How to systematically approach the evaluation of a randomized controlled trial

Dr Indranil Mallick
Tata Medical Center, Kolkata
What is a randomized controlled trial?

A study design that **randomly** assigns participants into an **experimental** group or a **control** group.

As the study is conducted, the only expected difference between the control and experimental groups in a randomized controlled trial (RCT) is the intervention being studied.
You *don’t* need to be an expert or statistician to interpret an RCT.
The Question

- Control Arm
- Experimental Arm
- Endpoints

Is the question correct and clinically meaningful?

The Methods

- Patient Selection
- Randomization
- Sample size

Are the methods appropriate?

The Results

- Control Arm Rx
- Exp Arm Rx

Analysis and Results

Are the results correctly analyzed?

Discussion

- Context

Are the results explainable and merit implementation?
The Question

Is the question correct and clinically meaningful?

Control Arm

Endpoints

Experimental Arm

A wrong question cannot have a right answer

Is the control treatment the current standard of care?

Is the experimental treatment logical, safe and implementable?

Are the endpoints clinically meaningful?
Is the control treatment the current standard of care?

- Unless the control arm represents the current standard of care, the trial may not provide a clinically meaningful answer.
- Check for details:
  - Drug dose schedules
  - Radiation volumes, dose-fractionation, techniques
  - Surgical details
- The control arm may need updating during the course of the study if standard of care changes.
The Question

Is the experimental treatment logical, safe and implementable?

- Is there a biological/clinical justification in using this experimental arm?
- Is there Phase I/II data that suggests safety/efficacy?
- Is the treatment schedule consistent with known usage?
  - Drug dose schedules
  - Radiation volumes, dose-fractionation, techniques
  - Surgical details
Are the endpoints valid and clinically meaningful?

- **Clinically meaningful endpoints** – overall survival and quality of life.
- **Surrogate endpoints often do not correlate with OS** – response rates, disease-free survival, progression free survival, biochemical control, metastasis free survival.
- **Toxicity endpoints** – valid only if reported by patients
- **Secondary endpoints** – only hypothesis generating
Long term results of RTOG 91-11
The Methods

- Is the patient selection criteria externally valid?
- Is the sample size appropriately calculated?
- Is the process of randomization robust?
- Was the treatment processes as specified and quality assured?
The Methods

Is the patient selection criteria reflective of common practice?

Inclusion criteria – is it accommodating the range of stages that matter?

Exclusion criteria – is it excluding a lot of patients with comorbidities?

Is this the true population where you are looking to use this new treatment?
Is the sample size appropriately calculated?

Calculated for the primary endpoint. Depends upon:

- The relative difference expected (Hazard Ratio, or likelihood of event in exp vs control arm)
- The specified type I and type II errors/power
- Duration of recruitment and follow up
Is the process of randomization +/- blinding robust?

Look for the randomization method, especially in smaller single-institution RCTs. Confirm allocation concealment. Is the randomization stratified using important variables? Or randomization with minimization?

Blinding reduces biases in reporting, assessment and surveillance.
The importance of treatment QA is underestimated especially for multi-institutional studies:

- Pathological/molecular characteristics
- Radiation treatment planning
- Surgical techniques/training
- Drug storage/administration/PD-PK studies

Was the treatment processes as specified and quality assured?
What happens to patients in the control arm if they fail?

Are the patients offered the standard salvage therapy (if necessary, with the therapy in the experimental arm – ‘crossover’)?

How soon are they offered salvage therapy?
The Results

What are the patient and disease characteristics?

Does the analysis accounts for all patients?

Is the treatment compliance and toxicity profile reported?

Is it an intention-to-treat analysis?
### The Results

#### Analysis and Results

**What are the patient and disease characteristics?**

- Is there a balance between the arms in terms of stage and risk factors of recurrence? (if there is no stratification in randomization)
- Do they reflect the usual distribution in your practice?

<table>
<thead>
<tr>
<th></th>
<th>Control (n=1029)</th>
<th>Radiotherapy (n=1032)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at randomisation (years)</strong></td>
<td>68 (63-73)</td>
<td>68 (63-73)</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>37-86</td>
<td>45-87</td>
</tr>
<tr>
<td><strong>WHO performance status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>732 (71%)</td>
<td>734 (71%)</td>
</tr>
<tr>
<td>1-2</td>
<td>297 (29%)</td>
<td>298 (29%)</td>
</tr>
<tr>
<td><strong>Pain from prostate cancer</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>820 (81%)</td>
<td>844 (83%)</td>
</tr>
<tr>
<td>Present</td>
<td>198 (19%)</td>
<td>170 (17%)</td>
</tr>
<tr>
<td>Missing data</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td><strong>Previous notable health issues</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>67 (7%)</td>
<td>57 (6%)</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>29 (3%)</td>
<td>30 (3%)</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>5 (&lt;1%)</td>
<td>8 (1%)</td>
</tr>
<tr>
<td>Angina</td>
<td>46 (4%)</td>
<td>51 (5%)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>408 (40%)</td>
<td>440 (43%)</td>
</tr>
<tr>
<td>Missing data</td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

