
ΜΠΣ Βιοστατιστικής

Μάθημα: Κλινικές Δοκιμές

Αθήνα, Εαρινό Εξάμηνο 2015

**Καταληκτικά σημεία
(Endpoints)**

Ουρανία Δαφνή
Καθηγήτρια, Διευθύντρια
Εργαστήριο Βιοστατιστικής
Πανεπιστήμιο Αθηνών

Διευθύντρια
Frontier Science Foundation-Hellas (FSF-H)

Καταληκτικά σημεία (Endpoints)

Ορισμός και χρήση των καταληκτικών σημείων:

- DFS: Disease Free Survival
- RFS: Relapse Free Survival
- PFS: Progression Free Survival
- OS: Overall Survival
- ORR: Objective Response Rate
- Surrogate Markers (PFS, ORR κτλ.)

Σκοποί και endpoints

- ❑ Οι κύριοι σκοποί μιας κλινικής δοκιμής πρέπει πάντα να περιορίζονται σε μετρήσιμα endpoints και ποσοτικοποιημένες υποθέσεις:
- ❑ Π.χ.
 - Μείωση του ρυθμού θνησιμότητας
 - Μείωση του ποσοστού παρενεργειών
 - Μείωση του μεγέθους μιας μετρήσιμης μεταβλητής, όπως η πίεση.

Στοιχεία που πρέπει να ληφθούν υπόψη

- Κλινική σημασία
- Η χρονική περίοδος όπου θα αξιολογηθούν οι ασθενείς (π.χ. θνησιμότητα 30 ημερών)
- Το σύνολο των γεγονότων που προσδιορίζουν ένα *endpoint* (πχ. TTF είναι ένα σύνολο από διαφορετικά γεγονότα)
- Οι διαδικασίες μέτρησης των *endpoints* (π.χ. Εργαστηριακές αναλύσεις, μεθοδολογίες κτλ.)
- Ο βαθμός αντικειμενικότητας / υποκειμενικότητας της μέτρησης και η δυνατότητα να αναπαραχθεί με συνέπεια σε διαφορετικά άτομα

Αριθμός endpoints

- Οι μελέτες έχουν συνήθως μία κύρια υπόθεση και ένα σχετικό *endpoint*. Ο σχεδιασμός της μελέτης μπορεί να απομακρυνθεί από αυτή την αρχή με ένα επιστημονικό σκεπτικό
- Γιατί μόνο ένα (ή μερικές φορές 2) *endpoint*:
 - Διογκωμένο σφάλμα τύπου I με πολλαπλά τεστ
 - Πολλά *endpoints* απαιτούν προσαρμογή του σφάλματος τύπου I, με αποτέλεσμα να είναι δύσκολο να εντοπιστούν διαφορές (χαμηλή ισχύ)
- Συνήθως ορίζονται ένα ή δύο πρωτεύοντα *endpoints* και διάφορα δευτερεύοντα. Πρέπει όμως να γίνει κατανοητό ότι η συνολική αξιολόγηση της μελέτης θα γίνει μόνο με το πρωτεύον *endpoint* (και αυτό δεν αλλάζει)

Hard Endpoints

- Ένας χρυσός κανόνας είναι να χρησιμοποιούνται *endpoints* τα οποία δεν επιδέχονται ερμηνείες ως προς τη μέτρησή τους
 - Ο θάνατος είναι απόλυτα αντικειμενικός
 - Άλλα γεγονότα όπως υποτροπή, εργαστηριακές εξετάσεις κτλ. είναι λιγότερο αντικειμενικά αλλά θεωρούνται αξιόπιστα.
- Όταν τα *endpoints* περιέχουν κρίση τότε πρέπει να είμαστε σκεπτικοί σε σχέση με την μεροληψία της αξιολόγησης.
 - Π.χ. Ελάττωση του πόνου, ποιότητα ζωής
 - Οι τυφλές μελέτες δίνουν πολλές φορές λύση

Είδη endpoint

- **Συνεχείς μετρήσεις:** πίεση του αίματος.
- **Δίτιμα ή διχότομα:** υποτροπή εντός 30 ημερών.
- **Ονομαστικά:** Το αποτέλεσμα ανήκει σε μια κατηγορία από ένα σύνολο μη διατεταγμένων κατηγοριών (ιστολογία).
- **Διατεταγμένα:** Το αποτέλεσμα ανήκει σε μια κατηγορία από ένα σύνολο διατεταγμένων κατηγοριών (τοξικότητα).
- **Σύνολα (Counts):** Ο αριθμός των δερματικών εντοπίσεων.
- **Επιβίωση ή χρόνος ως το γεγονός:** το *endpoint* στις περισσότερες μελέτες θανατηφόρων ασθενειών.

Sample Size and Power

- Sample size requirements should be considered early in the planning phase of a study
- Trials should have sufficient power to detect differences between groups considered to be of clinical interest
- Sample size calculations are an essential part of planning

Sample Size and Power

- The underlying theme of sample size considerations is precision
- For hypothesis testing, the objectives are:
 - a) Testing at a specified Type I error rate
 - b) Specified power to detect an effect size chosen for clinical interest
- For estimation, the goal is a confidence interval of a specified width

Conclusions in Hypothesis Testing (H_0 =null hypothesis)

As we have discussed, we can characterize the possible outcomes of a hypothesis test as follows:

	True State	
	<u>H_0 True</u>	<u>H_0 False</u>
<u>Test Result</u>		
Reject H_0	Type I Error	No Error
Do Not Rej H_0	No Error	Type II Error

Error Rates

- For a given test procedure and sample size, we denote the probability of a Type I error by α and the probability of a Type II error by β . The **power** of the study is the probability of concluding that H_a is true when it is actually true.

$$\text{Power} = 1 - \beta$$

- Power is always relative to a specific treatment difference, and increases as the difference between treatments increases.

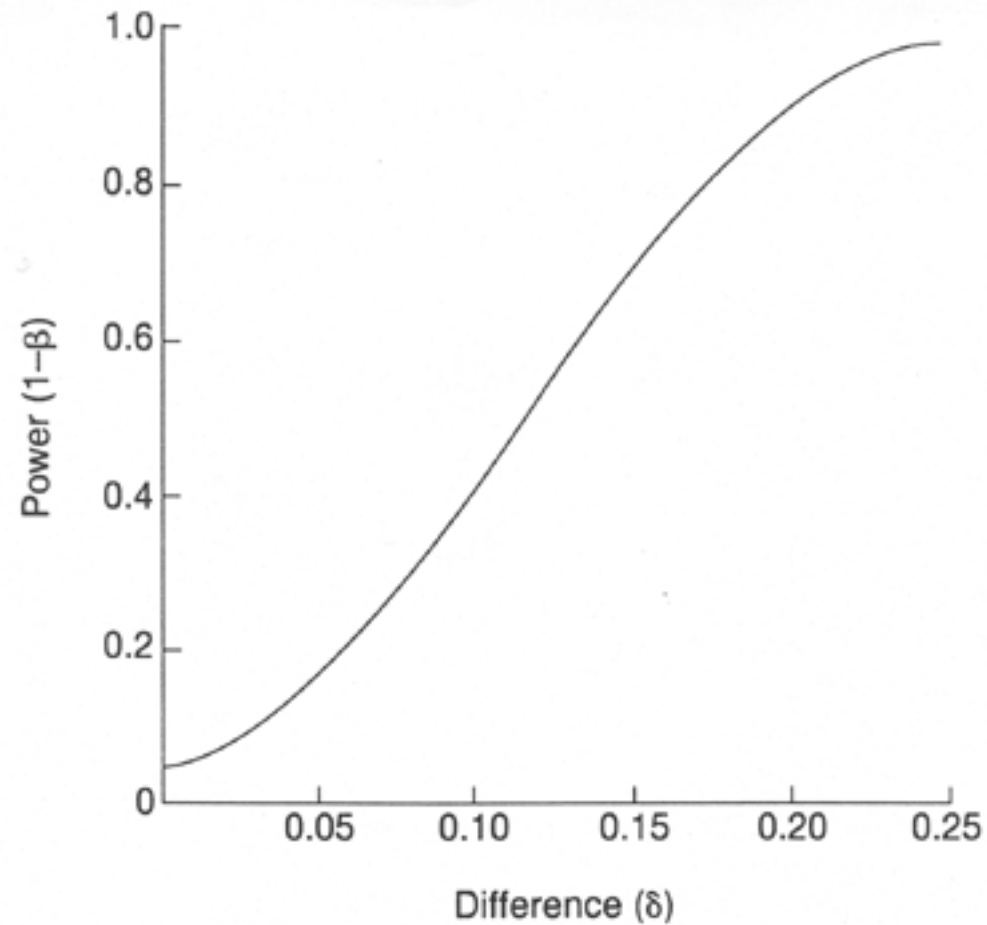


Fig. 7-1 A power curve for increasing differences (δ) between the control group rate of 0.5 and the intervention group rate with a one-sided significance level of 0.05 and a total sample size ($2N$) of 200.

-
- Ideally, the (specific) alternative hypothesis chosen for determination of sample size is the smallest difference of clinical importance between treatments
 - Why ? Because we would not want to conduct a trial and fail to detect a difference of clinical importance

 - Given a null and an alternative hypothesis, choose the sample size to achieve specified values for α (usually 0.05) and β (usually 0.10 or 0.20, but sometimes even less)

Detailed Illustration of the Principles Using Measured Outcomes

- Suppose that the endpoint is a measurement, and that the statistic for testing the null hypothesis will be based on the difference in **sample** means in the two treatment groups (denoted 1 and 2)

true population means: μ_1 and μ_2

population difference: $\Delta = \mu_2 - \mu_1$

$$H_0: \Delta = 0 \quad H_a: \Delta < 0 \text{ or } \Delta > 0$$

Example (continued)

- Observations:

- Random samples of sizes n_1 and n_2 from the two populations representing treatments 1 and 2

Example: measure blood pressure for n_1 persons randomly assigned to treatment 1 and to n_2 persons randomly assigned to treatment 2

Example (continued)

- If \bar{Y}_1 and \bar{Y}_2

are the sample means, and H_0 is true, then $(\bar{Y}_2 - \bar{Y}_1)$ will have mean 0 and variance

$$\sigma_D^2 = \sigma^2(1/n_1 + 1/n_2)$$

where σ^2 is the variance in each group.

As in many sample size calculations, we assume that $(\bar{Y}_2 - \bar{Y}_1)$ is normally distributed.

Example (continued)

- The test statistic will be of the form

$$T = \frac{\bar{Y}_2 - \bar{Y}_1}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{D}{\sigma_D}$$

and will be normally distributed with mean 0 and variance 1 under H_0 .

- Assume σ^2 known for this discussion

Example (continued)

- We will reject H_0 if $|T| > z_\alpha$
- For $\alpha = 0.05$ (2-sided), need to take $z_\alpha = 1.96$

- Now consider the alternative hypothesis

$$H_a: \Delta = \Delta_a > 0$$

and assume that the variance of individual observations remains σ^2 under H_a

Distribution of T Under H_a

- If $(\mu_2 - \mu_1) = \Delta_a (>0)$, then

$$E(\bar{Y}_2 - \bar{Y}_1) = \Delta_a$$

$$\text{Var}(\bar{Y}_2 - \bar{Y}_1) = \sigma^2(1/n_1 + 1/n_2) = \sigma_D^2,$$

the same variance as under H_0

Illustration (Continued)

- If H_a holds, then

$$T = \frac{\bar{Y}_2 - \bar{Y}_1}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

will have mean Δ_a/σ_D and variance 1

- Assume that $n_1 = n_2$

Then $\sigma_D^2 = 2\sigma^2 / n$ that

$$\Delta_a / \sigma_D = \Delta_a \sqrt{n} / \sigma \sqrt{2}$$

Δ_a / σ_D is called the noncentrality parameter.

Notice that Δ_a / σ_D increases as the square root of the sample size

Notice also, that when H_a is true,

$$T - \Delta_a / \sigma_D \sim N(0, 1)$$

Distribution of Test Statistic:

under H_0

Under H_a

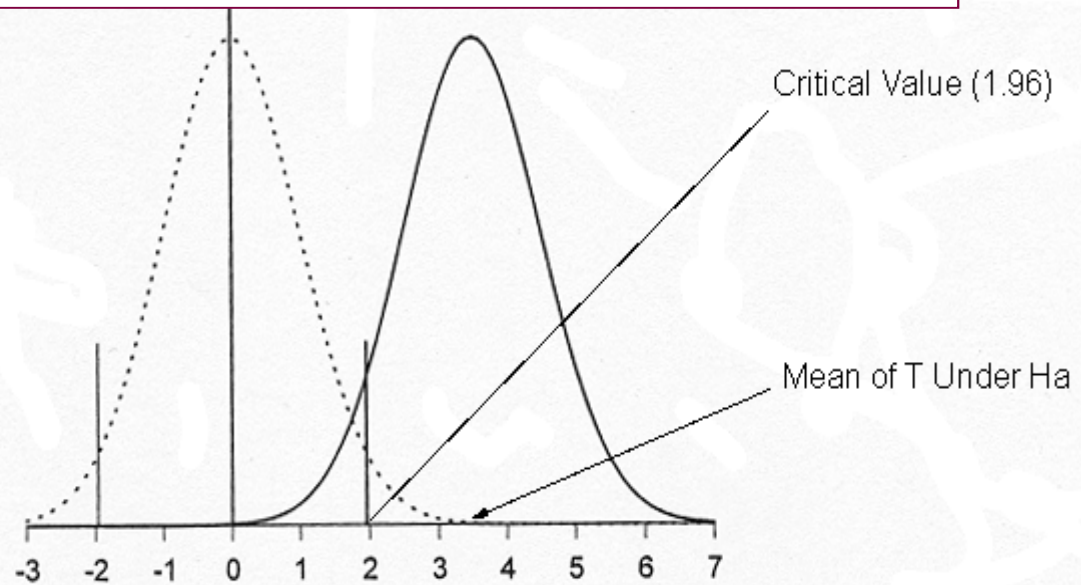


Figure 5.2 Distributions of a Treatment Effect under the Null and Alternative Hypotheses.

