Limitations of randomised trials. Is the king naked?

Luigia Scudeller (Italy)
Transparency declaration

– In my 1° and 3° life: Infectious Diseases physician
– In my 2° and 4° (current) life: Clinical Epidemiology with strong interest in
  – Research quality
  – Training of young professionals
  – (nice food)
– No financial interests

– My original set for this presentation included 124 slides…
Foreground

- 1948
- modern era of randomized controlled trials
- and not a woman among the researchers ;-(

BRITISH MEDICAL JOURNAL
LONDON SATURDAY OCTOBER 30 1948

STREPTOMYCIN TREATMENT OF PULMONARY TUBERCULOSIS
A MEDICAL RESEARCH COUNCIL INVESTIGATION

The following gives the short-term results of a controlled investigation into the effects of streptomycin on one type of pulmonary tuberculosis. The inquiry was planned and directed by the Streptomycin in Tuberculosis Trials Committee, composed of the following members: Dr. Geoffrey Marshall (chairman), Professor J. W. S. Blacklock, Professor C. Cameron, Professor N. B. Capon, Dr. R. Cruickshank, Professor J. H. Gaddum, Dr. F. R. G. Heaf, Professor A. Bradford Hill, Dr. L. E. Houghton, Dr. J. Clifford Hoyle, Professor H. Raistrick, Dr. J. G. Scadding, Professor W. H. Tytler, Professor G. S. Wilson, and Dr. P. D’Arcy Hart
- 1948
- First Republican Parliament elected in Italy
- Participation was 92%
- For the first time, women were allowed to vote
Outline of the presentation

1. RCTs are «republicans», not «kings»
2. Limitations
   – Sometimes you can’t use them
   – Sometimes you can, but should not use them
   – Sometimes, even if you should and can use them, they do not «work»
3. Conclusions
1. RCTs are «republicans», not «kings»
the question being asked determines the appropriate research architecture, strategy, and tactics to be used—not tradition, authority, experts, paradigms, or schools of thought.

The issue is which way of answering the specific question before us provides the most valid, useful answer.

Find the foot fitting the glass slipper.
Clinical questions and study designs

<table>
<thead>
<tr>
<th>Most common type of questions:</th>
<th>Type of study:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis</strong>&lt;br&gt;how to select and interpret diagnostic tests</td>
<td>prospective, blind comparison to a gold standard or cross-sectional</td>
</tr>
<tr>
<td><strong>Therapy</strong>&lt;br&gt;how to select treatments that do more good than harm and that are worth the efforts and costs of using them</td>
<td>randomized controlled trial &gt; cohort study</td>
</tr>
<tr>
<td><strong>Prognosis</strong>&lt;br&gt;how to estimate the patient’s likely clinical course over time (based on factors other than the intervention) and anticipate likely complications of disease</td>
<td>cohort study &gt; case control &gt; case series</td>
</tr>
<tr>
<td><strong>Harm/Etiology</strong>&lt;br&gt;how to identify causes for disease (including iatrogenic forms)</td>
<td>cohort &gt; case control &gt; case series</td>
</tr>
</tbody>
</table>
Observational research, randomised trials, and two views of medical science.

The first trial ever reported

–Old Testament, Daniel 1:8–16

8 Daniel decided not to eat the king's food or drink his wine because that would make him unclean.[…]

10 but Ashpenaz said to Daniel, "I am afraid of my master, the king. He ordered me to give you this food and drink. If you begin to look worse than other young men your age, the king will see this. Then he will cut off my head because of you."
The first trial ever reported

–Old Testament, Daniel 1:8–16

12 Daniel said to the guard, "Please give us this test for ten days: Don't give us anything but vegetables to eat and water to drink.
13 After ten days compare how we look with how the other young men look who eat the king's food. See for yourself and then decide how you want to treat us, your servants."
14 So the guard agreed to test them for ten days.
15 After ten days they looked healthier and better fed than all the young men who ate the king's food.
16 So the guard took away the king's special food and wine, feeding them vegetables instead.

P: young men
I: vegetable and water
C: king’s food and wine
O: looks
What methodological lessons can we learn from this trial?

– Clinically relevant, well-defined, pre-specified, questions
– Choice of intervention and control group (clinical equipoise)
– Clinically important outcome
– Testing, but avoiding bias
– Trial results inform clinical practice

What we can not?
– Confounding by indication not taken into account
Question 1...
Interactive question

- Research question: in BSI by Enterococcus spp, does antibiotic monotherapy improves 90-day survival when compared with combination?
- When do you prefer RCTs over observational studies in your research practice?

- Indicate just your PREFERRED reason

1) When I want to eliminate confounding
2) When I want to infer a causal effect
3) When I want to avoid treatment assignment bias
4) When I want to avoid selection bias
5) When I want an «objective» measurement of outcome
Schulz KF, Grimes DA.
Generation of allocation sequences in randomised trials: chance, not choice.
Lancet. 2002 Feb 9;359(9305):515-9

It eliminates bias in treatment assignment
Comparisons of different forms of health interventions can be misleading unless investigators take precautions to ensure that their trial comprises unbiased comparison groups relative to prognosis. In controlled trials of prevention or treatment, randomisation produces unbiased comparison groups by avoiding selection and confounding biases. Consequently, comparison groups are not prejudiced by selection of particular patients, whether consciously or not, to receive a specific intervention. The notion of avoiding bias includes eliminating it from decisions on entry of participants to the trial, as well as eliminating bias from the assignment of participants to treatment, once entered. Investigators need to properly register each participant immediately on identification of eligibility for the trial, but without knowledge of the assignment. The reduction of selection and confounding biases underpins the most important strength of randomisation. Randomisation prevails as the best study design for study of small or moderate effects.\(^6\)

It facilitates blinding (masking) of the identity of treatments from investigators, participants, and assessors, including the possible use of a placebo\(^7\)
Such manoeuvres reduce bias after random assignment, and would be difficult, perhaps even impossible, to implement if investigators assigned treatments by a non-random scheme.

