebo or no treatment)
  Procedure: statistical test

- **Equivalence**
  Objective of showing that the response to two treatments differs by an amount which is clinically unimportant
  Procedure: Construct a confidence interval (θ_L, θ_U) for the difference; if $(\theta_L, \theta_U) \subseteq (-\varepsilon, \varepsilon)$ [*equivalence margins*] equivalence is demonstrated

- **Non-inferiority**
  Objective of showing that E is not clinically inferior to C
  Procedure: Construct a confidence interval (θ_L, θ_U) for the difference; if $(\theta_L > -\varepsilon$ [*non-inferiority margin*] non-inferiority is demonstrated

7

Hypothesis tests for a superiority (inferiority) trial

- Reject $H_0^+(0)$ to demonstrate superiority

  $\mathbf{H_0^+(0) : \theta = 0}$

  vs

  $\mathbf{H_1^+(0) : \theta > 0}$

- Reject $H_0^-(0)$ to demonstrate inferiority

  $\mathbf{H_0^-(0) : \theta = 0}$

  vs

  $\mathbf{H_1^-(0) : \theta < 0}$

- Reject $H_0(0)$ to demonstrate superiority or inferiority (two-sided test)

  $\mathbf{H_0(0) : \theta = 0}$

  vs

  $\mathbf{H_1(0) : \theta \neq 0}$

8

Type I and Type II errors

Two kinds of error occur when testing hypotheses

Type I error The null hypothesis is rejected when it is true

Type II error The null hypothesis is not rejected when it is false

α = P{type I error}

β = P{type II error}

$1-\beta$ = power

9

Choice of the primary response variable

CV – CAD	**Mortality**
CV – HT	**Blood pressure, CV mortality & morbidity**
CV – CHF	**Exercise tolerance, n. of hospitalizations, mortality and morbidity**
Oncology	**Objective response, response duration, time to tumor recurrence, mortality**
Rheumatic D	**Pain-function endpoints**
Alzheimer's D	**Cognitive and functional questionnaires**

CV, cardiovascular; CAD, coronary artery disease; HT, hypertension; CHF, chronic heart failure; D, disease.

10

Normally distributed data

e.g. blood pressure, blood analytes, questionnaire scores

Parallel group, two treatments, equal allocation, known common variance

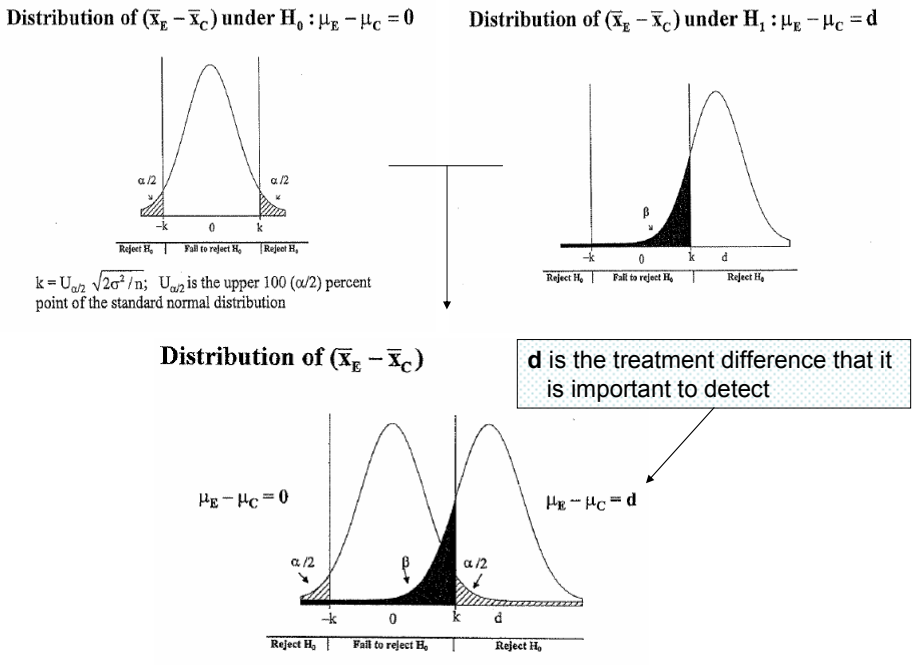
$(\bar{x}_E - \bar{x}_C)$ is an estimate of $(\mu_E - \mu_C)$

$(\bar{x}_E - \bar{x}_C) \sim N((\mu_E - \mu_C), 2\sigma^2 / n)$

where $\bar{x}_E$ and $\bar{x}_E$ are the sample means for E and C and σ^2 is the variance between patients on the same treatment

$(\bar{x}_E - \bar{x}_C)$ is the test statistic

11



12

$a)\ k = U_{\alpha/2}\sqrt{2\sigma^2/n}$

$b)\ P(\text{reject } H_0 \text{ with } E > C \mid (\mu_E - \mu_C) = d) = 1-\beta$

