ESMO
GOOD SCIENCE
BETTER MEDICINE
BEST PRACTICE

UNDERSTANDING CLINICAL TRIAL STATISTICS

Prepared by Urania Dafni, Xanthi Pedeli, Zoi Tsourti

DISCLOSURES

- Urania Dafni has reported no conflict of interest
- Xanthi Pedeli has reported no conflict of interest
- Zoi Tsourti has reported no conflict of interest

KEY POINTS

- Randomization – Stratification
- Superiority *vs.* non-inferiority
- Stopping boundaries
- Planned *vs.* post-hoc analysis
- Kaplan-Meier plots, medians
- Forest plots
- Waterfall plots

STUDY DESIGN

- Experiments answer a scientific question by isolating the intervention and the outcome from extraneous influences
- *What are the goals?*
  - Eliminate systematic error (**Bias**)
    - any effect rendering the observed results not representative of the treatment effect
  - Minimize random error (**Precision**)
    - inaccuracy of results due to sampling
  - Ensure the **generalizability** of study results
- *Study Design is the methodology for achieving these goals*
  - eliminate bias ➲ **randomization**
    - ➲ and stratification, blinding, choice of design
  - minimize random error ➲ establish a sample size sufficient to achieve study goals

RANDOMIZATION

Fundamental Principle in comparing interventions:

Groups must be alike in all important aspects and only differ in the intervention each group receives.

- *Randomization:*

Each patient has the same chance of receiving any of the study treatments.

Benefits

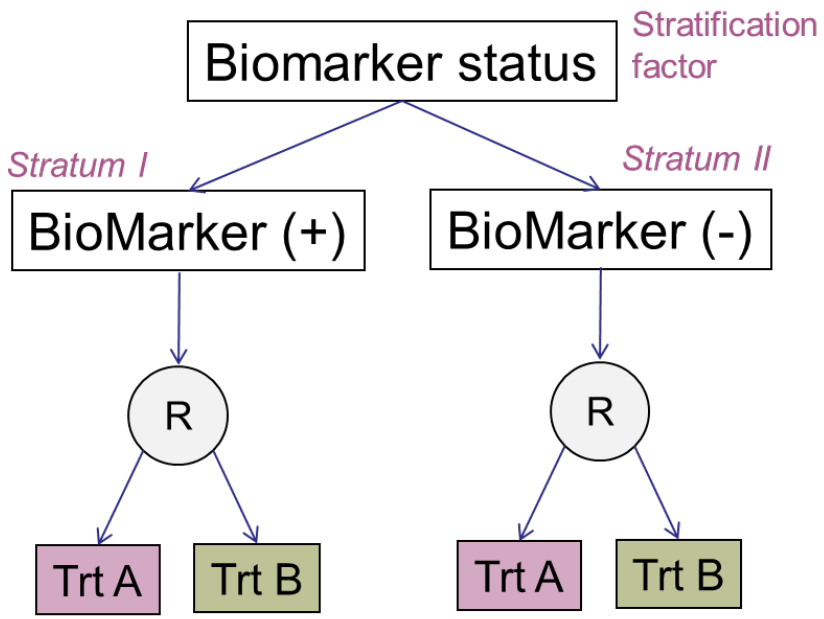
- Comparability is achieved
- Eliminates systematic bias
- Balances both known and unknown factors
- Randomized groups are "similar on the average"

Limitations

- "Similar on average" does not guarantee balanced groups
- Ethical Issues in some cases (usually bad designs)
- Interference with doctor/patient relationship
- Administrative Complexity

STRATIFICATION IN RANDOMIZED TRIALS

- Overall cohort of patients is partitioned in homogeneous subgroups (strata)
- Patients are randomized to treatment arms within each stratum



