Using GRADE to determine the quality of evidence and strength of recommendations

Gordon Guyatt
Plan

- Questions for you
- GRADE background
- Certainty in estimates
- Evidence profiles
- Strength of recommendation
- Exercise in applying GRADE
Who are you?

- Clinicians
- Experience systematic review guideline panels
- Use of grading systems, experience
- How will you be using GRADE
  - As a user of systematic reviews, guidelines
  - As a systematic review author
  - As a guideline panelist
Grading good idea, but which grading system to use?

- Many available
  - Australian National and MRC
  - Oxford Center for Evidence-based Medicine
  - Scottish Intercollegiate Guidelines (SIGN)
  - US Preventative Services Task Force
  - American professional organizations
    - AHA/ACC, ACCP, AAP, Endocrine society, etc....
Result

Confusion
Common international grading system?

- GRADE (Grades of recommendation, assessment, development and evaluation)
- International group
  - Australian NMRC, SIGN, USPSTF, WHO, NICE, Oxford CEBM, CDC, CC
- ~ 40 meetings over last 19 years
  - (~10 – 120 attendants)
2004 BMJ, first description

2008 BMJ six part series
  - for guideline users

2010-19, 26 part series
  - For systematic review authors, HTA practitioners, guideline developers
Where does GRADE certainty apply?

- Evidence regarding therapeutic interventions
- Evidence regarding screening interventions
- Evidence regarding diagnostic impact
- Evidence regarding diagnostic accuracy
- Evidence regarding prognosis

- Rating for a single study?
- Rating for a body of evidence
- Both single study and body of evidence
>110 organizations have adopted GRADE
Figure 1. Belief and confidence: a two-dimensional weather report. (Reprinted by permission from the Wall Street Journal).
What are we grading?

- Two components
- Certainty/confidence in estimate of effect adequate to support decision (quality of body of evidence)
  - high, moderate, low, very low
- Strength of recommendation
  - strong and weak
Confidence/Certainty in evidence (quality of evidence)
Apparent disagreement, true agreement
Apparent disagreement, true agreement

No confidence → Very Low → Low → Moderate → High → Totally confident
Structured question

- Patients:
  - Women considering breast cancer screening
  - Age 40-9; 50 to 74; ≥ 75
  - No risk genetic mutation, chest radiation

- Intervention
  - film mammography

- Alternative
  - no screening
Need to define all patient-important outcomes and evaluate their importance

- **Desirable consequences**
  - *Reduction in breast cancer mortality*

- **Undesirable consequences**
  - False positive screening results - anxiety
  - Invasive procedures from positive results
  - Complications of invasive procedures
  - Unnecessary diagnosis and treatment
Determinants of confidence

- RCTs start high
- Observational studies start low
- What can lower confidence?
  - Risk of bias
  - Inconsistency
  - Indirectness
  - Imprecision
  - Publication bias
Risk of Bias - RCTs

- Well established
  - Concealment
  - Intention to treat principle observed
  - Blinding
  - Completeness of follow-up

- More recent
  - Selective outcome reporting bias
  - Stopping early for benefit
Inconsistency – happy with these results?

Relative Risk (95% CI)

- 0.73 (0.49, 1.07)
- 0.74 (0.59, 0.94)
- 0.76 (0.51, 1.12)
- 0.71 (0.56, 0.90)
- 0.73 (0.61, 0.88)
What about these?

Relative Risk (95% CI)

- 0.44 (0.30, 0.65)
- 0.45 (0.36, 0.60)
- 1.25 (0.84, 1.84)
- 1.17 (0.92, 1.49)
- 0.73 (0.61, 0.88)
What criteria were you using?

- Similarity of point estimates
  - less similar, less happy

- Overlap of confidence intervals
  - less overlap, less happy
test for heterogeneity
what is the p-value?

what is the null hypothesis for the test for heterogeneity?

Ho: RR1 = RR2 = RR3 = RR4

p=0.99 for heterogeneity
Heterogeneous

Test for heterogeneity
What is the p-value?

Relative Risk (95% CI)

- 0.44 (0.30, 0.65)
- 0.45 (0.36, 0.60)
- 1.25 (0.84, 1.84)
- 1.17 (0.92, 1.49)
- 0.73 (0.61, 0.88)

P-value for heterogeneity < 0.001
Why are we pooling?
What is the $I^2$?