Example (continued)

- Note that, to achieve good power under H_0 , Δ_a must be about 3 times as large as σ_D
- Since σ_D decreases as n_1 and n_2 increase, we achieve this by increasing the sample size and, thereby reducing σ_D .
- Assume now that $n_1 = n_2$ (denote this n)

Example (continued)

- Then, under H_a , power is given by

$$P(\text{reject } H_0 \mid H_a) = P(|T| > z_\alpha \mid H_a)$$

$$\begin{aligned} &\sim P_a(T > z_\alpha) = P_a(T - \Delta_a/\sigma_D > z_\alpha - \Delta_a/\sigma_D) \\ &= 1 - \Phi(z_\alpha - \Delta_a/\sigma_D) \quad [\Phi = N(0,1) \text{ cdf}] \end{aligned}$$

- To achieve Type II error = β , we must have

$$1 - \beta = 1 - \Phi(z_\alpha - \Delta_a/\sigma_D), \text{ or}$$

$$\beta = \Phi(z_\alpha - \Delta_a/\sigma_D)$$

Example (continued)

- This will be true if the argument of Φ is $-z_\beta$, that is

$$z_\alpha - \Delta_a/\sigma_D = -z_\beta \text{ (Piantadosi, page 163)}$$

Rearranging this expression, we have

$$\Delta_a^2/\sigma_D^2 = (z_\alpha + z_\beta)^2$$

Μέγεθος Δείγματος

$$n = \frac{2(Z_{\alpha} + Z_{\beta})^2 \sigma^2}{\Delta a^2}$$

(n= sample size for each treatment,
so that total sample size=2n)

Specific Example

- Suppose that we choose $\alpha = 0.05$ and $\beta = 0.20$, so $Z_\alpha + Z_\beta = 1.96 + .84 = 2.8$
- Then, to achieve the desired power of 80% to detect a standardized difference of

$\Delta_a/\sigma = 0.25$, we require

$$n = 2*2.8^2/(0.25)^2 = 32*7.84 = 251$$

patients in each group

n depends inversely on $(\Delta_a/\sigma)^2$

Sample Sizes Per Group for
 $\alpha = 0.05$ and $\beta = 0.20$

<u>Δ_a/σ</u>	<u>n</u>
0.25	251
0.50	63
0.75	28
1.00	16

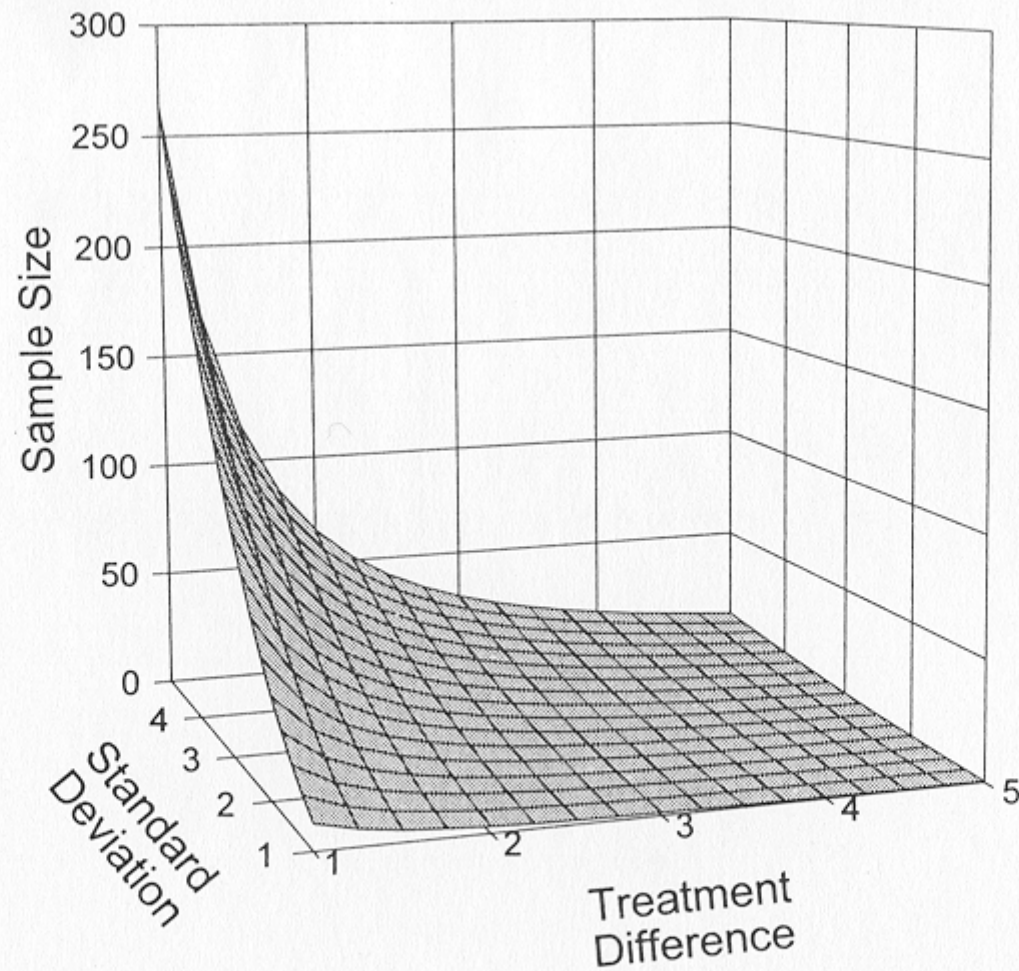


Figure 8.2 Sample size as a function of standard deviation and treatment difference.

Choosing the Type I Error Rate

- Common use of 0.05
- 1-sided versus 2-sided tests
- multiple hypotheses
- values smaller than 0.05 occasionally used for conservatism
- values larger than 0.05 rarely used

Choosing the Power ($=1-\beta$)

- 0.80 or 0.90 most commonly used
- larger values sometimes used
- when experimental constraints limit sample size, the power is reduced but the Type I error not increased

Unequal Sample Size

If $n_1 = rn_2$ ($r=n_1/n_2$)

$$n_2 = \frac{r + 1}{r} \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{\Delta_a^2}$$

- Unbalanced allocation is less efficient

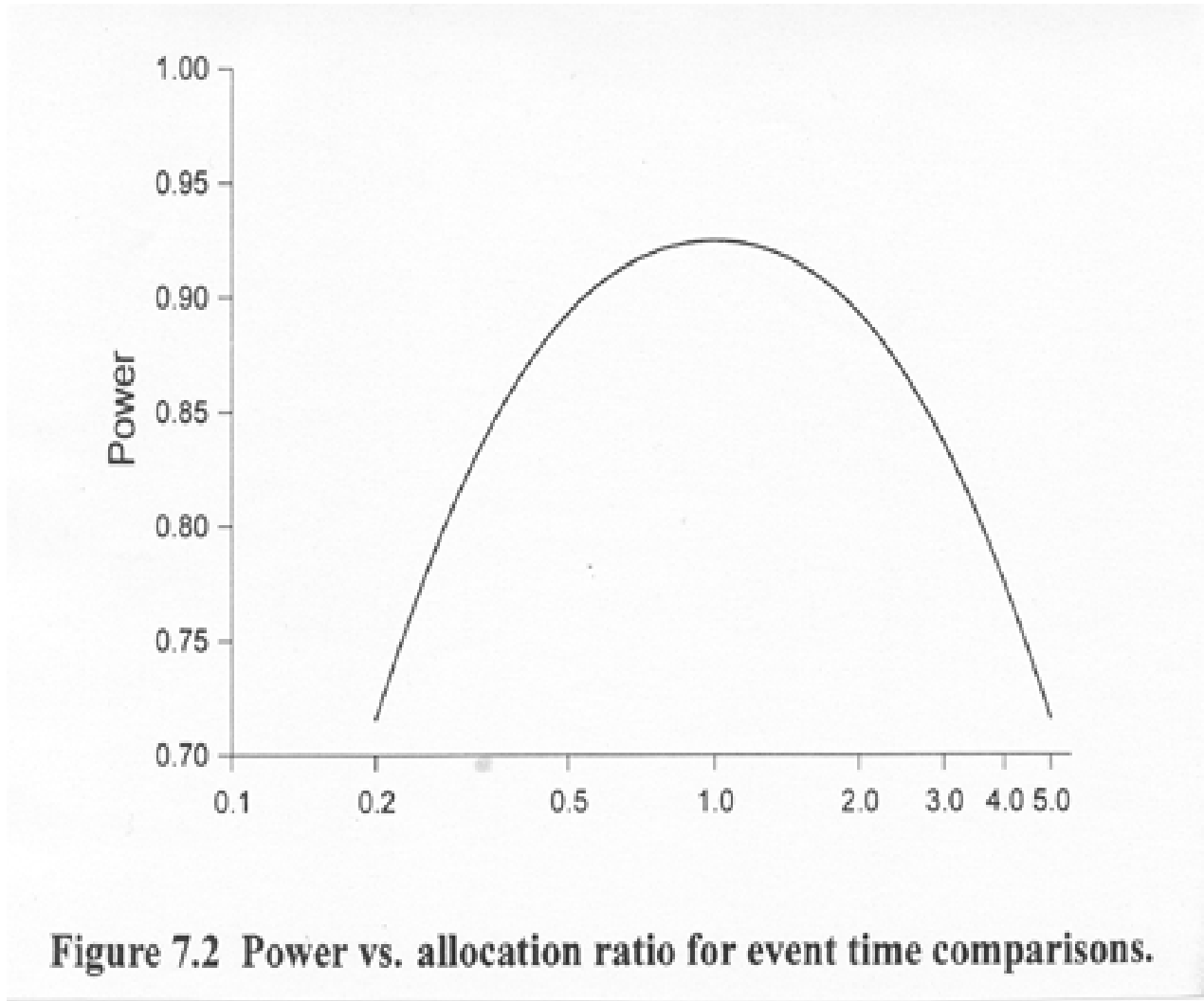


Figure 7.2 Power vs. allocation ratio for event time comparisons.

Loss-to-Follow-Up

(i.e., when patients leave study before schedule and response is not available)

■ Issues:

1. Are observations missing at random? If not, the missing data might induce bias
2. If **random**, then the missing observations reduce the effective sample size.
3. If a fraction p of the observations are randomly missing in each group, the **n must be increased to $n/(1-p)$** to achieve the desired net sample size
e.g. $n=100$, $p=.20$, then $n/.8 = 125$

Nonadherence (a.k.a. noncompliance)

- If n patients are enrolled in each group and a fraction, p , of patients in the treatment group do not comply with therapy and have the same response as controls, the actual observed treatment difference will be

$$(1-p)*\Delta_a + p*0 = (1-p)*\Delta_a$$

(this assumes all patients are analyzed as randomized)

- The required sample size in the treatment arm must be increased from n to $n/(1-p)^2$. e.g. $p=.20$ gives $\sim 1.5n$.

