Randomized Controlled Trials
Randomised Controlled Trials: the Basics

- An RCT seeks to measure and compare the outcomes of two or more clinical interventions.
- One intervention is regarded as the standard of comparison or control.
- Participants receive the interventions in random order to ensure similarity of characteristics at the start of the comparison.
Randomised Controlled Trials: the Basics

• Randomisation can be achieved through a variety of procedures
• Individuals, groups, and the order in which measurements are obtained can all be randomised
• RCTs cannot answer all clinical questions
3 Questions in the Evaluation of Clinical Evidence

• Are the results valid?
• What are the results?
• Will these results help me care for my patients?
Validity Issues
Are the Results Valid?

• Randomized?
• Allocation concealment?
• All participants accounted for?
  – Follow up complete
• Intention to treat
• Blinding?
• Groups similar at onset?
• Groups equally treated?

Measures to get rid of bias
Random Allocation

• All participants have the same chance of being assigned to each of the study groups
• Allocation is not determined by the investigators, the clinicians, or the study participants
• Aim: well balanced groups
• If sample size is sufficient the 2 groups will be equal for both known and unknown variables and characteristics
Allocation Concealment
Complete Follow-up

• Follow up with loss of more than 25% imply a serious threat to the validity
• This is especially true if the loss to follow up is significantly more in one of the study arms
Intention-to-Treat

- Subjects are analyzed in the treatment group to which they were randomized to, irrespective of receiving and completing the treatment or not.
Blinding
Results?
What Are the Results?

• Experimental event rate (EER)
• Control event rate (CER)
• Risk
• Relative risk (RR)
• Relative risk reduction (RRR)
• Absolute risk reduction (ARR)
• Numbers needed to treat (NNT)
### Key Terminology

<table>
<thead>
<tr>
<th></th>
<th>Outcome</th>
<th>Risk of Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treated (Y)</strong></td>
<td>a</td>
<td>Y=a/(a+b)</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>EER</td>
</tr>
<tr>
<td><strong>Control (X)</strong></td>
<td>c</td>
<td>X=c/(c+d)</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>CER</td>
</tr>
</tbody>
</table>
Key Terminology

- \( RR = \frac{Y}{X} \quad \text{EER} / \text{CER} \)
- \( RRR = 1 - RR = 1 - \frac{Y}{X} \times 100\% \)
  \( \text{CER} - \text{EER} / \text{CER} \)
- \( ARR = X - Y \quad \text{EER} - \text{CER} \)
- \( NNT = \frac{1}{ARR} = \frac{1}{(X - Y)} \)
What Are the Results? Cont.

- How large was the treatment effect?
  - ARR vs RRR
- How precise was the estimate of the treatment effect?
  - CI
  - P-value
Randomized Controlled Trial

107 men and women
15 to 30 years of age
Pulmonary tuberculosis

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Alive</th>
<th>Dead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streptomycin + Bed rest</td>
<td>51</td>
<td>4</td>
</tr>
<tr>
<td>Placebo + Bed rest</td>
<td>38</td>
<td>14</td>
</tr>
</tbody>
</table>
Results of RCT

Mortality at 6 months

RR = EER/CER
RRR = CER – EER/CER

EER: 4 of 55 (7.3%)
CER: 14 of 52 (26.9%)
ARR: 4 of 55 (7.3%)

Streptomycin + Bedrest
Placebo + Bedrest
<table>
<thead>
<tr>
<th>End Point</th>
<th>Cumulative Incidence</th>
<th>Absolute Risk Reduction (95% CI)</th>
<th>Relative Risk (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Radical-Prostatectomy Group</td>
<td>Watchful-Waiting Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>total no.</td>
<td>% (95% CI)</td>
<td>total no.</td>
<td>% (95% CI)</td>
</tr>
<tr>
<td>Disease-specific mortality</td>
<td>30</td>
<td>2.3 (1.2 to 4.6)</td>
<td>50</td>
<td>4.3 (2.6 to 7.1)</td>
</tr>
<tr>
<td>At 5 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At 10 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distant metastases</td>
<td>50</td>
<td>9.6 (6.5 to 14.2)</td>
<td>79</td>
<td>14.9 (11.2 to 19.8)</td>
</tr>
<tr>
<td>At 5 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At 10 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local progression</td>
<td>64</td>
<td>8.1 (5.7 to 11.6)</td>
<td>149</td>
<td>27.2 (22.8 to 32.3)</td>
</tr>
<tr>
<td>At 5 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At 10 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall mortality</td>
<td>83</td>
<td>19.2 (15.0 to 24.6)</td>
<td>106</td>
<td>44.3 (38.8 to 50.5)</td>
</tr>
<tr>
<td>At 5 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At 10 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Analysis of the cumulative incidence was performed with the method of Kalbflieisch and Prentice and relative risks were calculated with the use of the Cox proportional-hazards model. The absolute risk reduction and relative risk were calculated compared with watchful waiting. Gray’s test was used to determine P values. The mean follow-up periods were 11.8 years in the radical-prostatectomy group and 8.8 years in the watchful-waiting group. CI denotes confidence interval.
Survival Analysis
Survival Analysis