$P((\bar{x}_E - \bar{x}_C) > k \mid (\mu_E - \mu_C) = d) = 1-\beta$

$$P\left(\frac{(\bar{x}_E - \bar{x}_C) - d}{\sqrt{2\sigma^2/n}} > \frac{k-d}{\sqrt{2\sigma^2/n}}\right) = 1-\beta$$

$(k-d)/\sqrt{2\sigma^2/n} = -U_\beta$

$U_{\alpha/2} - d/\sqrt{2\sigma^2/n} = -U_\beta$

$$n = 2\sigma^2\left(\frac{U_{\alpha/2} + U_\beta}{d}\right)^2$$

$U_{\alpha/2}$ is the upper 100(α/2) percent point of the standard normal distribution
(replace $U_{\alpha/2}$ by U_α for a one-sided test)

U_β is the upper 100(β) percent point of the standard normal distribution

13

Calculation based on the standardized difference

σ^2 is usually unknown:

a) consider the standardized difference $d^*=d/\sigma$

$$n = 2\left(\frac{U_{\alpha/2} + U_\beta}{d^*}\right)^2$$

14

Calculation based on the t-test

σ^2 is usually unknown:

b) it is estimated at the time of analysis and the t test is performed

$$T = (\bar{x}_E - \bar{x}_C) / \sqrt{2s^2 / n}$$

where s^2 is the pooled sample variance

T has a non central t distribution with (2n-2) df and non-centrality parameter

$$(\mu_E - \mu_C) / \sqrt{2\sigma^2 / n}$$

- Iterative procedure using standardized differences (*Guenther 1975, 1977*)
- Approximate formula (*Guenther, 1981*)

$$n = 2\sigma^2 \left(\frac{U_{\alpha/2} + U_\beta}{d} \right)^2 + 0.25(U_{\alpha/2})^2$$

15

General approach to normally distributed data
Independence of w on θ

θ is the measure of the advantage of E over C

$\hat{\theta}$ is an estimate of θ such that asymptotically

$$\hat{\theta} \sim N((\theta, 1/w)$$

θ_R is the treatment difference that it is important to detect

$$w = \left(\frac{U_{\alpha/2} + U_\beta}{\theta_R} \right)^2$$

$$\theta = \mu_E - \mu_C \ , \quad \hat{\theta} = \bar{x}_E - \bar{x}_C \ , \quad \theta_R = d \ , \quad w = n / 2\sigma^2$$

$$\downarrow$$

$$n = 2\sigma^2 \left(\frac{U_{\alpha/2} + U_\beta}{d} \right)^2$$

16

Unequal allocation

n_E patients on E

n_C patients on C $\quad (n_E : n_C = r:1)$

$$\theta = \mu_E - \mu_C \;,\; \hat{\theta} = \bar{x}_E - \bar{x}_C \;,\; \theta_R = d \longrightarrow \hat{\theta} \sim N((\theta, \frac{\sigma^2}{n_E} + \frac{\sigma^2}{n_C}) \longrightarrow w = \frac{n_E n_C}{\sigma^2 (n_E + n_C)}$$

This gives

$$n_E = (r+1)\sigma^2 \left(\frac{U_{\alpha/2} + U_\beta}{d} \right)^2$$

$$n_C = \left(\frac{r+1}{r} \right) \sigma^2 \left(\frac{U_{\alpha/2} + U_\beta}{d} \right)^2$$

$$N = n_E + n_C = \frac{(r+1)^2}{r} \sigma^2 \left(\frac{U_{\alpha/2} + U_\beta}{d} \right)^2$$

17

Parameter	Value	Loss
n. groups	1	--
	2	x4
$U_{\alpha/c}$	1.6449	--
	1.9600	x1.27 (80% power) x1.23 (90% power)
U_β (*)	0.8416	--
	1.2816	x1.34 (5% α, two sided)
d*	1	--
	0.5	x4
	0.25	x16

(*) x1.70 combining two sided α=0.05 and 1-β=0.90

18

Binary data (1)

e.g. cured/diseased, alive/dead, success/failure …

Parallel group, two treatments	Treatment E	Treatment C	Total
Model:			
Probability of success	p_E	p_C	
Data:			
Successes	S_E	S_C	S
Failures	F_E	F_C	F
Total	n_E	n_C	t

If $\theta = (p_E - p_C)$ (**probability difference**)

then $\hat{\theta} = (\hat{p}_E - \hat{p}_C) = \frac{S_E}{n_E} + \frac{S_C}{n_C} \sim N\left(\theta, \frac{p_E(1-p_E)}{n_E} + \frac{p_C(1-p_C)}{n_C}\right)$

19

Binary data (2)

If $\theta = \log_e\left\{\frac{p_E(1-p_C)}{p_C(1-p_E)}\right\}$ (**log-odd ratio**)

then $\hat{\theta} \sim N\left(\theta, \frac{1}{n_E p_E(1-p_E)} + \frac{1}{n_C p_C(1-p_C)}\right)$

If $\theta = \eta(p_E) - \eta(p_C)$ *where* $\eta(p) = 2\sin^{-1}\left(\sqrt{p}\right)$ (**arcsine transformation**)

then $\hat{\theta} \sim N\left(\theta, \frac{1}{n_E} + \frac{1}{n_C}\right)$ *[the arcsine transformation is variance stabilizing]*