It permits the use of probability theory to express the likelihood that any difference in outcome between treatment groups merely indicates chance
Randomization to avoid confounding

– Controlling for known and unknown confounders
Randomization to avoid bias

– bias = prejudice
bias = distortion of true effect
Ceftazidime-Avibactam Is Superior to Other Treatment Regimens against Carbapenem-Resistant *Klebsiella pneumoniae* Bacteremia

Ryan K. Shields,a,c M. Hong Nguyen,a,c Liang Chen,d Ellen G. Press,a Brian A. Potoski,a,c,e Rachel V. Marini,c Yohei Doi,a,c Barry N. Kreiswirth,d Cornelius J. Clancy,a,b,f

We conducted a retrospective study of UPMC patients with CR-Kp bacteremia between January 2009 and February 2017 who received ≥3 days of treatment. CR-Kp was defined by resistance to any carbapenem (1); only the first episode of CR-Kp bacteremia was included. Clinical success was defined at 30 days as survival, resolution of signs and symptoms of infection, sterilization of blood cultures within 7 days of treatment initiation, and absence of recurrent infections.
Clinical success was achieved more frequently among patients treated with a regimen including C-A than with other regimens ($P = 0.006$), including those comprised of $\geq 2$ in vitro active agents (44% [12/27]; $P = 0.02$). By multivariable logistic regression, primary bacteremia (odds ratio [OR], 4.50; 95% confidence interval [CI], 1.53 to 13.21; $P = 0.006$) and receipt of C-A (OR, 8.64; 95% CI, 1.61 to 43.39; $P = 0.01$) were independent predictors of clinical success (Table 3).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Treatment group</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C-A (n = 13)</td>
<td>CB+AG (n = 28)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient demographics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n [%])</td>
<td>7 (54)</td>
<td>18 (64)</td>
</tr>
<tr>
<td>Age (median [range])</td>
<td>68 (32-81)</td>
<td>57 (32-87)</td>
</tr>
<tr>
<td>Underlying disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes (n [%])</td>
<td>4 (31)</td>
<td>9 (32)</td>
</tr>
<tr>
<td>Chronic liver disease (n [%])</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Chronic respiratory disease (n [%])</td>
<td>6 (33)</td>
<td>6 (20)</td>
</tr>
<tr>
<td>Immunocompromised (n [%])</td>
<td>8 (38)</td>
<td>13 (47)</td>
</tr>
<tr>
<td>Solid-organ transplant recipient (n [%])</td>
<td>3 (23)</td>
<td>11 (44)</td>
</tr>
<tr>
<td>Severity of illness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICU at time of bacteremia (n [%])</td>
<td>6 (48)</td>
<td>13 (52)</td>
</tr>
<tr>
<td>RRT (n [%])</td>
<td>2 (15)</td>
<td>7 (25)</td>
</tr>
<tr>
<td>PtT bacteremia score (median [range])</td>
<td>4 (1-5)</td>
<td>4 (0-9)</td>
</tr>
<tr>
<td>APACHE II score (median [range])</td>
<td>20 (18-33)</td>
<td>17 (9-26)</td>
</tr>
</tbody>
</table>
Limitations

2a. Sometimes you can’t use them
Figure 2: View of the defendants standing in the dock during the International Military Tribunal in Nuremberg
Reproduced with permission from the USHMM.

WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects

Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964 and amended by the:
- 29th WMA General Assembly, Tokyo, Japan, October 1975
- 35th WMA General Assembly, Venice, Italy, October 1983
- 41st WMA General Assembly, Hong Kong, September 1989
- 48th WMA General Assembly, Somerset West, Republic of South Africa, October 1996
- 52nd WMA General Assembly, Edinburgh, Scotland, October 2000
- 53rd WMA General Assembly, Washington DC, USA, October 2002 (Note of Clarification added)
- 55th WMA General Assembly, Tokyo, Japan, October 2004 (Note of Clarification added)
- 59th WMA General Assembly, Seoul, Republic of Korea, October 2008
- 64th WMA General Assembly, Fortaleza, Brazil, October 2013

Lancet Oncol 2007; 8: 1139–46
In 1972, Jean Heller of the Associated Press reported on a 40-year-old research study that had followed black Alabama sharecroppers, some of whom had syphilis. The revelation of deception, withholding of appropriate treatment, and other unethical practices exploded into the Tuskegee scandal. Tuskegee led to the National Research Act of 1974, which authorized the Department of Health, Education, and Welfare (now the Department of Health and Human Services [HHS]) to augment government policies for protecting human research subjects.