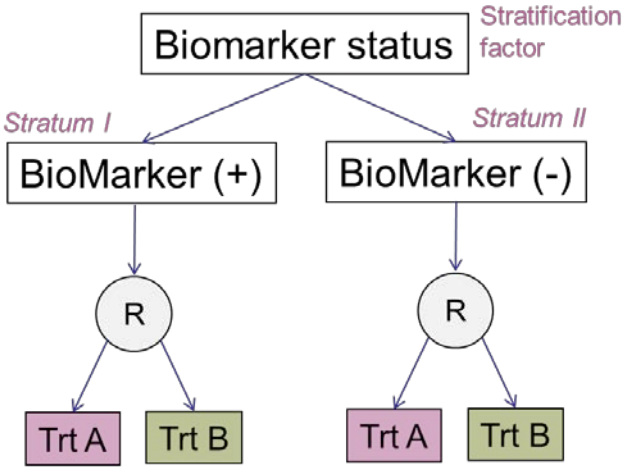
Benefits

- Equal allocation of strata of patients to each treatment arm
- Reduction of random error (variability of effect estimates)

Limitations

- Requires prior knowledge for possible stratification factors
- Too many stratification factors lead to the opposite result (imbalance instead of balance)

STRATIFICATION IN RANDOMIZED TRIALS



- *Possible stratification factors:*

Prognostic or predictive factors, *e.g.*
biomarker status,
previous treatment,
patient's baseline characteristics

- ***Implications for analysis:***

  Examine (stratification factor x treatment effect) interaction

  - ***Significant Interaction*** ➲ **predictive factor**

    Separate evaluation of the treatment effect is performed in each stratum

  - ***Non-significant Interaction*** ➲ **prognostic factor**:

    Comparison of treatments can take place in the overall population (adjusting for the stratification factor)

SUPERIORITY *VS*. NON-INFERIORITY TRIALS

Superiority trial

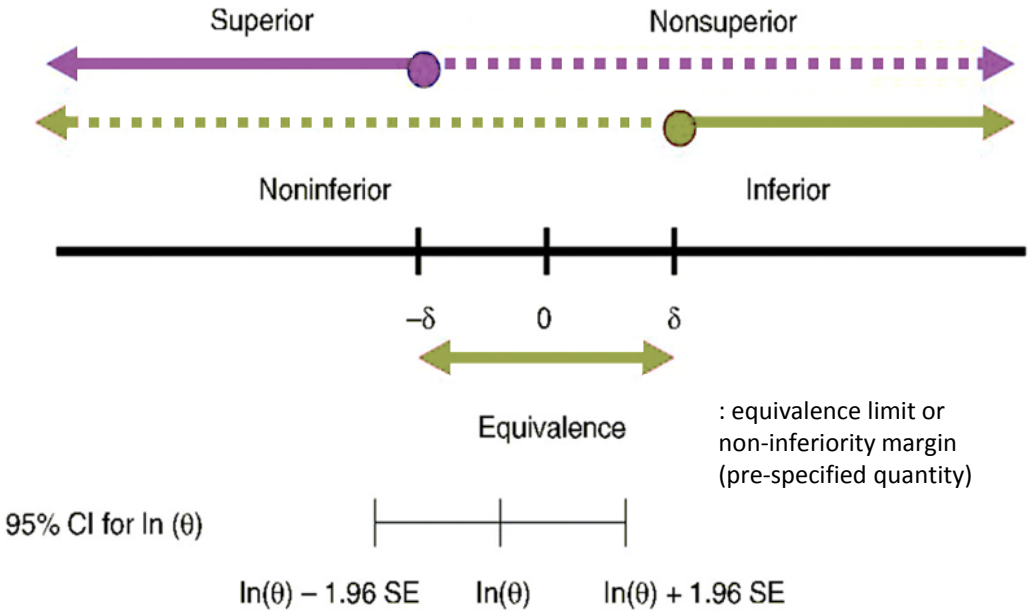
"Is the new treatment better than the standard one?"

H_0:

"No effect" or "no difference" in the clinical effect of the two treatments

Reject H_0

Prove superiority



: equivalence limit or non-inferiority margin (pre-specified quantity)

PROOF OF EQUIVALENCE

should not be confused with

FAILURE TO REJECT the null hypothesis in a superiority trial

Equivalence *or (Non-inferiority) trial*

"Is the new treatment as good as the standard one?"