**T category at randomisation**

- T0: 0 (0%) vs 2 (<1%)
- T1: 12 (1%) vs 12 (1%)
- T2: 84 (9%) vs 89 (9%)
- T3: 585 (62%) vs 603 (63%)
- T4: 260 (28%) vs 246 (26%)
- TX: 88 vs 80

**N category at randomisation**

- N0: 345 (36%) vs 344 (36%)
- N+: 620 (64%) vs 620 (64%)
- NX: 64 vs 68
Does the analysis accounts for all patients?

Important differences in the proportion of patients who are lost to follow up and analyzed – poor quality of a study.
Is the treatment compliance and toxicity profile reported?

### Table 2. Adverse Events.*

<table>
<thead>
<tr>
<th>Event</th>
<th>Nivolumab plus Chemotherapy (N = 176)</th>
<th>Chemotherapy Alone (N = 176)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Any Grade</td>
<td>Grade 3 or 4</td>
</tr>
<tr>
<td>Adverse events of any cause — no. (%)†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>163 (92.6)</td>
<td>72 (40.9)</td>
</tr>
<tr>
<td>Leading to discontinuation of treatment</td>
<td>18 (10.2)</td>
<td>10 (5.7)</td>
</tr>
<tr>
<td>Serious</td>
<td>30 (17.0)</td>
<td>19 (10.8)</td>
</tr>
<tr>
<td>Treatment-related adverse events — no. (%)†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>145 (82.4)</td>
<td>59 (33.5)</td>
</tr>
<tr>
<td>Leading to discontinuation of treatment</td>
<td>18 (10.2)</td>
<td>10 (5.7)</td>
</tr>
<tr>
<td>Serious</td>
<td>21 (11.9)</td>
<td>15 (8.5)</td>
</tr>
<tr>
<td>Death§</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Surgery-related adverse events — no./total no. (%)§</td>
<td>62/149 (41.6)</td>
<td>17/149 (11.4)</td>
</tr>
</tbody>
</table>
Is it an intention-to-treat analysis?

- **ITT analysis** - once randomized, always analysed in the randomized group
- Regardless of their
  - adherence with the entry criteria
  - treatment they actually received
  - subsequent withdrawal from treatment or deviation from the protocol
- Non ITT analyses removes the benefit of the balancing provided by randomization.
  - **Per-protocol analysis**: only those patients who were treated according to protocol
  - **As-treated analysis**: analysed on the basis of the treatment they actually received.
The Results

Looking at results

• Focus on the primary endpoint
• Look at the hazard ratio – point estimate – likelihood of relative benefit
• Look at the confidence intervals – estimate of precision of the point estimate
• Look at the Kaplan Meier curve do the curves truly reflect a difference.
• Look at the p-value last
• Look at secondary endpoints in context
• Focus on the primary endpoint
• Look at the hazard ratio – point estimate – likelihood of relative benefit
• Look at the confidence intervals – estimate of precision of the point estimate
• Look at the Kaplan Meier curve do the curves truly reflect a difference.
• Look at the p-value last
• Look at secondary endpoints in context
Are the right conclusions being drawn?

Statistical significance vs. clinical significance

Have the conflicts of interest been reported?

‘Tolerable safety profile’ ‘Standard of care’

Are differences with other studies explained?

Requires subject matter expertise
Statistical significance vs. clinical significance

Erlotinib Plus Gemcitabine Compared With Gemcitabine Alone in Patients With Advanced Pancreatic Cancer: A Phase III Trial of the National Cancer Institute of Canada Clinical Trials Group

Malcolm J. Moore, David Goldstein, John Hamm, Arie Figer, Joel R. Hecht, Steven Gallinger, Heather J. Au, Pawel Murawa, David Walde, Robert A. Wolff, Daniel Campos, Robert Lim, Keyue Ding, Gary Clark, Theodora Voskoglou-Nomikos, Mieke Ptasynski, and Wendy Parulekar

Results
A total of 569 patients were randomly assigned. Overall survival based on an intent-to-treat analysis was significantly prolonged on the erlotinib/gemcitabine arm with a hazard ratio (HR) of 0.82 (95% CI, 0.69 to 0.99; P = .038, adjusted for stratification factors; median 6.24 months vs 5.91 months). One-year survival was also greater with erlotinib plus gemcitabine (23% vs 17%; P = .002).
Practice points

• Don’t interpret an RCT (or any other clinical study) by the abstract alone

• Critical analysis is a systematic process and an essential skill – read, practice, discuss and argue

• Read editorials and letters to editor in the journals

• Don’t take studies on their face value