STATISTICS IN CLINICAL TRIALS: How to interpret published data correctly?

Adam Svobodník

This presentation is financially supported by GlaxoSmithKline
Information sources
Information sources: Books

**Design of Clinical Trials**

- Clinical Trials: A Methodological Perspective. Piantadosi S. *ISBN 0471163937*
- Clinical Trials: A Practical Approach. Pocock SJ. *ISBN 0471901555*
- Clinical Trials: Design, Conduct, and Analysis. Meinert CL. *ISBN 0195035682*
- Clinical Trials in Oncology. Green S, Benedetti J, Crowley J. *ISBN 1584883022*
- Guide to clinical Trials. Spilker B. *ISBN 0881677671*

**Data Analysis in Clinical Trials**

- Analyzing survival Data from Clinical Trials and Observational Studies. Marubini E, Valsecchi MG. *ISBN 0471939870*
- Biostatistics in Clinical Trials. Redmond C, Colton T. *ISBN 0471822116*
- Statistical Methods for Clinical Trials. Norleans MX. *ISBN 0824704673*

**Sample size estimates**

- Handbook of Sample Size Guidelines for Clinical Trials. Shuster JJ. *ISBN 0849335426*

**Data management**

- Management of Data in Clinical Trials. McFadden E. *ISBN 047130316X*
Basic terminology
Combination therapy with dutasteride and tamsulosin in men with moderate-to-severe benign prostatic hyperplasia and prostate enlargement: the CombAT (Combination of Avodart® and Tamsulosin) trial rationale and study design

Paul Siami a,*, Claus G. Roehrborn b, Jack Barkin c, Ronaldo Damiao d, Marek Wyczolkowski e, Annette Duggan f, Kim Major-Walker g, Betsy B. Morrill h on behalf of the CombAT study group
Abstract

Background: Combination therapy with dutasteride and tamsulosin provides significantly greater benefit than either monotherapy for various patient-reported outcomes in men with moderate-to-severe lower urinary tract symptoms (LUTS) due to benign prostatic hyperplasia (BPH) and prostatic enlargement.

Objective: To investigate whether combination therapy is more effective than either monotherapy in reducing the relative risk for acute urinary retention (AUR), BPH-related surgery, and BPH clinical progression over 4 yr in men at increased risk of progression.

Design, setting, and participants: The Combination of Avodart® and Tamsulosin (CombAT) study was a 4-yr, multicenter, randomised, double-blind, parallel-group study in 4844 men ≥50 yr of age with a clinical diagnosis of BPH, International Prostate Symptom Score ≥12, prostate volume ≥30 cm³, prostate-specific antigen 1.5–10 ng/ml, and maximum urinary flow rate (Qmax) >5 and ≤15 ml/s with minimum voided volume ≥125 ml.

Intervention: Oral daily tamsulosin, 0.4 mg; dutasteride, 0.5 mg; or a combination of both.

Measurements: The 4-yr primary end point was time to first AUR or BPH-related surgery. Secondary end points included BPH clinical progression, symptoms, Qmax, prostate volume, safety, and tolerability.

Results and limitations: Combination therapy was significantly superior to tamsulosin monotherapy but not dutasteride monotherapy at reducing the relative risk of AUR or BPH-related surgery. Combination therapy was also significantly superior to both monotherapies at reducing the relative risk of BPH clinical progression. Combination therapy provided significantly greater symptom benefit than either monotherapy at 4 yr. Safety and tolerability of combination therapy was consistent with previous experience with dutasteride and tamsulosin monotherapies, with the exception of an imbalance in the composite term of cardiac failure among the three study arms. The lack of placebo control is a study limitation.

Conclusions: The 4-yr CombAT data provide support for the long-term use of dutasteride and tamsulosin combination therapy in men with moderate-to-severe LUTS due to BPH and prostatic enlargement.


© 2009 European Association of Urology. Published by Elsevier B.V. All rights reserved.
Randomization

Methodology and process of random (or pseudorandom) assignment of subjects to two or more treatment arms.
When was randomization used for the first time in clinical research?

1. 447 BC
2. 1447
3. 1878
4. 1924
5. 1946
6. 2008
Interim analysis

Interim analysis is analysis of the data at one or more time points prior to the official close of the study with the intention of possibly terminating the study early.

O'Brien-Flemming interim monitoring boundaries for the primary endpoint are based on predetermined number of planned interim analysis with overall type error of $\alpha=0.05$. 

![Graph showing interim monitoring boundaries](image-url)
Interim analysis:

1. is generally not recommended to be used in clinical trials
2. is recommended to be used in clinical trials if proper rules are respected
3. affects study power but not probability of false positive result
4. is not recommended to be used in triple-blind clinical trials
5. is recommended only for multicentric clinical trials
6. is recommended for clinical trials with 500 and more subjects
7. is not used in clinical trials with overall survival as primary endpoint
The Effects of Combination Tamsulosin on Clinical Outcomes of Prostatic Hyperplasia: 4-Year Results

Claud G. Roehrborn\textsuperscript{a,\*}, Paul Siami\textsuperscript{b}, Jackie Lam\textsuperscript{c}, Indrani Nandy\textsuperscript{e}, Betsy B. Morrill\textsuperscript{e}, R. Paul Thapar\textsuperscript{e} on behalf of the CombAT Study Group

2.3. Study end point and statistical analyses

The primary end point at 4 yr was time to first event of AUR or BPH-related prostatic surgery, defined as the number of days from the date of first dose of randomised study drug to the date of the initial event. The proportion of subjects experiencing AUR or BPH-related surgery was a supportive end point to the primary analysis. To address multiplicity, secondary end points were analysed in a predefined hierarchy (Table 1). Additionally, all primary and secondary end points defined and initially tested at 2 yr were included as secondary end points at 4 yr and analysed according to the hierarchy at year 2 [10]: We report IPSS, $Q_{\text{max}}$, and prostate volume outcomes in this paper.

