Clinical Trial Design and Execution

Wim P Ceelen

GI Surgery - UZ Gent
Did investigator assign exposure/intervention?

Yes

Experimental study

Random allocation

Yes

Randomized controlled trial

No

Non-randomized controlled trial

No

Experimental study

Comparison group?

Yes

Analytical study

No

Descriptive study

Direction?

Exposure → Outcome

Outcome → Exposure

Exposure and outcome at the same time

Cohort study

Case-control study

Cross-sectional study
- RCT’s consist of < 5% of all published biomedical literature
- 30% of all RCT’s are methodologically sound
STREPTOMYCIN TREATMENT OF PULMONARY TUBERCULOSIS
A MEDICAL RESEARCH COUNCIL INVESTIGATION

Results at End of Six Months

Four of the 55 S patients (7%) and 14 of the 52 C patients (27%) died before the end of six months. The difference between the two series is statistically significant; the probability of it occurring by chance is less than one in a hundred.
48.6% (69/142) of the RCTs had a high risk of bias in at least one domain

Ivers BMJ 2013
Obstacles for conducting RCT’s in Surgery

- Organisational
  - Lack of funding
  - Lack of cooperation
  - Traditional master-student model
  - Lack of regulation
- Methodological
  - Lack of standardisation
    - Skill dependence
    - Learning curve effects
  - Impossibility to ‘blind’
  - Lack of equipoise
Alternative study designs

• Non Randomized
  – Case-matching studies, propensity score matching
  – Controlled interrupted-time series designs
  – Step-wedge designs
• Modified randomized
  – ‘Pragmatic’ trials
  – expertise-based, third party
  – Tracker trials
  – Phase 2S transition from database to randomised clinical trial
  – Feasibility randomised clinical trial
  – Explanatory randomised clinical trial
Interrupted time series

Outcome

Time

Slope pre

Intervention introduced

Change in level

Slope post
A combined teamwork training and work standardisation intervention in operating theatres: controlled interrupted time series study

Lauren Morgan,¹ Sharon P Pickering,² Mohammed Hadi,² Eleanor Robertson,¹ Steve New,³ Damian Griffin,² Gary Collins,⁴ Oliver Rivero-Arias,⁵,⁶ Ken Catchpole,⁷ Peter McCulloch¹
The figure displays box plots comparing the Total Team Oxford NOTECHS II Score (max score = 98) pre-intervention and post-intervention for the control and active groups. The box plots indicate that the active group shows a larger variation in scores compared to the control group post-intervention.
Stepped wedge design

- Involve sequential roll-out of an intervention to participants (individuals or clusters) over a number of time periods
- By the end of the study, all participants will have received the intervention
- Particularly relevant where it is predicted that the intervention will do more good than harm
Shaded cells represent intervention periods
Blank cells represent control periods
Each cell represents a data collection point
Effect of the World Health Organization Checklist on Patient Outcomes

A Stepped Wedge Cluster Randomized Controlled Trial

Arvid Steinar Haugen, MSc,*† Eirik Sofieeland, MD, PhD,* Stian K. Almeland, MD,‡ Nick Sevdalis, PhD,§ Barthold Vonen, MD, PhD,¶ Geir E. Eide, PhD,||** Monica W. Nortvedt, PhD,†† and Stig Harthug, MD, PhDikki

Ann Surg 2014
<table>
<thead>
<tr>
<th>Clusters of surgical specialties</th>
<th>Urology</th>
<th>General</th>
<th>Neuro</th>
<th>Cardio-thoracic</th>
<th>Orthopedic</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>SSC</td>
<td>SSC</td>
<td>SSC</td>
<td>SSC</td>
<td>SSC</td>
</tr>
<tr>
<td>June</td>
<td>SSC</td>
<td>SSC</td>
<td>SSC</td>
<td>SSC</td>
<td>SSC</td>
</tr>
</tbody>
</table>