• This is a method frequently utilized as outcome

• The time to an event e.g
  – Time until response
  – Time until recurrence in cancer
  – Time to death
  – Time to cardiovascular event
Survival Analysis

End of Study

Calendar Time

Entry time  × Event  ○ Censored

Study Duration
## Survival Analysis

<table>
<thead>
<tr>
<th>time</th>
<th>n.risk</th>
<th>n.event</th>
<th>survival</th>
<th>std.err</th>
<th>l. 95% CI</th>
<th>u. 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>21</td>
<td>3</td>
<td>0.857</td>
<td>0.0764</td>
<td>0.720</td>
<td>1.000</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>1</td>
<td>0.807</td>
<td>0.0869</td>
<td>0.653</td>
<td>0.996</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>1</td>
<td>0.753</td>
<td>0.0963</td>
<td>0.586</td>
<td>0.968</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>1</td>
<td>0.690</td>
<td>0.1068</td>
<td>0.510</td>
<td>0.935</td>
</tr>
<tr>
<td>16</td>
<td>11</td>
<td>1</td>
<td>0.627</td>
<td>0.1141</td>
<td>0.439</td>
<td>0.896</td>
</tr>
<tr>
<td>22</td>
<td>7</td>
<td>1</td>
<td>0.538</td>
<td>0.1282</td>
<td>0.337</td>
<td>0.858</td>
</tr>
<tr>
<td>23</td>
<td>6</td>
<td>1</td>
<td>0.448</td>
<td>0.1346</td>
<td>0.249</td>
<td>0.807</td>
</tr>
</tbody>
</table>
Survival Curve

\[ S(1.3) = \frac{3}{5} \]
Kaplan Meier Curve

Log rank test or the Cox proportional Hazards model determines if the difference between the 2 lines is significant or due to chance.
B

Cumulative Incidence of Death from Any Cause (%)

Years of Follow-up

No. at Risk

Radical prostatectomy 347 343 332 284 210 118
Watchful waiting 348 341 326 279 198 104

P = 0.04

Watchful waiting
Radical prostatectomy
Will Results Help in Caring for My Patients

• Can it be applied to my patients?
  – Are my patient similar to that of the study?
    Age, sex, stage of disease etc.

• Were all clinically important outcomes considered?

• Are the likely benefits worth the potential harm?
What an RCT Can and Can’t Do

This design can reduce problems associated with
• Experimenter bias
• Patients’ expectations
• Compliance

This design cannot
• Substitute for a poor question
• Makeup for inadequate statistical analysis
• Guard against poor experimental technique
Ethical Problems of RCTs

• Use of placebos
• Randomization is an emotionally charged issue
  – Physician may not have full knowledge about patient’s treatment
  – Patients object to being treated like guinea pigs
Practical Problems of RCTs

• Patients will not accept randomization
  – Concept of equipoise
  – Newer is better

• Experience of the breast cancer study comparing high dose chemotherapy +/- bone marrow transplantation
RCT Study Designs
Parallel Design

Eligible Individuals

Randomize

Treatment Group

Control Group

Analyze
Cross-over Design

Eligible Individuals

Randomize

Treatment Group

Washout period

Control Group

Washout period

Control Group

Washout period

Treatment Group

Analyze

Analyze
Factorial Design

Eligible Individuals

Randomize

Drug A + Drug B

Drug A + Placebo B

Drug B + Placebo A

Placebo A + Placebo B
Special Considerations

Subgroup Analysis
Equivalence, Superiority and Non-inferiority studies
Endpoints
Subgroup Analysis
Subgroup Analysis

• Clinical trials represent a major investment by investigators and sponsors – therefore they attempt to gain the maximum from them

• Meaningful information from subgroup analysis is restricted by:
  – Multiplicity
  – Low statistical power
# Multiplicity

## Probability of at least one significant result at the 5% significance level

<table>
<thead>
<tr>
<th>Number of subgroups</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>0.14</td>
</tr>
<tr>
<td>5</td>
<td>0.23</td>
</tr>
<tr>
<td>10</td>
<td>0.40</td>
</tr>
<tr>
<td>20</td>
<td>0.64</td>
</tr>
</tbody>
</table>
Low Statistical Power

• Most studies enroll just enough participants to ensure that the **primary hypothesis** can be tested.