The intent-to-treat population was the primary population analysed, consisting of all subjects randomised to double-blind study treatment. The primary comparison was combination versus tamsulosin, for which the study was powered at 94%; a comparison of combination versus dutasteride was also performed. The primary analysis used a log rank test stratified by investigative site cluster. Superiority for combination versus tamsulosin and dutasteride was based on a two-sided $p$ value at $\alpha = 0.01$. The relative risk (hazard ratio) for the treatment effect and associated two-sided 95% confidence intervals were estimated using a Cox proportional hazards model with treatment as the only covariate and stratified by investigative site cluster.
**ITT and PP analysis**

**Intention-to-treat (ITT) analysis** is based on data of all randomized subjects regardless of:
- fulfilling of inclusion criteria
- taking medicine in accordance to randomization code
- compliance with study protocol
- premature withdrawal from study

**Per-protocol (PP) analysis** is based only on data of subjects compliant to study protocol.
Study population that reflects the best situation in clinical practice is:

1. per protocol population
2. safety set
3. randomized set
4. intent to treat population
Benign Prostatic Hyperplasia

The Effects of Combination Therapy with Dutasteride and Tamsulosin on Clinical Outcomes in Benign Prostatic Hyperplasia: 4-Year Results

Claus G. Roehrborn, Paul Siami, Indrani Nandy, Betsy B. Morrill, on behalf of the CombAT Study Group

2. Materials and methods

2.1. Study design

The design of the multinational, multicenter, randomised, double-blind, parallel-group CombAT study has been previously reported [9–11]. Briefly, eligible subjects were randomised to receive one of the following treatments orally once daily for a period of 4 yr: dutasteride 0.5 mg and tamsulosin 0.4 mg, dutasteride 0.5 mg and tamsulosin-matched placebo, or dutasteride-matched placebo and tamsulosin 0.4 mg. Details of AUR and BPH-related prostatic surgery episodes were recorded at every visit, and the occurrence of recurrent urinary tract infection or urosepsis and/or first episode of incontinence (overflow or urge) was assessed at baseline and every 3 mo. The International Prostate Symptom Score (IPSS) questionnaire (including question 8, BPH-related health status) was implemented at screening, baseline, and every 3 mo, and Q_max was measured at screening, baseline, and every 6 mo. Transrectal ultrasound (TRUS) was performed at screening and annually to document change in total prostate volume.
Parallel design

Subjects are randomized to receive one of the tested treatments and use only this treatment during the whole experiment.
Cross-over design

All subjects are exposed to all treatments tested in experiment. Randomization is performed only to assign different sequences of treatments applied.
The best design for clinical trial is:

1. parallel groups
2. cross-over
3. triple triangle
4. exponential
5. sequential
6. depends on individual study settings (pathology, study duration etc.)
The effects of dutasteride or tamsulosin alone and in combination on storage and voiding symptoms in men with lower urinary tract symptoms (LUTS) and benign prostatic hyperplasia (BPH): 4-year data from the Combination of Avodart and Tamsulosin (CombAT) study

Francesco Montorsi¹, Claus Roehrborn², Javier Torrealba³, Ton A. Roeleveld⁴, Jean-Charles Alimi⁵, Paul A. Wilt⁶, Timothy H. Wilson⁷

¹Università Vita Salute San Raffaele, Milan, Italy; ²University of Texas Southwestern Medical Center, Dallas, Texas, USA; ³Hospital Costa Del Sol, Marbella, Spain; ⁴Aarhus University Hospital, Aarhus, Denmark; ⁵Alkmaar, The Netherlands; ⁶Clinique Sainte Marguerite, Hyères, France

Accepted for publication 25 January 2011

Data were analysed for the intention-to-treat population, which comprised all patients who were randomized into the double-blind treatment groups after the 4-week placebo run-in period. In the present post hoc 4-year analysis, the mean changes from baseline in IPSS storage and voiding subscales were summarised by treatment group using the last-observation-carried-forward approach. The mean changes in storage and voiding subscores were determined at each follow-up assessment using a general linear model with adjustments for treatment, investigative site cluster and baseline IPSS (storage or voiding). Pairwise treatment comparisons from the general linear model were carried out at \( \alpha = 0.05 \). These analyses were conducted outside of the protocol-specified endpoint hierarchy, with no formal adjustments for multiplicity of tests. For the present analyses, comparisons of combined therapy vs each monotherapy, and of dutasteride monotherapy vs tamsulosin monotherapy, were performed. These comparisons were performed for all patients and for patients...
No missing values in longitudinal study
LOCF: principle
LOCF: Bias in data display

**Figure 1** Mean Pain Scores (Non LOCF) Over Time

**Figure 2** Mean Pain Scores (LOCF) Over Time
Combination therapy with dutasteride and tamsulosin in men with moderate-to-severe benign prostatic hyperplasia and prostate enlargement: the CombAT (Combination of Avodart® and Tamsulosin) trial rationale and study design

Paul Siami a,*, Claus G. Roehrborn b, Jack Barkin c, Ronaldo Damiao d, Marek Wyczolkowski e, Annette Duggan f, Kim Major-Walker g, Betsy B. Morrill h on behalf of the CombAT study group

1.2. Study design

The CombAT trial is a multicenter, randomized, double-blind, parallel-group study designed to investigate whether combination therapy with dutasteride and tamsulosin is superior to each monotherapy in improving symptoms and long-term clinical outcomes in men with moderate-to-severe symptoms of BPH. The study is being conducted in Europe, North America, Latin America, and Asia Pacific.