November 4, 2015, at NEJM.org.
Experimental infection of human volunteers

Meta Roestenberg, Marie-Astrid Hoogerwerf, Daniela M Ferreira, Benjamin Mordmuller, Maria Yazdanbakhsh

Controlled human infection (CHI) trials, in which healthy volunteers are experimentally infected, can accelerate the development of novel drugs and vaccines for infectious diseases of global importance. The use of CHI models is expanding from around 60 studies in the 1970s to more than 120 publications in this decade, primarily for influenza, rhinovirus, and malaria. CHI trials have provided landmark data for several registered drugs and vaccines, and have generated unprecedented scientific insights. Because of their invasive nature, CHI studies demand critical ethical review according to established frameworks. CHI-associated serious adverse events are rarely reported. Novel CHI models need standardised safety data from comparable CHI models to facilitate evidence-based risk assessments, as well as funds to produce challenge inoculum according to regulatory requirements. Advances such as the principle of controlled colonisation, the expansion of models to endemic areas, and the use of genetically attenuated strains will further broaden the scope of CHI trials.

Lancet Infect Dis 2018; 18: e312–22
Published Online June 8, 2018
http://dx.doi.org/10.1016/S1473-3099(18)30177-4

![Graph showing number of clinical trials per year from 1900 to 2017](image-url)
Trials are complex

http://www.trialforge.org/pathway/
Feasibility Studies
Feasibility Studies are pieces of research done before a main study. They are used to estimate important parameters that are needed to design the main study. For instance:

- standard deviation of the outcome measure, which is needed in some cases to estimate sample size,
- willingness of participants to be randomised,
- willingness of clinicians to recruit participants,
- number of eligible patients,
- characteristics of the proposed outcome measure and in some cases feasibility studies might involve designing a suitable outcome measure,
- follow-up rates, response rates to questionnaires, adherence/compliance rates, ICCs in cluster trials, etc.
RCTs are expensive

THE PHARMA AD

WHAT DOES IT TAKE TO MAKE ONE MEDICINE?

MORE THAN A BILLION POUNDS!

THE REALITY

WHAT DOES IT TAKE TO MAKE ONE MEDICINE?

FROM OUR EXPERIENCE: AS LITTLE AS 114 MILLION POUNDS!

https://msfaccess.org
Limitations

2b. Sometimes you can, but should not use them
When are randomised trials unnecessary? Picking signal from noise

The relation between a treatment and its effect is sometimes so dramatic that bias can be ruled out as an explanation. Paul Glasziou and colleagues suggest how to determine when observations speak for themselves.
Clinical equipoise

Equipoise and the Ethics of Clinical Research

Benjamin Freedman, Ph.D.

I suggest an alternative concept of equipoise, which would be based on present or imminent controversy in the clinical community over the preferred treatment. According to this concept of "clinical equipoise," the requirement is satisfied if there is genuine uncertainty within the expert medical community — not necessarily on the part of the individual investigator — about the preferred treatment. (N Engl J Med 1987; 317: 141-5.)
The concept of clinical equipoise and its relevance to infectious disease clinical trials

Matthias Maiwald1,2,3

DOI: https://doi.org/10.1016/j.pathol.2018.12.134

Abstract

Evidence-based medicine places big emphasis on randomised clinical trials and systematic reviews as the best evidence sources. However, this may happen at the expense of other evidence sources, such as laboratory research. The relevant concept is the ‘evidence pyramid’, whereby evidence sources are ranked according to quality. However, ‘lower-level’ evidence often provides valuable information, and any drug development progresses through stages, such as compound discovery, laboratory testing, preclinical testing, and clinical trials. The concept of clinical equipoise is a principle in research ethics that stipulates that there should be genuine uncertainty over which treatment is more effective, before starting a trial. If one treatment is already known to be superior, a trial may not be ethically justified. In the absence of prior trials, an important source to know about such imbalance is ‘lower-level’ evidence. I will provide examples from infectious disease research where clinical equipoise likely did not exist because other evidence sources indicated superiority of one treatment. I will discuss examples from antiseptic trials, antimicrobial treatment trials, and emerging infectious diseases trials. I will discuss ethical and research rigour implications, highlight potential approaches to address the problem, and call for a more holistic approach to evidence assessment.
What is clinical equipoise?

Chlorhexidine-alcohol versus povidone-iodine for pre-operative skin preparation: A systematic review and meta-analysis

Piras Ayoub, Michael Quincke, Ronan Conway, Arnold Hill

SSI - relative risk

<table>
<thead>
<tr>
<th>Study</th>
<th>RR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paochareon (2009)</td>
<td>0.62 (0.21, 1.88)</td>
<td>5.42</td>
</tr>
<tr>
<td>Sistla (2010)</td>
<td>0.74 (0.38, 1.43)</td>
<td>15.07</td>
</tr>
<tr>
<td>Darouiche (2010)</td>
<td>0.59 (0.41, 0.85)</td>
<td>49.10</td>
</tr>
<tr>
<td>Pati (2013)</td>
<td>0.62 (0.39, 0.99)</td>
<td>30.41</td>
</tr>
<tr>
<td>Bibbo (2005)</td>
<td>(Excluded)</td>
<td>0.00</td>
</tr>
<tr>
<td>Saltzman (2009)</td>
<td>(Excluded)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Overall (I-squared = 0.0%, \( p = 0.955 \))

