Experimental designs: Randomized controlled trials

F. Farrokhhyar, MPhil, PhD
Objectives

- To understand the design features of randomized controlled trials (RCT)
- To learn how to minimize biases in RCTs.
- To understand the methodological challenges of conducting RCTs of surgical interventions.
What distinguishes experimental studies from observational studies?

- Its ability to control for confounding factors, i.e., tumor grade, age, tumor size, etc.
- In a prospective cohort study, one might measure all potential known confounding factors, but...
- One cannot control for unknown and unanticipated confounder factors.
“Explain to me how comparing apples and oranges is fruitless.”
Randomization

Control Group

Investigational Group

Randomization
Benefits of randomization?

- Randomization gives everyone equal chance of receiving experimental or control treatment. This eliminates *selection* bias.

- The process of randomization controls for the *known* and *unknown* confounding factors.

- It provides comparable groups so that the differences in the outcomes at the end of the trial can be attributed to the treatments under investigation.
Randomized controlled trials

- Randomized controlled trials are quantitative, comparative, and controlled experiments in which investigators study a series of individuals who receive a randomly assigned treatment.

- Randomized controlled trials are the most rigorous design to determine a cause-and-effect relationship.

- They play a central role in providing a basis for evidence-based practice and clinical decision-making.

- They are appropriate only for mature research questions.
Basic design of a RCT

Reference population

Eligibility criteria

Study population

Eligibility criteria

Randomization

New treatment

Follow-up

Improve

Do not improve

Current treatment

Follow-up

Improve

Do not improve
In a well-designed RCT:

- Potential sources of random error and systematic errors (bias) are controlled.
  - Random error is the same for study groups (1:1).
  - Random error creates “noise” or “variation” in the data and results in underestimating the association between experimental treatment and outcome.
  - Random error can be reduced by increasing the number of observations.
  - Systematic error or bias is a reproducible inaccuracy that is different for study groups
    - Systematic error deviates the results of the study from the truth.
    - Once occurred, it is impossible to estimate the extent or the direction.
Potential sources of bias in RCTs

- **Selection bias** – systematic differences in prognostic factors between study groups.
- **Performance bias** – systematic differences in care provided to study groups.
- **Detection bias** – systematic differences in outcome assessment between study groups.
- **Attrition bias** – systematic differences in withdrawal from the study or loss to follow-up between study groups.
Methods to minimize bias or systematic error

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Randomization</td>
</tr>
<tr>
<td>2</td>
<td>Concealment of allocation</td>
</tr>
<tr>
<td>3</td>
<td>Blinding</td>
</tr>
<tr>
<td>4</td>
<td>Complete follow-up</td>
</tr>
<tr>
<td>5</td>
<td>Intention-to-treat approach</td>
</tr>
</tbody>
</table>

Methods to minimize random error

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Sample size calculation or power analysis</td>
</tr>
</tbody>
</table>
1- Randomization

- Methods of generating randomization sequence
  - Quasirandom methods - date of admission, chart number
  - True random methods - toss of coin, toss of dice, random number table,
    
    **Computer generated randomization sequence**

- Timing of randomization is very important to be close to when the intervention is performed.
  - if there is a lag time between randomization and surgical intervention, an eligible patient might die, experience outcome or become ineligible.
Different methods of randomization

- Most commonly used methods are:
  - Unrestricted or simple randomization
  - Restricted blocked randomization – fixed block sizes vs. random blocks sizes
  - Stratified randomization
Simple randomization

- Each eligible patient will be randomly assigned to one of the treatments until the required numbers are assigned.

\[
\begin{array}{c}
\text{=} \quad \text{control (10)} \\
\text{=} \quad \text{treatment (16)}
\end{array}
\]

- One major disadvantage is that there could be substantial imbalances in numbers of patients in treatment groups (and in prognostic factors), particularly in smaller trials (<100).

- Such imbalances do not cause the statistical tests to be invalid, but they reduce the ability to detect the true differences and lead to the loss of trial’s credibility.
Blocked randomization

- Blocked randomization is used to ensure balanced groups.
- The block size is selected by study statistician/researcher.
  - Smaller increments should be used for small trials to ensure balanced groups.
  - Block size should be a multiple of the number of treatment groups. (i.e. with 2 treatment groups, block size of 4, 6, 8)
- One block size is selected, all possible balanced combinations of assignment are calculated for the chosen block size.
- Then, one of the block combinations is randomly chosen to determine the patient’s assignment into the treatment groups.
- This process of random block selection is repeated until all patients are randomly assigned into the study.
Stratified randomization

- Before planning the randomization process, we need to determine if there are any prognostic factors that are strongly associated with patient response or outcome measure.