Following screening for inclusion, eligible subjects received placebo tamsulosin and placebo dutasteride orally for 4 weeks, to minimize any contribution of the placebo effect to the study results. After this single-blind, placebo run-in period, subjects were randomized in a 1:1:1 ratio, in accordance with a computer-generated randomization schedule, to the double-blind phase and will receive one of the following treatments orally for 208 weeks:
Hypotheses in clinical trials

Equivalence testing

Non-inferiority testing

Superiority testing

Better
Meta-analysis refers to the analysis of analyses... the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. (Glass, 1976, p3)

Meta-analysis techniques are needed because only summary statistics are typically available in the literature.

Problems with meta-analysis:
- „publication bias“ („funnel shape“)
- multiple results from one population
- heterogeneous assessment of efficiency and safety in different studies
Meta-analysis: general approach

- Finding and selecting studies
- Identifying and coding study characteristics
- Analyzing the studies selected
- Reporting meta-analytic findings

Paper? submit 

Journal/conference
Basic classification of economical analysis:

- "Cost-minimization" analyses (CMA)
- "Cost-effectiveness" analyses (CEA)
- "Cost-utility" analyses (CUA)
- "Cost-benefit" analyses (CBA)

The main objective of pharmacoeconomical analyses in clinical trials is to compare two or more treatments from the view of costs and benefits.
Calculating QALYs: an example

<table>
<thead>
<tr>
<th>Intervention</th>
<th>QALYs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention A: Four years in health state 0.75</td>
<td>3 QALYs</td>
</tr>
<tr>
<td>Intervention B: Four years in health state 0.5</td>
<td>2 QALYs</td>
</tr>
<tr>
<td>Additional number of QALYs generated by A</td>
<td>1 QALY</td>
</tr>
<tr>
<td>Intervention</td>
<td>£/QALY at 1990 prices</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Cholesterol testing and diet therapy (all adults aged 40–69)</td>
<td>220</td>
</tr>
<tr>
<td>Neurosurgical intervention for head injury</td>
<td>240</td>
</tr>
<tr>
<td>GP advice to stop smoking</td>
<td>270</td>
</tr>
<tr>
<td>Neurosurgical intervention for subarachnoid haemorrhage</td>
<td>490</td>
</tr>
<tr>
<td>Antihypertensive treatment to prevent stroke (ages 45–64)</td>
<td>940</td>
</tr>
<tr>
<td>Pacemaker implantation</td>
<td>1,100</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>1,180</td>
</tr>
<tr>
<td>Valve replacement for aortic stenosis</td>
<td>1,410</td>
</tr>
<tr>
<td>Cholesterol testing and treatment (all adults aged 40–69)</td>
<td>1,480</td>
</tr>
<tr>
<td>Docetaxel (as opposed to paclitaxel) in treatment of recurrent metastatic breast cancer</td>
<td>1,890*</td>
</tr>
<tr>
<td>CABG (left main-vessel disease, severe angina)</td>
<td>2,090</td>
</tr>
<tr>
<td>Kidney transplantation</td>
<td>4,710</td>
</tr>
<tr>
<td>Breast cancer screening</td>
<td>5,780</td>
</tr>
<tr>
<td>Heart transplantation</td>
<td>7,840</td>
</tr>
<tr>
<td>Cholesterol testing and treatment incrementally (all adults aged 25–39)</td>
<td>14,150</td>
</tr>
<tr>
<td>Home haemodialysis</td>
<td>17,260</td>
</tr>
<tr>
<td>CABG (one-vessel disease, moderate angina)</td>
<td>18,830</td>
</tr>
<tr>
<td>Hospital haemodialysis</td>
<td>21,970</td>
</tr>
<tr>
<td>Erythropoietin treatment for anaemia in dialysis patients (assuming 10% reduction in mortality)</td>
<td>54,380</td>
</tr>
<tr>
<td>Addition of interferon-α2b to conventional treatment in newly diagnosed multiple myeloma</td>
<td>55,060*</td>
</tr>
<tr>
<td>Neurosurgical intervention for malignant intracranial tumours</td>
<td>107,780</td>
</tr>
<tr>
<td>Erythropoietin treatment for anaemia in dialysis patients (assuming no increase in survival)</td>
<td>125,290</td>
</tr>
</tbody>
</table>

* Adjusted to 1990 prices using Hospital and Community Health Service Pay and Prices Index, Unit Costs of Health and Social Care, PPSSRU, 1996. (2,431 ^200.7 x 155.6 = 1,890). ^ Translated into 1990 prices, as above.
QALYs are used in:

1. cost-minimization analysis
2. survival analysis
3. cost-effectiveness analysis
4. oncology trials
5. cost-utility analysis
6. cost-benefit analysis
Multivariate statistical methods

- Methods of common evaluation of the effect of several factors (type of treatment, stage of disease, genetic factors, markers, etc.) on the treatment efficiency and safety

- In clinical trials are typically not used for primary endpoint analysis

- Are used mainly on exploratory purposes

- Could be used to verify whether different distribution of prognostic factors in treatment arms could affect the differences in outcomes between treatment arms
Subgroup analyses

- Analyses of efficiency or/and safety on subgroup of subjects defined by specific entry criteria (eg. sex, age).

