Rapid Critical Appraisal of Controlled Trials

Carl Heneghan
Dept of Primary Health Care
University of Oxford

November 23rd 2009
Five steps in EBM

1. Formulate an answerable question
2. Track down the best evidence
3. Critically appraise the evidence for:
   - Relevance
   - Validity
   - Impact (size of the benefit)
   - Applicability
4. Integrate with clinical expertise and patient values
5. Evaluate our effectiveness and efficiency
   - keep a record; improve the process
A CHECKLIST FOR APPRAISING RANDOMIZED CONTROLLED TRIALS

1. Was the objective of the trial sufficiently described?
2. Was a satisfactory statement given of the diagnostic criteria for entry to the trial?
3. Were concurrent controls used (as opposed to historical controls)?
4. Were the treatments well defined?
5. Was random allocation to treatments used?
6. Was the potential degree of blindness used?
7. Was there a satisfactory statement of criteria for outcome measures? Was a primary outcome measure identified?
8. Were the outcome measures appropriate?
9. Was a pre-study calculation of required sample size reported?
10. Was the duration of post-treatment follow-up stated?
11. Were the treatment and control groups comparable in relevant measures?
12. Were a high proportion of the subjects followed up?
13. Were the drop-outs described by treatment and control groups?
14. Were the side-effects of treatment reported?
15. How were the ethical issues dealt with?
16. Was there a statement adequately describing or referencing all statistical procedures used?
17. What tests were used to compare the outcome in test and control patients?
18. Were 95% confidence intervals given for the main results?
19. Were any additional analyses done to see whether baseline characteristics (prognostic factors) influenced the outcomes observed?
20. Were the conclusions drawn from the statistical analyses justified?
Victims of DVT are told that they can’t sue

(Except if it’s in Australia)

By Andrew Lew
The Daily Telegraph

DVT (Deep Vein Thrombosis) is a blood clot that forms in the veins of the leg. It can be a serious condition, and in some cases, it can lead to fatal complications. The condition is common in people who sit for long periods, such as flight passengers.

In Australia, DVT is covered by the Civil Aviation Safety Authority (CASA) regulations, which provide guidelines for airlines and other transport operators on how to manage the risks associated with DVT. These guidelines include the use of graduated compression stockings, regular leg exercises, and in-flight reminders.

However, in the UK, DVT is not covered by the same regulations, and passengers are not protected by the same legal framework as in Australia. This means that passengers in the UK are at risk of not being able to sue if they suffer from DVT as a result of their flight.

The case of the DVT victim is concerning not only because of the potential harm to the individual but also because it highlights the need for stronger regulations to protect passengers on flights. The lack of legal protection in the UK means that passengers are left vulnerable, and it is essential to address this issue to ensure the safety and well-being of all passengers.
Clinical Question

In people who take long-haul flights does wearing graduated compression stockings prevent DVT?

Page 71 and 95 in your books
Causes of an “Effect” in a controlled trial

• Who would consider wearing stockings on a long haul flight?
QUESTION:

Participants

Intervention Group (IG) & Comparison Group (CG)

Outcome
QUESTION:

Participants

Intervention Group (IG) & Comparison Group (CG)

Outcome
QUESTION:
Participants

Intervention Group (IG) & Comparison Group (CG)

Outcome

VALIDITY
Recruitment
QUESTION:

Participants

Intervention Group (IG) & Comparison Group (CG)

Outcome

VALIDITY

Recruitment

Allocation concealment comparable groups
QUESTION:

Participants

Intervention Group (IG) & Comparison Group (CG)

Outcome

VALIDITY

Recruitment

Allocation
- concealment
- comparable groups

Maintenance
- treated equally
- compliant
QUESTION:

Participants

Intervention Group (IG) & Comparison Group (CG)

Outcome

VALIDITY

Recruitment

Allocation concealment comparable groups

Maintenance
• Treated equally
• Compliant

Measurements blind? OR objective?
Appraisal checklist - RAMMbo

Study biases

1. Recruitment
   • Who did the subjects represent?

2. Allocation
   • Was the assignment to treatments randomised?
   • Were the groups similar at the trial’s start?

3. Maintenance
   • Were the groups treated equally?
   • Were outcomes ascertained & analysed for most patients?

4. Measurements
   • Were patients and clinicians “blinded” to treatment? OR
   • Were measurements objective & standardised?