Test for overall effect: \( Z = 3.61 \) (\( p = 0.000 \))

NOTE: Weights are from random effects analysis

Favours Chlorhexidine-Alcohol

Favours Povidone-Iodine
Chlorhexidine-Alcohol Versus Povidone-Iodine for Cesarean Antisepsis (CAPICA)

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT02202577

Recruitment Status: Completed
First Posted: July 29, 2014
Last Update Posted: July 28, 2016

Study Design

- Study Type: Interventional (Clinical Trial)
- Actual Enrollment: 932 participants
- Allocation: Randomized
- Intervention Model: Parallel Assignment
- Masking: None (Open Label)
- Primary Purpose: Treatment
- Official Title: Chlorhexidine-Alcohol Versus Povidone-Iodine for Surgical Site Antisepsis Prior to Cesarean Delivery
- Study Start Date: March 2014
- Actual Primary Completion Date: July 2016
- Actual Study Completion Date: July 2016

Should the investigators perform a systematic review of the literature before designing a RCT?
Chlorhexidine–Alcohol Compared with Povidone–Iodine Preoperative Skin Antisepsis for Cesarean Delivery: A Systematic Review and Meta-Analysis

Mary Catherine Tolcher, MD, MSc
Megan D. Whitham, MD
Sherif A. El-Nashar, MBBS
Steven L. Clark, MD

Should these trials have been performed?
To embark on research without reviewing systematically evidence of what is already known, particularly when the research involves people or animals, is unethical, unscientific, and wasteful.

No new studies without prior systematic review of existing evidence
Efficient production, updating and dissemination of systematic reviews

1. Clarke M, Brice A, Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. PloS One 2014; 9: e102670.
2c. Sometimes, even if you should and can use them, they do not «work»
Some issues

– […]
– Generalizability
  – rRCT?
– Internal validity
– Surrogate endpoints
– Long-term harms
– COS
– Noninferiority
– Small sample size
– Studies are not published
– […]
External validity: Generalizability

Figure 1. Hypothetical selection process of patients entered into a clinical trial (read left to right) and generalization of results from randomized patients to other patients (read right to left). Multiplication of area of first two boxes by $10^9$ and $10^{9-8}$ for patients with disorder and the subset of patients appropriate for treatment indicates order of magnitude of respective cohorts.

### Adapting clinical guidelines to take account of multimorbidity

*BMJ* 2012; 345 doi: https://doi.org/10.1136/bmj.e6341 (Published 04 October 2012)

Percentage of patients with the row condition who also have the column condition

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Coronary heart disease</th>
<th>Hypertension</th>
<th>Heart failure</th>
<th>Stroke/transient ischaemic attack</th>
<th>Atrial fibrillation</th>
<th>Diabetes</th>
<th>Chronic obstructive pulmonary disease</th>
<th>Painful condition</th>
<th>Depression</th>
<th>Dementia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coronary heart disease</strong></td>
<td>52%</td>
<td>14%</td>
<td>13%</td>
<td>11%</td>
<td>22%</td>
<td>24%</td>
<td>27%</td>
<td>51%</td>
<td>5%</td>
<td>21%</td>
</tr>
<tr>
<td><strong>Hypertension</strong></td>
<td>18%</td>
<td>57%</td>
<td>59%</td>
<td>61%</td>
<td>37%</td>
<td>55%</td>
<td>23%</td>
<td>24%</td>
<td>41%</td>
<td>41%</td>
</tr>
<tr>
<td><strong>Heart failure</strong></td>
<td>10%</td>
<td>16%</td>
<td>16%</td>
<td>13%</td>
<td>23%</td>
<td>23%</td>
<td>20%</td>
<td>22%</td>
<td>19%</td>
<td>19%</td>
</tr>
<tr>
<td><strong>Stroke/transient ischaemic attack</strong></td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td><strong>Atrial fibrillation</strong></td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
</tr>
<tr>
<td><strong>Diabetes</strong></td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
</tr>
<tr>
<td><strong>Chronic obstructive pulmonary disease</strong></td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
</tr>
<tr>
<td><strong>Painful condition</strong></td>
<td>17%</td>
<td>17%</td>
<td>17%</td>
<td>17%</td>
<td>17%</td>
<td>17%</td>
<td>17%</td>
<td>17%</td>
<td>17%</td>
<td>17%</td>
</tr>
<tr>
<td><strong>Depression</strong></td>
<td>23%</td>
<td>23%</td>
<td>23%</td>
<td>23%</td>
<td>23%</td>
<td>23%</td>
<td>23%</td>
<td>23%</td>
<td>23%</td>
<td>23%</td>
</tr>
<tr>
<td><strong>Dementia</strong></td>
<td>12%</td>
<td>12%</td>
<td>12%</td>
<td>12%</td>
<td>12%</td>
<td>12%</td>
<td>12%</td>
<td>12%</td>
<td>12%</td>
<td>12%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage who only have the row condition*</th>
<th>Mean No of conditions in people aged ≤65 years with row condition</th>
<th>Mean No of conditions in people aged ≥65 years with row condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.8%</td>
<td>3.4</td>
<td>4.4</td>
</tr>
<tr>
<td>21.9%</td>
<td>2.5</td>
<td>3.6</td>
</tr>
<tr>
<td>2.8%</td>
<td>3.9</td>
<td>5.6</td>
</tr>
<tr>
<td>6.0%</td>
<td>3.6</td>
<td>4.8</td>
</tr>
<tr>
<td>6.5%</td>
<td>3.3</td>
<td>5.0</td>
</tr>
<tr>
<td>17.6%</td>
<td>2.9</td>
<td>6.5</td>
</tr>
<tr>
<td>14.3%</td>
<td>2.8</td>
<td>4.5</td>
</tr>
<tr>
<td>12.7%</td>
<td>3.1</td>
<td>4.3</td>
</tr>
<tr>
<td>25.4%</td>
<td>2.6</td>
<td>4.9</td>
</tr>
<tr>
<td>5.3%</td>
<td>4.1</td>
<td>4.6</td>
</tr>
</tbody>
</table>