- **Advantages**: Potential of finding subgroup of patients with significantly better/worse outcomes – targeting of treatment

- **Disadvantages**: Increase of risk of false positive results

- When performing subgroup analyses respecting of guidelines for such analysis is required (prospectively defined subgroups etc.)
## Risk ratio: Example of study

<table>
<thead>
<tr>
<th>Death</th>
<th>Treatment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Active drug</td>
<td>Placebo</td>
</tr>
<tr>
<td>Yes</td>
<td>110 (a)</td>
<td>165 (b)</td>
</tr>
<tr>
<td>No</td>
<td>1935 (c)</td>
<td>1857 (d)</td>
</tr>
<tr>
<td>Total</td>
<td>2045 (n_1)</td>
<td>2022 (n_2)</td>
</tr>
</tbody>
</table>

Proportion of deaths:

- \(p_1 = \frac{a}{n_1} \times 100\) for Active drug: 5.38\%
- \(p_2 = \frac{b}{n_2} \times 100\) for Placebo: 8.16\%
- \(p = \frac{a + b}{n} \times 100\) for Total: 6.76\%
Risk ratio: Calculation

<table>
<thead>
<tr>
<th>Death</th>
<th>Treatment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Active drug</td>
<td>Placebo</td>
</tr>
<tr>
<td>Yes</td>
<td>110 [a]</td>
<td>165 [b]</td>
</tr>
<tr>
<td>No</td>
<td>1935 [c]</td>
<td>1857 [d]</td>
</tr>
<tr>
<td>Total</td>
<td>2045 [n_1]</td>
<td>2022 [n_2]</td>
</tr>
<tr>
<td>Proportion of deaths</td>
<td>5.38% ( p_1 = \frac{a}{n_1} \times 100 )</td>
<td>8.16% ( p_2 = \frac{b}{n_2} \times 100 )</td>
</tr>
</tbody>
</table>

**Risk ratio**

The **risk ratio** is the ratio of the risks in the active drug treatment group compared to the placebo group. The **risk ratio** is often abbreviated to \( RR \), and is also sometimes called the relative **risk**:

\[
RR = \frac{p_1}{p_2} = \frac{a/n_1}{b/n_2} = \frac{a/(a+c)}{b/(b+d)}
\]

For the MI trial, the \( RR = 0.66 \) (5.38% / 8.16%), meaning that the risk of death for the patients in the active drug treatment group is only 66% of the risk in the placebo group. Equivalently, we could say that the drug treatment is associated with a 34% (100% – 66%) reduction in mortality at 30 days.
### Odds ratio: Calculation

A third measure of treatment effect is the odds ratio (OR). The odds of an outcome event are calculated as the number of events divided by the number of nonevents. For example, in the active treatment arm in the MI trial, the number of deaths is 110 and the number of survivals is 1,935, so the odds of death are $110 / 1935 = 0.057$. If the odds of an event are $>1$, the event is more likely to happen than not. In particular, the odds of an event that is certain to happen are infinite, and the odds of an impossible event are zero. The OR is calculated by dividing the odds in the active treatment group ($a / c$) by the odds in the placebo group ($b / d$):

$$OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}$$

For the MI trial data, the OR is calculated as $(110 \times 1857) / (1936 \times 165) = 0.64$, meaning that the odds of deaths after MI in the drug group are 64% of the odds in the placebo group. Clinical trials typically look at treatments that reduce the proportion of patients with an event or, equivalently, have an OR of $<1$. In these cases, a percentage reduction in the OR is often quoted instead of the OR itself. For the preceding OR, we can say that there is a 36% (100% – 64%) reduction in the odds of deaths in the active treatment group.
Absolute and relative risk reduction, NNT

Absolute risk reduction (ARR)
The difference between the control group’s event rate (CER, eg. 10%) and the experimental group’s event rate (EER eg. eg. 20%): \( \text{ARR} = 10\% \).

Relative risk reduction (RRR)
The percent reduction in events in the treated group compared to the control group event rate. For the example above: \( 10\% / 20\% = 0.5 = 50\% \).

Number needed to treat (NNT)
The number of patients that must be treated to prevent one adverse outcome or for one patient to benefit. The NNT is the inverse of the ARR; \( \text{NNT} = 1/\text{ARR} \).
The use of many significance tests in a single study, e.g. many subgroup analyses, the use of many testing methods, analysis of many variables, and serial measurements (time course) of the same variable can increase the probability of obtaining false-positive results. Care should therefore be exercised in the study design and in the interpretation of $P$ values in such a situation.
1) *Look for other simpler endpoints*
The best approach for avoiding multiplicity of testing is to use one or only a few statistics instead of multiple ones wherever possible.

2) *Give a conservative interpretation and do not use the term `significant'.*
Another approach is to indicate exact $P$ values and to interpret each result conservatively without the term `significant' or `not significant'.