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993
How were the patients recruited?
Randomization
Volunteers were randomized by sealed envelope to one of two groups.

Passengers were randomly allocated to one of two groups: one group wore class-I below-knee graduated elastic compression stockings, the other group did not.
• Take out the envelopes
• Sign the back
• You have now consented to the trial
• Please open your envelopes now
<table>
<thead>
<tr>
<th></th>
<th>Blue Bunnies</th>
<th>Pink Bunnies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argued with your boss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Been to New York</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ensuring Allocation Concealment

**BEST** – most valid technique
- Central computer randomization

**DOUBTFUL**
- Envelopes, etc

**NOT RANDOMIZED**
- Date of birth, alternate days, etc
Were the groups similar at the trial’s start?

By chance a greater proportion of women were included in the stocking group.

### Results

Volunteers were excluded before randomisation if they did not fulfil the entry requirements or could not attend hospital for investigation both before and after travel (figure). Thus, 231 of 479 volunteers were randomised. 27 passengers were unable to attend for subsequent ultrasound investigation because of ill-health (three), change of travel plans, or inability to keep appointments (24). Two who

<table>
<thead>
<tr>
<th></th>
<th>No stockings</th>
<th>Stockings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>116</td>
<td>115</td>
</tr>
<tr>
<td>Age (years)</td>
<td>62 (56-68)</td>
<td>61 (56-66)</td>
</tr>
<tr>
<td>Number of women (%)</td>
<td>61 (53%)</td>
<td>81 (70%)</td>
</tr>
<tr>
<td>Number with varicose veins</td>
<td>41</td>
<td>45</td>
</tr>
<tr>
<td>Days of stay</td>
<td>17 (13-32)</td>
<td>16 (13-27)</td>
</tr>
<tr>
<td>Hours flying time</td>
<td>22 (18-36)</td>
<td>24 (19-35)</td>
</tr>
<tr>
<td>Haemoglobin (g/L)</td>
<td>142 (133-149)</td>
<td>140 (133-147)</td>
</tr>
<tr>
<td>WBC ($\times 10^9$/L)</td>
<td>5-9 (5-0-7.3)</td>
<td>6-0 (5-0-6.9)</td>
</tr>
<tr>
<td>Packed cell volume</td>
<td>0.44 (0.42-0.47)</td>
<td>0.44 (0.41-0.46)</td>
</tr>
<tr>
<td>Platelets ($\times 10^9$/L)</td>
<td>240 (206-272)</td>
<td>242 (219-290)</td>
</tr>
<tr>
<td>Number FVL positive</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Number PGM positive</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Median (interquartile range) shown, unless otherwise indicated. WBC=white blood cells. FVL=factor V Leiden. PGM=prothrombin gene mutation.

**Table 1: Characteristics of study groups**
Appraisal checklist - **RAMMbo**

**Study biases**

1. **Recruitment**
   - Who did the subjects represent?

2. **Allocation**
   - Was the assignment to treatments randomised?
   - Were the groups similar at the trial’s start?

3. **Maintenance**
   - Were outcomes ascertained & analysed for most patients?
   - Were the groups treated equally?

4. **Measurements**
   - Were patients and clinicians “blinded” to treatment? OR
   - Were measurements objective & standardised?

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993
Effects of non-equal treatment

- Apart from actual intervention - groups should receive identical care!
  - Trial of Vitamin E in pre-term infants (1949)
  - Vit E "prevented" retrolental fibroplasia

Rx: Give placebo in an identical regime, and a standard protocol
Equal treatment in DVT study?

**Table 3: All drugs taken by volunteers who attended for examination before and after air travel***

<table>
<thead>
<tr>
<th>Drug Type</th>
<th>No stockings</th>
<th>Stockings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Hormone replacement therapy</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Thyroxine</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Antihypertensives, including diuretics</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Antieptic ulcer drugs</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>

*Includes additions to usual drugs.