Sensitivity to Assumptions

- Investigators are sometimes surprised that small changes in assumptions cause large changes in n
- n depends on the square of effect size
- A 10% Decrease in effect size yields a $>20\%$ increase in sample size [$1/.9^2$]
- Errors in assumptions can also diminish power
- Dropouts, noncompliance can also diminish power, allow for this in sample size calculation

Summary: for measured outcomes

- Determine sample size ratio $r=n_1/n_2$ (usually 1)
- Fix Type I error (e.g., 0.05)
- Pick sample size to achieve desired power (e.g., 0.80)
- Sometimes: increase calculated sample size to account for loss in power from noncompliance or losses to follow-up

Sample Size for Binary Outcomes

(e.g.: did MI patients survive 30 days or not)

- Population quantities are probabilities of ‘success’ for 2 treatments: p_1 vs. p_2
- Population treatment difference usually measured by risk ratio (RR) or odds-ratio (OR):

$$RR = p_1/p_2$$

$$OR = [p_1/(1-p_1)] / [p_2/(1-p_2)]$$

- For both metrics, null hypothesis that $p_1=p_2$ corresponds to a value of 1.
- Alternatives to H_0 correspond to values other than 1
- Different tests statistics than with measured outcomes (usually Fisher’s exact test or χ^2 test)
- However, same ideas to determine sample size

Time to Event Responses

- In many clinical trials, the primary response variable is the time, T , to an event; e.g.:

T =time to death, T =time to relapse, T =time to infection

- Usually like to express distribution of T in terms of ‘survival distribution’ $S(t)=1-c.d.f.$

$$S(t)=P(T>t)=P(\text{‘alive’ at } t)$$

- Also examine ‘hazard function’ $h(t)= -S'(t)/S(t)$

$$h(t) \sim P(\text{fail at } t \mid \text{haven't failed before } t)$$

- To compare 2 treatment groups, common to measure difference in terms of ‘risk ratio’ $h_2(t)/h_1(t)$ to assume this is constant in t (‘proportional hazards’)

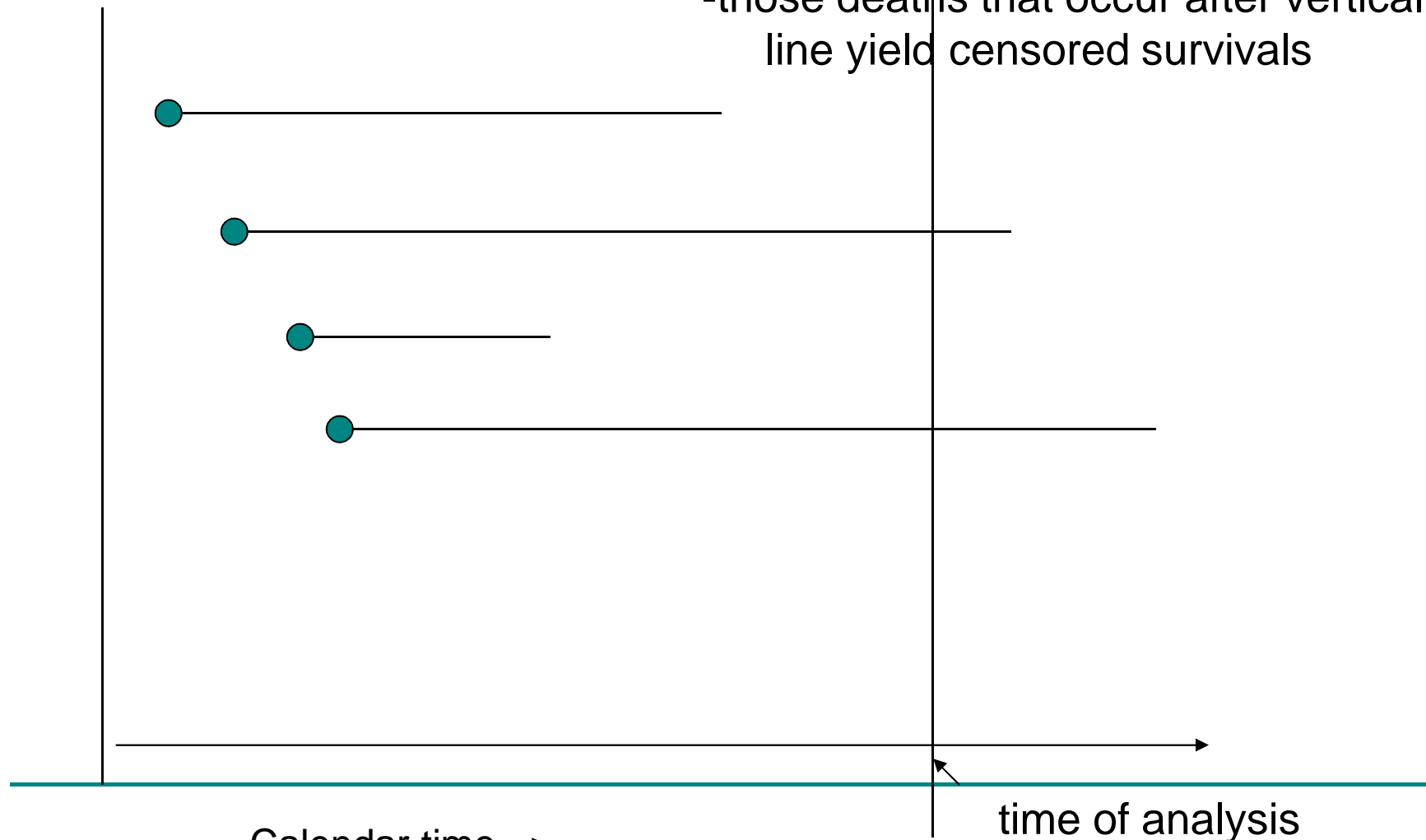
$$RR= h_2(t)/h_1(t)$$

Time to Event Response (continued)

- Note that null hypothesis corresponds to $RR=1$
- Values of $RR > 1$ mean risk of 'death' in group 2 is higher than in group 1. $RR < 1$ means risk in group 2 is lower than in group 1
- Complication to analyses: censored data
 - Some patients have not yet failed by the time we end our trial and must analyze the data. E.g., a patient was randomized 17 weeks ago but has not yet failed. For him/her, $T = "17+ "$. That is, we know only that $T > 17$.

Example of Staggered Entry into a Trial:

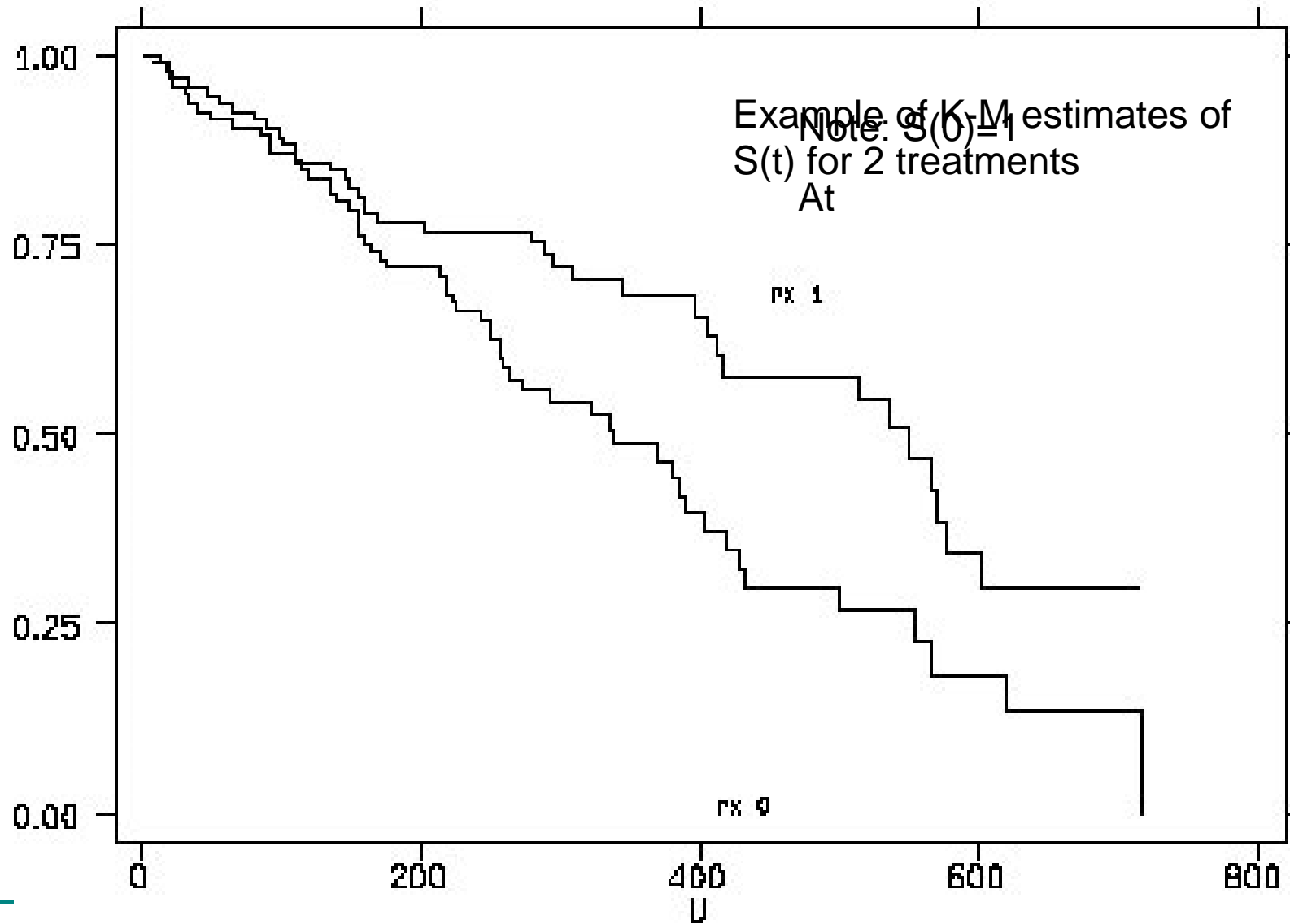
- each horizontal line is a patient
- circle denotes time of enrollment
- length of line denotes survival time
- those deaths that occur after vertical line yield censored survivals



Time to Event Response (continued)

- Statistical methods have been developed to analyze such censored data. Most common:
 - To estimate $S(t)$ for a treatment group:
 - Kaplan-Meier estimator (nonparametric)
 - To test equality of survival distributions of 2 groups:
 - Logrank test (nonparametric)
 - To estimate RR:
 - Cox's proportional hazards model (semiparametric)

Kaplan-Meier survival estimates, by rx



Time to Event Responses (continued)

- Key result: the ‘power’ of the test to compare treatment groups depends on the number of ‘failures’ (deaths) that are observed during the trial, not on the number of patients:

Thus:

- a smaller trial in which each patient is followed for longer can have the same power as a larger trial in which patients are followed less time

Example: If survival is exponentially distributed, a trial with $\frac{1}{2}$ the sample size, but double the patient follow-up will have the same power as one with twice the sample size but $\frac{1}{2}$ the follow up.

Time to Event Responses (continued)

- Implications for design: need to consider both sample size and average follow-up (AFU) per patient
- Example: with staggered entry trial in which all patients are followed until end of trial:

W₁ = duration of enrollment period

R = rate of enrollment (patients/ unit time)

W₂ = additional follow-up of all patients after last one enrolled

Then: total accrual: $N = n_1 + n_2 = W_1 * R$

average FU: $AFU = \frac{1}{2} W_1 + W_2$

Design parameters (suppose R fixed): W_1, W_2

~> different combinations of (W_1, W_2) will give the same power

Time to Event Responses (continued)

Example: survival in advanced lung cancer

- Treatment 1=control (standard treatment)
 - assume median exponential survival = 6 months
- Treatment 2=new treatment
 - assume RR=1/2 (new treatment halves risk => doubles median exponential survival)
- 2-sided Type I error=0.05; power=0.80
- Accrual rate=30 patients / month; equal allocation
- The following designs give the desired power:
 - (a) W1= 3 months accrual, W2= 15.62 mos additional FU
 - (b) W1= 6 months accrual, W2= 2.98 mos additional FU
 - (c) W1= 8 months accrual, W2= 0.34 months additional FU

Time to Event Responses (continued)

<u>Design</u>	<u>W1</u>	<u>W2</u>	<u>N</u>	<u>Avg FU</u>	<u>Max FU</u>
(a)	3	15.62	90	17.13	18.62
(b)	6	2.98	180	5.98	8.98
(c)	8	.34	240	4.34	8.34

Note the ‘other’ differences among these designs (that have equal power):

- ❑ Design (a) will take 18+ months to complete compared to many fewer months for designs (b) and (c)
- ❑ Design (a) will tell us about the survival function $S(t)$ for much longer values of t
~> a tradeoff !!
- ❑ Costs might also differ substantially

Different designs reveal more/less of the estimated survival function $S(t)$

- Recall the Kaplan Meier estimator, and suppose this represents the results from this study.
 - With design (a), the maximum follow-up is 18.63 months, or about 570 days. At the time of the analysis we would observe the first 570 days of the estimator on slide 36
 - With designs (b) and (c), we would see only the first 270 days and 250 days
- These differences could be important!!

Sources for Sample Size Determination

- Most sample sizes are based on a normal approximation to the test statistic (an exception is Fisher's exact test)
- Details are sometimes a bit complicated
- We will illustrate this using the package EAST

Ορισμοί Endpoints

- **Overall Survival**
- **Disease-Free Survival**
- **Time to Progression and Progression-Free Survival**
- **Time to Treatment Failure**

**Guidance for Industry, Clinical Trial Endpoints
for the Approval of Cancer Drugs and
Biologics**

*U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
April 2005, Clinical/Medical*

Overall Survival

- ❑ Overall survival is defined as the time from randomization until death from any cause, and is measured in the intent to treat (ITT) population.

Disease-Free Survival

- ❑ Disease-free survival (DFS) is usually defined as the time from randomization until recurrence of tumor or death from any cause.

Time to Progression and Progression-Free Survival

- ❑ Time To Progression (TTP)
 - the time from randomization until objective tumor progression
- ❑ Progression-Free Survival (PFS)
 - the time from randomization until objective tumor progression or death

Time to Treatment Failure

- ❑ Time to treatment failure (TTF) is a composite endpoint measuring time from randomization to discontinuation of treatment for any reason (including progression of disease, treatment toxicity, and death).