3) *Controlling the level of significance*
A simple *ad hoc* method is to use the Bonferroni correction. The idea behind this is that if one were conducting $n$ significance tests, then to obtain an overall type I error rate of $\alpha$, one would only declare any one of them `significant' if the $P$ value were smaller than $\alpha / n$; for example, the significance level is set at 0.01 for 5 subgroup analyses.
Classification of clinical trials
PROCESS OF NEW DRUG DEVELOPMENT

- **Laboratory experiments**
- **Preclinical testing**
- **Clinical trials:**
  - Phase I
  - Phase II
  - Phase III
- **Registration and approval from regulatory authorities**
- **Clinical trials:** Phase IV

10-15 YEARS
Objectives:

- Assessment of basic product pharmacokinetic parameters in humans
- Estimation of the maximal tolerated dose MTD (cytostatics etc.)
- Evaluation of Adverse Events (AE)
- Dose finding study

Study subjects:

- 12-20
- Mostly healthy volunteers
- Not vulnerable subjects

Design:

- Ideal design enable exact evaluation of the "dose – response" curve
- Because of ethical reasons, adaptive designs are used: dosage for subsequent subject is based on the response of previous subject
- The first dose used is based on results of preclinical testing (animal testing)
PHASE 1 – example of subjects’ enrollment

3 subjects, initial dose

- Adverse events NOT present
  - NO
  - Another 3 subjects, initial dose
    - AE at most in 1 subject
      - NO
      - End of study
    - YES
      - Another 3 subjects, higher dose

- YES
  - Another 3 subjects, higher dose
1 subject, initial dose

Adverse events
NOT present

NO

Another subject, lower dose

AE not present in 2 subjects

NO

Another subject, lower dose

YES

Another subject, higher dose

NO

End of study

AE present in 2 subjects

YES

Another subject, higher dose

NO

Another subject, lower dose
“PHASE 2“ clinical trials

- **Objectives:**
  - Verification of the treatment effectiveness
  - Evaluation of tolerance and safety
  - Decision whether Phase III trials will be performed

- **Study subjects:**
  - 20 – 200
  - Number of study subjects
    - fixed
    - enrollment by groups
    - sequential (including evaluation of response of each subject continuously)

- **Design:**
  - Randomization is rare
  - Experiments with one arm
  - Treatment effectiveness and safety is compared to known products or placebo
"PHASE 3" clinical trials

- **Objectives:**
  - Comparison of the effectiveness and safety of the test product with placebo or other type of control (active treatment control)
  - Getting data for regulatory authorities
  - "Cost – effectiveness" analyses

- **Study subjects:**
  - 100 – 1 000 number of subjects
    - fixed
    - enrollment by groups
    - sequential (including evaluation of response of each subject continuously)

- **Design:**
  - Parallel
  - "Cross – over"
  - Factorial
  - Randomization
Comparison of cross-over vs. parallel design

ADVANTAGES:
• elimination of between-subject variability of symptoms
• no need for large samples
• fewer ethical problems
• subjects are able to express their preference for one of the compounds being given

DISADVANTAGES:
• carryover effect
• time effect due to spontaneously evolving symptoms
Objectives:

- Verification of product characteristics in „real settings“
- Detailed analyses of adverse events
- Evaluation of QoL
- Changes in dosage
- „Cost-effectiveness“ studies

Design

- **Descriptive studies** (analysis of existing databases)

- „**Cross - sectional**“ studies (analysis of structured sample of patients)

- „**Case - control**“ studies (retrospective studies with selected paired control groups)

- **Cohort studies** (retrospective or prospective comparison of selected cohort with control group)
Safety is usually assessed:

1. only in Phase I trials
2. only in Phase II trials
3. only in Phase III trials
4. only in Phase IV trials
5. only in Phase I and Phase II trials
6. only in Phase I and Phase III trials
7. in Phase I – IV trials
Study designs

**Descriptive studies**
- designed to describe occurrence of disease by time, place and person

**Experimental (intervention studies)**
- Investigator intentionally alters one or more factors to study the effects of so doing

**Uncontrolled trials**
- experimental trials without control or comparison groups (e.g. phase I/II clinical trials)

**Randomized (RCTs)**
- interventions allocated randomly (all participants or clusters have the same chance of being allocated to each of the study groups)

**Controlled trials**
- trials with control groups (e.g. phase III trials)
- controlled trials can be clinical trials (unit of randomization is an individual) or community trials (unit of randomization is a community or cluster)

**Quasi-randomized**
- allocation done using schemes such as: according to date of birth (odd or even), number of the hospital record, date at which they are invited to participate in the study (odd or even), or alternatively into the different study groups

**Non-randomized**
- allocation to different groups done arbitrarily (without any underlying random process)

**Analytic studies**
- designed to examine etiology and causal associations

**Non-experimental (observational studies)**
- Does not involve intervention; investigator observes without intervention other than to record, count, and analyze results

- Cohort (retrospective and prospective)
- Case-control
- Cross-sectional
- Ecological

**Classification of experiments**
- Prevalence surveys
- Case-series
- Surveillance data
- Analyses of routinely collected data (registries, mortality data, etc.)
Fundamentals of biostatistics for clinical trials
Benign Prostatic Hyperplasia

The Effects of Combination Therapy with Dutasteride and Tamsulosin on Clinical Outcomes in Men with Symptomatic Benign Prostatic Hyperplasia: 4-Year Results from the CombAT Study

Claus G. Roehrborn\textsuperscript{a,}*, Paul Siami\textsuperscript{b}, Jack Barkin\textsuperscript{c}, Ronaldo Damião\textsuperscript{d}, Kim Major-Walker\textsuperscript{e}, Indrani Nandy\textsuperscript{e}, Betsy B. Morrill\textsuperscript{e}, R. Paul Gagnier\textsuperscript{e}, Francesco Montorsi\textsuperscript{f}

on behalf of the CombAT Study Group

\textsuperscript{a} Department of Urology, UT Southwestern Medical Center, Dallas, Texas, USA
\textsuperscript{b} Deaconess Clinic, Evansville, Indiana, USA
\textsuperscript{c} Department of Urology, University of Toronto, Toronto, Ontario, Canada