$p=0.99$ for heterogeneity

$I^2=0\%$

Relative Risk (95% CI)

- 0.73 (0.49, 1.07)
- 0.74 (0.59, 0.94)
- 0.76 (0.51, 1.12)
- 0.71 (0.56, 0.90)
- 0.73 (0.61, 0.88)
What is the $I^2$?
Consistency of results

- Judgment of consistency
- Variation in size of effect
- Overlap in confidence intervals
- Statistical significance of heterogeneity
- $I^2$
Homogenous

If this result, what next?

Relative Risk (95% CI)

- 0.73 (0.49, 1.07)
- 0.74 (0.59, 0.94)
- 0.76 (0.51, 1.12)
- 0.71 (0.56, 0.90)
- 0.73 (0.61, 0.88)

$p=0.99$ for heterogeneity

$I^2=0\%$
Heterogeneous

If this result, what next?

Relative Risk (95% CI)

- 0.44 (0.30, 0.65)
- 0.45 (0.36, 0.60)
- 1.25 (0.84, 1.84)
- 1.17 (0.92, 1.49)
- 0.73 (0.61, 0.88)

p-value for heterogeneity < 0.001
I²=89%
Heterogeneity

- Look for explanation: Where?
  - Patients
  - Interventions
  - Comparators
  - Outcomes
  - Risk of bias
- No good explanation? What to do?
- Decrease confidence in effect estimates
Relative Risk with 95% CI for Vitamin D Non-vertebral Fractures

Favors Vitamin D Favors Control

Chapuy et al, (1994) 0.79 (0.69, 0.92)
Lips et al, (1996) 1.10 (0.87, 1.39)
Dawson-Hughes et al, (1997) 0.46 (0.24, 0.88)
Pfeifer et al, (2000) 0.48 (0.13, 1.78)
Meyer et al, (2002) 0.92 (0.68, 1.24)
Chapuy et al, (2002) 0.85 (0.64, 1.13)
Trivedi et al, (2003) 0.67 (0.46, 0.99)

Pooled Random Effect Model
0.82 (0.69 to 0.98)
p = 0.05 for heterogeneity, I² = 53%
Relative Risk with 95% CI for Vitamin D (Non-Vertebral Fractures, Dose >400)

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Relative Risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapuy</td>
<td>1994</td>
<td>0.79 (0.69, 0.92)</td>
</tr>
<tr>
<td>Dawson-Hughes</td>
<td>1997</td>
<td>0.46 (0.24, 0.88)</td>
</tr>
<tr>
<td>Pfeifer</td>
<td>2000</td>
<td>0.48 (0.13, 1.78)</td>
</tr>
<tr>
<td>Chapuy</td>
<td>2002</td>
<td>0.85 (0.64, 1.13)</td>
</tr>
<tr>
<td>Trivedi</td>
<td>2003</td>
<td>0.67 (0.46, 0.99)</td>
</tr>
</tbody>
</table>

Random Effects Estimate: p=0.26 for heterogeneity, I²=24%
Relative Risk with 95% CI for Vitamin D
(Non-Vertebral Fractures, Dose = 400)

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Relative Risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lips</td>
<td>1996</td>
<td>1.10 (0.87, 1.39)</td>
</tr>
<tr>
<td>Meyer</td>
<td>2002</td>
<td>0.92 (0.68, 1.24)</td>
</tr>
</tbody>
</table>

Random Effects Estimate: p=0.35 for heterogeneity, I²=0%

1.03 (0.86, 1.24)
Credibility of Subgroup Analysis

- Within-study comparison?
- Unlikely chance
- A priori hypothesis, direction specified
- One of small number hypotheses
- Biologically compelling
Within and between study

Situation 1
- Study 1 includes only men
  - RR of outcome with treatment X: 0.5
- Study 2 includes only women
  - RR of outcome with treatment X: 1.0

Situation 2
- Study 1 includes both men and women
  - RR of outcome with treatment X in men: 0.5
  - RR of outcome with treatment X in women: 1.0
Within and between study

Situation 1
- Study 1 includes only men
  - RR of outcome with treatment X: 0.5
- Study 2 includes only women
  - RR of outcome with treatment X: 1.0

Possible explanations?
- Men were older, sicker, etc.
- Study 1 used different doses
- Study 1 failed to blind, high LFUP, etc.
- Chance
- Treatment x really does benefit men not women
Within and between study

Within study
- Study 1 treatment x benefits men
- Study 1 treatment x fails to benefit women

Possible explanations?
- Men were older, sicker – No
- Study 1 used different doses – No
- Study 1 failed to blind, high LFUP, etc. – No
- Chance
- Treatment x really does benefit men not women

Within-study much stronger than between
Believe sub-group analysis high vs low dose vitamin?