20

General approach to normally distributed data
Dependence of w on θ

Consider the parameterization $\underline{\Phi}$

e.g. for binary data $\underline{\Phi}' = (p_E, p_C)$

If w is dependent on θ, through $\underline{\Phi}$, then the estimate of θ satisfies

$$\hat{\theta} \sim N((\theta, \frac{1}{w(\underline{\Phi}_0)}) \quad under \ H_0 : \theta = 0; \quad \hat{\theta} \sim N((\theta, \frac{1}{w(\underline{\Phi}_R)}) \quad under \ H_1 : \theta = \theta_R$$

θ_R satisfies

$$\theta_R = \frac{U_{\alpha/2}}{\sqrt{w(\underline{\Phi}_0)}} + \frac{U_\beta}{\sqrt{w(\underline{\Phi}_R)}}$$

from which n can be deduced

form of $w(\underline{\Phi}_0)$ under H_0	under H_1	Probability difference
$w(\underline{\Phi}_0)$	$w(\underline{\Phi}_0)$	**Lachin (1981)** $n = \frac{(U_{\alpha/2} + U_\beta)^2}{(p_{ER} - p_{CR})^2}\{2\bar{p}_R(1-\bar{p}_R)\}$
$w(\underline{\Phi}_0)$	$w(\underline{\Phi}_R)$	**Fleiss (1981)** $n = \frac{(U_{\alpha/2}\sqrt{2\bar{p}_R(1-\bar{p}_R)} + U_\beta\sqrt{p_{ER}(1-p_{ER}) + p_{CR}(1-p_{CR})})^2}{(p_{ER} - p_{CR})^2}$
$w(\underline{\Phi}_R)$	$w(\underline{\Phi}_R)$	**Pocock (1983)** $n = \frac{(U_{\alpha/2} + U_\beta)^2}{(p_{ER} - p_{CR})^2}\{p_{ER}(1-p_{ER}) + p_{CR}(1-p_{CR})\}$

Note that $T^2 = \frac{(\hat{p}_E - \hat{p}_C)^2}{\hat{p}(1-\hat{p})\left(\frac{1}{n_E} + \frac{1}{n_C}\right)}$ **is Pearson's chi-square statistic**

form of $w(\underline{\Phi}_0)$ under H_0	H_1	
$w(\underline{\Phi}_0)$	$w(\underline{\Phi}_0)$	**Log-odds ratio** $n = \frac{2}{\bar{p}_R(1-\bar{p}_R)} \times \left(\frac{U_{\alpha/2}+U_\beta}{\theta_R}\right)^2$
$w(\underline{\Phi}_0)$	$w(\underline{\Phi}_R)$	$n = \left[U_{\alpha/2}\sqrt{\frac{2}{\bar{p}_R(1-\bar{p}_R)}} + U_\beta\sqrt{\frac{1}{p_{ER}(1-p_{ER})}+\frac{1}{p_{CR}(1-p_{CR})}}\right]^2 \Big/ \theta_R^2$
		Arcsine transformation $n = \frac{2(U_{\alpha/2}+U_\beta)^2}{\theta_R^2}$

23

Which sample size formula is most accurate?

- **All are approximate**
  (data are discrete, approximations are continuous)
- **Sample size formula depends on test statistic used**
  (e.g. normally distributed, binary, survival time)

	α=0.05, 1-β=0.90, p_{ER}=0.9, p_{CR}=0.8		
Form of $w(\underline{\Phi}_0)$ H0 H1	Probability difference	Log-odds ratio	Difference in arcsine
$w(\underline{\Phi}_0)$ $w(\underline{\Phi}_0)$	267	250	261
$w(\underline{\Phi}_0)$ $w(\underline{\Phi}_R)$	266	261	
$w(\underline{\Phi}_0)$ $w(\underline{\Phi}_R)$	262		

24

Continuity corrections

- **Methods for continuity correction have been proposed**
  Kramer and Greenhouse (1959)
  Casagrande, Pike and Smith (1978) [sample size given by the generalized hypergeometric distribution]
- **Some authors argue persuasively against the use of continuity corrections and Fisher's exact test as well**
  (Lydersen et al, 2009)
- **The use of unconditional tests (and consistent sample size calculation) is also recommended when only one margin of the frequency table is fixed**
  (Lydersen et al, 2009)