Σύγκριση των endpoints (1)

Endpoint	Regulatory Nature of Evidence	Assessment	Some Advantages	Some Disadvantages
Overall Survival	Clinical benefit	<ul style="list-style-type: none">• Randomized studies needed• Blinding not essential	<ul style="list-style-type: none">• Universally accepted direct measure of benefit• Easily measured• Precisely measured	<ul style="list-style-type: none">• Requires larger studies• Requires longer studies• Potentially affected by crossover therapy• Does not capture symptom benefit• Includes noncancer deaths
Disease-Free Survival	Surrogate for accelerated approval or regular approval*	<ul style="list-style-type: none">• Randomized studies needed• Blinding preferred	<ul style="list-style-type: none">• Considered to be clinical benefit by some• Needs fewer patients and shorter studies than survival	<ul style="list-style-type: none">• Not a validated survival surrogate in most settings• Not precisely measured; subject to assessment bias• Various definitions exist

*Adequacy as a surrogate endpoint for accelerated approval or regular approval is highly dependent upon other factors such as effect size, effect duration, and benefits of other available therapy. See text for details.

Σύγκριση των endpoints (2)

Endpoint	Regulatory Nature of Evidence	Assessment	Some Advantages	Some Disadvantages
Progression Free Survival (PFS)	Surrogate for accelerated approval or regular approval*	<ul style="list-style-type: none">• Randomized studies needed• Blinding preferred• Blinded review recommended	<ul style="list-style-type: none">• Activity measured in responding and stable tumors• Usually assessed prior to change in therapy• Less missing data than for symptom endpoints• Assessed earlier and in smaller studies compared with survival	<ul style="list-style-type: none">• Various definitions exist• Not a direct measure of benefit• Not a validated survival surrogate• Not precisely measured compared with survival• Is subject to assessment bias• Frequent radiologic studies are needed• Data are voluminous and complex compared to survival

*Adequacy as a surrogate endpoint for accelerated approval or regular approval is highly dependent upon other factors such as effect size, effect duration, and benefits of other available therapy. See text for details.

Σύγκριση των endpoints (3)

Endpoint	Regulatory Nature of Evidence	Assessment	Some Advantages	Some Disadvantages
Objective Response Rate (ORR)	Surrogate for accelerated approval or regular approval*	<ul style="list-style-type: none">• Single-arm or randomized studies can be used• Blinding preferred in comparative studies	<ul style="list-style-type: none">• Can be assessed in single-arm studies	<ul style="list-style-type: none">• Not a direct measure of benefit• Usually reflects drug activity in a minority of patients• Data are moderately complex compared to survival
Complete Response (CR)	Surrogate for accelerated approval or regular approval*	<ul style="list-style-type: none">• Single-arm or randomized studies can be used• Blinding preferred in comparative studies	<ul style="list-style-type: none">• Durable CRs represent obvious benefit in some settings (see text)• Can be assessed in single-arm studies	<ul style="list-style-type: none">• Few drugs produce high rates of CR• Data are moderately complex compared to survival

*Adequacy as a surrogate endpoint for accelerated approval or regular approval is highly dependent upon other factors such as effect size, effect duration, and benefits of other available therapy. See text for details.

Υποκατάστατοι δείκτες έκβασης (*Surrogate Endpoints*)

Ουρανία Δαφνή
Καθηγήτρια, Διευθύντρια
Εργαστήριο Βιοστατιστικής
Πανεπιστήμιο Αθηνών

Διευθύντρια
Frontier Science Foundation-Hellas (FSF-H)

Surrogate Endpoints

- Ένα Surrogate Endpoint είναι ένα αποτέλεσμα το οποίο χρησιμοποιείται στην θέση του κύριου αποτελέσματος
 - Οστεοπόρωση: οστική πυκνότητα ως surrogate του κατάγματος από πτώση.
 - HIV: Αλλαγή στο viral load ως surrogate της υποτροπής.
- Κύρια πλεονεκτήματα είναι το κόστος, ο χρόνος και η ευκολία μέτρησης.
- Κρίσιμο σημείο είναι η αξιοπιστία της αξιολόγησης της επίδρασης της θεραπείας στο κλινικό αποτέλεσμα βάσει της επίδρασης στο surrogate.
- Π.χ. Αν ένα φάρμακο αυξάνει την οστική πυκνότητα θα μειώσει τον κίνδυνο κατάγματος;

Οι Surrogates είναι Disease-Specific

<u>Νόσος</u>	<u>Endpoint</u>	<u>Surrogate</u>
HIV	AIDS ή θάνατος	CD4 Count
CA	Υποτροπή	CR - PR
Prostate CA	Υποτροπή	PSA Level
CVD	Stroke MI	BP Lipid Level
Glaucoma	Vision Loss	Interocular Pressure
Osteoporosis	Fractures	Bone Min. Density

Επιθυμητές ιδιότητες των Surrogates

- Να μετρούνται εύκολα χωρίς πολύπλοκες διαδικασίες
- Κοντά ή μέρος της αιτιακής διαδικασίας
- Αποδίδει το ίδιο συμπέρασμα για το αποτέλεσμα της θεραπείας όπως και το endpoint

Οι μελέτες που χρησιμοποιούν surrogates μπορεί να είναι παραπλανητικές

- CAST (Cardiac Arrhythmia Suppression Trial)
 - Το encainide και το flecanide μείωσαν τις αρρυθμίες (surrogates), οπότε θεωρήθηκε ότι θα μείωναν την θνησιμότητα.
 - Αργότερα διαπιστώθηκε ότι αύξαναν τον ξαφνικό θάνατο.
- Το Milrinone βελτίωσε τις αιμοδυναμικές παραμέτρους του CHF αλλά οδήγησε σε αυξημένη μακροπρόθεσμη νοσηρότητα και θνησιμότητα.
- Η θεραπεία με Fluoride για την οστεοπόρωση αύξησε την οστική μάζα αλλά οδήγησε σε αυξημένο ρυθμό καταγμάτων.

Πλαίσιο αξιολόγησης ενός υποψήφιου surrogate (1)

- Y= κλινικό outcome (π.χ. AIDS υποτροπή)
- X= ομάδα θεραπείας (π.χ. AZT ή placebo)
- Z= υποψήφιο surrogate (π.χ. Μεταβολή στα CD4)

- Ας υποθέσουμε ότι τα δεδομένα δείχνουν ότι:
 - X συνδέεται με Y (AZT μειώνει την AIDS υποτροπή)
 - X συνδέεται με Z (AZT βελτιώνει τα CD4)
 - Z συνδέεται με Y (υψηλότερα CD4 συνδέονται με μικρότερο ρίσκο για AIDS υποτροπή)

- Αυτό σημαίνει ότι το Z (CD4) είναι αξιόπιστο surrogate για το Y (AIDS υποτροπή); **ΟΧΙ!!**

Πλαίσιο αξιολόγησης ενός υποψήφιου surrogate (2)

- Απαραίτητη συνθήκη: $P(Y | X, Z) = P(Y | Z)$
Ερμηνεία: η επίδραση του X (θεραπεία) στο Y (κλινικό αποτέλεσμα) είναι εξ' ολοκλήρου το αποτέλεσμα της επίδρασης του X στο Z . Οπότε αν λάβουμε υπόψη το Z , το να λάβουμε υπόψη και το X δεν προσθέτει πληροφορία.
- Στην πράξη η αξιολόγηση των surrogates είναι δύσκολη
 - ❑ Χρειαζόμαστε δεδομένα για το Y και το Z από όσους ασθενείς που έλαβαν και όσους δεν έλαβαν θεραπεία.
 - ❑ Η στατιστική αξιολόγηση γίνεται με πολύπλοκες μεθόδους ειδικά για surrogates που μελετώνται διαχρονικά (CD4, PSA, BMD).

Αξιολόγηση ενός υποψήφιου surrogate

- Παράδειγμα: Θνησιμότητα ασθενών με AIDS
 - Y: θάνατος
 - X: AZT versus Placebo
 - Z: διαχρονική μεταβολή στα CD4
- Το AZT βελτιώνει τα CD4 και μειώνει το ρίσκο του θανάτου. Όμως η θετική επίδραση του AZT στα CD4 εξηγεί ένα μικρό ποσό της βελτίωσης της επιβίωσης.
- Τα CD4 δεν είναι αξιόπιστο surrogate του θανάτου.
- Reference: Wulfsohn & Tsiatis, Biometrics, 1985.

Surrogate Markers: Ανακεφαλαίωση

- Δύσκολο να αξιολογήσουμε αν ένας δείκτης είναι surrogate στο πνεύμα που περιγράψαμε.
- Συχνά δεν είναι ρεαλιστικό ότι ένας δείκτης θα είναι «πλήρης» surrogate.
- Αν ένας δείκτης είναι ένα από τα πιθανά «αιτιακά μονοπάτια» για το κλινικό αποτέλεσμα τότε:
 - Μια θεραπεία με θετικό κλινικό αποτέλεσμα μπορεί να έχει μικρή/καμία επίδραση στον surrogate.
 - Μια θεραπεία με ελάχιστο κλινικό αποτέλεσμα μπορεί να έχει μεγάλη επίδραση στον surrogate.

Προσοχή λοιπόν στην ερμηνεία των surrogates

Περίληψη

- Η επιλογή των endpoints είναι κρίσιμη σε μια κλινική δοκιμή ή και οποιαδήποτε μελέτη.
- Το endpoint πρέπει να έχει κλινικό νόημα αλλά ταυτόχρονα να είναι σαφώς ορισμένο και μετρήσιμο.
- Η ερμηνεία των αποτελεσμάτων γίνεται στα πλαίσια των endpoints.
- Surrogate endpoints/ δείκτες μπορεί να είναι πολύ χρήσιμα, αλλά πρέπει να είμαστε προσεκτικοί με τους περιορισμούς σχετικά με τα συμπεράσματα για την επίδραση της θεραπείας στο κλινικό αποτέλεσμα.

Η ανταπόκριση ως surrogate marker της αποτελεσματικότητας της θεραπείας (επιβίωση)

- Αυθαιρεσία στον ορισμό της ανταπόκρισης
 - RECIST 33% → ανταπόκριση αλλά 28% όχι
- Μέτρο βιολογικής δράσης αλλά όχι οφέλους για τον ασθενή
 - Ανταπόκριση και τοξικότητα
- Συχνά παρατηρούνται μεγάλες διαφορές στις δημοσιευμένες μελέτες
- Λάθη μέτρησης και αξιολόγησης



Κανόνες πρόωρου τερματισμού *(Early stopping rules)*

Ουρανία Δαφνή
Καθηγήτρια, Διευθύντρια
Εργαστήριο Βιοστατιστικής
Πανεπιστήμιο Αθηνών

Διευθύντρια
Frontier Science Foundation-Hellas (FSF-H)

Επιτάχυνση Συμπερασμάτων μέσω Σχεδιασμού Κλινικής Δοκιμής

- Σχεδιασμός με Δυνατότητα Πρώιμης Διακοπής της Μελέτης
 - Ακολουθιακός Σχεδιασμός και Ανάλυση
 - Δίνει τη δυνατότητα να διακοπεί η μελέτη εάν υπάρχει απόδειξη διαφοράς αποτελεσματικότητας σε ενδιάμεση ανάλυση βάσει της διαθέσιμης πληροφορίας (ή δεν υπάρχει)
 - Στοχαστική Περικοπή – Δεσμευμένη Ισχύς - Futility
 - Δίνει τη δυνατότητα να διακοπεί η μελέτη εάν η συνέχισή της έως τέλους δεν θα οδηγήσει σε συμπέρασμα διαφοράς στην αποτελεσματικότητα

Interim Monitoring Κλινικών Δοκιμών

- Είναι σύνηθες να εξετάζουμε περιοδικά τα ενδιάμεσα αποτελέσματα μιας δοκιμής (π.χ. ετήσια)
- Η εξέταση διεξάγεται από μια ανεξάρτητη επιτροπή: Data and Safety Monitoring Committees ("DSMC") (ή Data Mon. Cmte and Data & Safety Mon. Board)
- Η DSMC προτείνει είτε να τροποποιηθεί η μελέτη είτε να διακοπεί πρόωρα είτε να συνεχίσει έως την ολοκλήρωσή της βάσει του αρχικού σχεδιασμού

Σκεπτικό για το Interim Monitoring

- Οι μελέτες διαρκούν συνήθως πολλά έτη
- Ενδέχεται νεότερες πληροφορίες να είναι διαθέσιμες για την αποτελεσματικότητα της θεραπείας
- Επίσης μπορεί να διαπιστωθεί ότι το μέγεθος του δείγματος της μελέτης δεν είναι δυνατόν να επιτευχθεί

Εξαιτίας λοιπόν της νεότερης πληροφορίας είναι δόκιμο να διεξάγονται ενδιάμεσες αναλύσεις για να διαπιστωθεί αν:

1. Είναι ακόμη ηθικό να συνεχίσουμε τη συγκεκριμένη θεραπεία
2. Είναι δυνατόν η μελέτη να επιτύχει τους επιστημονικούς της στόχους

Αιτίες πρόωρου τερματισμού κατά την Interim Analysis

- Οι θεραπείες είναι κατά τεκμήριο διαφορετικές
- Οι θεραπείες είναι κατά τεκμήριο μη διαφορετικές
- Εύρεση μη αποδεκτών επιπέδων τοξικότητας και ΑΕ
- Ο ρυθμός εισαγωγής στη μελέτη (accrual) είναι τόσο αργός που η μελέτη δεν είναι πια εφικτή
- Εξωτερικές πληροφορίες καθιστούν τη μελέτη μη χρήσιμη ή μη ηθική
- Φτωχή διεξαγωγή δεν επιτρέπει στη μελέτη να επιτύχει τους στόχους της
- Απάτη ή μη συμμόρφωση