- Unlikely chance \[ p = 0.006 \]
- Consistent across studies \( \text{yes} \)
- Small # a priori direction right \( \text{yes} \)
- Biologically compelling \( \text{yes} \)
- Within-study comparison \( \text{no} \)
Credibility of sub-group analysis

no way

sure thing

0

100
Confidence judgments: Directness

- Populations
  - Older, sicker or more co-morbidity
- Interventions
  - Warfarin in trials vs clinical practice
- Comparators
  - Standard care
- Outcomes
  - Important versus surrogate outcomes
  - Glucose control versus CV events
Directness

Interested in A versus B
available data A vs C, B vs C
Imprecision

- Small sample size
  - Small number of events

- Wide confidence intervals
  - Uncertainty about magnitude of effect

- How do you decide what is too wide?
Atrial fib at risk of stroke
Anticoagulants increases serious gi bleeding
  3% per year
1,000 patients 1 less stroke
  30 more bleeds for each stroke prevented
1,000 patients 100 less strokes
  3 strokes prevented for each bleed

Where is your threshold?
  How many strokes in 100 with 3% bleeding?
Imprecision – additional problem

- Small trials, large effect
  - Likely to be overestimate

- Analogy to stopping early

- Lack of prognostic balance

- Solution: optimal information size
  - # of pts from conventional sample size calculation
  - specify control group risk, $\alpha$, $\beta$, $\Delta$
High likelihood could lower quality

When to suspect

- Number of small studies
- Industry sponsored
Funnel Plot

Precision of estimate of treatment effect

Magnitude of the effect size
Publication Bias

Precision

Outcome Measure

Favor Intervention  Favor Control
Funnel Plot
Fish oil on mortality

Fig 4 | Funnel plot for assessment of publication bias for death from cardiac causes in 11 included studies reporting data on this outcome
What can raise confidence?

- Clinicians: no RCTs, high certainty?
- Large magnitude can rate up one level
  - Very large two levels
- Common criteria
  - Everyone used to do badly
  - Almost everyone does well
  - Quick action
- Hip replacement for hip osteoarthritis
Childhood lymphoblastic leukemia

Risk for CNS malignancies 15 years after cranial irradiation

No radiation: 1% (95% CI 0% to 2.1%)
12 Gy: 1.6% (95% CI 0% to 3.4%)
18 Gy: 3.3% (95% CI 0.9% to 5.6%).
## Certainty assessment criteria

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Confidence in estimates</th>
<th>Lower if</th>
<th>Higher if</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized trials</td>
<td>High</td>
<td>Risk of bias</td>
<td>Large Effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td>+1 Very large</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>Inconsistency</td>
<td>Dose response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Evidence of a gradient</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Indirectness</td>
<td>All plausible confounding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Would reduce a demonstrated effect or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td>+1 would suggest a spurious effect when results show no effect</td>
</tr>
<tr>
<td>Observational studies</td>
<td>Low</td>
<td>Imprecision</td>
<td>Publication bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>-1 Likely</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td>-2 Very likely</td>
</tr>
<tr>
<td></td>
<td>Very Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Trading off

What do patients/clinicians need to know
  - Relative risk reduction?
  - Absolute risk difference?

Why do meta-analyses always report relative?
Constant Relative Risk With Varying Risk Differences

- **Population 1**
  - RR 0.67
  - RD 10%

- **Population 2**
  - RR 0.67
  - RD 3.3%

- **Population 3**
  - RR 0.67
  - RD 1%
What do patients/clinicians need to know
  - Relative risk reduction?
  - Absolute risk difference?

Why do meta-analyses always report relative?

Body of evidence
  - How do we get risk difference?
How to get absolute?

- Meta-analysis get pooled relative risk
- Obtain baseline risk and multiply
- BR 10%, RRR 50%, RD 5%
## Quality Assessment

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. of patients (studies)</th>
<th>Risk of Bias</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality</th>
<th>Relative Risk (95% CI)</th>
<th>p-value</th>
<th>Illustrative risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoster episodes</td>
<td>38,546 (1)</td>
<td>No serious risk</td>
<td>only one study</td>
<td>Direct</td>
<td>Precise</td>
<td>Undetected</td>
<td>High</td>
<td>not reported</td>
<td></td>
<td>11.12 per 1,000 pt-years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.42 (difference 5.7 per 1,000 pt-years (p&lt; 0.001))</td>
</tr>
<tr>
<td>Post-herpetic neuralgia</td>
<td>38,546 (1)</td>
<td>No serious risk</td>
<td>only one study</td>
<td>Direct</td>
<td>Precise</td>
<td>Undetected</td>
<td>High</td>
<td>not reported</td>
<td></td>
<td>1.38 per 1,000 pt-years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.46 (difference 0.92 per 1,000 pt-years (p&lt; 0.001))</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>38,546 (1)</td>
<td>No serious risk</td>
<td>only one study</td>
<td>Direct</td>
<td>Precise</td>
<td>Undetected</td>
<td>High</td>
<td>Not reported</td>
<td></td>
<td>13 per 1,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19 (difference 6 per 1,000)</td>
</tr>
</tbody>
</table>