THE LANCET • Vol 357 • May 12, 2001
Follow-up in DVT study?

• 200 of 231 analyzed (87%)

• 27 were unable to attend for subsequent ultrasound

• 2 were excluded from analysis because they were upgraded to business class

• 2 were excluded from analysis because they were taking anticoagulants

See figure on page 96

Scurr et al, Lancet 2001; 357:1485-89
Losses-to-follow-up

How many is too many?

“5-and-20 rule of thumb”

- 5% probably leads to little bias
- >20% poses serious threats to validity

Depends on outcome event rate and comparative loss rates in the groups

Loss to follow-up rate should not exceed outcome event rate and should not be differential
How important are the losses?

- Equally distributed?
  - **Stocking group**: 6 men, 9 women - 15
  - **No stocking group**: 7 men, 9 women - 16

- Similar characteristics?
  - No information provided
Intention-to-Treat Principle

Maintaining the randomization

**Principle:**

Once a patient is randomized, s/he should be analyzed in the group randomized to - even if they discontinue, never receive treatment, or crossover.

**Exception:** If patient is found on BLIND reassessment to be ineligible based on pre-randomization criteria.
Appraisal checklist

Study biases

1. Recruitment
   • Who did the subjects represent?

2. Allocation
   • Was the assignment to treatments randomised?
   • Were the groups similar at the trial’s start?

3. Maintainence
   • Were outcomes ascertained & analysed for most patients?
   • Were the groups treated equally?

4. Measurements
   • Were patients and clinicians “blinded” to treatment? OR
   • Were measurements objective & standardised?

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993
Measures in DVT study?

- Blood was taken from all participants before travel
- All participants had US once before travel (30 had US twice)
- All participants were seen within 48 hr of return flight, were interviewed and completed a questionnaire, had repeat US

Scurr et al, Lancet 2001; 357:1485-89
Measurement Bias - minimizing differential error

- Blinding – Who?
  - Participants?
  - Investigators?
  - Outcome assessors?
  - Analysts?

- Most important to use "blinded" outcome assessors when outcome is not objective!

- Papers should report WHO was blinded and HOW it was done

Schulz and Grimes. Lancet, 2002
Evaluation
Most passengers removed their stockings on completion of their journey. The nurse removed the stockings of those passengers who had continued to wear them. A further duplex examination was then undertaken with the technician unaware of the group to which the volunteer had been randomized.
Appraisal checklist

Study biases
1. Recruitment
   • Who did the subjects represent?
2. Allocation
   • Was the assignment to treatments randomised?
   • Were the groups similar at the trial’s start?
3. Maintainence
   • Were the groups treated equally?
   • Were outcomes ascertained & analysed for most patients?
4. Measurements
   • Were patients and clinicians “blinded” to treatment? OR
   • Were measurements objective & standardised?
5. Placebo Effect
6. Chance
7. Real Effect

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993
Placebo effect

Trial in patients with chronic severe itching

Itching score

Treatment vs no treatment for itching
Placebo effect

Trial in patients with chronic severe itching

Placebo effect - attributable to the expectation that the treatment will have an effect
Appraisal checklist

**Study biases**

1. **Recruitment**
   - Who did the subjects represent?
2. **Allocation**
   - Was the assignment to treatments randomised?
   - Were the groups similar at the trial’s start?
3. **Maintainence**
   - Were the groups treated equally?
   - Were outcomes ascertained & analysed for most patients?
4. **Measurements**
   - Were patients and clinicians “blinded” to treatment? OR
   - Were measurements objective & standardised?
5. **Placebo Effect**
6. **Chance**
7. **Real Effect**

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993
Two methods of assessing the role of chance

• **P-values** *(Hypothesis Testing)*
  - use statistical test to examine the ‘null’ hypothesis
  - associated with “p values” - if p<0.05 then result is statistically significant

• **Confidence Intervals** *(Estimation)*
  - estimates the range of values that is likely to include the true value
P-values (Hypothesis Testing) - in DVT study