Data and Safety Monitoring Committees

- Σύνθεση
 - Ιατροί, βιοστατιστικοί, άλλοι επιστήμονες (μερικές φορές), εκπρόσωποι ασθενών (μερικές φορές)
- Σκοπός: διερεύνηση κατά πόσον το συμφέρον των ασθενών εξυπηρετείται εξετάζοντας περιοδικά τα δεδομένα της μελέτης όσον αφορά
 - την ποιότητα διεξαγωγής της μελέτης και τις αρχικές υποθέσεις
 - την ασφάλεια και τοξικότητα
 - την αποτελεσματικότητα
- Ενέργειες: (μετά από την interim analysis)
 - προτείνεται η συνέχιση της μελέτης με ή χωρίς τροποποίηση
 - προτείνεται ο πρόωρος τερματισμός

Data and Safety Monitoring Committees

- Πολλά έχουν γραφτεί για τη μη στατιστική πλευρά του interim monitoring.
- Ερωτήματα όσον αφορά όμως την καθαρά στατιστική πλευρά δηλαδή την αξιολόγηση της αποτελεσματικότητας:
 - Περιοδικότητα εξέτασης των δεδομένων
 - Κριτήρια που πρέπει να επιβληθούν ούτως ώστε να περιοριστεί το σφάλμα τύπου I (false positive rate) στο αρχικό επίπεδο (π.χ. 0.05) δεδομένου ότι διεξάγουμε ελέγχους πολλαπλές φορές
 - Επιλογή στατιστικών διαδικασιών πρόωρου τερματισμού
 - A. Εάν διαφαίνεται υπεροχή της μιας θεραπείας έναντι
 - B. Εάν δεν αναμένεται υπεροχή καμίας έναντι των υπολοίπων (futility) με αρκετά υψηλή πιθανότητα
 - Πώς θα πρέπει να παρακολουθούνται οι μελέτες ισοδυναμίας;

Κύριες προσεγγίσεις Κανόνων Διακοπής

- group sequential methods
 - fixed monitoring schedules
 - rules based on Type I spending functions

Γιατί θα πρέπει να ανησυχούμε για τις πολλαπλές ενδιάμεσες αναλύσεις;

- Αν ελέγχουμε τη μηδενική υπόθεση (ισότητας των θεραπειών) πολλαπλές φορές με τα συνήθη κριτήρια ($\alpha=0.05$) η συνολική πιθανότητα σφάλματος τύπου I αυξάνεται
- Παράδειγμα: Ας υποθέσουμε ότι ελέγχουμε 10 φορές τη μηδενική υπόθεση. Τότε αν η μηδενική υπόθεση είναι όντως αληθινή (οι θεραπείες δεν διαφέρουν) η συνολική πιθανότητα σφάλματος τύπου I θα είναι $1 - (1-0.05)^{10}$ δηλαδή

Σφάλμα τύπου I = 0.40

=> 40% πιθανότητα να συμπεράνουμε εσφαλμένα ότι οι θεραπείες διαφέρουν!

- Πρέπει λοιπόν να λάβουμε υπόψη τους πολλαπλούς ελέγχους λόγω των ενδιάμεσων αναλύσεων

Κοινές παράμετροι

- Οι επίσημοι κανόνες διακοπής βασίζονται στο κύριο καταληκτικό σημείο (primary efficacy endpoint)
- Τα δεδομένα πρέπει να είναι διαθέσιμα αθροιστικά
- Η παρακολούθηση της ασφάλειας και της τοξικότητας είναι δυνατόν να είναι πιο ευέλικτη

Group Sequential Methods

- Στις περισσότερες μελέτες δεν είναι δυνατόν να αναλύουμε τα στοιχεία κάθε φορά που εισέρχεται ένας ασθενής:
 - Κάτι τέτοιο δεν είναι δυνατόν λειτουργικά
 - Η ανταπόκριση δεν είναι άμεση οπότε μπορεί να μην είναι διαθέσιμη πριν την εισαγωγή στη μελέτη του επόμενου ασθενή
- Ο πλέον συχνός τρόπος ανάλυσης στις Τυχαιοποιημένες Κλινικές Δοκιμές:
 - Αναλύουμε τα στοιχεία σε προκαθορισμένα μεγέθη εισαγωγής (π.χ. 50, 100, 150, ... ασθενείς έχουν εισαχθεί) και συγκρίνουμε το **test statistic**, σε κάθε **interim** ανάλυση, με τα όρια διακοπής που έχουμε ορίσει

Σχεδιασμοί Μελετών με Ενδιάμεση Παρακολούθηση

- Μηδενική Υπόθεση:

H_0 : Θεραπεία A = Θεραπεία B

(π.χ., $\pi_1 = \pi_2$), όπου π πιθανότητα πλήρους ανταπόκρισης (CR)

- Χρόνοι των K αναλύσεων: T_1, T_2, \dots, T_K .

- Ενδιάμεσες αναλύσεις στους χρόνους T_1, T_2, \dots, T_{K-1}

- Τελική ανάλυση στο χρόνο T_K

- Σε κάθε ανάλυση υπολογίζουμε τη συνήθη στατιστική συνάρτηση ελέγχου Z_k (π.χ., διαφορά μεταξύ των εκτιμώμενων πιθανοτήτων του event) και ελέγχουμε κατά πόσον ξεπερνάει κάποιο προ-αποφασισμένο όριο (stopping boundary)

Σχεδιασμοί Μελετών με Ενδιάμεση Παρακολούθηση

- Μηδενική Υπόθεση:

H_0 : Θεραπεία A = Θεραπεία B

- Έστω:

- Z_k η στατιστική συνάρτηση ελέγχου για τη σύγκριση των θεραπειών κατά την διάρκεια της k -όστης ενδιάμεσης ανάλυσης, και
- B_1, B_2, \dots, B_K τα όρια διακοπής βάσει των οποίων οδηγούμαστε σε διακοπή της μελέτης:

-Αν $|Z_k| > B_k$ διακόπτουμε τη μελέτη προς όφελος της θεραπείας A

-Αν $|Z_k| \leq B_k$ συνεχίζουμε στην επόμενη ($k+1$) ανάλυση

□ ...

- Συνεχίζουμε μέχρι το τέλος της μελέτης (K -οστή ανάλυση) - Αν η μελέτη ολοκληρωθεί χωρίς να υπερβούμε τα όρια δεν απορρίπτουμε την μηδενική υπόθεση H_0

Σχεδιασμοί Μελετών με Ενδιάμεση Παρακολούθηση

■ Αρα:

- *Αν $|Z_1| > B_1$, διακόπτουμε τη μελέτη προς όφελος της θεραπείας A (αν $Z_1 < -B_1$) ή της θεραπείας B (αν $Z_1 > B_1$) - διαφορετικά συνεχίζουμε στην 2^η ανάλυση*
- *Αν $|Z_2| > B_2$, διακόπτουμε τη μελέτη προς όφελος της θεραπείας A (αν $Z_2 < -B_2$) ή της θεραπείας B (αν $Z_2 > B_2$) - διαφορετικά συνεχίζουμε στην 3^η ανάλυση*
- ...
- *Συνεχίζουμε μέχρι το τέλος της μελέτης (K-οστή ανάλυση) - Αν η μελέτη ολοκληρωθεί χωρίς να υπερβούμε τα όρια δεν απορρίπτουμε την μηδενική υπόθεση H_0*

Σχεδιασμοί Μελετών με Ενδιάμεση Παρακολούθηση

- Χρειάζεται επαρκές μέγεθος δείγματος για την επίτευξη της επιθυμητής ισχύς για καθορισμένη εναλλακτική υπόθεση καθώς και για την επίτευξη του επιθυμητού σφάλματος τύπου I (συνήθως 5%).
- Το σφάλμα τύπου I σε αυτό το πλαίσιο ορίζεται ως:
 $P(\text{Συμπεραίνουμε διαφορά των θεραπειών} \mid H_0)$.
- Στους ακολουθιακούς σχεδιασμούς, συμπεραίνουμε διαφορά μεταξύ των ομάδων αν
 - $|Z_1| > B_1$ (σταματάμε στην 1^η ενδιάμεση ανάλυση),
 - $|Z_1| < B_1$ αλλά $|Z_2| > B_2$ (σταματάμε στην 2^η ενδιάμεση ανάλυση),
 - ...
 - $|Z_1| < B_1, |Z_1| < B_2, \dots, |Z_{k-1}| < B_{k-1}, |Z_K| > B_K$ (τελική ανάλυση)

Προσδιορισμός των ορίων διακοπής B_1, \dots, B_K

- Θέλουμε να επιλέξουμε B_1, \dots, B_K τέτοια ώστε αν ισχύει η H_0 , η συνολική πιθανότητα να απορρίψουμε την H_0 , είναι ίση με το προ-αποφασισμένο σφάλμα τύπου I, α .

$$P[\text{να απορρίψουμε την } H_0 \text{ σε μια από τις } K \text{ αναλύσεις}] = \alpha$$

- Αν συμβολίσουμε ως

$$\begin{aligned} & P[\text{να απορρίψουμε την } H_0 \text{ στην } k \text{ ανάλυση}] = \\ & = P[|Z_k| > B_k, \text{ δεδομένου ότι } |Z_1| < B_1, |Z_2| < B_2, \dots, |Z_{k-1}| < B_{k-1}] \\ & = \pi_k, \end{aligned}$$

υπάρχουν πολλοί τρόποι επιλογής του π_k

- Καθορίζουν τον ρυθμό με τον οποίο «δαπανούμε» το σφάλμα τύπου I (0.05)

Προσδιορισμός των ορίων διακοπής B_1, \dots, B_K

- Π.χ., ας υποθέσουμε ότι συγκρίνουμε την πιθανότητα ανταπόκρισης μεταξύ 2 θεραπειών και θα διενεργηθούν 4 αναλύσεις (3 ενδιάμεσες και η τελική)
- Ο Pocock ήταν ο πρώτος που ακολούθησε αυτή την προσέγγιση και επέλεξε ίσα π_k
- Ενδεχομένως, η πιο δημοφιλής μέθοδος που χρησιμοποιείται σήμερα είναι η μέθοδος των O'Brien & Fleming:

‘Ξοδεύουν’ πολύ μικρό κομμάτι του σφάλματος τύπου I στις πρώτες ενδιάμεσες αναλύσεις κάνοντας δυσκολότερο να σταματήσει κανείς στην αρχή. Μικρές τροποποιήσεις απαιτούνται στο τέλος

Κρίσιμες Τιμές - Ονομαστικές P values: 4 Αναλύσεις, $\alpha = 0.05$ Αμφίπλευρος Κανόνας

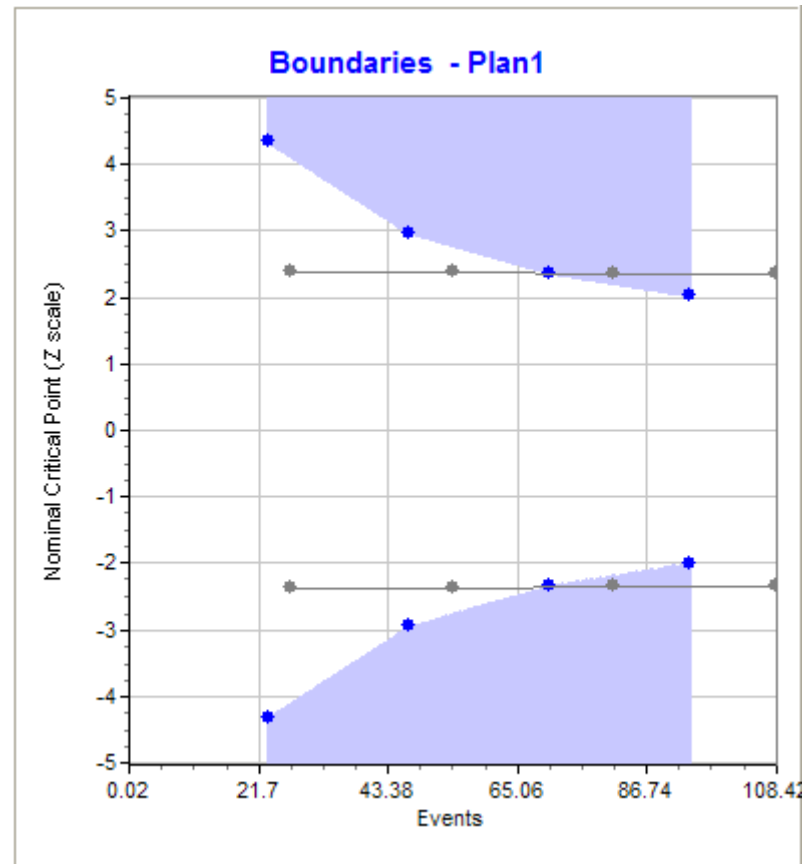
	<u>Look</u>	<u>Pocock</u>	<u>O'Brien-Fleming</u>	
Analysis	CV	P	CV	P
1	2.36	.016	4.08	.000005
2	2.36	.016	3.22	.0013
3	2.36	.016	2.28	.0228
4	2.36	.016	2.04	.0417

Και οι δύο προσεγγίσεις αντιστοιχούν σε συνολική πιθανότητα σφάλματος τύπου I: 5%