Zoster vaccine
## Beta blockers in non-cardiac surgery

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of participants (studies)</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Publication Bias</th>
<th>Quality</th>
<th>Relative Effect (95% CI)</th>
<th>Absolute risk difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocardial infarction</td>
<td>10,125 (9)</td>
<td>No serious limitations</td>
<td>No serious limitations</td>
<td>No serious limitations</td>
<td>No serious limitations</td>
<td>Not detected</td>
<td>High</td>
<td>0.71 (0.57 to 0.86)</td>
<td>1.5% fewer (0.7% fewer to 2.1% fewer)</td>
</tr>
<tr>
<td>Mortality</td>
<td>10,205 (7)</td>
<td>No serious limitations</td>
<td>No serious limitations</td>
<td>No serious limitations</td>
<td>Imprecise</td>
<td>Not detected</td>
<td>Moderate</td>
<td>1.23 (0.98 – 1.55)</td>
<td>0.5% more (0.1% fewer to 1.3% more)</td>
</tr>
<tr>
<td>Stroke</td>
<td>10,889 (5)</td>
<td>No serious limitations</td>
<td>No serious limitations</td>
<td>No serious limitations</td>
<td>Serious limitations</td>
<td>Not detected</td>
<td>Moderate</td>
<td>1.67 (1.00 – 2.80)</td>
<td>0.3% more (0 more to 1.5% more)</td>
</tr>
</tbody>
</table>
Most systems just use evidence about primary benefit outcome

But what about others (risk)?

What to do?

Options
- Ignore all but primary
- Lowest of any outcome
- Some blended approach
- Lowest of critical outcomes
Strength of Recommendation

- Strong recommendation
  - Benefits clearly outweigh risks/hassle/cost
  - Risk/hassle/cost clearly outweighs benefit

- What can downgrade strength?
  - Low confidence in estimates
  - Close balance between up and downsides
Aspirin after myocardial infarction
- 25% reduction in relative risk
- side effects minimal, cost minimal
- benefit obviously much greater than risk/cost

Anticoagulants in low risk atrial fibrillation
- anticoagulants reduce stroke vs ASA by 50%
- but if risk only 1% per year, ARR 0.5%
- increased bleeds by 1% per year
Strength of Recommendations

Aspirin after MI – do it

Anticoagulant rather than ASA in Afib:
-- probably do it
-- probably don’t do it
Determinants of strength of recommendation

<table>
<thead>
<tr>
<th>Factor</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance between desirable and undesirable effects</td>
<td>The larger the difference between the desirable and undesirable effects, the higher the likelihood that a strong recommendation is warranted. The narrower the gradient, the higher the likelihood that a weak recommendation is warranted</td>
</tr>
<tr>
<td>Quality of evidence</td>
<td>The higher the quality of evidence, the higher the likelihood that a strong recommendation is warranted</td>
</tr>
<tr>
<td>Values and preferences</td>
<td>The more values and preferences vary, or the greater the uncertainty in values and preferences, the higher the likelihood that a weak recommendation is warranted</td>
</tr>
<tr>
<td>Costs (resource allocation)</td>
<td>The higher the costs of an intervention—that is, the greater the resources consumed—the lower the likelihood that a strong recommendation is warranted</td>
</tr>
</tbody>
</table>

Additional criteria in evidence to decision frameworks;
Importance of the problem
Acceptability
Feasibility
Equity
Significance of strong vs weak

- Variability in patient preference
  - strong, almost all same choice (> 90%)
  - weak, choice varies appreciably

- Interaction with patient
  - strong, just inform patient
  - weak, ensure choice reflects values

- Use of decision aid
  - strong, don’t bother; weak, use the aid

- Quality of care criterion
  - strong, consider; weak, don’t consider
Flavanoids for Hemorrhoids

- Venotonic agents
- Popularity
  - 90 venotonics commercialized in France
  - None in Sweden and Norway
  - France 70% of world market
- Possibilities
  - French misguided
  - Rest of world missing out
Systematic review

- 14 trials, 1432 patients
- Key outcome
  - Risk not improving/persistent symptoms
  - 11 studies, 1002 patients, 375 events
  - RR 0.4, 95% CI 0.29 to 0.57
- Minimal side effects
- Is France right?
- What is the certainty of evidence?
What can lower confidence?