H_0:

"Different effect" or "difference" in the clinical effect of the two treatments

Reject H_0

Prove **equivalence** or **non-inferiority**

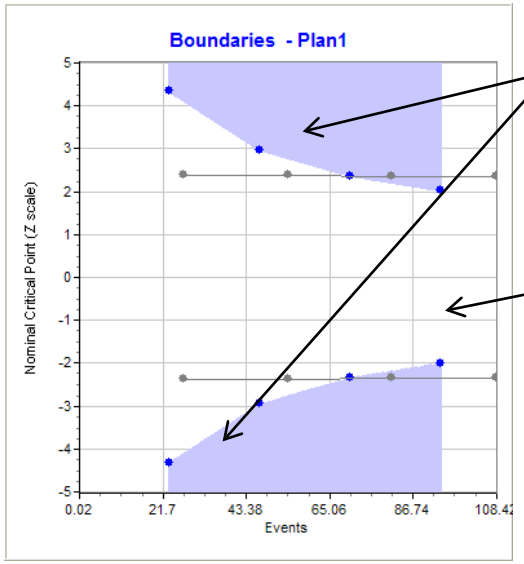
Lesaffre E, Superiority, Equivalence and Non-Inferiority Trials. Bulletin of the NYU Hospital for Joint Diseases, 2008; 66(2):150-4.

Zee BCY, Planned Equivalence or Noninferiority Trials Versus Unplanned Noninferiority Claims: Are they equal? J Clin Oncol, 2006; 24: 1026-1028, Reprinted with permission ©2006 American Society of Clinical Oncology, All rights reserved

DETERMINING THE STOPPING BOUNDARIES

- The Group Sequential Approach: Repeated significance testing
- Aim: Ability to stop the trial earlier with statistically significant conclusions without increasing the type I and type II errors
- Want to choose boundaries at different time points for interim analyses while keeping the overall desired type I and type II errors

Pocock / O'Brien-Fleming boundaries H_0 / H_1



Analysis	Pocock	O'Brien -Fleming
1.	0.16	0.000005
2.	0.16	0.0013
3.	0.16	0.0228
4.	0.16	0.0417

Critical values – nominal p-values:
4 analyses, α= 0.05, two-sided rule

Pocock: performs each test at the same nominal α level (spends α evenly)

O'Brien-Fleming: spends very little α during the initial analysis and keeps almost all of α for later during the final analysis

EaSt 6.3.1: a software package for the design and interim monitoring of group sequential clinical trials, Cytel Software Corporation, Cambridge, MA, 2011

SUBGROUP ANALYSIS: PLANNED *VS*. POST-HOC

- Problem of multiplicity

An important limitation of subgroup analysis: performing multiple subgroup comparisons can increase the risk of *false positive findings*.

- What is the difference between planned & post-hoc analysis?

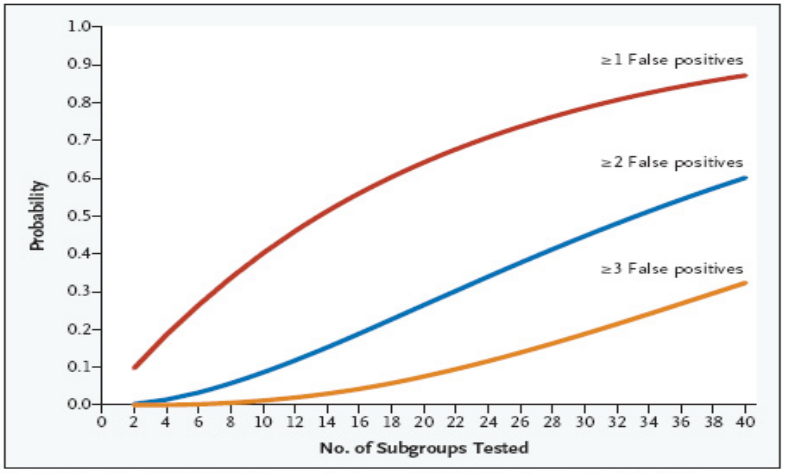
Planned analysis is predetermined and documented before any exploratory data analysis, while post-hoc is not.

Rui Wang MS, *et al.* Statistics in Medicine - Reporting of Subgroup, Analyses in Clinical Trials. N Engl J Med 2007; 357; 21: 2189-2194

From Lagakos SW, (2006). The Challenge of Subgroup Analyses — Reporting without Distorting, N Engl J Med, 2006; 354:1667-1669, Copyright ©2006 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

- **Why planned analysis is preferable to post-hoc?**

Post-hoc analysis increases the risk of approving drugs that have no beneficial effect (false positives), while with planned analysis one can control this error by limiting their number and adjusting for multiple comparisons.