1. Η μέθοδος του Pocock 'ξοδεύει' το 5% ομοιόμορφα
2. Η μέθοδος των O-F 'ξοδεύει' λίγο στις αρχικές αναλύσεις και στη συνέχεια διαθέτει ένα μεγάλο μέρος κατά την τελική ανάλυση (look 4)
3. (CV= κριτική τιμή της στατιστικής συνάρτησης ελέγχου Z)

Κρίσιμες Τιμές (Pocock , O'Brien-Fleming)

4 Αναλύσεις, $\alpha = 0.05$ Αμφίπλευρος Κανόνας



Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer

The NEW ENGLAND
JOURNAL of MEDICINE

ESTABLISHED IN 1812

OCTOBER 20, 2005

VOL. 353 NO. 16

Martine J. Piccart-Gebhart, M.D., Ph.D., Marion Procter, M.Sci., Brian Leyland-Jones, M.D., Ph.D., Aron Goldhirsch, M.D., Michael Untch, M.D., Ian Smith, M.D., Luca Gianni, M.D., Jose Baselga, M.D., Richard Bell, M.D., Christian Jackisch, M.D., David Cameron, M.D., Mitch Dowsett, Ph.D., Carlos H. Barrios, M.D., Günther Steger, M.D., Chiun-Shen Huang, M.D., Ph.D., M.P.H., Michael Andersson, M.D., Dr.Med.Sci., Moshe Inbar, M.D., Mikhail Lichinitser, M.D., István Láng, M.D., Ulrike Nitz, M.D., Hiroji Iwata, M.D., Christoph Thomssen, M.D., Caroline Lohrisch, M.D., Thomas M. Suter, M.D., Josef Rüschoff, M.D., Tamás Sütő, M.D., Ph.D., Victoria Greatorex, M.Sc., Carol Ward, M.Sc., Carolyn Straehle, Ph.D., Eleanor McFadden, M.A., M. Stella Dolci, and Richard D. Gelber, Ph.D., for the Herceptin Adjuvant (HERA) Trial Study Team

Results

Data were available for 1694 women randomly assigned to two years of treatment with trastuzumab, 1694 women assigned to one year of trastuzumab, and 1693 women assigned to observation. We report here the results only of treatment with trastuzumab for one year or observation. **At the first planned interim analysis (median follow-up of one year), 347 events (recurrence of breast cancer, contralateral breast cancer, second non breast malignant disease, or death) were observed: 127 events in the trastuzumab group and 220 in the observation group. The unadjusted hazard ratio for an event in the trastuzumab group, as compared with the observation group, was 0.54 (95 percent confidence interval, 0.43 to 0.67; P<0.0001 by the log-rank test, crossing the interim analysis boundary), representing an absolute benefit in terms of disease-free survival at two years of 8.4 percentage points. Overall survival in the two groups was not significantly different (29 deaths with trastuzumab vs. 37 with observation). Severe cardiotoxicity developed in 0.5 percent of the women who were treated with trastuzumab.**

Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer

Martine J. Piccart-Gebhart, M.D., Ph.D., Marion Procter, M.Sci., Brian Leyland-Jones, M.D., Ph.D., Aron Goldhirsch, M.D., Michael Untch, M.D., Ian Smith, M.D., Luca Gianni, M.D., Jose Baselga, M.D., Richard Bell, M.D., Christian Jackisch, M.D., David Cameron, M.D., Mitch Dowsett, Ph.D., Carlos H. Barrios, M.D., Günther Steger, M.D., Chiun-Shen Huang, M.D., Ph.D., M.P.H., Michael Andersson, M.D., Dr.Med.Sci., Moshe Inbar, M.D., Mikhail Lichinitser, M.D., István Láng, M.D., Ulrike Nitz, M.D., Hiroji Iwata, M.D., Christoph Thomssen, M.D., Caroline Lohrisch, M.D., Thomas M. Suter, M.D., Josef Rüschoff, M.D., Tamás Sütő, M.D., Ph.D., Victoria Greatorex, M.Sc., Carol Ward, M.Sc., Carolyn Strahle, Ph.D., Eleanor McFadden, M.A., M. Stella Dolci, and Richard D. Gelber, Ph.D., for the Herceptin Adjuvant (HERA) Trial Study Team

Statistical analysis

Enrollment of 4482 patients was planned to detect a 23 percent relative reduction in the risk of a disease-free-survival event with 80 percent power, with the use of a two-sided significance level of 2.5 percent for each comparison: two years of trastuzumab versus observation and one year of trastuzumab versus observation. A total of 951 disease-free-survival events were required for the final analysis. One interim efficacy analysis was planned after 475 events, with a specified significance level of $P \leq 0.001$ required, with the use of a sequential plan according to the O'Brien-Fleming boundary as implemented by Lan and DeMets. The independent data monitoring committee reviewed data on patient enrollment, deaths, compliance, and safety every six months and conducted the interim cardiac safety and efficacy reviews as preplanned. The efficacy analyses were conducted according to the intention-to-treat principle. Chi-square tests for categorical data and log-rank tests for time-to-event end points provided two-sided P values. Kaplan-Meier curves are presented. Cox proportional hazards regression analysis was used to estimate hazard ratios and 95 percent confidence intervals.

Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer

Martine J. Piccart-Gebhart, M.D., Ph.D., Marion Procter, M.Sci., Brian Leyland-Jones, M.D., Ph.D., Aron Goldhirsch, M.D., Michael Untch, M.D., Ian Smith, M.D., Luca Gianni, M.D., Jose Baselga, M.D., Richard Bell, M.D., Christian Jackisch, M.D., David Cameron, M.D., Mitch Dowsett, Ph.D., Carlos H. Barrios, M.D., Günther Steger, M.D., Chiun-Shen Huang, M.D., Ph.D., M.P.H., Michael Andersson, M.D., Dr.Med.Sci., Moshe Inbar, M.D., Mikhail Lichinitser, M.D., István Láng, M.D., Ulrike Nitz, M.D., Hiroji Iwata, M.D., Christoph Thomssen, M.D., Caroline Lohrisch, M.D., Thomas M. Suter, M.D., Josef Rüschoff, M.D., Tamás Sütő, M.D., Ph.D., Victoria Greatorex, M.Sc., Carol Ward, M.Sc., Carolyn Strahle, Ph.D., Eleanor McFadden, M.A., M. Stella Dolci, and Richard D. Gelber, Ph.D., for the Herceptin Adjuvant (HERA) Trial Study Team

Statistical analysis

Enrollment of **4482 patients was planned to detect a 23 percent relative reduction in the risk of a disease-free-survival event with 80 percent power**, with the use of a **two-sided significance level of 2.5 percent for each comparison**: two years of trastuzumab versus observation and one year of trastuzumab versus observation. A total of 951 disease-free-survival events were required for the final analysis. One interim efficacy analysis was planned after 475 events, **with a specified significance level of $P \leq 0.001$** required, with the use of a sequential plan according to the O'Brien-Fleming boundary as implemented by Lan and DeMets. The independent data monitoring committee reviewed data on patient enrollment, deaths, compliance, and safety every six months and conducted the interim cardiac safety and efficacy reviews as preplanned. The efficacy analyses were conducted according to the intention-to-treat principle. **Chi-square tests for categorical data and log-rank tests for time-to event end points provided two-sided P values. Kaplan-Meier curves are presented. Cox proportional hazards regression analysis was used to estimate hazard ratios and 95 percent confidence intervals.**

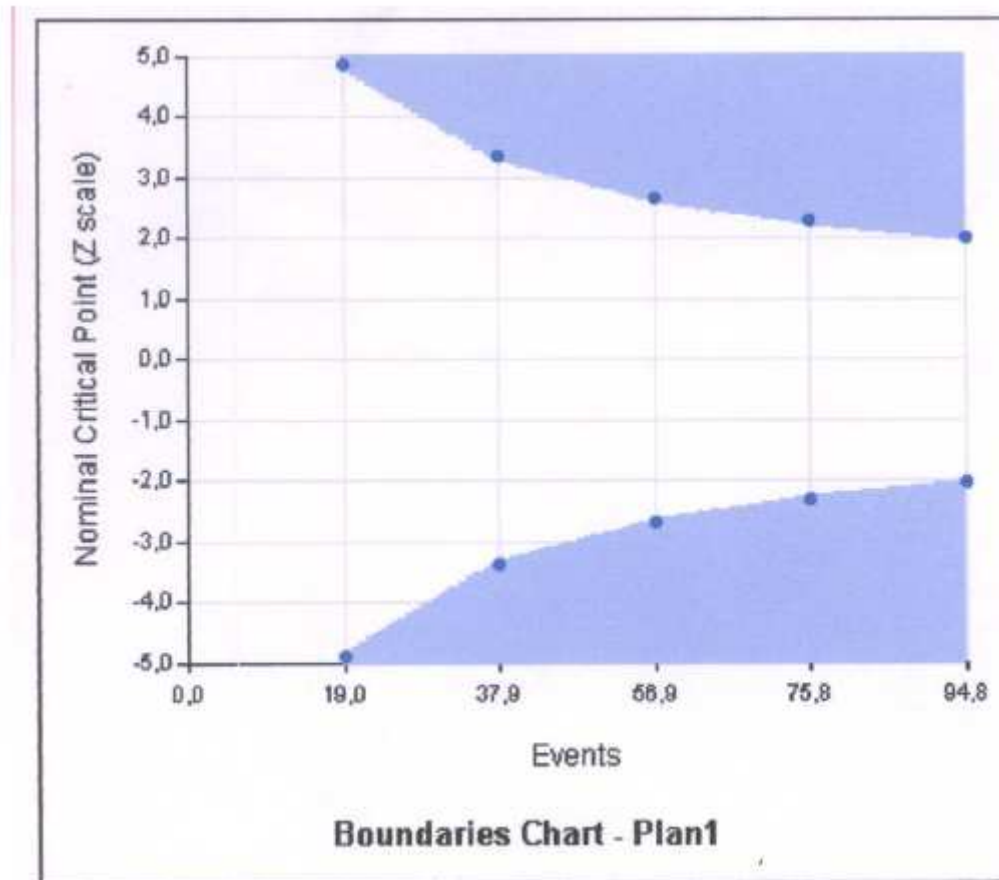
Επιλέγοντας κανόνα

- Και οι δύο δίνουν την ίδια συνολική πιθανότητα σφάλματος τύπου I
- Οι κανόνες του Pocock **δεν χρησιμοποιούνται συχνά** γιατί είναι πιο ευαίσθητοι στην αρχή όταν η ισχύς είναι μικρή
- Οι κανόνες των O'Brien-Fleming χρησιμοποιούνται συχνότερα. Επιτρέπουν την πρόωρη διακοπή αλλά απαιτούν ισχυρή απόδειξη
 - Δεν απαιτείται επομένως μεγάλη διόρθωση του α στο τέλος της μελέτης
- Φυσικά υπάρχουν και ενδιάμεσες επιλογές

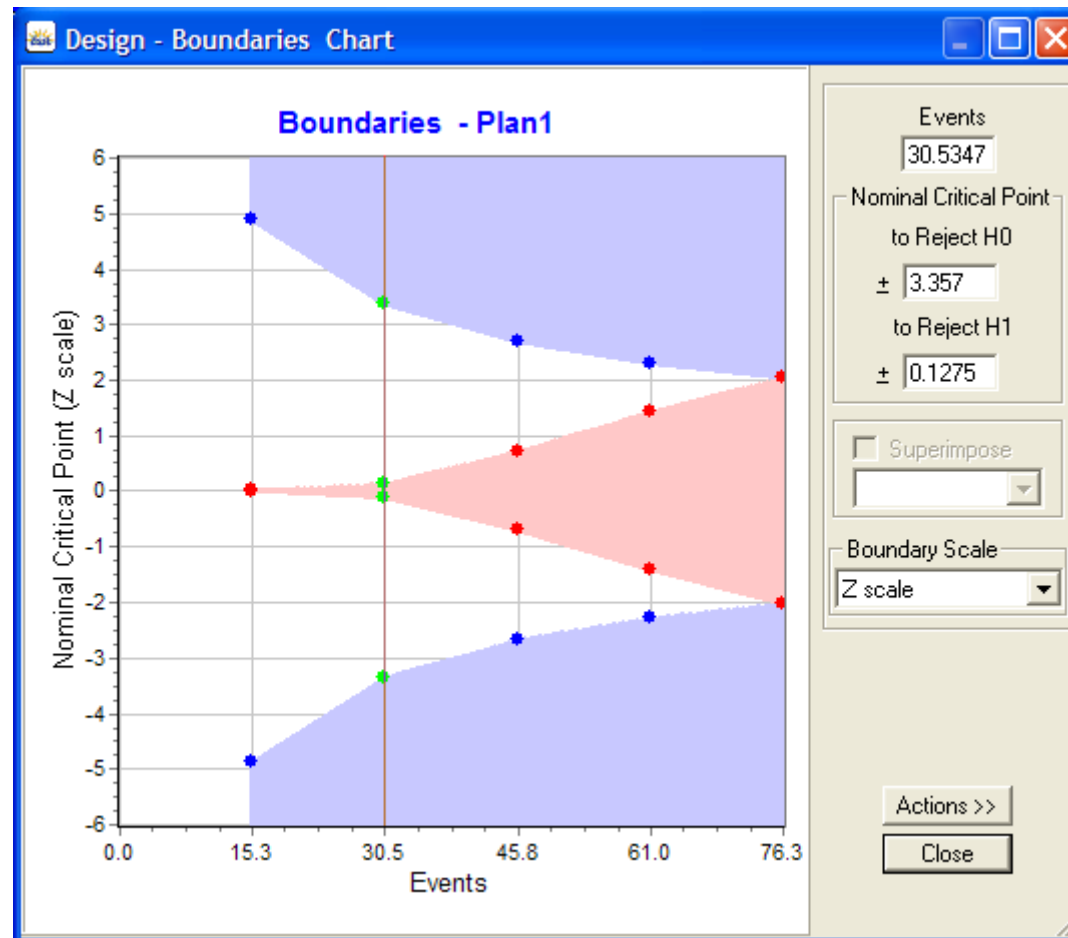
Άλλες δυνατότητες

- Οι χρονικές στιγμές των ενδιάμεσων αναλύσεων δεν χρειάζεται να ισαπέχουν όσον αφορά τον αριθμό των ασθενών
- Συνήθως έχουμε κανόνες διακοπής για την απόρριψη της H_0 , αλλά είναι δυνατόν να έχουμε ένα σχεδιασμό για την απόρριψη της H_0 ή κάποιας H_1
- Είναι δυνατόν να υιοθετήσουμε σχεδιασμό που να επιτρέπει τον πρόωρο τερματισμό για λόγους «ματαιότητας» - “futility”, δηλαδή όταν η πιθανότητα να δούμε σημαντική διαφορά εάν συνεχιστεί η μελέτη έως την προγραμματισμένη ολοκλήρωσή της είναι αρκετά μικρή