Table 2 – Baseline demographics and patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>Combination ($n = 1610$)</th>
<th>Dutasteride ($n = 1623$)</th>
<th>Tamsulosin ($n = 1611$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age ± SD, yr</td>
<td>66.0 ± 7.05</td>
<td>66.0 ± 6.99</td>
<td>66.2 ± 7.00</td>
</tr>
<tr>
<td>White ethnicity (%)</td>
<td>1421 (88)</td>
<td>1433 (88)</td>
<td>1405 (87%)</td>
</tr>
<tr>
<td>Mean total IPSS score ± SD, points</td>
<td>16.6 ± 6.35</td>
<td>16.4 ± 6.03</td>
<td>16.4 ± 6.10</td>
</tr>
<tr>
<td>Prostate volume, cm\textsuperscript{3}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean total ± SD</td>
<td>54.7 ± 23.51</td>
<td>54.6 ± 23.02</td>
<td>55.8 ± 24.18</td>
</tr>
<tr>
<td>Median total</td>
<td>48.9</td>
<td>48.4</td>
<td>49.6</td>
</tr>
<tr>
<td>Mean transition zone ± SD</td>
<td>27.7 ± 20.20</td>
<td>30.3 ± 21.02</td>
<td>30.5 ± 24.47</td>
</tr>
<tr>
<td>Mean serum PSA ± SD, ng/ml</td>
<td>4.0 ± 2.05</td>
<td>3.9 ± 2.06</td>
<td>4.0 ± 2.08</td>
</tr>
<tr>
<td>Mean Q\textsubscript{max} ± SD, ml/s</td>
<td>10.9 ± 3.61</td>
<td>10.6 ± 3.57</td>
<td>10.7 ± 3.66</td>
</tr>
<tr>
<td>Mean postvoid residual volume ± SD, ml</td>
<td>68.2 ± 66.12</td>
<td>67.4 ± 63.49</td>
<td>67.7 ± 65.14</td>
</tr>
<tr>
<td>Sexually active (%)</td>
<td>1176 (73)</td>
<td>1189 (73)</td>
<td>1164 (72%)</td>
</tr>
<tr>
<td>Previous α-blocker use (%)</td>
<td>805 (50)</td>
<td>820 (51)</td>
<td>819 (51%)</td>
</tr>
<tr>
<td>Previous 5α-reductase inhibitor use (%)</td>
<td>171 (11)</td>
<td>188 (12)</td>
<td>172 (11%)</td>
</tr>
</tbody>
</table>

IPSS = International Prostate Symptom Score; PSA = prostate-specific antigen; Q\textsubscript{max} = maximum urinary flow rate; SD = standard deviation.

* In a subset of 656 men.
Types of data in clinical trials

- **Continuous data**
  - How many times?
- **Interval data**
  - How much?
- **Ordinal data**
  - Higher, lower?
- **Binary data**
  - Is equal?

Questions "Yes/No"

Continuous data

Discrete data

Questions "Yes/No"
Types of analyses in clinical trials

Descriptive analyses

Representative sampling

Variability

RESULTS

Hypotheses testing

Selection of subjects

RANDOMIZATION

Arm A

Measurement of parameter X

Variability of X in arm A

RESULTS

Arm B

variability of X in arm B

RESULTS
Descriptive statistics

- Continuous data
  - MEAN
  - Continuous data

- Interval data
  - MEDIAN

- Ordinal data
  - MODUS
  - Discrete data

- Binary data
Central statistics computation: possible variable distributions

X: Concentration of marker in blood sample
Central statistics computation: possible variable distributions

Median = 50% quantile
MAX – MIN = range
Modus = the most frequent value
Descriptive statistics: Interval estimates

Interval width is determined by:
  a) Sample size
  b) Variability
  c) Required accuracy

Distribution of $x$ in target population

Sample of $n=10$ for estimation of mean

Sample of $n=100$ for estimation of mean

$\varphi(x)$

$\mu$

$-3s$

$+3s$

$\mu$

$\frac{-3s}{\sqrt{10}}$

$\frac{+3s}{\sqrt{10}}$

$\frac{-3s}{\sqrt{100}}$

$\frac{+3s}{\sqrt{100}}$

$\varphi(x)$

$\varphi(x)$
Hypotheses testing
How to define hypothesis in clinical trial?

Null hypothesis $H_0$

usually “no difference” or what is NOT supposed in target population

Alternative hypothesis $H_1$

usually “exist difference” or what IS supposed to be true in target population
### Possible errors in hypothesis testing

#### True situation in population

<table>
<thead>
<tr>
<th>Conclusion</th>
<th>H₁ (difference exists)</th>
<th>H₀ (no difference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject H₀</td>
<td>O.K.</td>
<td>Type I error, α error</td>
</tr>
<tr>
<td>(difference exists)</td>
<td>Power ( (1 - \beta) )</td>
<td></td>
</tr>
<tr>
<td>Do not reject H₀</td>
<td>Type II error, β error</td>
<td>O.K.</td>
</tr>
<tr>
<td>(no difference)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Type I error** - relative frequency (probability) of rejection of H₀ if the same clinical trial is repeated many times and in reality (in population) H₀ is true.

**Type II error** - relative frequency (probability) of failing to reject H₀ if the same clinical trial is repeated many times and in reality (in population) H₀ is false.