Probability That Multiple Subgroup Analyses Will Yield at Least One (Red), Two (Blue), or Three (Yellow) False Positive Results.

KAPLAN-MEIER PLOTS: *THE CURVE THAT CHANGED THE WORLD*

What is a Kaplan-Meier (KM) plot?

- Graphical tool for presenting survival
- Useful for comparison between groups

Advantages

- Model-free
- Takes censoring into account
- Unbiased: Censoring affects precision but not accuracy
- Median is read directly from the plot

Limitations

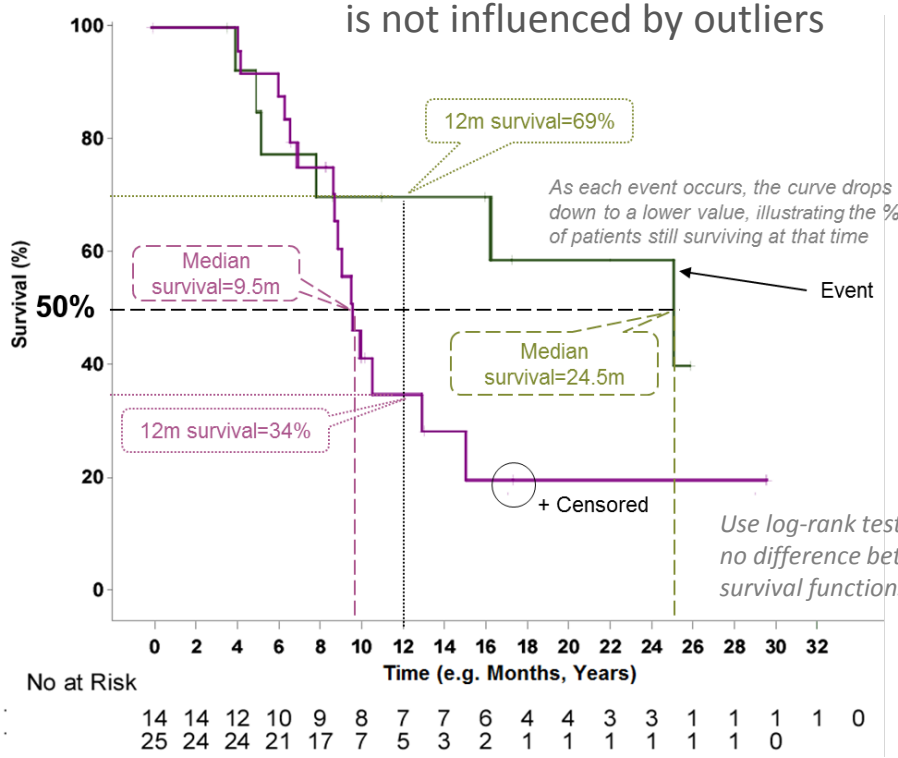
- Mainly descriptive
- No control for covariates
- Requires categorical predictors
- No time-varying variables

Kaplan EL, Meier P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assn. **53**:457–481.