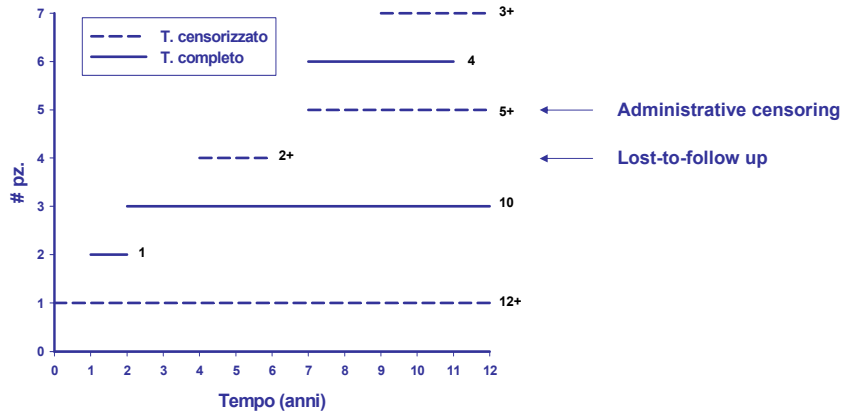
25

Survival data

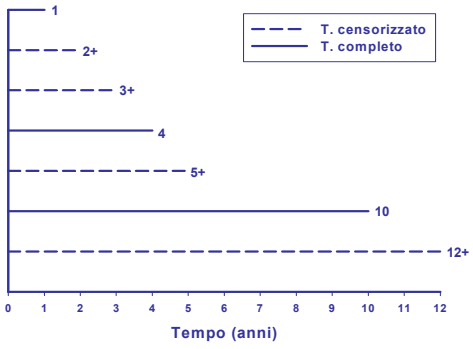
e.g. time to death, time to cancer recurrence, event time

- **Survival analysis methods are designed for studies in which subjects are entered into a trial and followed until a specified event occurs, they are lost to follow-up, or study ends**
- **For subjects who are lost to follow-up or are free from the event of interest at the end of the study, time to the event is known only to be larger than some follow-up interval (censored observations)**

26



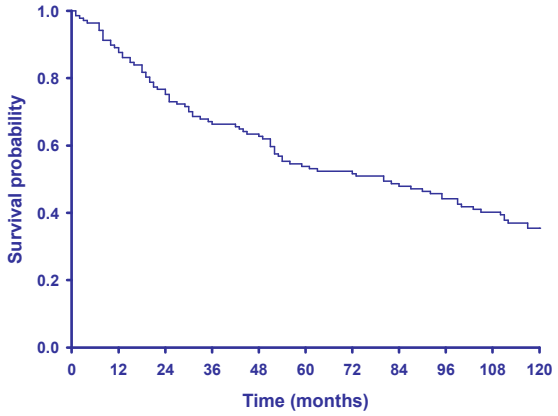
27



t	δ
1	1
2	0
3	0
4	1
5	0
10	1
12	0

28

Kaplan-Meier survival curve



29

Survival and hazard functions

If X is a survival time, with distribution function F and density function f, then the **survivor function** of X is

$$S(t) = P(X > t) = 1 - F(t)$$

and the **hazard function** of X is

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \le X < t + \delta t \mid X > t)}{\delta t} \right\} = f(t) / S(t)$$

Thus

$$h(t) = -\frac{d}{dt}\{\log S(t)\} \quad and \quad S(t) = \exp\left\{ -\int_0^t h(u)du \right\} \quad or \quad \Lambda(t) = -\log S(t)$$

30

Proportional Hazard (PH) Model

$$h_E(t) = \Psi h_C(t)$$

Ψ is the hazard ratio

The treatment difference can be measured by the **log-hazard ratio**

$$\theta = -\log_e\left\{\frac{h_E(t)}{h_C(t)}\right\} = -\log_e(\Psi)$$

or, equivalently, by the difference between complementary log transformations of the survivor functions

$$\theta = -\log_e\{-\log_e S_E(t)\} + \log_e\{-\log_e S_C(t)\}$$

31

Required total number of deaths

$$D = 4\left(\frac{U_{\alpha/2} + U_{\beta}}{\theta_R}\right)^2$$ (*Schoenfeld, 1981*)

$$D = 4\left(\frac{e^{-\theta_R} + 1}{e^{-\theta_R} - 1}\right)^2 (U_{\alpha/2} + U_{\beta})^2$$ (*Freedman, 1982*)

32

Translation from events into sample size and trial duration

- Method 1: each patient is followed up for exactly F time

A proportion $1-\frac{1}{2}\{S_{CR}(F)+S_{ER}(F)\}$ of all study patients will die

The <u>total number</u> of patients required to yield D deaths is

$$n=\frac{D}{1-\frac{1}{2}\{S_{CR}(F)+S_{ER}(F)\}}$$

33

Translation from events into sample size and trial duration

- Method 2: patients are recruited over R time and followed up to the end of the study

- Assumptions
  1. Exponential survival times, constant hazard λ (λ_{CR} and λ_{ER})
  $$S(t)=\exp(-\lambda t)$$
  2. Entry is uniform at rate *a* per time unit
- P(dead at analysis) [*P_{DC} and P_{DE} are similarly defined*]

$$P_{DC}=1-\exp(-\lambda_{CR}F)(1-\exp(-\lambda_{CR}R))/(\lambda_{CR}R)$$

$$\longrightarrow \quad \frac{aR}{2}\left[P_{DC}+P_{DE}\right]$$

Choose a, R and F to make this D deaths

34

Loss to follow-up

Extension of Method 2:

- Assumptions 1 and 2 as before plus
  3. Times to loss to follow-up are exponentially distributed with constant hazards Φ_C and Φ_E

$$P_{DC} = \frac{\lambda_C}{\lambda_C + \Phi_C}\left\{1 - \frac{1}{R(\lambda_C + \Phi_C)}(\exp(-(\lambda_C + \Phi_C)F)(1 - \exp(-(\lambda_C + \Phi_C)R)))\right\}$$

P_{DE} is similarly defined for the experimental group

a, R and F are chosen as before

For a review see Bonacina et al, Statistica Applicata (1993) and references therein

35

nQuery example (D calculated with the Schoenfeld formula)

STT1-1

Two group test of equal exponential survival (n large), no dropouts

	1	2	3
Test significance level, α	0.050	0.050	0.050
1 or 2 sided test?	2	2	2
Length of accrual period	3.00	3.00	3.00
Maximum length of followup	5.00	5.00	5.00
Group 1 exponential parameter, λ_1	0.0446	0.1022	0.1833
Group 2 exponential parameter, λ_2	0.0713	0.1386	0.2408
Hazard ratio, $h=\lambda_1 / \lambda_2$	0.626	0.737	0.761
Power (%)	80	80	80
n per group	411	505	414
Total number of events required, E	143	337	421

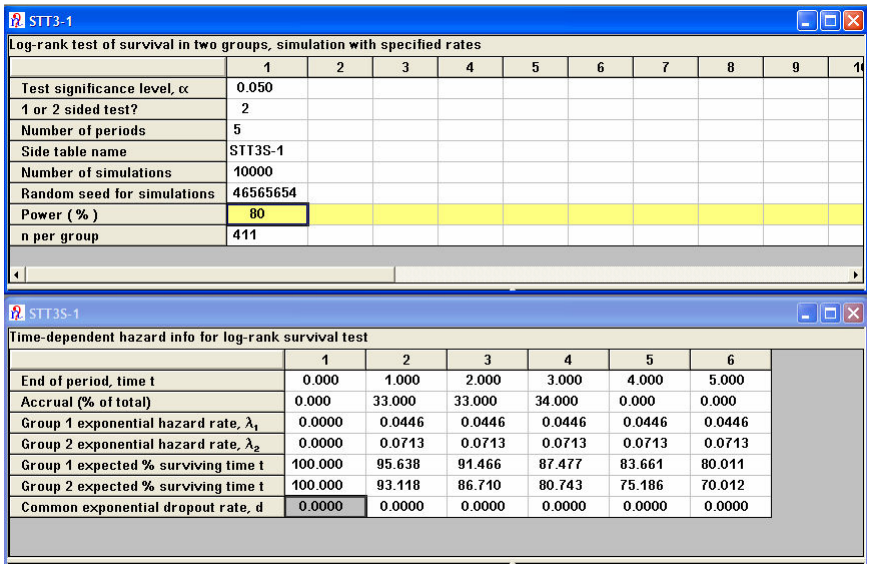
STT00-1

Conversion to alternate rates for exponential survival curves

	1	2	3
Time t	5.0	5.0	5.0
Group 1 proportion π_1 at time t	0.800	0.600	0.400
median survival	15.531	6.785	3.782
exponential parameter, λ_1	0.0446	0.1022	0.1833
Group 2 proportion π_2 at time t	0.700	0.500	0.300
median survival	9.717	5.000	2.879
exponential parameter, λ_2	0.0713	0.1386	0.2408
Hazard ratio, $h=\ln(\pi_1)/\ln(\pi_2)=med_2/med_1=\lambda_1/\lambda_2$	0.626	0.737	0.761

36

Sample size calculation by simulation (nQuery)



STT3-1

Log-rank test of survival in two groups, simulation with specified rates

	1	2	3	4	5	6	7	8	9	10
Test significance level, α	0.050									
1 or 2 sided test?	2									
Number of periods	5									
Side table name	STT3S-1									
Number of simulations	10000									
Random seed for simulations	46565654									
Power (%)	80									
n per group	411									

STT3S-1

Time-dependent hazard info for log-rank survival test

	1	2	3	4	5	6
End of period, time t	0.000	1.000	2.000	3.000	4.000	5.000
Accrual (% of total)	0.000	33.000	33.000	34.000	0.000	0.000
Group 1 exponential hazard rate, λ_1	0.0000	0.0446	0.0446	0.0446	0.0446	0.0446
Group 2 exponential hazard rate, λ_2	0.0000	0.0713	0.0713	0.0713	0.0713	0.0713
Group 1 expected % surviving time t	100.000	95.638	91.466	87.477	83.661	80.011
Group 2 expected % surviving time t	100.000	93.118	86.710	80.743	75.186	70.012
Common exponential dropout rate, d	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

37

Equivalence / non-inferiority trials

Types of treatment comparison

- **Superiority (Inferiority)**
  Objective of showing that the response to the experimental treatment (E) is superior (inferior) to the control (C; active, placebo or no treatment)
  Procedure: statistical test

- **Equivalence**
  Objective of showing that the response to two treatments differs by an amount which is clinically unimportant
  Procedure: Construct a confidence interval (θ_L, θ_U) for the difference; if $(\theta_L, \theta_U) \subseteq (-\varepsilon, \varepsilon)$ [*equivalence margins*] equivalence is demonstrated