* Percentage who do not have one of 39 other conditions in the full count
Differences in implementation of HIV/AIDS clinical research in developed versus developing world: an evidence-based review on protease inhibitor use among women and minorities

E Seminari MD, A De Silvestri MSc, L Scudeller MD, V Scotti MA and C Tinelli MD


Test for trend $p = 0.012$
Spearman rho = 0.38 ($p = 0.009$)
Making trials matter: pragmatic and explanatory trials and the problem of applicability.


Abstract

Randomised controlled trials are the best research design for decisions about the effect of different interventions but randomisation does not, of itself, promote the applicability of a trial’s results to situations other than the precise one in which the trial was done. While methodologists and trialists have rightly paid great attention to internal validity, much less has been given to applicability.

This narrative review is aimed at those planning to conduct trials, and those aiming to use the information in them. It is intended to help the former group make their trials more widely useful and to help the latter group make more informed decisions about the wider use of existing trials. We review the differences between the design of most randomised trials (which have an explanatory attitude) and the design of trials more able to inform decision making (which have a pragmatic attitude) and discuss approaches used to assert applicability of trial results.

If we want evidence from trials to be used in clinical practice and policy, trialists should make every effort to make their trial widely applicable, which means that more trials should be pragmatic in attitude.
The PRECIS-2 tool: designing trials that are fit for purpose

Kirsty Loudon, Shaun Treweek, Frank Sullivan, Peter Donnan, Kevin E Thorpe, Merrick Zwarenstein

Diagram:

- Eligibility: Who is selected to participate in the trial?
- Recruitment: How are participants recruited into the trial?
- Setting: Where is the trial being done?
- Organisation: What expertise and resources are needed to deliver the intervention?
- Flexibility: delivery: How should the intervention be delivered?
- Flexibility: adherence: What measures are in place to make sure participants adhere to the intervention?
- Follow-up: How closely are participants followed up?
- Primary outcome: How relevant is it to participants?
- Primary analysis: To what extent are all data included?
The Bacteriuria in Renal Transplantation (BiRT) Study: A Trial Comparing Antibiotics Versus no Treatment in the Prevention of Symptomatic Urinary Tract Infection in Kidney Transplant Recipients With Asymptomatic Bacteriuria (BiRT)

Eligibility - Who is selected to participate in the trial?

Primary analysis - To what extent are all data included?

Primary outcome - How relevant is it to participants?

Recruitment - How are participants recruited into the trial?

Setting - Where is the trial being done?

Follow-up - How closely are participants followed-up?

Organisation - What expertise and resources are needed to deliver the intervention?

Flexibility - What measures are in place to make sure participants adhere to the intervention?

Flexibility - How should the intervention be delivered?
Big data and black-box medical algorithms

W. Nicholson Price

New machine-learning techniques are already making their way into practice.

European Commission › Strategy › Digital Single Market › Reports and studies ›

Digital Single Market

REPORT / STUDY | 8 April 2019

Ethics guidelines for trustworthy AI

Following the publication of the draft ethics guidelines in December 2018 to which more than 500 comments were received, the independent expert group presents today their ethics guidelines for trustworthy artificial intelligence.
Question 2...
Interactive question

You are designing a very pragmatic, phase IV, randomized controlled clinical trial, comparing two oral antibiotic treatments for acute upper airways infections in the community setting.

–Do you think informed consent would be required?
  1. No, it would disrupt the clinical encounter and would not be truly pragmatic
  2. No, but I would design it as cluster-randomized trial
  3. Yes, autonomy and patient decision-making are essential
  4. Yes, it is a regulatory requirement
Low risk pragmatic trials do not always require participants’ informed consent

BMJ 2019; 364, doi: https://doi.org/10.1136/bmj.l1092 (Published 27 March 2019)
Cite this as: BMJ 2019;364:l1092

Rafael Dal-Ré, senior investigator1, Cristina Avendaño-Solà, head2, Brigitte Bloechl-Daum, professor3, Anthonius de Boer, professor4, Stefan Eriksson, associate professor5, Uwe Fuhr, professor6, Søren Holm, professor7, Stefan K James, professor8, Robert J Mentz, investigator9, Emilio Perucca, professor10, Frits R Rosendaal, professor11, Shaun Treweek, professor12

Author affiliations

Correspondence to: R Dal-Ré rafael.dalre@quironsalud.es

Clinical trial regulations should remove unnecessary obstacles for the conduct of pragmatic trials assessing the comparative effectiveness of medicines posing no or minimal risk to participants, say Rafael Dal-Ré and colleagues
Internal validity