- Such factors should be considered for stratification.

- Stratified randomization prevents an imbalance between treatment groups for factors that influence treatment responsiveness.

- It maintains credibility and internal validity.

- In large multi-centre trials, each centre is considered for stratification to prevent imbalance and also because centers usually vary with respect to patient population, and quality and type of care given to patients.

- Blocking and stratifying should be taken into account during data analyses.
Concealment of allocation relates to what happens before randomizing patients into the study and eliminates *selection* bias.

It is possible to conceal the randomization sequence in every RCT.

The benefit of randomization is lost if individuals are aware of which treatment participants will receive.

Most optimal method is to have the randomization process independently administered:

i.e. 24-hour telephone randomization line, web-based randomization service, or hospital pharmacy

Convenient and least optimal method is use of envelopes.

Use of envelopes is highly susceptible to corruption and should be avoided.

**Schulz & Grimes, 2002**
Use of envelopes

- If envelopes must be used, the potential for abuse must be minimized.
- To maintain concealment, the envelopes should be:
  - Opaque, sealed, signed and serially numbered.
  - Prepared for each block in a systematic fashion
  - Opened sequentially and one block at a time.
  - Opened sequentially within each block
  - Before opening the envelope, patient identification number and name should be written on the envelope and on the cue card once opened
  - Kept in a locked and secure place.
  - An individual independent of study should perform randomization.
3- Blinding

- Blinding relates to what happens after randomizing patients into the study and minimizes *performance* bias and *detection* bias.

- Bias occurs when individuals are aware of which intervention participants have received and treat them differently.

- Every effort should be made to blind as many involved individuals as possible.

- Blinding individuals involved in the trial prevents systematic imbalances in effective outcome evaluations and concomitant interventions.

- Blinding is easily achieved in drug trials with placebo but it is problematic in trials of surgical interventions when:
  - interventions result in different incisions and scars
  - surgical technique is compared to medical management.
Blinding in surgical trials

- Surgeons most often cannot be blinded.
- Participants, care providers and data collectors can often be blinded.
- Data analyst can always be blinded.
- Blinding outcome assessors is most important. It protects a trial against the differential assessment of the outcomes especially when primary outcome is subjective.
- The outcome assessors might assess the outcomes differently if they are aware of treatment allocation.
- Surgical researchers rarely incorporate blinding in surgical trials.
- Karanicolas *et al* (2008) systematically reviewed trials in orthopaedic trauma and concluded that less than 10% of trials blinded outcome assessors.
To optimize blinding in surgical trials

- Blinding outcome assessors may be achieved using simple and creative techniques:
  - Using placebo surgery or sham surgery if ethical concerns are justified.
  - Concealing incisions or scars with large dressing
  - Using two or more independent assessors unaware of the treatment allocation to assess the outcomes.
  - Altering digital radiographs or images.
4- Complete follow-up

- Complete follow-up for participating patients avoids attrition bias and maintains balanced groups.

- Failure to account for all patients at the end of the study presents a major threat to internal validity.

- Patients who attend follow-up visits are usually different from those who do not.

- Losses to follow-up is greater and differential when:
  - No treatment is required after surgical treatment, especially when a longer follow-up is required.
  - Concomitant interventions (i.e., Physiotherapy) are required for one arm but not the other.
  - Patients are not blinded to treatment allocation.

- Some approaches are suggested to enhance complete follow-up (Schulz 2002, Thoma 2008) at the stage of design and the conduct of the trial.

- Different methods are used to handle losses to follow-up at data analysis.
5- Intention-to-treat principle

- Intention-to-treat approach analyzes participants in the groups to which they were randomized, irrespective of the treatment they have received.

- It maintains the benefits of randomization and provides the least biased assessment of the efficacy of the treatment.

- Analyzing data by the treatment patients have received (per protocol analysis) introduces prognostic imbalances between study groups, and causes the loss of the benefits conferred by randomization.

- In surgical trials, some patients are switched from new treatment to control treatment for practical reasons i.e. surgeon’s limited experience or patient’s condition.