- **Non-inferiority**
  Objective of showing that E is not clinically inferior to C
  Procedure: Construct a confidence interval (θ_L, θ_U) for the difference; if $\theta_L > -\varepsilon$ [*non-inferiority margin*] non-inferiority is demonstrated

38

Hypothesis tests for an equivalence (non-inferiority) trial

$H_0^+(-\varepsilon) : \theta = -\varepsilon$
vs
$H_1^+(-\varepsilon) : \theta > -\varepsilon$

and

$H_0^-(\varepsilon) : \theta = \varepsilon$
vs
$H_1^-(\varepsilon) : \theta < \varepsilon$

If reject $H_0^+(-\varepsilon)$ with $p \le \alpha/2$ and reject $H_0^-(\varepsilon)$ with $p \le \alpha/2$ then the $100(1-\alpha)\%$ confidence interval (θ_L, θ_U) will lie in the interval $(-\varepsilon, \varepsilon)$ [equivalence]

If reject $H_0^+(-\varepsilon)$ with $p \le \alpha$ then the $100(1-2\alpha)\%$ confidence interval (θ_L, θ_U) will lie above $-\varepsilon$ [non-inferiority]

39

Null and alternative hypothesis (1)

a) $H_0^+(\theta_1) : \theta = \theta_1$ vs $H_1^+(\theta_1) : \theta > \theta_1$ ($\theta=\theta_3$ under the alternative)
b) $H_0^-(\theta_2) : \theta = \theta_2$ vs $H_1^-(\theta_2) : \theta < \theta_2$ ($\theta=\theta_4$ under the alternative)

Non-inferiority
Sample size needs to satisfy

$$(\theta_3 - \theta_1) = \frac{U_\alpha}{\sqrt{w(\underline{\Phi}_1)}} + \frac{U_\beta}{\sqrt{w(\underline{\Phi}_3)}}$$

with $\theta_1 = -\varepsilon$, $\theta_3 = \pm\tau$ (typically $\tau=0$)

See the EMEA guideline on the choice of the non-inferiority margin (EMEA/CPMP/EWP/2158/99)

40

Null and alternative hypothesis (2)

a) $H_0^+(\theta_1) : \theta = \theta_1$ vs $H_1^+(\theta_1) : \theta > \theta_1$ ($\theta=\theta_3$ under the alternative)
b) $H_0^-(\theta_2) : \theta = \theta_2$ vs $H_1^-(\theta_2) : \theta < \theta_2$ ($\theta=\theta_4$ under the alternative)

Equivalence
Sample size needs to satisfy

$$(\theta_3 - \theta_1) = \frac{U_{\alpha/2}}{\sqrt{w(\underline{\Phi}_1)}} + \frac{U_{\beta/2}}{\sqrt{w(\underline{\Phi}_3)}}$$

and

$$(\theta_2 - \theta_4) = \frac{U_{\alpha/2}}{\sqrt{w(\underline{\Phi}_2)}} + \frac{U_{\beta/2}}{\sqrt{w(\underline{\Phi}_4)}}$$

with $\theta_1 = -\varepsilon$, $\theta_2 = \varepsilon$, $\theta_3 = \theta_4 = \pm\tau$ (typically $\tau=0$)
Choose the larger sample size

41

Design and practical issues

- **Sample size review**
- **Studies with planned interim analyses**
- **Choice of θ_R**
- **Unequal allocation**
- **Choice of the primary response variable**
- **Non-compliance**
- **Drop-outs**
- **Phase II trials**

42

Sample size review

PROCEDURE

1. **Guess values of nuisance parameters**
   (σ^2, Φ, rates of recruitment, drop-out and protocol violation)
2. **Calculate total sample size t_0**
3. **Collect data from t_1 patients ($t_1=\pi t_0$, e.g. $\pi=1/2$)**
4. **Use data so far to estimate nuisance parameters**
5. **Use these new estimates to recalculate sample size, t_1+t_2**
6. **Recruit remaining t_2 patients**
7. **Perform final analysis with all t_1+t_2 patients**
   (or possibly stop if t_2 large)

43

Some considerations about sample size reviews

1. **Reviews preserve power**
2. **Reviews do not require special analysis**
   Type I error rate not meaningfully affected (*Kieser & Friede*, 2000)
3. **Do not allow a reduction from the original sample size**
   (*Wittes and Brittain,* 1990)
4. **Limit the increase in the total sample size**
   (*Gould*, 1992)
5. **Methods exist to preserve blindness**
   (*Zucker et al*, 1999; *Gould & Shih*, 1992; *Friede & Kieser*, 2002)

44

Studies with planned interim analyses

- Obtain early evidence regarding treatment efficacy through repeated significance testing

Group sequential methods

- Early stop under the null hypothesis
  - Pocock's test
  - O'Brien and Fleming test
- Early stop under the null or the alternative hypothesis
  - Inner wedge test
  - Conditional power