Revised Cochrane risk-of-bias tool for randomized trials (RoB 2)

Edited by Julian PT Higgins, Jelena Savović, Matthew J Page, Jonathan AC Sterne on behalf of the ROB2 Development Group

9 October 2018

Dedicated to Professor Douglas G Altman, whose contributions were of fundamental importance to development of risk of bias assessment in systematic reviews

Surrogate outcomes

– Or «of research fit to clinical purposes»
What is a surrogate endpoint

– Ideally, you would always want to employ a «hard» clinical endpoint
– But often for some reason you can not…
– … then, sometimes, you can use surrogate endpoints

– Is it a valid surrogate?
“...a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions. Molecular, histologic, radiographic, or physiologic characteristics are types of biomarkers. A biomarker is not an assessment of how an individual feels, functions, or survives.”

“...an endpoint that is used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives. A surrogate endpoint does not measure the clinical benefit of primary interest in and of itself, but rather is expected to predict that clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. From a US regulatory standpoint, surrogate endpoints and potential surrogate endpoints can be characterized by the level of clinical validation: validated surrogate endpoint, reasonably likely surrogate endpoint, candidate surrogate endpoint.”
Defining Surrogate Endpoints for Clinical Trials in Severe Falciparum Malaria

Attanee Jeeapant, Hugh W. Kingston, Katherine Plewes, Richard J. Maude, Josh Hanson, Trent Herdman, Stijne Leopold, Thatsanun Ngerseng, Prakaykaew Charunwatthana, Nguyen Hoan Phu, Aniruddha Ghose, Mahtab Uddin Hasan, Caterina I. Fanello, Md Abul Faiz, Tran Tinh Hien, Nicholas P. J. Day, Nicholas J. White, Arjen M. Dondorp

Conclusions

The relative changes in plasma lactate concentration assessed at 8 or 12 hours after admission are valid surrogate endpoints for severe malaria studies on antimalarial drugs or adjuvant treatments aiming at improving the microcirculation. Measures of coma recovery are not valid surrogate endpoints for mortality.
674 participants, genital HSV shedding was detected on 17% of days, and genital lesions on 10% of days.

- Within the same session, HSV shedding rates strongly correlated with lesion rates \( (\rho = 0.61, P < .0001) \).
- Relative reduction in the recurrence rate was 72% \( (P = .041) \) for recipients of the antiviral agent pritelivir as compared to recipients of placebo.
  - but it decreased to 21% \( (P = .75) \) after adjustment for HSV shedding rate.
- When evaluating valacyclovir and acyclovir, adjustment for the HSV shedding rate also led to a reduced association of these antivirals with the recurrence rate.
- Overall, 40%–82% of the antiviral effect on recurrences was explained by its effect on HSV shedding.

**Conclusion.** HSV genital shedding measured by PCR analysis in swab specimens self-collected daily is an appropriate surrogate outcome for genital herpes lesions because it is in the causal pathway to recurrences.
Regulating vaccines at the FDA: development and licensure of Zika vaccines

Marion F. Gruber and Philip R. Krause

In this regard, the clinical development program and regulatory strategy for a Zika vaccine must be tailored to the particular vaccine under investigation and may differ from that of other Zika vaccines. In addition, it will be influenced by several factors including characteristics of the vaccine, available nonclinical and clinical data for the particular vaccine and/or related vaccines, proposed indication, target population, and availability of an immune correlate of protection or a surrogate end point reasonably likely to predict clinical benefit.
Direct-acting antivirals for chronic hepatitis C (Review)

Authors’ conclusions

The evidence for our main outcomes of interest come from short-term trials, and we are unable to determine the effect of long-term treatment with DAAs. The rates of hepatitis C morbidity and mortality observed in the trials are relatively low and we are uncertain as to how DAAs affect this outcome. Overall, there is very low quality evidence that DAAs on the market or under development do not influence serious adverse events. There is insufficient evidence to judge if DAAs have beneficial or harmful effects on other clinical outcomes for chronic HCV. Simeprevir may have beneficial effects on risk of serious adverse event. In all remaining analyses, we could neither confirm nor reject that DAAs had any clinical effects. DAAs may reduce the number of people with detectable virus in their blood, but we do not have sufficient evidence from randomised trials that enables us to understand how SVR affects long-term clinical outcomes. SVR is still an outcome that needs proper validation in randomised clinical trials.
Assessing the validity of surrogate endpoints in the context of a controversy about the measurement of effectiveness of hepatitis C virus treatment

Claudia C Dobler,¹ Rebecca L Morgan,² Yngve Falck-Ytter,³ Victor M Montori,⁴ M Hassan Murad¹

Evidence-based medicine is to view results based on surrogate endpoints as less certain than results based on long term, final, patient-important outcomes and rate them as ‘lower quality evidence’.

