Interpreting the Results of Patient-reported Outcomes

Gordon Guyatt
McMaster University
Plan

- Something about you
- What is a patient–important outcome?
- The problem of interpretability
- How to interpret results in individual studies
  - The minimal important difference
- How to interpret results in systematic reviews and meta–analyses
Something about you

- Experience conducting a systematic review
  - If not, plans
- Experience leading a systematic review
  - If not, plans
- Experience continuous outcomes in SR?
- Experience different measures for some outcome
  - Came to find out
Clinical Outcomes Assessment -- Sources and Examples

Biomarkers
- Cholesterol (coronary disease)
- C-reactive protein (inflammation)

Clinician-Reported
- Global impression of severity
- Performance status
- Forced expiratory volume

Observer-Reported
- Cough
- Activity level
- Sleep

Patient-Reported
- Symptoms
- Function
- Quality of life

Survival and Major Morbid Events
Patient-Reported Outcomes (PRO)

- **PRO**: Any report directly from patients, without interpretation by anyone else, about how they function or feel in relation to health condition and therapy (from diaries, questionnaires, interviews, etc.)
- Term PRO requires construct be specified, i.e., respiratory symptoms, physical function, reduction in pain severity
- Almost invariably important to patients
- What PROs have you seen in the literature?
Plan

- What is a patient-important outcome?
- *The problem of interpretability*
- How to interpret results in individual studies
  - The minimal important difference
- How to interpret results in systematic reviews and meta-analyses
Interpretability: The problem

- Mean score for treatment group improves 5 points on the PRO measure, no change in control
- Is this trivial, large, or somewhere between?
- Statistically significant – does that help?
Effect of alefacept on quality of life in 553 patients with psoriasis

Alefacept improved mean Dermatology Quality of Life Scale scores compared with placebo: 4.4 vs. 1.8 at 2 weeks after the last dose (P<0.0001)

Magnitude of Effect?
- trivial, small but important, large?
Plan

- What is a patient-important outcome?
- The problem of interpretability
- *How to interpret results in individual studies*
  - *The minimal important difference*
- How to interpret results in systematic reviews and meta-analyses
Randomized trial of lung volume reduction surgery

- Severe emphysema over inflated
- Reducing lung volume may improve mechanical properties
- RCT of 55 pts followed for 1 year
- Key QOL CRQ
  - Dyspnea, fatigue, motional function
- 1.5 point difference: recommend surgery?
- What could investigators do to help?
Minimally important difference

- Smallest change that patients would consider important

- Approaches
  - Patient scenarios
  - Between-patient ratings
  - Within-patient ratings
    - *Global ratings of change*
      - *Are you the same, a little better, a lot better*
Establishing Interpretability

- Chronic Respiratory Questionnaire (CRQ)
- Chronic Heart Failure Questionnaire (CHQ)
- 20 Items
  - Dyspnea
  - Fatigue
  - Emotional Function
- Evidence for validity, responsiveness
1 EXTREMELY SHORT OF BREATH
2 VERY SHORT OF BREATH
3 QUITE A BIT SHORT OF BREATH
4 MODERATE SHORTNESS OF BREATH
5 SOME SHORTNESS OF BREATH
6 A LITTLE SHORTNESS OF BREATH
7 NOT AT ALL SHORT OF BREATH
Establishing Interpretability

- Clinical impression
  - MID 0.5 per question
- 31 patients respiratory rehab program
  - before, 2, 6, 12, and 24 weeks after
- 24 CAL patients in bronchodilator trial
- 20 patient with CHF in digoxin trial
Global rating of change

Overall, has there been any change in your shortness of breath since the last time you saw us?

1 WORSE
2 ABOUT THE SAME
3 BETTER
1 ALMOST THE SAME, HARDLY ANY WORSE AT ALL
2 A LITTLE WORSE
3 SOMEWHAT WORSE
4 MODERATELY WORSE
5 A GOOD DEAL WORSE
6 A GREAT DEAL WORSE
7 A VERY GREAT DEAL WORSE
### All Trials Combined Mean Change per Question