**TIMELINE**

Aug 2009 to June 2010
Study protocol

Totally laparoscopic versus conventional ileoanal pouch procedure – design of a single-centre, expertise based randomised controlled trial to compare the laparoscopic and conventional surgical approach in patients undergoing primary elective restorative proctocolectomy- LapConPouch-Trial

Dalibor Antolovic, Peter Kienle, Hanns-Peter Knaebel, Jan Schmidt, Carsten N Gutt, Jürgen Weitz, Moritz Koch, Markus W Büchler and Christoph M Seiler*
Evaluation of Devices and Interventions: IDEAL
<table>
<thead>
<tr>
<th></th>
<th>1 Idea</th>
<th>2a Development</th>
<th>2b Exploration</th>
<th>3 Assessment</th>
<th>4 Long-term study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Proof of concept</td>
<td>Development</td>
<td>Learning</td>
<td>Assessment</td>
<td>Surveillance</td>
</tr>
<tr>
<td>Number and types of patients</td>
<td>Single digit; highly selected</td>
<td>Few; selected</td>
<td>Many; may expand to mixed; broadening indication</td>
<td>Many; expanded indications (well defined)</td>
<td>All eligible</td>
</tr>
<tr>
<td>Number and types of surgeons</td>
<td>Very few; innovators</td>
<td>Few; innovators and some early adopters</td>
<td>Many; innovators, early adopters, early majority</td>
<td>Many; early majority</td>
<td>All eligible</td>
</tr>
<tr>
<td>Output</td>
<td>Description</td>
<td>Description</td>
<td>Measurement; comparison</td>
<td>Comparison; complete information for non-RCT participants</td>
<td>Description; audit, regional variation; quality assurance; risk adjustment</td>
</tr>
<tr>
<td>Intervention</td>
<td>Evolving; procedure inception</td>
<td>Evolving; procedure development</td>
<td>Evolving; procedure refinement; community learning</td>
<td>Stable</td>
<td>Stable</td>
</tr>
<tr>
<td>Method</td>
<td>Structured case reports</td>
<td>Prospective development studies</td>
<td>Research database; explanatory or feasibility RCT (efficacy trial); diseased based (diagnostic)</td>
<td>RCT with or without additions/ modifications; alternative designs</td>
<td>Registry; routine database (eg. SCOAP, STS, NSQIP); rare-case reports</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Proof of concept; technical achievement; disasters; dramatic successes</td>
<td>Mainly safety; technical and procedural success</td>
<td>Safety; clinical outcomes (specific and graded); short-term outcomes; patient-centred (reported) outcomes; feasibility outcomes</td>
<td>Clinical outcomes (specific and graded); middle-term and long-term outcomes; patient-centred (reported) outcomes; cost-effectiveness</td>
<td>Rare events; long-term outcomes; quality assurance</td>
</tr>
<tr>
<td>Ethical approval</td>
<td>Sometimes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Examples</td>
<td>NOTES video</td>
<td>Tissue engineered vessels</td>
<td>Italian D2 gastrectomy study</td>
<td>Swedish obese patients study</td>
<td>UK national adult cardiac surgical database</td>
</tr>
</tbody>
</table>


**Table: Stages of surgical innovation**
How to execute a RCT

1. Design a protocol
2. Design a CRF (database)
3. Register your study: EC, NCT (clinicaltrials.gov)
4. Recruit patients
5. Analyze and publish the results
How to design a RCT

1. Define a research question
2. Define a relevant primary endpoint
3. Define the baseline risk, incidence, hazard...
4. Define the expected treatment effect
5. Calculate the sample size
6. Define method and timing of randomization
7. Ensure blinding of intervention
What is a good research question?

**Feasible** (answerable with a robust method)

**Interesting**

**Novel**

**Ethical**

**Relevant**

FINER criteria
Was the outcome appropriate?

