Why this talk?

- You will be seeing a lot of GRADE
- Exemplifies three key principles of EBHC
  - need for systematic reviews of best evidence
  - hierarchy of evidence
  - need for values and preferences
- If you understand GRADE you understand how to use evidence to inform practice
Plan

- GRADE background
- Two steps
  - certainty in estimates (quality of evidence)
  - strength of recommendation
- Evidence profiles
- An exercise in applying GRADE
GRADE (Grades of recommendation, assessment, development and evaluation)

International group

- Australian NMRC, SIGN, USPSTF, WHO, NICE, Oxford CEBM, CDC, CC

~ 40 meetings over last 16 years
>100 organizations have adopted GRADE
What are we grading?

Two components

Strength of recommendation: Strong and weak (conditional)
Determinants of confidence

- RCTs start high
- Observational studies start low
- What can lower confidence?
## Determinants of confidence

- **Bias**
  - study design and implementation
    - concealment, blinding, loss to follow-up
  - publication bias

- **Imprecision**
  - wide confidence intervals

- **Indirectness**
  - patients, interventions
  - outcomes
  - indirect comparisons
Consistency of results

- Variation in size of effect
- Overlap in confidence intervals
- Statistical significance of heterogeneity
- $I^2$
What can raise confidence?

- 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
## 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></td>
<td>-2 Very serious</td>
<td>All plausible confounding</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Indirectness</td>
<td>+1 Would reduce a demonstrated effect or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 would suggest a spurious effect when results show no effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td></td>
</tr>
<tr>
<td>Observational studies</td>
<td>Very Low</td>
<td>Imprecision</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Publication bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Likely</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very likely</td>
<td></td>
</tr>
</tbody>
</table>
## 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>Possibly inconsistent</td>
<td>No serious limitations</td>
<td>Imprecise</td>
<td>Not detected</td>
<td>Moderate or low</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>No serious limitations</td>
<td>Not detected</td>
<td>High</td>
<td>2.21 (1.37 – 3.55)</td>
<td>0.5% more (0.2% more to 1.3% more)</td>
</tr>
</tbody>
</table>
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
Risk/Benefit tradeoff

- 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 1.5% per year
Aspirin after MI – do it

Anticoagulants vs than ASA in low risk Afib
  -- probably do it
  -- probably don’t do it
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
Venotonic agents
- increase venous return

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 quality 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
Review: Phlebotonics for hemorrhoids
Comparison: 01 Venotonics vs placebo
Outcome: 08 Overall improvement: no improvement/some improvement

<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><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>11.18</td>
<td>0.65 [0.36, 1.17]</td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td></td>
<td></td>
<td>29.36 [0.37, 0.77]</td>
</tr>
<tr>
<td>Test for heterogeneity: Chi² = 6.92, df = 2 (P = 0.03), I² = 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>
</tbody>
</table>

| **02 Up to four weeks** |                         |          |                  |
| Annoni F               | -1.6094 (0.7073)         | 4.50     | 0.20 [0.05, 0.80]|
| Clyne MB               | -0.9943 (0.3983)         | 8.94     | 0.37 [0.17, 0.81]|
| Pirard J               | -1.1712 (0.3086)         | 10.94    | 0.31 [0.17, 0.57]|
| Thanapongsathorn       | -1.1087 (1.1098)         | 2.18     | 0.33 [0.04, 2.91]|
| Thorp                  | 0.2624 (0.3291)          | 10.46    | 1.30 [0.68, 2.48]|
| Titapan                | -0.8916 (0.3691)         | 9.56     | 0.41 [0.20, 0.85]|
| Wijayanegara           | -0.5978 (0.1375)         | 14.97    | 0.55 [0.42, 0.72]|
| **Subtotal (95% CI)** |                         |          | 61.54 [0.48, 0.72]|
| Test for heterogeneity: Chi² = 13.87, df = 6 (P = 0.03), I² = 56.7% |
| Test for overall effect: Z = 3.57 (P = 0.0004) |

| **03 Further than four weeks** |                         |          |                  |
| Godeberg                | -1.7719 (0.3906)         | 9.10     | 0.17 [0.08, 0.37]|
| **Subtotal (95% CI)** |                         |          | 9.10 [0.17, 0.37]|
| Test for heterogeneity: not applicable |
| Test for overall effect: Z = 4.54 (P < 0.00001) |

| **Total (95% CI)** |                         |          |                  |
| Test for heterogeneity: Chi² = 28.66, df = 10 (P = 0.001), I² = 65.1% |
| Test for overall effect: Z = 5.14 (P < 0.00001) |

![Forest plot indicating treatment effects and statistical significance](attachment:image.png)
Publication bias?

- Size of studies
  - 40 to 234 patients, most around 100
- All industry sponsored
What can lower confidence?

- **Risk of bias**
  - lack of detail re concealment
  - questionnaires not validated

- **Inconsistency**
  - almost all show positive effect, trend
  - heterogeneity $p < 0.001$; $I^2$ 65.1%

- **Indirectness**

- **Imprecision**
  - RR $0.4$, 95% CI $0.29$ to $0.57$

- **Publication bias**
  - 40 to 234 patients, all industry sponsored
Is France right?

- Recommendation
  - yes
  - no against use

- Strength
  - strong
  - weak
Clinicians, policy makers need summaries
- confidence in estimates
- strength of recommendations

Explicit rules
- transparent, informative

GRADE
- simple, transparent, systematic
- increasing wide adoption
- captures all key elements of EBM approach