Ο'Brien-Fleming όρια διακοπής H_0



Ο'Brien-Fleming όρια διακοπής H_0 / H_1



“Futility” Κανόνες Διακοπής

- Ας υποθέσουμε ότι σε μια μελέτη σύγκρισης μιας νέας θεραπείας με την καθιερωμένη θεραπεία, η νέα θεραπεία φαίνεται να είναι υποδεέστερη. Ακόμα και αν δεν είναι στατιστικά σημαντικά χειρότερη ίσως να είναι πολύ απίθανο να δούμε ότι η νέα θεραπεία είναι καλύτερη (H_1)

- Παράδειγμα: εξετάζουμε αν η νέα θεραπεία βελτιώνει την επιβίωση σε ασθενείς με καρκίνο του πνεύμονα. Στη 2η από τις 3 προκαθορισμένες αναλύσεις η διάμεση επιβίωση είναι:

καθιερωμένη θεραπεία : 9.5 μήνες

νέα θεραπεία: 7.1 μήνες

Μήπως είναι απίθανο να δούμε ότι η νέα θεραπεία είναι καλύτερη κατά την τελική ανάλυση; Ποιο το όφελος να συνεχίσουμε μια μελέτη αν η νέα θεραπεία θα αντικαθιστούσε την καθιερωμένη θεραπεία μόνο αν βελτίωνε την επιβίωση;

“Futility” Κανόνες Διακοπής

- Προσοχή! Αν σταματήσουμε δεν ισχυριζόμαστε ότι η νέα θεραπεία είναι καλύτερη ή ισοδύναμη, οπότε δεν χρειάζεται να ανησυχούμε για το σφάλμα τύπου I και II. Παρόλα αυτά δεν θέλουμε να σταματήσουμε αν υπάρχει αρκετή πιθανότητα να δούμε διαφορά στην τελική ανάλυση
- Δεν υπάρχει τρόπος να τυποποιήσουμε στατιστικά αυτό το σκεπτικό. Μια λύση είναι να υπολογίσουμε:

P [η μελέτη να δείξει ένα στατιστικά σημαντικό αποτέλεσμα υπέρ της νέας θεραπείας, με βάση τα μέχρι τότε δεδομένα και κάτω από την υπόθεση ότι το πραγματικό αποτέλεσμα της νέας θεραπείας είναι σύμφωνα με τις υποθέσεις του σχεδιασμού]

Αν αυτή η πιθανότητα είναι πολύ μικρή (π.χ. <10%) τότε υπάρχουν βάσιμες ενδείξεις για πρόωρο τερματισμό λόγω ‘futility’

Παράδειγμα 1: πρόωρη διακοπή μιας συγκριτικής μελέτης

- Πρόωρη διακοπή (αποτελεσματικότητα) κλινικής δοκιμής σε ενδιάμεση ανάλυση Hammer et al (NEJM, 1997)
- Πλαίσιο: Πριν από 10 χρόνια στους HIV ασθενείς χορηγούνταν συνήθως 2 ‘reverse transcriptase inhibitor’ φάρμακα (AZT, 3TC, ddI, ddC). Όμως ένας νέος πολλά υποσχόμενος τύπος φαρμάκου ‘protease inhibitor’ είχε πρόσφατα αναπτυχθεί. Ο σκοπός της μελέτης ήταν να εξετάσει αν η προσθήκη του inhibitor (indinavir) σε 2 reverse transcriptase φάρμακα θα βελτίωνε την αποτελεσματικότητα.
- Το κοινό και τα MME περίμεναν με ανυπομονησία τα αποτελέσματα. Αυτό άσκησε πίεση στους ερευνητές και τη DSMB

Παράδειγμα 1 (συνέχεια)

- Σχεδιασμός:
 - Ασθενείς με HIV
 - Πρωτεύων σκοπός: χρόνος ως το AIDS ή το θάνατο (όποιο εμφανιστεί πρώτο)
 - Οι ασθενείς τυχαιοποιούνταν σε :
 - AZT+3TC (standard θεραπεία)
 - AZT+3TC+Indinavir (νέα θεραπεία)
 - Μελέτη υπεροχής: είναι το AZT+3TC+IDV καλύτερο από το AZT+3TC;
 - Μέγεθος δείγματος: $n=1156$
 - 2 προγραμματισμένες interim αναλύσεις (O'Brien-Fleming spending function)

Παράδειγμα 1: Αποτελέσματα interim ανάλυσης

- 1η Interim ανάλυση: τα στοιχεία ήταν υπέρ του AZT+3TC+IDV, αλλά το «stopping boundary» δεν ξεπεράστηκε.

=> συνέχιση μελέτης.

- 2η Interim ανάλυση :

	<u>n</u>	<u># θανάτων ή είχαν αναπτύξει AIDS</u>
<u>AZT+3TC:</u>	579	63
<u>AZT+3TC+IDV:</u>	577	33

Εκτιμώμενο relative risk (95% CI): 0.5 (0.33-0.76)

Έλεγχος της H_0 : $p=0.001$

Τα υπόλοιπα study endpoints ήταν επίσης υπέρ του AZT+3TC+IDV

=> Η μελέτη σταμάτησε.

Παράδειγμα 1: Ζητήματα

- Κατά την 1η Interim ανάλυση ασκήθηκε πίεση από το κοινό και τα ΜΜΕ να σταματήσει η μελέτη καθώς στοιχεία από μελέτες Φάσης I και II έδειχναν υποσχόμενα αποτελέσματα την ώρα που ασθενείς που λάμβαναν reverse transcriptase φάρμακα ανέπτυσσαν AIDS και πέθαιναν
- Επομένως, κατά την 1η Interim ανάλυση όταν υπήρχαν ενδείξεις υπέρ του AZT+3TC+IDV η DSMB σίγουρα ένοιωσε την πίεση. Ευτυχώς όμως δεν υποχώρησε και έτσι κατέστη δυνατόν να ποσοτικοποιηθεί το όφελος και να υπάρξουν αδιαμφισβήτητα αποτελέσματα

Παράδειγμα 2: Πρόωρη διακοπή για futility

- Reference: Hall et al, NEJM, 1998
- Πλαίσιο: Η progressive multifocal leukoencephalopathy (PML) είναι μια σοβαρή νευρολογική κατάσταση ασθενών με HIV, και η επιβίωση μετά τη διάγνωση της PML είναι πολύ σύντομη (~ 3 μήνες). Σύμφωνα με κάποιες ανέκδοτες εκθέσεις το φάρμακο cytarabine θα βοηθούσε, μόνο που είχε σοβαρά side effects.
- Η μελέτη: HIV ασθενείς με PML τυχαιοποιούνταν σε :
 - Antiviral θεραπεία (control group) versus
 - Antiviral θεραπεία +cytarabine (IV)
 - Antiviral θεραπεία +cytarabine (intrathecal)
- Πρωτεύων στόχος: επιβίωση
- Μέγεθος δείγματος : 90 (30 ανά σκέλος)

Παράδειγμα 2 (συνέχεια)

- 1η Interim ανάλυση : 18 μήνες από την έναρξη. Η DSMB πρότεινε τη συνέχιση
- 2η Interim ανάλυση : Όταν 50% του αναμενόμενου αριθμού θανάτων είχε εμφανιστεί. Μέχρι τότε, n=57 από το σύνολο των 90 ασθενών είχε εισαχθεί

	<u>n</u>	<u>#θάνατοι</u>	
❑ Antivirals (control group)	18	14	
❑ Antivirals + IV cytarabine	20	14	(p=.85)
❑ Antivirals +intra cytarabine	19	14	

- Περισσότερα *side effects* στους ασθενείς που λάμβαναν cytarabine
- Η DSMB αναρωτήθηκε αν η συνέχιση της μελέτης θα είχε αρκετή πιθανότητα να δείξει υπεροχή των ομάδων cytarabine στην επιβίωση με επιπλέον 33 ασθενείς.

Παράδειγμα 2 (συνέχεια)

- Ο στατιστικός της μελέτης υπολόγισε:

P[η cytarabine, με δεδομένα τα αποτελέσματα για τους 57 ασθενείς, να δείξει στατιστικά σημαντική υπεροχή στην επιβίωση]

- Η δεσμευμένη πιθανότητα (με βάσει τα μέχρι τότε στοιχεία) υπολογίστηκε κάτω από πολλές υποθέσεις σε σχέση με την αληθινή συνάρτηση πιθανοτήτων επιβίωσης.
- Πρόκειται για μια ανάλυση σκοπιμότητας. Αν συνεχίσουμε τη μελέτη θα μπορέσουμε να παρατηρήσουμε υπεροχή του cytarabine αν αυτή όντως υφίσταται;

Παράδειγμα 2 (συνέχεια)

- Ανάλογα με τις υποθέσεις των πιθανοθεωρητικών μοντέλων οι εκτιμώμενες πιθανότητες ήταν μεταξύ 0.4% -15%.
- Επομένως υπήρχε πολύ μικρή πιθανότητα να δουν διαφορά υπέρ του cytarabine με τη συνέχιση της μελέτης
 - => Η DSMB πρότεινε τον τερματισμό για ‘futility’
- ‘Futility’ δε σημαίνει ότι η cytarabine είναι ή δεν είναι καλύτερη από το control. Απλώς η συνέχιση της μελέτης είναι μάταιη



Εξαιρέσεις Παράδειγμα Αξιολόγησης

Ουρανία Δαφνή
Καθηγήτρια, Διευθύντρια
Εργαστήριο Βιοστατιστικής
Πανεπιστήμιο Αθηνών

Διευθύντρια
Frontier Science Foundation-Hellas (FSF-H)

Παράδειγμα Αξιολόγησης

A Low-Carbohydrate as Compared with a Low-Fat Diet in Severe Obesity

Frederick F. Samaha, M.D., Nayyar Iqbal, M.D., Prakash Seshadri, M.D.,
Kathryn L. Chicano, C.R.N.P., Denise A. Daily, R.D., Joyce McGrory, C.R.N.P.,
Terrence Williams, B.S., Monica Williams, B.S., Edward J. Gracely, Ph.D.,
and Linda Stern, M.D.

N Engl J Med 2003;348:2074-81.

Επίδραση «μη-συμμόρφωσης» και «απόσυρσης»

- Στόχος της Κλινικής Δοκιμής : η σύγκριση μίας Δίαιτας με χαμηλούς υδατάνθρακες (low-carbohydrate) και μίας Δίαιτας χαμηλού λίπους (low-fat) όσον αφορά στην απώλεια βάρους σε σοβαρά υπέρβαρα άτομα
- Σχεδιασμός:
 - Πληθυσμός: Σοβαρά υπέρβαρα άτομα (μέσο βάρος=131 kg)
 - Τυχαιοποίηση 132 ατόμων σε δίαιτα Low-Carb έναντι Low-Fat για 6 μήνες. Και στις δύο ομάδες ήταν διαθέσιμη Συμβουλευτική Υποστήριξη.
 - Κύρια Έκβαση: Απώλεια Βάρους μετά από 6 μήνες.