• Therefore subgroup analysis will only have power to detect **substantially larger** effects on the same endpoint.

• If a specific subgroup is of interest the study can be powered to detect the outcome in this subgroup (**a priori**).

• Proper subgroup analysis will test the hypothesis whether the treatment effect in a subgroup is different from that of the overall population.
Superiority, Non-inferiority and Equivalence Trials
Superiority Trials

- Designed with the expectation that the one drug is better than a control or placebo
- Designed to rule out equality between treatments
- Remember the lack of evidence of a difference ≠ evidence of a lack of difference
Interpretation of Superiority Trials

If the CI include the null value, it can on only be stated that the drug is not superior to the other.

CI does not include the null value.
Equivalence Trials

• Goal is to rule out all differences of clinical importance between two treatments
• Establishing equivalence requires a determination of what specifically constitutes a clinical important difference
• Decision of what constitutes a clinical important difference should be made at the trial design
Equivalence Trials

• Advance stating of **numerical limits** for allowable difference between 2 treatments is essential

• Any difference within the limits will indicate that the 2 treatments are equivalent

• **90% CI** are usually used in equivalence trials
Interpretation of Equivalence Trials

90% CI completely within limits, therefore equivalent

90% CI crosses the limit therefore not equivalent

Predefined Delta

Lower limit -10

Upper limit +10
Non-inferiority Trials

- Aim is to indicate that a treatment is at least as effective and not worse than the existing standard treatment.
- The lower limit of effectivity is set in advance:
  - According to minimally important effect (subjective) or
  - Based on previous trials.
Non-inferiority Trials

- Sample size is smaller than superiority trials but considerably larger than placebo controlled trials
- Both non-inferiority and superiority can be assessed in the same trial, but limits should be set in advance
Interpretation of Non-inferiority Trials

90% CI includes the lower limit thus can be inferior

90% CI does not cross the lower limit, thus non-inferiority confirmed
Endpoints

Surrogate endpoints
Composite endpoints
Primary & secondary
“Primary” and “Secondary”

• Primary Endpoints
  ▪ These endpoints define the disease in the sense that an experimental drug that does not show superiority over placebo for all of these endpoints is not a viable treatment for the disease under study

• Secondary Endpoints
  ▪ These endpoints, although not considered primary, are considered important to prescribing physicians in helping to identify the ideal treatment for each of their patients
Surrogate Endpoints

A *surrogate endpoint* does not directly measure any clinical benefit to patient, it only predicts the outcome.
Surrogate Endpoint

• Why do we use surrogate endpoint?
  – Can be measured earlier
  – Convenient or less invasive
  – Can be measured more frequently
  – Can accelerate the approval process

• Advantages:
  – May reduce the size of clinical trials
  – May shorten the duration of clinical trials
  – May reduce the cost of clinical trials
Relationships Between Treatment, Surrogate and Clinical Outcome
Surrogate Endpoints in Practice

• The surrogate must be in the **causal pathway** of the disease process
• An intervention’s entire effect on the clinical outcome of interest should be **fully captured** by the surrogate
• We need to assess more than a single study to decide on the adequacy of a surrogate
Guideline in Reviewing Medical Literature on Surrogate Endpoint

- Is there a strong, independent, consistent association between the surrogate and clinical endpoint?
- Is there evidence from randomized trials in other drug classes that improvement in the surrogate has consistently led to improvement in the outcome?
- Is there evidence from randomized trials in the same drug classes that improvement in the surrogate has consistently led to improvement in the outcome?
- What were the results? How large, precise, and lasting was the treatment effect?
Composite Endpoints

- A combination of measurements is viewed as a single primary endpoint.
- Used in studies of diseases that occur rarely or that may require a sample size so large that it cannot be completed in a reasonable time.
- Components of composite endpoints must be related to a common pathophysiology and be closely related in the degree of severity.
Common Composite Endpoints

- **CV**: First event of MI, Stroke, CABG, Hospitalization, Death
- **Diabetic Nephropathy**: Decreased Renal Function, End Stage Renal Disease, Death
- **Oncology**: Progression or Death
Problems with Composite Endpoints

- Conflicting results
- The combination of less important events can lead to the overestimation of the clinical importance thereof
- All composite endpoints should have the same pathophysiological mechanism.
Composite Endpoints: Components

• How to interpret components?
  – Significant in one and weak in others
  – None significant, but all in right direction
  – Should you analyze components individually?