- Risk of bias
  - Lack of detail re concealment
  - Questionnaires not validated
- Indirectness – no problem
- Inconsistency, need to look at the results
## Study or sub-category

<table>
<thead>
<tr>
<th>Study or sub-category</th>
<th>RR (random) log[RR] (SE)</th>
<th>Weight %</th>
<th>RR (random) 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>01 Up to seven days</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chauvenet</td>
<td>-0.8916 (0.2376)</td>
<td>12.67</td>
<td>0.41 [0.26, 0.65]</td>
</tr>
<tr>
<td>Cospite</td>
<td>-2.2073 (0.6117)</td>
<td>5.51</td>
<td>0.11 [0.03, 0.36]</td>
</tr>
<tr>
<td>Thanapongsathorn</td>
<td>-0.4308 (0.2985)</td>
<td>11.16</td>
<td>0.65 [0.36, 1.17]</td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-29.36</td>
<td></td>
<td>0.37 [0.18, 0.77]</td>
</tr>
<tr>
<td>Test for heterogeneity: $\chi^2 = 6.92$, df = 2 (P = 0.03), $I^2 = 71.1%$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: $Z = 2.67$ (P = 0.008)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>02 Up to four weeks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annoni F</td>
<td>-1.6094 (0.7073)</td>
<td>4.50</td>
<td>0.20 [0.05, 0.80]</td>
</tr>
<tr>
<td>Clyne MB</td>
<td>-0.9943 (0.3983)</td>
<td>8.94</td>
<td>0.37 [0.17, 0.81]</td>
</tr>
<tr>
<td>Pirard J</td>
<td>-1.1712 (0.3086)</td>
<td>10.94</td>
<td>0.31 [0.17, 0.57]</td>
</tr>
<tr>
<td>Thanapongsathorn</td>
<td>-1.1087 (1.1098)</td>
<td>2.16</td>
<td>0.33 [0.04, 2.91]</td>
</tr>
<tr>
<td>Thorp</td>
<td>0.2624 (0.3291)</td>
<td>10.46</td>
<td>1.30 [0.68, 2.48]</td>
</tr>
<tr>
<td>Titapan</td>
<td>-0.8916 (0.3691)</td>
<td>9.56</td>
<td>0.41 [0.20, 0.85]</td>
</tr>
<tr>
<td>Wijayanegara</td>
<td>-0.5978 (0.1375)</td>
<td>14.97</td>
<td>0.55 [0.42, 0.72]</td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td></td>
<td></td>
<td>61.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.48 [0.32, 0.72]</td>
</tr>
<tr>
<td>Test for heterogeneity: $\chi^2 = 13.87$, df = 6 (P = 0.03), $I^2 = 56.7%$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: $Z = 3.57$ (P = 0.0004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>03 Further than four weeks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Godeberg</td>
<td>-1.7719 (0.3906)</td>
<td>9.1C</td>
<td>0.17 [0.08, 0.37]</td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.1C</td>
<td></td>
<td>0.17 [0.08, 0.37]</td>
</tr>
<tr>
<td>Test for heterogeneity: not applicable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: $Z = 4.54$ (P &lt; 0.00001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100.0C</td>
<td></td>
<td>0.40 [0.29, 0.57]</td>
</tr>
<tr>
<td>Test for heterogeneity: $\chi^2 = 28.66$, df = 10 (P = 0.001), $I^2 = 65.1%$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: $Z = 5.14$ (P &lt; 0.00001)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Favours treatment  Favours control**
Publication Bias

- Size of studies
  - 40 to 234 patients, most around 100
- All industry sponsored
Review: Phlebotonics for hemorrhoids
Comparison: 01 Venotonics vs placebo
Outcome: 08 Overall improvement: no improvement/some improvement
What can lower certainty?

- Risk of bias
  - Lack of detail re concealment
  - Questionnaires not validated
- Inconsistency
  - Almost all show positive effect, trend
  - Heterogeneity p < 0.001; I² 65.1%
- Indirectness
- Imprecision
  - RR 0.4, 95% CI 0.29 to 0.57
- Publication bias
  - 40 to 234 patients, most around 100
Is France right?

- Recommendation
  - Yes
  - No against use

- Strength
  - Strong
  - Weak
Conclusion

- Systematic review, HTA need quality evidence
- Guideline need recommendation strength
- GRADE very widely increasingly used
- Transparent, explicit to quality, strength
- Do you tweet?
@EBCPMcMaster
Follow us on Twitter