- The complications in these patients are usually higher than others.
Learning curve for a new surgical treatment

If we fail to control for learning curve, the treatment effect may be biased toward the null hypothesis.
To minimize the effect of learning curve

- The following might be considered in the design of surgical trials:
  - General training in the area
  - Pre-specified number of cases in life-time
  - Number of cases in the year preceding the trial
  - Outcomes consistent with good clinical practice
  - Assessing skills (e.g., videotapes, direct observation, quality scores for specimens).
  - Expertise-based design
Expertise-based trials

- Patients are randomized to different surgeons, some of whom deliver only the control treatment and some only the experimental treatment.
- The advantage of expertise-based design is that it minimizes the effect of learning curve.
- This design is preferable for large trials. However, there are major drawbacks.
  - Multiple surgeons are needed who are strongly in favour of each treatment. i.e. each participating centre should have expert surgeons doing each type of surgery
  - Adds cost
  - Increases complexity of the randomization.
Steps in design of a RCT — PICOT

- Consider all aspects of ethics and confidentiality.
  - Can RCT design answer the research question?
  - Is random allocation of patients to study treatments ethical, practical and possible?
- Define reference population – external validity
- Define study participants – set inclusion and exclusion criteria
  - High incidence - high risk for outcome
  - Likely to benefit and not be harmed
  - Likely to adhere
- Define intervention treatment – should have some evidence of effectiveness.
- Define control treatment – placebo or standard treatment
- Define outcome measures; primary and secondary – patient-important
- Define follow-up time
Steps in design of a RCT --- methods

- Search for systematic reviews/meta-analyses
- Plan a feasibility or pilot study
- Plan randomization method
- Plan methods of concealing random allocation sequence
- Plan methods of blinding the treatment allocation
- Plan follow-up visits and how to maximize compliance
- Plan the measurement of appropriate baseline variables (for confounding and/or interaction effect)
- Calculate sample size
- Plan statistical analyses, subgroup and/or sensitivity analyses.
- Write study protocol.
Importance of pilot studies

- Pilot studies do not guarantee the success of the main trial but they do increase the likelihood of success. They help in:
  - Assessing whether the research protocol is realistic and workable
  - Assessing the feasibility of full-scale trial.
  - Assessing the likely success of proposed recruitment approaches
  - Identifying logistical problems which might occur using proposed methods.
  - Further developing the research question and proposed research methods
  - Estimating variability in outcomes to help determine sample size
  - Determining what resources, such as financial support and research staff, are needed for a full-scale trial.
  - Convincing funding bodies that the main trial is feasible and worth supporting.
Example of an RCT – rationale

Dancey Al., Cheerma M; Thomas SS; Plast Reconstr Surg. 2010 May;125(5):1309-17

The extended latissimus dorsi is a workhorse flap and plays an important role in breast reconstruction. Unfortunately, seromas at the flap donor site are a frustrating problem complicating many procedures. The purpose of this study was to evaluate the efficacy of a combination of fibrin sealant versus limited quilting sutures at reducing seroma formation in patients requiring latissimus dorsi breast reconstruction.

Methods: This was a prospective, randomized trial. Randomization was achieved by means of computer-generated numbers. A sealed envelope was opened in OR to reveal the treatment allocation. Patients were randomized to receive either quilting sutures only (group 1) or a combination of fibrin sealant plus marginal quilting sutures (group 2). Patients and researchers were blinded to randomization. The primary outcome was the proportion of symptomatic seroma. Twenty-six patients were enrolled in the study, and all were followed for a period of 6 months.

Is this a well-designed trial? Is it controlled for random and systematic errors?
Example of an RCT

We plan to conduct a multi-centre randomized controlled trial to compare the proportion of symptomatic seroma between fibrin sealant and quilting sutures.

Clinical hypothesis – symptomatic seroma is less frequent in fibrin sealant than quilting sutures surgery.

Methods: Consecutive patients requiring latissimus dorsi breast reconstruction will be recruited. Patients will randomly receive either fibrin sealant or quilting sutures treatment. Random blocked randomization sequence will be computer generated for each centre using block sizes of 4, 6, and 8. Web-based randomization method will be used to randomize eligible patients after obtaining informed consent. Outcome assessors, care-givers and patients will be blinded to the type of treatment using same wound dressing. Patients will be followed up for 6 months and symptomatic seroma will be determined in each group.