**Power** - relative frequency (probability) of rejection of H₀ (and accepting H₁) if in reality (in population) H₀ is false.
Type I error (a) – interpretation

- Type I error is interpreted as the probability of rejecting the null hypothesis \( H_0 \) if this hypothesis is true

- Error of the false-positive result

- Mostly is \( \alpha \) set to level of 5%

- \( P < 0.05 \)........we reject the null hypothesis
- \( P \geq 0.05 \)........we accept the null hypothesis

**Negative result of the study (\( P > 0.05 \)) needs to be interpreted in the context of the clinical significant difference and sample size (study power)!!!!!!!!!!!!!!!!**
Type II error is the probability of not-rejecting the null hypothesis $H_0$ if this hypothesis is not true

- Error of the false-negative result
- Mostly is $\beta$ set to level of 10-20%
- Power of the statistical test is: $1-\beta$

The power of statistical test is the probability of proving difference in experiment if the difference really exists in reality.
Concept of p-value

α level of significance
- have to be chosen before the statistical test is performed. This is the probability of incorrectly rejecting the null hypothesis when it is actually true. Traditional values of α are 0.05, 0.01 or 0.001.

p-value level of significance
- could be approximately interpreted as the probability that the observed result is due to chance alone if H_0 hypothesis is true. p-value is calculated after the statistical test, if p-value is less than α => the null hypothesis is rejected

Statistical conclusion is based only on predefined α level ! (conclusion is affected only by fact if p-value is less or higher than predefined α level).
Variability of data could be characterized by:

1. mean
2. median
3. modus
4. standard deviation
5. geometric mean
Benign Prostatic Hyperplasia

The Effects of Combination Therapy with Dutasteride and Tamsulosin on Clinical Outcomes in Men with Symptomatic Benign Prostatic Hyperplasia: 4-Year Results from the CombAT Study

Claus G. Roehrborn a,*, Paul Siadak b, Indrani Nandy e, Betsy B. Morris a, for the CombAT Study Group.

a Department of Urology, UT Southwestern Medical Center, Dallas, TX, USA
b Deaconess Clinic, Evansville, Indiana, USA
c Department of Urology, University of Toronto, Toronto, ON, Canada
d Urology Department, State University of Rio de Janeiro, Brazil
e Research and Development, GlaxoSmithKline, Research Triangle Park, NC, USA
f Department of Urology, Universita Vita Salute San Raffaele, Milan, Italy

Fig. 3 – Kaplan-Meier estimates of time to the first episode of acute urinary retention or benign prostatic hyperplasia-related prostatic surgery.
### Application of survival analysis

<table>
<thead>
<tr>
<th>„Lifetime data“ in clinical research</th>
</tr>
</thead>
<tbody>
<tr>
<td>- „Overall Survival (OS)“</td>
</tr>
<tr>
<td>- „Time to Progression (TTP)“</td>
</tr>
<tr>
<td>- „Time to Treatment Failure (TTF)“</td>
</tr>
<tr>
<td>- „Duration of Response“</td>
</tr>
<tr>
<td>- „Relapse Free Survival“</td>
</tr>
<tr>
<td>- others</td>
</tr>
</tbody>
</table>

| „Reliability studies“ in industry    |
Most common usage of survival analysis in clinical research

- Description of survival in one group of patients and estimation of survival parameters (e.g. median survival)

- Comparison of survival in two or more groups of patients with different types of treatment

- Multivariate analysis of prognostic factors and its impact on survival
### Example of survival data

**Data of patients with angina pectoris in study with 15 years of follow-up (Mayo Clinic, Gehan 1969)**

<table>
<thead>
<tr>
<th>Survival time [years]</th>
<th>Number of patients known to survive at beginning of interval</th>
<th>Number of patients lost to follow up</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>2418</td>
<td>0</td>
</tr>
<tr>
<td>1-2</td>
<td>1962</td>
<td>39</td>
</tr>
<tr>
<td>2-3</td>
<td>1697</td>
<td>22</td>
</tr>
<tr>
<td>3-4</td>
<td>1523</td>
<td>23</td>
</tr>
<tr>
<td>4-5</td>
<td>1329</td>
<td>24</td>
</tr>
<tr>
<td>5-6</td>
<td>1170</td>
<td>107</td>
</tr>
<tr>
<td>6-7</td>
<td>938</td>
<td>133</td>
</tr>
<tr>
<td>7-8</td>
<td>722</td>
<td>102</td>
</tr>
<tr>
<td>8-9</td>
<td>546</td>
<td>68</td>
</tr>
<tr>
<td>9-10</td>
<td>427</td>
<td>64</td>
</tr>
<tr>
<td>10-11</td>
<td>321</td>
<td>45</td>
</tr>
<tr>
<td>11-12</td>
<td>233</td>
<td>53</td>
</tr>
<tr>
<td>12-13</td>
<td>146</td>
<td>33</td>
</tr>
<tr>
<td>13-14</td>
<td>95</td>
<td>27</td>
</tr>
<tr>
<td>14-15</td>
<td>59</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
</tr>
</tbody>
</table>
Example of censored data in clinical trial

Model clinical trial:

- Number of patients: 4 pts
- Primary endpoint: Overall Survival (OS)
- Accrual Period: 12 months
- Minimal follow-up: 24 months
Example of censored data in clinical trial

- **Accrual**
  - Patient 1: Death after 01/01/2001
  - Patient 2: Lost to follow up after 01/01/2001
  - Patient 3: Death after 01/01/2001
  - Patient 4: Withdrawn after 31/12/2002