• Question may be:
  – Does the drug do something? Vs. What does the drug do?

• Number of components may impact interpretation
Composite Endpoints: Components

- Is the composite a measure of the disease (individual components do not fully measure the disease) or is it for convenience of analysis?
  - Sparse events
  - Competing risk
  - Multiplicity

- Are the events surrogates for other events or surrogates for something else?
  - CV events are an outcome of underlying disease
  - Diabetic nephropathy increasing severity of disease
Analysis of RCT
A Bit of Statistics

P-value
Nil-hypothesis
Statistical significance
Confidence interval
Error
The Magical P-value

Yeah!!!
P-value

• The probability that the finding is due to chance
• If $p < \alpha$ (threshold value) we assume that the finding is not due to chance (statistical significant)
• The threshold value can be selected, but according to convention 0.05 is accepted
• If $p = 0.05$ then there are still a 5% probability that the finding is due to chance
### Null Hypothesis

The null hypothesis states that no difference exists or that the finding is due to chance.

<table>
<thead>
<tr>
<th>COMMON LANGUAGE</th>
<th>STATISTICAL STATEMENT</th>
<th>CONVENTIONAL TEST THRESHOLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Statistically significant”</td>
<td>The null hypothesis was rejected.</td>
<td>$P &lt; 0.05$</td>
</tr>
<tr>
<td>“Unlikely due to chance”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Not significant”</td>
<td>The null hypothesis could not be rejected.</td>
<td>$P &gt; 0.05$</td>
</tr>
<tr>
<td>“Due to chance”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Statistical Significance

Statistical significance ($p<0.05$) depends on 3 factors:

– **Magnitude** of the effect
– **Number of observations** – “big studies make small differences significant”
– **Spread** of the data

Statistical significance $\neq$ clinical importance
Chance (random error) is much less of a threat than bias (systematic error)
Confidence Interval

- CI is the range of certainty of the sample where-in the true population central value lies.
95% CI

• This states that we are 95% confident that the true population mean (RR, OR, etc.) lies between these values.

• CI are influenced by:
  – The number of observations (the closer the sample size to the population size, the narrower the CI).
  – The spread in the data.
Sample Size vs. CI

Number of coin tosses

- 500 heads, 500 tails
- 50 heads, 50 tails
- 25 heads, 25 tails
- 5 heads, 5 tails
- 1 head, 1 tail
CI vs. p-value

If the 95% CI includes no difference between groups, then the \( P \) value is > 0.05.

If the 95% CI does not include no difference between groups, then the \( P \) value is < 0.05.

OR = 1, RR = 1
Difference in means or proportions = 0
Error

<table>
<thead>
<tr>
<th>STUDY CONCLUSION</th>
<th>“TRUTH”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIFFERENCE</td>
</tr>
<tr>
<td>“Positive” study</td>
<td>True positive</td>
</tr>
<tr>
<td>(significant difference)</td>
<td></td>
</tr>
<tr>
<td>“Negative” study</td>
<td>Type II error</td>
</tr>
<tr>
<td>(no significant difference)</td>
<td></td>
</tr>
</tbody>
</table>
Error

• Type 1 error can be reduced by lowering the significant threshold \( (\alpha) \) from 0.05 to e.g. 0.01.

• Type 2 error is a function of sample size and can be reduced by increasing the sample size.
Relationship Between the CI and Type 2 Error

The narrower the CI the less likely a type 2 error.
Reporting of RCT

Consort Statement
Threats to Internal Validity

- **History** - people change during the course of a study. Sometimes they get better, sometimes worse, and sometimes they die.
- **Testing** - subjects can often learn to interact with the testing instruments.
- **Instrumentation** - the results often depend on the quality of the test instruments.
- **Compound effects** - are all other factors accounted for?
Warning!

Acting like a Doctor can threaten the internal validity of your study