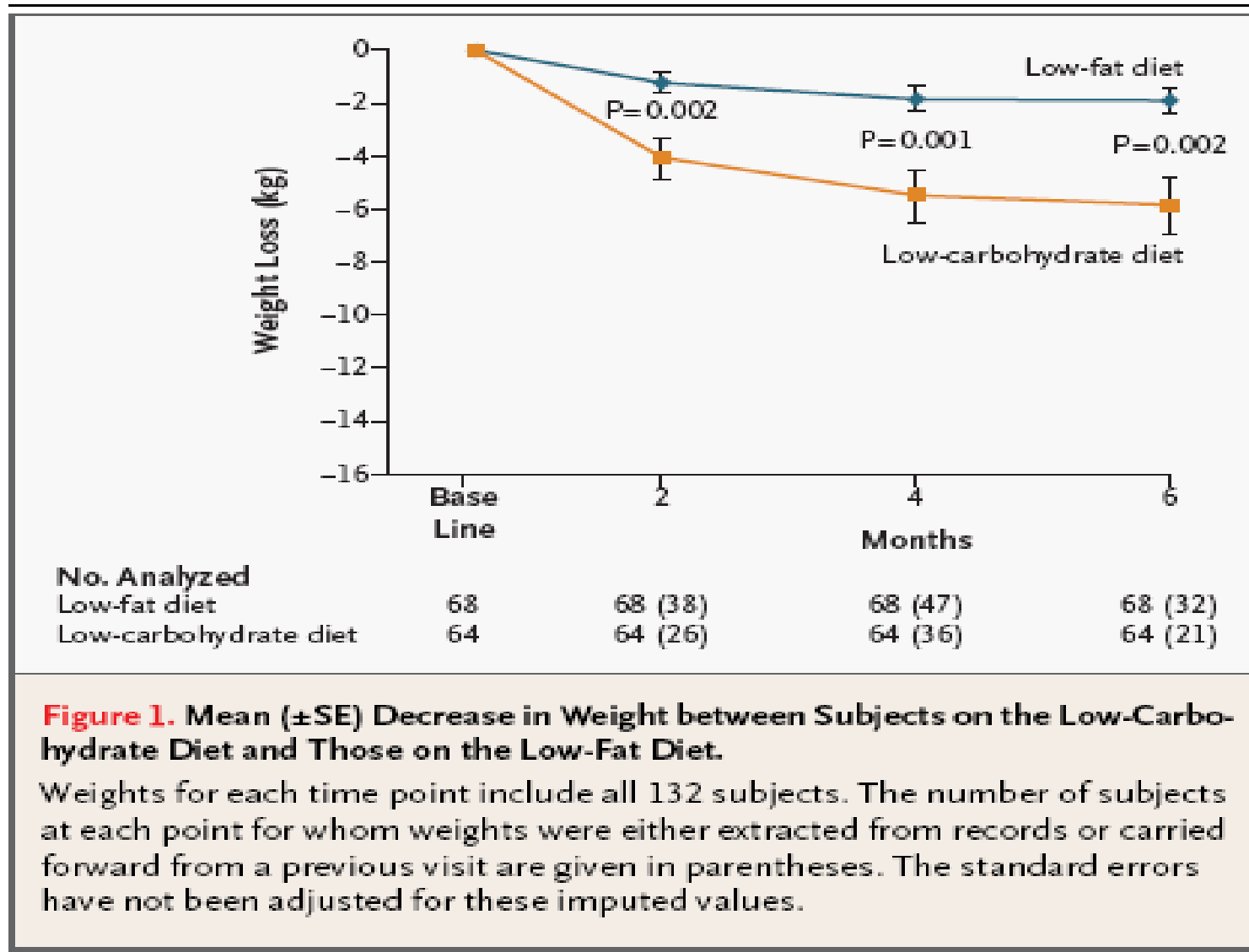
■ Απώλειες/Αποσύρσεις Ατόμων κατά τη διάρκεια της Μελέτης (Dropouts)

	<u>Low-Carb</u>	<u>Low-Fat</u>
# τυχαιοποιημένων	64	68
Dropouts:		
έως τον 1 μήνα:	25%	38%
έως τους 3 μήνες:	27%	44%
έως τους 6 μήνες:	33%	47%
# ολοκλήρωσαν 6 μήνες:	43/64	36/68

Πώς αναλύονται τα δεδομένα σε αυτή την περίπτωση;

- Για 53 άτομα δεν υπάρχουν τα βάρη στους 6 μήνες.
- Σύγκριση Απώλειας Βάρους μεταξύ των ατόμων που ολοκλήρωσαν τους 6 μήνες στη μελέτη;

- Τι χρησιμοποιήθηκε για «βάρος στους 6 μήνες»
 - Βάρος στους 6 μήνες όπως μετρήθηκε στους 79 που ολοκλήρωσαν τη μελέτη
 - Εκτιμήσεις βάρους εκτός μελέτης για 29 άτομα (external sources)
 - Για τους υπόλοιπους 24 ασθενείς, χρησιμοποιήθηκε το τελευταίο διαθέσιμο βάρος τους στη μελέτη πριν την απόσυρσή τους (πχ, βάρος στον 1 μήνα), (LOCF, “last observation carried forward”).
- Αποτελέσματα αυτής της Ανάλυσης:
 - Low Carb: 5.8 kg απώλεια βάρους κατά μέσο όρο
 - Low Fat: 1.9 kg απώλεια βάρους κατά μέσο όρο (p=0.002)



-
- Προβλήματα με αυτή την Ανάλυση;
 - Οι απώλειες είναι πιο πιθανό άτομα που δεν ακολούθησαν τη Δίαιτα. Η χρήση του πιο πρόσφατου διαθέσιμου βάρους τους μπορεί να εισαγάγει μεροληψία στα αποτελέσματα (τα δεδομένα στους 6 μήνες να φαίνονται καλύτερα από ό,τι είναι). Δεν γνωρίζουμε αν ισχύει αυτό γιατί δεν έχουμε τα πραγματικά δεδομένα

 - Ακόμη και αν η p-value (0.002) ήταν έγκυρη, είναι αυτή η διαφορά ‘κλινικά σημαντική’?
 - 5.8 έναντι 1.9 kg διαφορά σε ασθενείς που το μέσο βάρος τους ήταν 131 kg στην αρχή της μελέτης;

Προβλήματα Σχεδιασμού:

- Διάρκεια Μελέτης:

Είναι οι 6 μήνες αρκετός χρόνος για να δεί κανείς εάν μία δίαιτα όντως δουλεύει; Το μεγαλύτερο πρόβλημα μιας δίαιτας είναι να παραμείνει ο ασθενής στη δίαιτα για μεγάλο χρονικό διάστημα.

- Τελική Έκβαση:

Είναι η μέση απώλεια βάρους ένα κατάλληλο μέτρο αποτελεσματικότητας; Δεν μας δείχνει πόσα άτομα πέτυχαν μία «ικανοποιητική» απώλεια βάρους.

- Εναλλακτική Τελική Έκβαση:

Απώλεια της τάξης του 10% του βάρους ενός ατόμου

- Παρακολούθηση(Follow-Up):

- Η μελέτη δεν είχε σχεδιαστεί για να μετράται το βάρος στους 6 μήνες και σε ασθενείς που έχουν αποχωρήσει.

- Κάποιες φορές, κανείς δεν μπορεί να αποφύγει τέτοιου είδους απώλειες αλλά μπορεί να σχεδιάσει τη μελέτη ούτως ώστε να μετρά το βάρος στους 6 μήνες ακόμη και σε αυτούς που αποχωρούν νωρίτερα.

ΣΥΜΠΕΡΑΣΜΑ

- ❑ Με τόσο μικρού μεγέθους απώλεια βάρους και τόσο μεγάλο αριθμό ασθενών που δεν παρέμειναν στην αντίστοιχη θεραπεία, ίσως το πραγματικό συμπέρασμα είναι ότι καμία από τις δύο δίαιτες δεν είναι ικανοποιητικά αποτελεσματική

ΑΠΟΤΕΛΕΣΜΑ

- ❑ Μη ξεκάθαρο Μήνυμα για μεταφορά στην Κλινική Πράξη

Επιφυλάξεις

- Μη σαφής παρουσίαση πληροφορίας για όλα τα στάδια της μελέτης
- ITT ανάλυση; Εξαιρέσεις
- Πολλαπλοί έλεγχοι/στατιστικές μέθοδοι
- Συμπεράσματα που βασίζονται σε υπο-ομάδες ασθενών
- Γενίκευση σε πληθυσμό που δεν αντιπροσωπεύεται πλήρως στο δείγμα (Ήταν οι ασθενείς υπό μελέτη παρόμοιοι με τους δικούς σας;)

Συμπεράσματα

- Η Ποιότητα μιας Κλινικής Δοκιμής εξαρτάται από τη σχεδίαση, τον τρόπο διεξαγωγής της, την ανάλυσή της και τη δυνατότητα γενίκευσης των αποτελεσμάτων της.
- Η Δημοσίευση πρέπει:
 - να μεταδίδει ξεκάθαρα την πληροφορία για όλα αυτά τα στάδια της μελέτης
 - **χωρίς ενδοιασμό** να αναφέρεται σε όλα τα πιθανά προβλήματα σε κάθε ένα από αυτά τα στάδια
 - Να αναφέρει τους περιορισμούς που αυτά επιφέρουν στην ερμηνεία και τα συμπεράσματά της.

Κριτική Ανάγνωση Δημοσιεύσεων Κλινικών Δοκιμών

■ Περίληψη

- Είναι περιεκτική η περίληψη όσον αφορά το σχεδιασμό της μελέτης και των κυρίων αποτελεσμάτων;
- Κάθε αναγραφόμενη τιμή περιέχει κάποιο μέτρο διασποράς (αν μη τι άλλο, το μέγεθος δείγματος);

■ Η εισαγωγή περιέχει

- Τη λογική του Σχεδιασμού της Μελέτης;
- Τους Σκοπούς της Μελέτης;

Κριτική Ανάγνωση Δημοσιεύσεων Κλινικών Δοκιμών

- Η Περιγραφή των Μεθόδων περιλαμβάνει
 - Ορισμό του πληθυσμού υπό μελέτη;
 - Ορισμό της θεραπείας και των τροποποιήσεων δοσολογίας (μία περίληψη είναι χρήσιμη);
 - Ορισμό διαγνωστικών κριτηρίων και μέτρων τελικής έκβασης (τυποποίηση);
 - Τεχνικές ανάλυσης;
 - Μηχανισμοί ελέγχου ποιότητας δεδομένων (εξωτερική επιθεώρησή;)

Κριτική Ανάγνωση Δημοσιεύσεων Κλινικών Δοκιμών

- Τα Αποτελέσματα περιλαμβάνουν:
 - Καταμέτρηση ασθενών;
 - Χαρακτηριστικά πληθυσμού ασθενών;
 - Θεραπεία που χορηγήθηκε(σε αντίθεση με αυτήν που είχε σχεδιαστεί);
 - Ελάχιστοι, μέγιστοι και μέσοι χρόνοι παρακολούθησης ασθενών;
 - Αν το πρωτόκολλο διεκόπη και γιατί;
 - Ανάλυση των δεδομένων όπως η μελέτη είχε σχεδιασθεί και εκτελεσθεί;

Κριτική Ανάγνωση Δημοσιεύσεων Κλινικών Δοκιμών

■ Αποτελέσματα

- Η ανάλυση έλαβε υπόψη της όλους τους γνωστούς προγνωστικούς παράγοντες (διότι οι προγνωστικοί παράγοντες συχνά κυριαρχούν των θεραπευτικών διαφορών);
- Η ανάλυση λαμβάνει υπόψη της πολλαπλές ενδιάμεσες αναλύσεις και υποομάδες (μήπως ψάχνουν υπερβολικά για διαφορές;)
- Παρουσιάζει όσο το δυνατόν περισσότερα δεδομένα;

Κριτική Ανάγνωση Δημοσιεύσεων Κλινικών Δοκιμών

■ Συζήτηση

- Παρουσίαση λογικών επιχειρημάτων, συμπερασμάτων και συνεπειών;
- Συμπεριλαμβάνονται και τα θετικά σημεία και οι αδυναμίες της μελέτης;
- Υπάρχει διαχωρισμός μεταξύ επισήμων συγκρίσεων και διερευνητικών αναλύσεων;

*Πώς να διαβάσουμε κλινικά περιοδικά:
Διαχωρισμός χρήσιμων από άχρηστες ή ακόμα και ζημιογόνες
θεραπείες (Sackett, D.L.)*

- Οι ασθενείς διαμοιράστηκαν στις θεραπείες τυχαία;
- Παρουσιάσθηκαν όλα τα σχετικά αποτελέσματα;
- Ήταν οι ασθενείς υπό μελέτη παρόμοιοι με τους δικούς σας;
- Εξετάστηκαν και στατιστικά και κλινικά κριτήρια;
- Είναι η θεραπευτική αγωγή δυνατή στην δική σας πραγματικότητα;
- Συμπεριελήφθησαν όλοι οι ασθενείς για την εξαγωγή των συμπερασμάτων της μελέτης;

Canadian Med Assoc J, May 1, 1981, Vol. 124, pp. 1156-62

*Κατευθυντήριες οδηγίες δημοσίευσης κλινικών δοκιμών
ογκολογίας: Ευθύνες των εκδοτών και ερευνητών (Zelen M.)*

- Πληθυσμός ασθενών
- Θεραπεία (που έχει δοθεί στην πραγματικότητα)
- Σχεδιασμός της μελέτης
- Καταμέτρηση ασθενών (registration system)
- Παρακολούθηση (follow-up)
- Ποιότητα Δεδομένων
- Τελικά σημεία και περικοπή παρατηρήσεων(endpoints)
- Μεροληψία μη τυχαιοποίησης της μελέτης
- Στατιστικοί μέθοδοι και Ισχύς
- Διακοπή της μελέτης
- Προκαταρκτικά έναντι τελικών αποτελεσμάτων

J Clin Onc, Vol. 1, No. 2, February, 1983