45

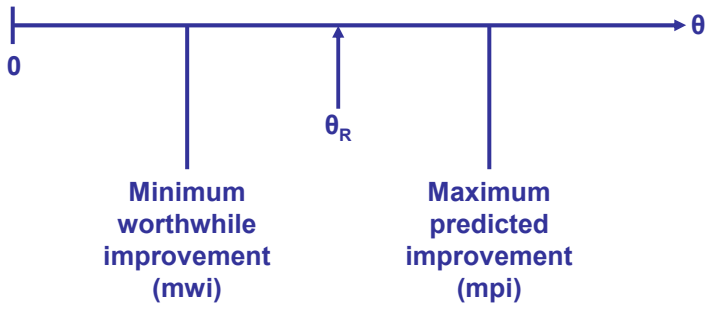
Group sequential methods

K	$C_P(K,\alpha)$	$R_P(K,\alpha,\beta)$	$C_B(K,\alpha)$	$R_B(K,\alpha,\beta)$
1	1.960	1.000	1.960	1.000
2	2.178	1.100	1.977	1.007
3	2.289	1.151	2.004	1.016
4	2.361	1.183	2.024	1.022
5	2.413	1.207	2.040	1.026

- Stop and reject H0 if:
  - $|Z_k| > C_P(K,\alpha)$
  - $|Z_k| > C_B(K,\alpha)/\sqrt{(k/K)}$ $\quad k=1,\ldots,K$
- First calculate the fixed sample size without interim analysis, then multiply it by $R_.(K,\alpha,\beta)$

46

Choice of θ_R



- If mwi > mpi give up!
- Choose θ_R somewhere between mwi and mpi
- θ_R is not the anticipated advantage, but the smallest advantage which is clinically important

47

Unequal allocation

$$N = n_E + n_C = \frac{(r+1)^2}{r}\sigma^2\left(\frac{U_{\alpha/2} + U_{\beta}}{d}\right)^2$$

r	$(r+1)^2/r$	% increase
1	4.0	--
2	4.5	12.5
3	5.3	33.3
4	6.3	56.3
5	7.2	80.0
10	12.1	302.5

48

Choice of the primary response variable

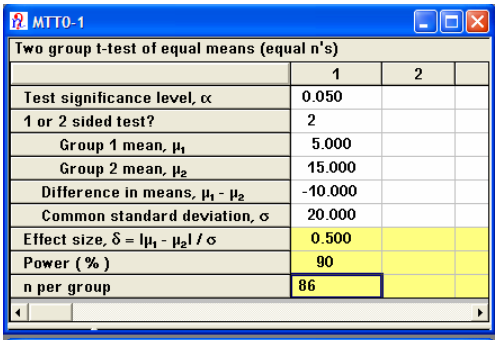
Typical options

- **Binary or normally distributed**
  e.g. % of patients with blood pressure drop ≥10 mmHg or actual blood pressure drop in mmHg?
- **Binary or survival**
  e.g. % of patients dying within one year or time to death?
- **Binary or ordered categories**
  e.g. Glasgow outcome score (good/moderate=success, severely disabled/vegetative or dead=failure)

Use the most informative form of data provided that the assumptions made are reasonable

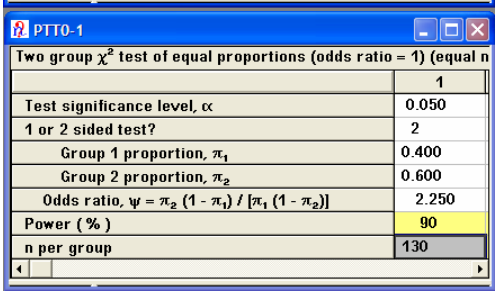
49

Example: binary or normally distributed



MTT0-1

Two group t-test of equal means (equal n's)	1	2
Test significance level, α	0.050	
1 or 2 sided test?	2	
Group 1 mean, μ_1	5.000	
Group 2 mean, μ_2	15.000	
Difference in means, $\mu_1 - \mu_2$	-10.000	
Common standard deviation, σ	20.000	
Effect size, $\delta = \lvert\mu_1 - \mu_2\rvert / \sigma$	0.500	
Power (%)	90	
n per group	86	



PTT0-1

Two group χ^2 test of equal proportions (odds ratio = 1) (equal n	1
Test significance level, α	0.050
1 or 2 sided test?	2
Group 1 proportion, π_1	0.400
Group 2 proportion, π_2	0.600
Odds ratio, $\psi = \pi_2 (1 - \pi_1) / [\pi_1 (1 - \pi_2)]$	2.250
Power (%)	90
n per group	130

Blood pressure reduction (mmHg)

μ_{CR}=5

μ_{ER}=15

σ=20

Reduction of ≥ 10mmHg

- 40% of patients on C
- 60% of patients on E

Sample size increases from 86 to 130 patients per group

50

Non-compliance

Suppose 100ε % of the experimental group do not take the treatment as intended. If θ_R is the difference due to the experimental treatment, then the difference between groups will be $\theta_R(1- \varepsilon)$

Recruit $n' = n/(1- \varepsilon)^2$

patients in each treatment group

[*If in addition some patients on C take the treatment, use* $(1-\varepsilon_C-\varepsilon_E)$]

51

Drop-outs

Patients withdraw from treatment or are lost to follow-up

Replacement

Suppose 100δ % drop-out and never provide a measure of primary efficacy, then

Recruit $n' = n/(1- \delta)$

patients in each treatment group

52

Phase II trials

- Comparative trials
- Non-comparative trials (one-sample)

- The one-sample design, commonly adopted in oncology, may be supported by several approaches:
- Frequentist
- Bayesian
- Decision theoretic

See Mariani e Marubini, 1996 and 2000.

