This session is going to be interactive. You will be breaking into groups to solve problems. So please sit with people with whom you will enjoy a conversation.

**My first abstract**

- Prolix
- Obfuscated
GRADE:
One cutting edge of EBHC

- GRADE background
- two steps
  - quality of evidence
  - strength of recommendation
- profiles and summary of findings
- an example in applying GRADE

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
  - confidence in estimates (quality of evidence)
  - strength of recommendation
- evidence profiles
- an exercise in applying GRADE

Proliferation of systems

Common international grading

- GRADE (Grades of recommendation, assessment, development and evaluation)
- international group
  - Australian NMRC, SIGN, USPSTF, WHO, NICE, Oxford CEBM, CDC, CC
- ~ 35 meetings over last 15 years
- The system – over 80 organizations thus far
What are we grading?

two components

strength of recommendation:
strong and weak

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

- if inconsistency, look for explanation
  - patients, intervention, outcome, methods

- judgment of consistency
  - variation in size of effect
  - overlap in confidence intervals
  - statistical significance of heterogeneity
  - $I^2$
Relative Risk with 95% CI for Vitamin D
Non-vertebral Fractures

Chapuy et al, (2002) 0.85 (0.64, 1.13)
Chapuy et al, (2002) 0.85 (0.64, 1.13)
Pooled Random Effect Model
0.82 (0.69 to 0.98)
p= 0.05 for heterogeneity, I^2=53%

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)

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
## Confidence assessment criteria

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Confidence in estimates</th>
<th>Lower if Risk of bias</th>
<th>Higher if Large effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right trial</td>
<td>High</td>
<td>-1 Serious</td>
<td>+1 Large</td>
</tr>
<tr>
<td>Right trial</td>
<td>Moderate</td>
<td>-2 Very serious</td>
<td>+2 Very large</td>
</tr>
<tr>
<td>Observational study</td>
<td>Low</td>
<td>Inconsistency</td>
<td>Dose response</td>
</tr>
<tr>
<td>Observational study</td>
<td>Low</td>
<td>-1 Serious</td>
<td>+1 Evidence of a gradient</td>
</tr>
<tr>
<td>Observational study</td>
<td>Low</td>
<td>-2 Very serious</td>
<td>All plausible confounding</td>
</tr>
<tr>
<td>Observational study</td>
<td>Very low</td>
<td>Imprecision</td>
<td>+1 Would reduce a demonstrated effect or</td>
</tr>
<tr>
<td>Observational study</td>
<td>Very low</td>
<td>-1 Serious</td>
<td>+1 Would suggest a spurious effect when</td>
</tr>
<tr>
<td>Observational study</td>
<td>Very low</td>
<td>-2 Very serious</td>
<td>results show no effect</td>
</tr>
</tbody>
</table>

## Beta blockers in non-cardiac surgery

<table>
<thead>
<tr>
<th>Quality Assessment</th>
<th>Summary of Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quality</td>
</tr>
<tr>
<td>Outcome</td>
<td>Number of participants (studies)</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>10,125 (9)</td>
</tr>
<tr>
<td>Mortality</td>
<td>10,205 (7)</td>
</tr>
<tr>
<td>Stroke</td>
<td>10,889 (5)</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

- warfarin in low risk atrial fibrillation
  - warfarin reduces stroke vs ASA by 50%
  - but if risk only 1% per year, ARR 0.5%
  - increased bleeds by 1% per year
Aspirin after MI – do it

Warfarin rather than ASA in low risk Afib
  -- probably do it
  -- probably don’t do it

---

Strength of recommendations

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
  - mechanism unclear, increase venous return

- popularity
  - 90 venotonics commercialized in France
  - none in Sweden and Norway
  - France 70% of world market

- possibilities
  - French misled
  - 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

<table>
<thead>
<tr>
<th>Study or sub-category</th>
<th>RR (random)</th>
<th>Weight</th>
<th>RR (random)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log[RR] (SE)</td>
<td>%</td>
<td>95% CI</td>
</tr>
<tr>
<td><strong>01 Up to 7 days</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chauvenet</td>
<td>-0.834 (0.2376)</td>
<td>12.67</td>
<td>0.41 [0.26, 0.65]</td>
</tr>
<tr>
<td>Cospite</td>
<td>-2.2073 (0.4177)</td>
<td>5.51</td>
<td>0.21 [0.09, 0.43]</td>
</tr>
<tr>
<td>Thanapongsathorn</td>
<td>-1.4304 (0.1985)</td>
<td>14.18</td>
<td>0.46 [0.34, 0.62]</td>
</tr>
<tr>
<td><strong>Total</strong> (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-2.9635 (1.4523)</td>
<td>100.00</td>
<td>0.48 [0.32, 0.72]</td>
</tr>
</tbody>
</table>

Test for heterogeneity: Chi² = 6.92, df = 2 (P = 0.03), I² = 71.1%
Test for overall effect: Z = 2.67 (P = 0.008)

| **02 Up to 4 weeks**  |             |        |             |
| Annoni F              | -1.6094 (0.7073) | 4.50  | 0.28 [0.13, 0.46] |
| Clyne MB              | -0.9943 (0.3993) | 8.94  | 0.37 [0.22, 0.55] |
| Pirard J              | -1.1722 (0.3966) | 10.94 | 0.31 [0.17, 0.57] |
| Thanapongsathorn      | -1.1887 (0.1498) | 12.18 | 0.33 [0.27, 0.53] |
| Thorp                 | 0.2424 (0.3292)  | 10.44 | 1.30 [0.44, 2.48] |
| Thapan               | -0.6916 (0.1963) | 9.54  | 0.42 [0.26, 0.65] |
| Wrayneagera          | -0.5978 (0.3777) | 14.97 | 0.55 [0.42, 0.72] |
| **Total** (95% CI)    |             |        |             |
|                       | -0.9618 (1.1093) | 100.00 | 0.44 [0.32, 0.72] |

Test for heterogeneity: Chi² = 13.87, df = 6 (P = 0.001), I² = 56.7%
Test for overall effect: Z = 3.57 (P < 0.0004)

| **03 Further than 4 weeks** |             |        |             |
| Goldberg              | -1.7719 (0.3966) | 9.15  | 0.27 [0.19, 0.37] |
| **Total** (95% CI)    |             |        |             |
|                       | -1.1310 (0.2169) | 100.00 | 0.40 [0.29, 0.52] |

Test for heterogeneity: Chi² = 28.65, df = 10 (P = 0.001), I² = 65.1%
Test for overall effect: Z = 5.14 (P < 0.0001)
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 confidence?

- detailed design and execution
  - 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
**Conclusion**

- 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