- Clinically relevant
- Reliable and reproducible
- Unique
  - Defined *a priori*
  - Multiple endpoints: more false positive results
- Available for all patients
Outcomes

• **Types of outcomes:**
  – Hard endpoints
  – Surrogate endpoints (valid and invalid)
  – Composite endpoints

• **Way outcomes are reported**
  – Relative risk/risk ratio
  – Hazard ratio
  – Odds ratio
  – Absolute risk (difference)
  – Number needed to treat
  – Effect size (difference)
Hard Endpoints

- Mortality
- Quality of Life
- Amputations, hearing loss, loss of vision
- Pain reduction/increase
Surrogate Outcomes

• Valid:
  – the marker is intermediate on the causal pathway between exposure and hard outcome AND the association between exposure and surrogate endpoint always results in the same association between the surrogate outcome and the hard endpoint
  – The association between the exposure and the surrogate has always the same extent and sign as that between the exposure and the hard endpoint
• Unvalid
  – The surrogate marker is associated with the exposure, but there is no causal association between the surrogate marker and the hard endpoint
Surrogate endpoints: examples

- Oncology trials: DFS, PFS, pCR as surrogate for OS
- LN harvest or amputation rate as surrogate for surgical quality in colorectal surgery
- Prognostic indicators are not always surrogate endpoints!
Oba. Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. JNCI 2013
Median survival

17.4 m
16.5 m

P=0.362

Mitoxantrone + Prednisone

P<0.0001

PSA response rate

0%
32%
48%

Weekly Docetaxel

Collette Eur J Cancer 2006
Composite endpoints

• An aggregate of different outcomes rather than one outcome

• Examples
  – ‘Overall complication rate’
  – MACE (major adverse cardiac events)
Selective outcome reporting

- Studies reporting positive or significant results are more likely to be published
- Outcomes that are statistically significant are more likely to be fully reported
- 40–62% of publications had at least one primary outcome changed, newly introduced or omitted compared to protocol [Dwan et al, PLoS ONE 2008]
Core outcome set

• An agreed standardised set of outcomes that should be measured and reported, as a minimum, in all clinical trials in specific areas of health or health care
Factors that determine Sample Size

- $\alpha$ value (probability of type I error) $0,05$
- Power $(1-\beta)$ $0,8$
- Size of estimated treatment effect
  - Large sample needed to detect small differences
  - Small sample size can only detect large effects → often underpowered for smaller, but clinically relevant effects
Blinding in surgical trials

- Investigator blinding impossible
- Patient blinding very difficult
  - Sham surgery: ethical?
  - Wound dressings covering entire abdomen
- Blinding of outcome assessors
Proper Randomization

- **Acceptable methods**
  - Central randomization using computer generated random number lists
  - Numbered opaque sealed envelopes
- **Unacceptable methods (pseudo-random)**
  - Alternation
  - Date of birth
  - Treatment date

→ *Violates Treatment Allocation Blinding*
Analysis of the results

- Should focus on the primary endpoint
- Avoid subgroup analyses
- Include CONSORT flowchart
- Use appropriate statistical tests
Target population

Patient sample

Selection bias

Intervention group

Control group

Randomization

Patient blind
Care provider blind

Performance bias

Exposed to intervention

Not exposed to intervention

Intention-to-treat analysis

Attrition bias

Follow-up

Follow-up

Outcome assessor blind

Detection bias

Outcomes

Outcomes
Probability That Multiple Subgroup Analyses Will Yield at Least One (Red), Two (Blue), or Three (Yellow) False Positive Results.
Multiple Outcome Measures: Bonferroni correction