- **Follow up**

**Calendar time**
- 01/01/2000
- 01/01/2001
- 01/01/2002
- 31/12/2002
Example of censored data in clinical trial

Follow-up

Patient 1

Patient 2

Patient 3

Patient 4

Follow-up

0 12 months 24 months 36 months

D=death, L-lost to follow-up, W-withdrawn
**Study objective:**
Analysis of the effect of maintenance therapy on prolongation of time to relapse in patients with acute lymphoblastic leukemia

**Study design:**
After CR, patients were randomised into two arms:
- Placebo
- 6-mercaptopurine (6-MP)

**Primary endpoint:**
Time to Progression (TTP)
### Study outcomes:

**A total of 42 patients randomised (1:1):**

**Placebo:** 21 of 21 patients with relapse

**6-MP:** 12 of 21 without relapse at the study end

#### Time in weeks:

**Placebo:** 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

**6-MP:** 6, 6, 6*, 7, 9*, 10, 10*, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35* (*censored observations)
Possible evaluation of study results:

1/ Comparison of mean or median of time to relapse
   - not possible, not available in censored patients

2/ Comparison of relapse rates in both groups
   (100% for placebo, 43% for 6-MP)
   - no information on prolongation of time to relapse
   - in censored patients it is possible, that relapse will occur in follow-up

Need for another method of analysis: Survival analysis
Basic principle of Kaplan-meier curve:

\[ \hat{\rho}(X) = \text{conditional survival probability of } X \text{ months} \]

\[ \hat{\rho} \text{ is calculated for each time interval separately} \]

\[ \hat{\rho} \text{ in different time intervals are independent} \]

\[ \hat{\rho}(1) \quad \hat{\rho}(2) \quad \hat{\rho}(3) \quad \hat{\rho}(4) \quad \hat{\rho}(5) \quad \hat{\rho}(6) \quad \hat{\rho}(7) \quad \hat{\rho}(8) \quad \hat{\rho}(9) \quad \hat{\rho}(10) \quad \hat{\rho}(11) \quad \hat{\rho}(12) \]

Time [months]
Basic principle of Kaplan-meier curve:

Conditional survival probability of 1 month after diagnosis:
\[ \hat{p}(1) = \frac{\text{Number of patients entering study} - \text{number of pts died during 1. month}}{\text{Number of patients entering study}} \]

Conditional survival probability of 6 month after diagnosis:
\[ \hat{p}(6) = \frac{\text{Number of pts „at risk“ in 6. month} - \text{number of pts died during 6. month}}{\text{Number of pts „at risk“ in 6. month}} \]

Cumulative survival probability 12 months after diagnosis:
\[ P(12) = \hat{p}(1) \times \hat{p}(2) \times \hat{p}(3) \ldots \times \hat{p}(12) \]
Calculation of survival in Placebo arm:

<table>
<thead>
<tr>
<th>t_{(j)}</th>
<th>d_j</th>
<th>n_j</th>
<th>p_j = (n_j - d_j) / n_j</th>
<th>P(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>19/21 = 0,905</td>
<td>0,905</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>17/19 = 0,895</td>
<td>0,905 x 0,895 = 0,810</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>16/17 = 0,941</td>
<td>0,810 x 0,941 = 0,762</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>14/16 = 0,875</td>
<td>0,762 x 0,875 = 0,667</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>12/14 = 0,857</td>
<td>0,667 x 0,857 = 0,571</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>4</td>
<td>8/12 = 0,667</td>
<td>0,571 x 0,667 = 0,381</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>2</td>
<td>6/8 = 0,750</td>
<td>0,381 x 0,750 = 0,286</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>2</td>
<td>4/6 = 0,667</td>
<td>0,286 x 0,667 = 0,191</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>1</td>
<td>3/4 = 0,750</td>
<td>0,191 x 0,750 = 0,143</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>1</td>
<td>2/3 = 0,667</td>
<td>0,143 x 0,667 = 0,095</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>1</td>
<td>1/2 = 0,500</td>
<td>0,095 x 0,500 = 0,048</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>1</td>
<td>0/1 = 0,000</td>
<td>0,048 x 0,000 = 0,000</td>
</tr>
</tbody>
</table>
Survival in Placebo arm

Cumulative Proportion Surviving

Time

Cumulative Proportion

Placebo
### Calculation of survival in arm with 6-MP:

<table>
<thead>
<tr>
<th>Time to progression</th>
<th>Number of censored</th>
<th>Number of relapses</th>
<th>Number of pts „at risk“</th>
<th>Conditionalsurvival probability</th>
<th>Cumulative survival probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{(j)}$</td>
<td>$d_j$</td>
<td>$n_j$</td>
<td>$p_j=(n_j-d_j)/n_j$</td>
<td>$P(t)$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>21</td>
<td>$18/21=0,857$</td>
<td>$0,857$</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>$16/17=0,941$</td>
<td>$0,857 \times 0,941 = 0,807$</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>$14/15=0,933$</td>
<td>$0,807 \times 0,933 = 0,753$</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>$11/12=0,917$</td>
<td>$0,753 \times 0,917 = 0,690$</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>$10/11=0,909$</td>
<td>$0,690 \times 0,909 = 0,628$</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>$6/7=0,857$</td>
<td>$0,628 \times 0,857 = 0,538$</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>$5/6=0,833$</td>
<td>$0,538 \times 0,833 = 0,448$</td>
</tr>
</tbody>
</table>
Comparison of survival in treatment arms