However, careful appraisal of the validity of a surrogate endpoint as a measure of the final, patient-important outcome is more useful than an automatic judgement. In this guide, we use a contemporary and currently highly debated example of the surrogate endpoint ‘sustained viral response’ (ie, viral eradication considered to represent successful treatment) in patients treated for chronic hepatitis C virus. We demonstrate how the validity of a surrogate endpoint can be critically appraised to assess the quality of the evidence (ie, the certainty in estimates) and the implications for decision-making.
Question 3…
Interactive question

You are designing a phase II clinical trial, on a new combination antibiotic therapy, for ventilator-associated pneumonia due to MDR

–Would you employ bacterial load from respiratory samples as primary endpoint?
  1. Yes
  2. No
  3. Only if validated
Many discrepancies were observed between microbiological outcomes and clinical responses in patients with severe pneumonia.

The discrepancies were noted primarily as failures to eradicate in association with good clinical responses.

Acinetobacter, Enterobacter, P. aeruginosa, and S. maltophilia continued to be isolated even when clinical improvements were observed, whereas other organisms were generally eradicated when good clinical responses were achieved.

Many organisms persisting along with good clinical responses were associated with the presence or development of resistance to the antibiotic used for treatment.

<table>
<thead>
<tr>
<th>Demographic data &amp; microbiological outcomes</th>
<th>Clinical outcome (end of treatment)</th>
<th>Clinical outcome (follow up)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cured</td>
<td>Failed</td>
</tr>
<tr>
<td>Number of patients</td>
<td>55</td>
<td>19</td>
</tr>
<tr>
<td>Age (SD) (yr)</td>
<td>68.5 (13.2)</td>
<td>67.3 (14.2)</td>
</tr>
<tr>
<td>Sex (male/female)</td>
<td>34:21</td>
<td>13.6</td>
</tr>
<tr>
<td>Height (SD) (cm)</td>
<td>170.9 (11.1)</td>
<td>168.4 (10.8)</td>
</tr>
<tr>
<td>Weight (SD) (kg)</td>
<td>74.8 (26.4)</td>
<td>71.2 (15.3)</td>
</tr>
<tr>
<td>Charlson Weighted Index (SD)</td>
<td>2.0 (1.5)</td>
<td>2.7 (2.3)</td>
</tr>
<tr>
<td>Trauma</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Operation/procedure</td>
<td>28</td>
<td>16</td>
</tr>
<tr>
<td>Steroid</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Hemodialysis</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Mechanical ventilation</td>
<td>49</td>
<td>18</td>
</tr>
<tr>
<td>Endotracheal tube/tracheostomy</td>
<td>49</td>
<td>19</td>
</tr>
<tr>
<td>Bacteremia</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Microbiological outcome (end of treatment)</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>All organisms eradicated</td>
<td>47</td>
<td>9</td>
</tr>
<tr>
<td>Microbiological outcome (follow up)</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>All or part of organisms eradicated</td>
<td>44</td>
<td>9</td>
</tr>
</tbody>
</table>
Development and application of agreed standardised sets of outcomes, known as a ‘core outcome set.’

Rather, there is an expectation that the core outcomes will be collected and reported to allow the results of trials and other studies to be compared, contrasted and combined as appropriate; and that researchers will continue to collect and explore other outcomes as well.

Standards are like toothbrushes. Everybody wants one but nobody wants to use anybody else’s.

http://www.comet-initiative.org/
Appropriate endpoints for evaluation of new antibiotic therapies for severe infections: a perspective from COMBACTE’s STAT-Net

Jean-François Timsit², Marlieke E. A. de Kraker¹, Harriet Sommer¹, Emmanuel Weiss³, Esther Bettiol³, Martin Wolkewitz¹, Stavros Nikolakopoulos⁷, David Wilson⁸ and Stephan Harbarth⁹, on behalf of the COMBACTE-Net consortium

**Purpose:** In this era of rising antimicrobial resistance, slowly refilling antibiotic development pipelines, and an aging population, we need to ensure that randomized clinical trials (RCTs) determine the added benefit of new antibiotic agents effectively and in a valid way, especially for severely ill patients. Unfortunately, universally accepted endpoints for the evaluation of new drugs in severe infections are lacking.

**Methods:** We review and discuss the current practices and challenges regarding endpoints in RCTs in this field and propose novel approaches.

**Results:** Usual endpoints actually recommended for drug development suffer from important flaws. Mortality requires large sample size and only partly related to the infectious process. Clinical cure rate is highly subjective in critically ill patients where symptoms may be related to other intercurrent events. Currently, composite endpoints, hierarchical nested designs, and competing risks analysis seem to be the most promising new tools for designing and analyzing clinical trials in this area, although they require further validation.