<table>
<thead>
<tr>
<th></th>
<th>SEMS</th>
<th>GJ</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>24</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Technical success</td>
<td>23 (96%)</td>
<td>23 (100%)</td>
<td>0.51</td>
</tr>
<tr>
<td>Clinical success</td>
<td>22 (92%)</td>
<td>13 (56%)</td>
<td>0.0067</td>
</tr>
<tr>
<td>Mean hospital stay (SD), d</td>
<td>3.0 (1.4)</td>
<td>24.2 (10.4)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Mean survival (SD), d*</td>
<td>96.2 (9.6)</td>
<td>70.2 (36.2)</td>
<td>0.0165*</td>
</tr>
<tr>
<td>Morbidity</td>
<td>4 (17%)</td>
<td>14 (61%)</td>
<td>0.0021</td>
</tr>
<tr>
<td>30-d mortality</td>
<td>0</td>
<td>7 (30%)</td>
<td>0.0039</td>
</tr>
<tr>
<td>Operating room and anesthesiologist</td>
<td>Never</td>
<td>Always</td>
<td></td>
</tr>
</tbody>
</table>

SEMS, Self-expanding metallic stent; GJ, gastrojejunostomy; SD, standard deviation.
*Difference in mean survival was nominally significant in a single test of hypothesis; however, Bonferroni correction for multiple testing of data removed that significance.

Adjusted p=0.05/6=0.0083
Appropriate Statistical Methods

- Small sample size $\rightarrow$ non parametric tests
- Use of SE instead of SD: misleading
- Reporting of *precision of the estimate* of the observed effect, in addition to a P value (confidence intervals)
- P values should be two sided
Appropriate Statistical Methods

- Actuarial survival curves: include numbers at risk, or confidence intervals
- Beware of Crossing survival curves
- Multivariable models
  - Specify why/how independent variables were entered
  - Include some measure of goodness of fit
  - Cox model: verify proportionality of hazards
  - Sufficient nr of events needed for valid model (at least 10 per variable entered)
|                                         | M linear                  | Logistic                  | Cox PH                   |
|                                         | F test                    | Likelihood ratio test    | Likelihood ratio test    |
| Independent variables significantly associated with outcome? | R²                        | Pearson goodness-of-fit test | Estimated versus Expected outcome |
| Quantitative assessment of how well model explains observed outcome |                          |                           |                          |
The resource centre for good reporting of health research studies

Library for health research reporting

The Library contains a comprehensive searchable database of reporting guidelines and also links to other resources relevant to research reporting.

Key reporting guidelines

- CONSORT
- STROBE
- PRISMA
- CRD4
- COREQ
- ENTREQ
- SQUIRE
- CARE
- NDPH
- SPIRIT
- PRISMA-P

Search for reporting guidelines

Visit the library for more resources

Toolkits

The EQUATOR Network works to improve the reliability and value of medical research literature by promoting transparent and accurate reporting of research studies.

Our Toolkits support different user groups, including:

- Authors
  Information and resources for authors

- Editors
  Information and resources for editors and peer reviewers

EQUATOR highlights

Edinburgh, UK 28-30 September 2016

The 2015 Research Waste / EQUATOR Conference will be held in Edinburgh, UK Save the dates: 28-30 September 2015 Venue: John McIntyre Conference Centre, Edinburgh. UK Conference aims (1) Review the progress made by research regulators, academic institutions, researchers, funders, and ... Read More

13/09/2014 - Videos now available from the scientific meeting in Paris: Improving reporting to decrease the waste of research

The 6th annual lecture, presentations and roundtable discussion were recorded and are now available to watch Read More

13/09/2014 - Interview with Iveta Simera about the EQUATOR Network

The plagiarism detection software company iTentivate recently

News

Editorial: Stronger post-publication culture is needed for better science
23/01/2015

Pioneers of transparency in health research
9/01/2015

What trials are really supposed to be – a succinct reminder from J. Ioannidis
8/01/2015

Guidelines for reporting multivariable prediction models for individual prognosis or diagnosis (TRIPOD) published
7/01/2015

PRISMA-P Statement for protocols published
7/01/2015

Sign-up to our newsletter

to keep up-to-date with the latest