The calculated sample size for an alpha error of 0.05, beta error of 0.8 and 9% clinically important difference in the proportion of symptomatic seroma was 187 per group. 14% symptomatic seroma was assumed for quilting sutures. We decided to recruit 400 patients to ensure 80% power throughout study: 200 in each group.

Is this the most optimal design?
Patients requiring latissimus dorsi breast construction

- Fibrin sealant: 200
  - Symptomatic seroma:
    - Yes = 10
    - No = 190
- Quilting sutures: 200
  - Symptomatic seroma:
    - Yes = 18
    - No = 182
Arrange our 2 x 2 table and estimate risk

<table>
<thead>
<tr>
<th></th>
<th>Outcome</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Experimental treatment</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>Control treatment</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td>N</td>
</tr>
</tbody>
</table>

\[
EER = \frac{a}{(a + b)} \\
CER = \frac{c}{(c + d)} \\
RR = \frac{EER}{CER} \\
RRR = (EER - CER) / CER \\
ARR = EER - CER \\
NNT = \frac{100}{ARR}
\]
Estimating risk

- **EER (Experimental Event Rate):** event rate in new treatment group
- **CER (Control Event Rate):** event rate in control group
- **RR (Relative Risk):** the ratio of the event rate in new treatment group to the event rate in control group.
- **RRR (Relative Risk Reduction):** It is the effect of new treatment that reduces the probability of the bad outcome. It is the difference in event rates in relative terms.
- **ARR (Absolute Risk Reduction):** the proportion of reduction in the outcome that can be attributed to the new treatment. It is the difference in event rates in arithmetic terms.
- **NNT (Number Needed to Treat):** number needed to treat to prevent one bad outcome. It is the reciprocal of the ARR.

number needed to benefit versus number needed to harm
Arranging data and estimating risk

Symptomatic seroma

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibrin sealant</td>
<td>10</td>
<td>190</td>
</tr>
<tr>
<td>Quilting sutures</td>
<td>18</td>
<td>182</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>372</td>
</tr>
</tbody>
</table>

\[ EER = \frac{a}{a+b} = \frac{10}{200} = .05 = 5\% \]
\[ CER = \frac{c}{c+d} = \frac{18}{200} = .09 = 9\% \]

\[ RR = \frac{EER}{CER} = \frac{0.05}{0.09} = .56 \ (0.26, 1.17) \]
\[ RRR = \frac{CER - EER}{CER} = \frac{0.09 - 0.05}{0.09} = .44 = 44\% \]

The risk of symptomatic seroma is twice lower in fibrin sealant than in quilting sutures.
The risk of symptomatic seroma is reduced by 44% from fibrin sealant compared to quilting sutures.
### More on risk ...

**Symptomatic seroma**

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fibrin sealant</strong></td>
<td>10</td>
<td>190</td>
<td>200</td>
</tr>
<tr>
<td><strong>Quilting sutures</strong></td>
<td>18</td>
<td>182</td>
<td>200</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>28</td>
<td>372</td>
<td>400</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
EER &= \frac{a}{a+b} = \frac{10}{200} = .05 = 5\% \\
CER &= \frac{c}{c+d} = \frac{18}{200} = .09 = 9\%
\end{align*}
\]

**Absolute risk reduction** = 9% - 5% = 4% (-0.01, 0.09%)

**Number Needed to benefit** = 100/4 = 25 (95% CI: 11 to 78)

Absolute benefit from fibrin sealant is 4%.

25 patients need to be treated with fibrin sealant to prevent one symptomatic seroma.
Assignment
Assignment—rationale and PICOT

Acute hospital acquired hyponatremia is common and clinically important. The contribution of common maintenance parenteral solutions (PMS) to its occurrence continues to be a debated subject in pediatrics. We aimed to evaluate the risk of acute hyponatremia following isotonic (0.9% NS) compared to hypotonic (0.45% NS) PMS, when administered for a maximum of 48 hours on post-operative pediatric patients.

**Design** = fully blinded randomized controlled trial.

**Patient population** = Surgical patients aged 6 months to 16 years with an expected post-operative stay of greater than 24 hrs were eligible if they did not meet any of the exclusion criteria.

**Interventions** = isotonic PMS and hypotonic PMS

**Primary outcome** = the proportion of acute hyponatremia.

**Secondary outcomes**: severe hyponatremia, hypernatremia, adverse events attributable to acute PNa changes, proportion who changed to open-label PMS, etc.