**Conclusion:** Regulatory authorities, pharmaceutical companies, and clinicians need to agree on the most appropriate clinical endpoints for severe infections to ensure efficient approval of new, effective antibiotic agents.
<table>
<thead>
<tr>
<th>Indication</th>
<th>Methodology</th>
<th>Major Findings, Conclusions, and Recommendations</th>
<th>Date</th>
</tr>
</thead>
</table>
| CABP            | Review of literature, analysis of modern-day clinical trial data submitted in kind by pharmaceutical sponsors | *Progressive improvement in four symptoms (cough, dyspnea, chest pain, and sputum production) during the first 48 h of therapy was sufficiently well documented that an early response endpoint measure could be proposed.*
To assess durability of response and other late events, supportive information should be obtained by assessing outcomes at a fixed time point after therapy has been completed. Such information could include a late response endpoint similar to the traditional test-of-cure endpoint. Although based on limited data and requiring further research, an early response endpoint can be used to anchor a non-inferiority trial for this indication. The early response endpoint is thus suggested for possible use by FDA in review of registration trials and approval of applications in CABP while further research into this area is conducted. |
| ABSSSI          | Review of literature, analysis of modern-day clinical trial data submitted in kind by pharmaceutical sponsors | *Control of lesion spread at 48 to 72 h after randomization was sufficiently well documented that an early response endpoint measure could be proposed.*
To assess sustained response and other late events, supportive information should be obtained by assessing outcomes at a fixed time point after therapy has been completed. Such information could include a late response endpoint similar to the traditional test-of-cure (TOC) endpoint but more clearly defined. Although incompletely validated under the proposed conditions of use and requiring further research, an early response endpoint can be used to anchor a non-inferiority hypothesis in a trial for this indication. Thus, the Project Team supports a primary endpoint based on early response in review of registration trials and approval of applications in ABSSSI while further research into outcomes at later time points in this area is conducted. |
| HABP            | Review of literature                                                          | *A clinically meaningful endpoint on symptom improvement plus survival for non-ventilated patients could be based on the historical data for community-acquired bacterial pneumonia for which there is a large treatment effect to day 7 of antibiotic drug therapy.*
*There were some concerns that mortality and other differences between HABP and VABP suggest these are different diseases, meaning that combining both in a single trial could raise methodological issues.* |
| VABP            | Review of literature                                                          | *Despite the potential clinical trial implementation feasibility issues that have been raised with current FDA HABP/VABP Guidance, including an all-cause mortality (ACM) endpoint, most (90%) participants are comfortable with ACM as an endpoint, specifically for VABP, if trial feasibility could be addressed by changing other parameters of study design.*
The outstanding questions for use of ACM relate to timing of its assessment, as well as to whether there are suitable intermediate clinical endpoints. One concern with ACM is its lower incidence in registraional trials versus “real life.” It is hypothesized that making exclusion criteria less restrictive, and thereby increasing the severity of illness in the enrolled population, has the potential to facilitate enrollment.
*"A number of candidate changes to other aspects of trial design (eg, primary analysis set) were identified as promising potential approaches to improving feasibility, while maintaining scientific validity.* |
Safety outcomes

– RCTs performs poorly at detecting AEs
Quinolone- and fluoroquinolone-containing medicinal products

Table of contents

- Overview
- Key facts
- Public hearing
- All documents

Overview

Disabling and potentially permanent side effects lead to suspension or restrictions of quinolone and fluoroquinolone antibiotics
Financial ties of principal investigators and randomized controlled trial outcomes: cross sectional study

Rosa Ahn,1 Alexandra Woodbridge,2 Ann Abraham,2 Susan Saba,3 Deborah Korenstein,4 Erin Madden,2 W John Boscadin,2,5 Salomeh Keyhani2,5

OBJECTIVE
To examine the association between the presence of individual principal investigators’ financial ties to the manufacturer of the study drug and the trial’s outcomes after accounting for source of research funding.

DESIGN
Cross sectional study of randomized controlled trials (RCTs).

CONCLUSIONS
Financial ties of principal investigators were independently associated with positive clinical trial results. These findings may be suggestive of bias in the evidence base.
Commentary

Patient and public involvement in infection clinical research

S. Grier 1,*, A. Gibson 2, A. Micciche 3, E. Berry 3, A. MacGowan 1

1) North Bristol NHS Trust, Department of Infection Sciences, Bristol, UK
2) University of the West of England, Department of Health and Social Sciences, Bristol, UK

Patient and public involvement (PPI) in research is not related to patients as participants in research. PPI involves people as partners in the research process, collaborating with researchers to benefit all aspects of a project from the application for funding, or preclinical stages of medicines development, through to the dissemination of research findings [1]. This is based on the supposition that people who have been patients or carers have an experience, understanding or insight of illness that researchers do not.
3. Conclusions
Why Most Clinical Research Is Not Useful

John P. A. Ioannidis\textsuperscript{1,2,*}

- Blue-sky research cannot be easily judged on the basis of practical impact, but clinical research is different and should be useful. It should make a difference for health and disease outcomes or should be undertaken with that as a realistic prospect.
- Many of the features that make clinical research useful can be identified, including those relating to problem base, context placement, information gain, pragmatism, patient centeredness, value for money, feasibility, and transparency.
- Many studies, even in the major general medical journals, do not satisfy these features, and very few studies satisfy most or all of them. Most clinical research therefore fails to be useful not because of its findings but because of its design.
- The forces driving the production and dissemination of nonuseful clinical research are largely identifiable and modifiable.
- Reform is needed. Altering our approach could easily produce more clinical research that is useful, at the same or even at a massively reduced cost.

PLOS Medicine | DOI:10.1371/journal.pmed.1002049
So… is the king naked?

– It is not a king, just «primum inter pares»
– It wears the good, old, plain clothes of «good science»
  – The answer is only ever as good as the question
  – Do your homework before designing a RCT
  – Do your homework when designing a RCT
  – Do your homework before embarking into a RCT

– Do your homework after you have done a RCT
Hundreds of thousands of people have taken part in clinical trials that have not published results. Make their contributions count.

Sign the petition

Medical is broken
We need your help to fix it

http://www.alltrials.net/

Around half of clinical trials have never been reported. This is the story of the campaign to find them—and to fix medicine.