53

Notation

x **binary response variable**
x=1 (success) with probability p
x=0 (failure) with probability (1-p)

N **number of patients**

X **cumulative number of successes observed over N patients**

$\hat{p}$ **X/N**

54

SS calculation in phase II trials

- **Based on exact tests for small samples**
  Sample sizes are usually small / asymptotic approximations may not be useful
- **Multiple-stage designs**
  Desirable to terminate the study as early as possible when the treatment is not effective

55

Fixed size one-sample design

Study parameters: p_0, p_1, α, β, N, C

Decision rule: H_0 rejected if X>C

Sample size calculation:
N e C are computed using the power function

$$R(p) = 1-B(C';p,N)$$

Algorithm:
Iterative search of N and C, such that $R(p_0) = \alpha' < \alpha$ and $1-R(p_1) = \beta' < \beta$

In practice, a small-sized N is used as the starting value; then the maximum value of X for which $\beta'<\beta$ is sought; if $\alpha'<\alpha$ at this value of X, N is retained and C is set equal to X, otherwise N is increased and the cycle is repeated

56

Sample size N and critical value C

			1−β=80%		1−β=90%	
α	p0	p1	C	N	C	N
0.05	0.05	0.20	3	27	4	38
	0.10	0.25	7	40	9	55
			…			
0.10	0.05	0.20	2	21	3	32
	0.10	0.25	5	31	6	40
			…			
0.05	0.05	0.25	2	16	3	25
	0.10	0.30	5	25	6	33
			…			
0.10	0.05	0.25	2	16	2	20
	0.10	0.30	3	18	4	25
			…			

57

Two-stage Simon's designs (1999)

Methodology

- Two stages
- Early discontinuation allowed only for inactivity
- Based on binomial probabilities, direct search of all designs (for a given p_0, p_1, α and β combination) fulfilling type I and type II error constraints
- Design optimization based on
  - expected N under H_0 (optimal design)
  - n1+n2 minimization (minimax design)

58

Simon's design: sample size N and critical value C

				Optimal design		Minimax design	
p_0	p_1	α	β	C_1/N_1	C/N	C_1/N_1	C/N
0.05	0.25	0.10	0.10	0/9	2/24	0/13	2/20
		0.05	0.20	0/9	2/17	0/12	2/16
		0.05	0.10	0/9	3/30	0/15	3/25
0.10	0.30	0.10	0.10	1/12	5/35	1/16	4/25
		0.05	0.20	1/10	5/29	1/15	5/25
		0.05	0.10	2/18	6/35	2/22	6/33
0.20	0.40	0.10	0.10	3/17	10/37	3/19	10/36
		0.05	0.20	3/13	12/43	4/18	10/33
		0.05	0.10	4/19	15/54	5/24	13/45
				...			

59

One-sample studies are dangerous!

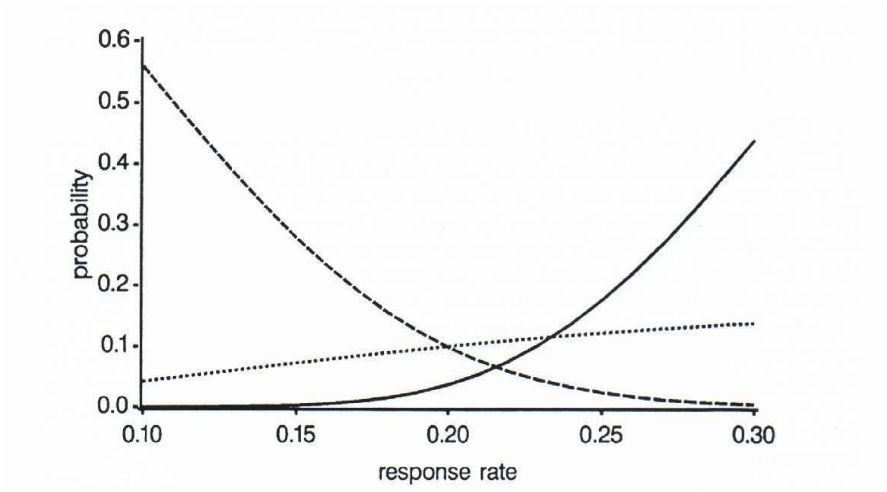


Figure 1. *Actual type I (continuous line) and type II (dashed line) error probabilities for the one-sample study and type II error probability (dotted line) for the controlled trial according to the true baseline rate.*

60