Median

50% of observations are below this value

It accommodates censoring and is not influenced by outliers

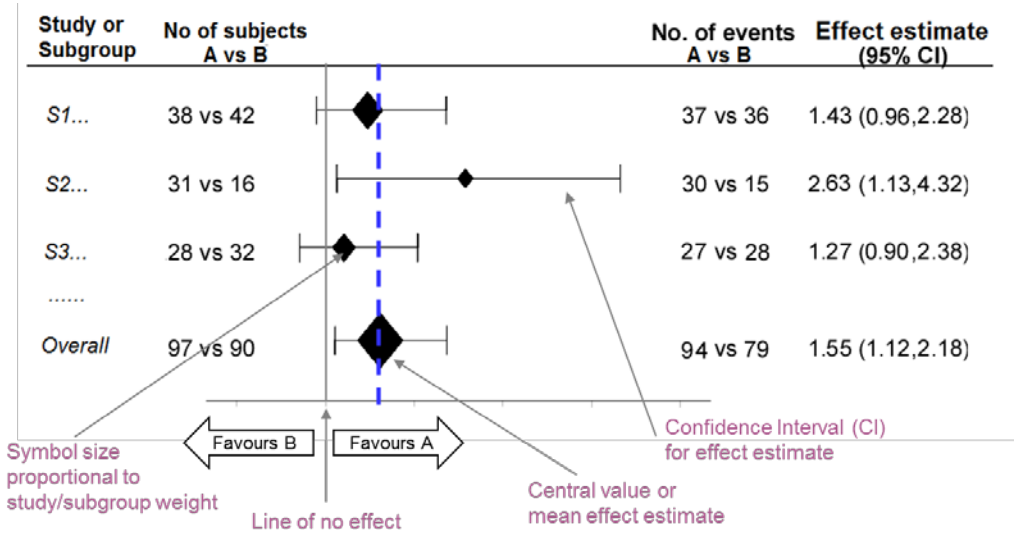


Attention: Do not over-interpret plateaus!

FOREST PLOTS: FOREST OF LINES

- A quick overview of multiple effect estimates from different studies or subgroups
- Presentation of the *relative strength* (and its variation) of effects of interest

- Common effect estimates:
  - Hazard Ratio,
  - Odds Ratio,
  - Relative Risk,
  - Mean difference,
  - Median survival
- It is used in:
  - Meta-analysis (originally)
  - Subgroup analysis (EMA, 2014)
  - Presentation of multivariate models, Sensitivity analysis, etc.



If "value of no effect" included in the CI
→
Effect not significant

Cochrane Handbook for Systematic Reviews of Interventions, Part 2, Chapter 11 (Updated 2011)
Lewis, S. and Clarke, M. (2001). 'Forest plots: trying to see the wood and the trees,' BMJ, 322, 1479-1480
European Medicines Agency (2014). Guidelines on the investigation of subgroups in confirmatory clinical trials

WATERFALL PLOTS

- A graphical illustration of a quantitative variable per subject.
- Commonly used in oncology clinical trials for response or treatment duration.

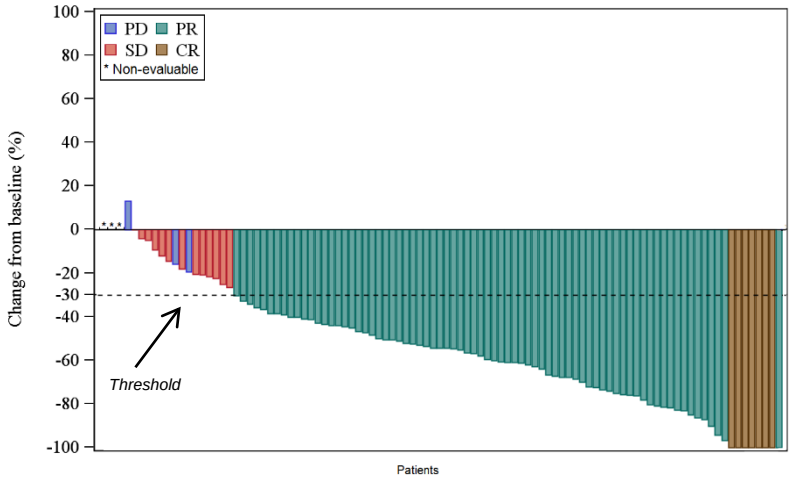
Advantages

- A novel efficacy measure for presenting the reduction in tumour burden for each subject
- Allows for a more detailed interpretation of stable disease as graded with RECIST

Limitations

- Can become intractable as a visualisation tool for large cohorts of patients
- Displays limited ability to portray randomization schemes other than 1:1

Example. *Waterfall plot of best % change from baseline in the sum of tumour diameters for targeted lesions*



- Each vertical bar represents an individual patient.
- Each colour represents key patient characteristic *e.g*. objective response or smoking status.
- The data are organised from worst to best (based on the parameters included) resembling a waterfall.

Gillespie TW, Understanding Waterfall Plots. J Adv Pract Oncol, 2012; 3(2): 106-111

Thank you!