- **Incidence of DVT**
  - Stocking group - 0
  - No Stocking group - 0.12

Risk difference $= 0.12 - 0 = 0.12$ ($P=0.001$)

The probability that this result would only occur by chance is

1 in 1000 $\rightarrow$ statistically significant
Confidence Intervals (Estimation)

- Incidence of DVT
  - Stocking group = 0
  - No Stocking group = 0.12

Risk difference = \((0.12 - 0) = 0.12\) (95% CI, 0.058 - 0.20)

The true value could be as low as 0.058 or as high as 0.20 - but is probably closer to 0.12

Since the CI does not include the ‘no effect’ value of ‘0’ \(\rightarrow\) the result is statistically significant
Appraisal checklist

**Study biases**

1. Recruitment
   - Who did the subjects represent?
2. Allocation
   - Was the assignment to treatments randomised?
   - Were the groups similar at the trial’s start?
3. Maintainence
   - Were the groups treated equally?
   - Were outcomes ascertained & analysed for most patients?
4. Measurements
   - Were patients and clinicians “blinded” to treatment? OR
   - Were measurements objective & standardised?
5. Placebo Effect
6. Chance
7. Real Effect

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993
Causes of an “Effect” in a controlled trial

- Who would now consider wearing stockings on a long haul flight?
M Clarke, S Hopewell, E Juszczak, A Eisinga, M Kjeldstrøm

Compression stockings for preventing deep vein thrombosis in airline passengers

Cochrane Database of Systematic Reviews 2006 Issue 4

- 10 RCTs (n = 2856)
- Seven included low or medium risk (n = 1548) and two included high risk participants (n = 1273).
- All flights > seven hours.
- Fifty of 2,637 participants in the trials of wearing stockings on both legs had a symptomless DVT; three wore stockings, 47 did not
  (OR 0.10, 95% CI 0.04 to 0.25, P < 0.00001).

- No deaths, pulmonary emboli or symptomatic DVTs were reported.
- Wearing stockings had a significant impact in reducing oedema (based on six trials).
- No significant adverse effects were reported.
## Compression stockings for preventing deep vein thrombosis in airline passengers

*Cochrane Database of Systematic Reviews 2006 Issue 4*

### Study Results

<table>
<thead>
<tr>
<th>Stockings</th>
<th>No stockings</th>
<th>Odds Ratio (Fixed)</th>
<th>Weight (%)</th>
<th>Odds Ratio (Fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LONFLIT 2</td>
<td>1/411</td>
<td>38.0</td>
<td>0.05 [0.01, 0.39]</td>
<td></td>
</tr>
<tr>
<td>LONFLIT 4 - Kendall 1</td>
<td>0/72</td>
<td>0.0</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>LONFLIT 4 - Kendall 2</td>
<td>0/60</td>
<td>5.1</td>
<td>0.19 [0.01, 4.12]</td>
<td></td>
</tr>
<tr>
<td>LONFLIT 4 - Scholl 1</td>
<td>0/179</td>
<td>9.3</td>
<td>0.11 [0.01, 2.03]</td>
<td></td>
</tr>
<tr>
<td>LONFLIT 4 - Scholl 2</td>
<td>0/130</td>
<td>7.2</td>
<td>0.14 [0.01, 2.71]</td>
<td></td>
</tr>
<tr>
<td>LONFLIT 4 - Traveno 1</td>
<td>0/97</td>
<td>0.0</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>LONFLIT 4 - Traveno 2</td>
<td>0/75</td>
<td>0.0</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>LONFLIT 5</td>
<td>2/178</td>
<td>14.2</td>
<td>0.28 [0.00, 1.37]</td>
<td></td>
</tr>
<tr>
<td>Scurr 2001</td>
<td>0/100</td>
<td>25.0</td>
<td>0.04 [0.00, 0.00]</td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI): 1314/1323

Test for heterogeneity chi-square=2.81 df=5 p=0.73 I²=0.0%

Test for overall effect z=4.92 p<0.00001
Thank you

Small groups