**Time** = hospital discharge
**Example of an RCT – design**

**Methods:**
Randomization sequence was computer generated by the study statistician. Stratified (admission wards [PCCU and SW]) random blocked randomization was used. The intervention was independently administered by the research pharmacist. Patients, medical and research staff, investigators and members of Data Safety Monitoring Committee were masked to the group assignment.

**Sample size** = We calculated that 206 patients were required to detect a 20% absolute difference in the rate of acute hyponatremia using a chi-squared test (assuming an intention-to-treat principle) with a two-sided alpha level of 0.05 and a statistical power of 80%. Assuming a 25% loss to follow up or inability to measure the primary outcome, our total sample size was augmented to 258 (129 patients per group using 1:1 allocation ratio).

**Statistical analysis** = The statistical analysis of the primary outcome was conducted using the generalized linear regression model using intention-to-treat principle, and then according to the treatment received. We used multiple imputations to handle missing data.
Example of an RCT – design

Sensitivity analyses were performed on patients for whom complete data for primary outcomes was available. The results are reported as proportion or mean difference with 95% confidence intervals (95% CI) with associated p-values. ...... An alpha of 0.05 was considered for statistical significance.

The data are summarized in the table. Answer the following questions.

<table>
<thead>
<tr>
<th></th>
<th>hyponatremia</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>28</td>
</tr>
<tr>
<td>no</td>
<td>100</td>
</tr>
<tr>
<td>Isotonic</td>
<td>128</td>
</tr>
<tr>
<td>Hypotonic</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>258</td>
</tr>
</tbody>
</table>
Questions?

1- Was the study powered to answer the research question? What type of error is introduced when sample size is not calculated?
   Yes, the sample size was calculated. Random error.

2- Was generating randomization sequence appropriate? How was this generated?
   Yes, by computer

3- Was the treatment allocation concealed from the individuals who were involved in the study? What method did the investigators use?
   Yes, they used hospital pharmacy for concealment.
Questions?

4- Was the outcome assessment and care given to patients unbiased? If yes, how did the investigators achieve this?

Yes, by using opaque bags to mask the study treatments.

5- Was there a high chance of losses to follow-up and why?

No, the follow-up time was up to discharge.

6- Was the appropriate method of statistical analysis performed and why?

Yes, regression analysis was used to account for stratification and blocked randomization, multiple imputation was used to handle and replace the missing data, sensitivity analysis was used on patients with complete data for primary outcome.
### Calculate EER, CER, RR and RRR

- **EER** = \( \frac{a}{a+b} = \frac{29}{128} = .22 = 22\% \)
- **CER** = \( \frac{c}{c+d} = \frac{53}{130} = .41 = 41\% \)
- **RR** = \( \frac{EER}{CER} = \frac{0.22}{0.41} = .54 \) (0.38, 0.81)
- **RRR** = \( \frac{CER - EER}{CER} = \frac{0.41 - 0.22}{0.41} = .46 = 46\% \)

<table>
<thead>
<tr>
<th></th>
<th>hyponatremia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>Isotonic</td>
<td>28</td>
</tr>
<tr>
<td>Hypotonic</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>82</td>
</tr>
</tbody>
</table>

The risk of hyponatremia is twice lower for isotonic PMS than it is for hypotonic PMS.
The risk of hyponatremia is reduced by 46% from isotonic PMS compared to hypotonic PMS.
### 9&10 - Calculate ARR and NNT

<table>
<thead>
<tr>
<th></th>
<th>hyponatremia</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isotonic</td>
<td>29</td>
<td>99</td>
<td>128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypotonic</td>
<td>53</td>
<td>77</td>
<td>130</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>82</td>
<td>176</td>
<td>258</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
EER = \frac{a}{a+b} = \frac{29}{128} = 0.23 = 23\
\]

\[
CER = \frac{c}{c+d} = \frac{53}{130} = 0.41 = 41\
\]

**Absolute benefit of isotonic** = 41% - 22% = 19% (7% to 29.5%)

**NNT to prevent one bad outcome** = \(\frac{1}{0.18} = 5.26 = 6\) (4, 13)

Absolute benefit from isotonic PMS is 18%.

6 patients need to be treated with isotonic PMS to prevent one hyponatremia.
11- What is the level of evidence and are the findings generalizable to reference population?

level 1, it might be generalizable to reference population. However, it is a single centre study and the findings might not apply to other centres.