<table>
<thead>
<tr>
<th></th>
<th>Unchanged</th>
<th>Small Important</th>
<th>Moderate</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dyspnea</strong></td>
<td>0.10</td>
<td><strong>0.43</strong></td>
<td>0.96</td>
<td>1.47</td>
</tr>
<tr>
<td><strong>Fatigue</strong></td>
<td>0.03</td>
<td><strong>0.46</strong></td>
<td>0.88</td>
<td>0.96</td>
</tr>
<tr>
<td><strong>Emotional Function</strong></td>
<td>0.02</td>
<td><strong>0.49</strong></td>
<td>0.81</td>
<td>0.86</td>
</tr>
</tbody>
</table>
Would you recommend surgery to your patients on the basis of these results?
Interpreting MID Results

- RCT respiratory rehabilitation in COPD
- Assume MID is 0.50 and patients mean improvement vs control is 0.25
- What is your conclusion about rehabilitation?
- Does this mean no one benefits?
- What if 0.6 – everyone benefits?
- If 0.25 mean change could mean:
  - 75% have 0 improvement
  - 25% have 1.0
  - NNT of 4
### Differences between rehabilitation and conventional care in CAL

<table>
<thead>
<tr>
<th>CRQ domain</th>
<th>Difference between groups</th>
<th>Estimated proportion better on rehabilitation</th>
<th>Estimated proportion better on conventional care</th>
<th>Proportion benefiting from rehabilitation</th>
<th>NNT for a single patient to benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>P value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyspnoea</td>
<td>0.60</td>
<td>0.0003</td>
<td>0.47</td>
<td>0.28</td>
<td>0.19</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.45</td>
<td>0.06</td>
<td>0.45</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Emotional function</td>
<td>0.40</td>
<td>0.001</td>
<td>0.47</td>
<td>0.17</td>
<td>0.30</td>
</tr>
</tbody>
</table>
Plan

- What is a patient-important outcome?
- The problem of interpretability
- How to interpret results in individual studies
  - The minimal important difference
- How to interpret results in systematic reviews and meta-analyses
  - When all studies use the same PROM
Meta-analysis

- Studies all use same or similar outcome
- Could give weighted mean difference in natural units
- Not intuitively interpretable to the audience

Solution
- MID if available
- Range of possible results if not
Would you recommend respiratory rehabilitation to your patients?

<table>
<thead>
<tr>
<th>CRQ</th>
<th>Point estimate (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyspnea</td>
<td>1.06 (0.85, 1.26)</td>
</tr>
<tr>
<td>Emotional Function</td>
<td>0.76 (0.52, 1.00)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.92 (0.71, 1.13)</td>
</tr>
<tr>
<td>Mastery</td>
<td>0.97 (0.74, 1.20)</td>
</tr>
<tr>
<td>Overall</td>
<td>0.94 (0.57, 1.32)</td>
</tr>
</tbody>
</table>
Alternative: dichotomize

- Rankin Stroke Scale
- Five levels
  - no symptoms
  - minor handicap
    - restriction in life style, can look after self
  - moderate handicap
    - restrict life style, prevent independent existence
  - moderately severe handicap
    - clearly prevent independence, no constant attention
  - severe handicap, require constant attention
Systematic review of RCTs of thrombolysis in acute stroke

- Use Rankin threshold 2 to 3
  - 2 minor handicap
  - 3 moderate handicap
  - proportion “dead or disabled”

- “Death or dependency”
  - odds ratio 0.84 (95% CI 0.75 to 0.95)
  - 4% absolute risk reduction
  - NNT 25
Flavanoids for Hemorrhoids

- **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

- **Key outcome**
  - Risk not improving/persistent symptoms
  - 11 studies, 1002 patients, 375 events
Phlebotonics for Hemorrhoids (Venotonics vs. Placebo)
Relative Risk (95%CI)

- Chauvenet 0.41 (0.26, 0.65)
- Cospite 0.11 (0.03, 0.36)
- Thanapongsathorn 0.65 (0.36, 1.17)
- Annoni 0.20 (0.05, 0.80)
- Clyne 0.37 (0.17, 0.81)
- Pirard 0.31 (0.14, 0.57)
- Thanapongsathorn 0.33 (0.04, 2.91)
- Thorp 1.30 (0.68, 2.48)
- Titapan 0.41 (0.20, 0.85)
- Wijayanegara 0.55 (0.42, 0.72)
- Godeberg 0.17 (0.08, 0.37)

Pooled Estimate (95%CI) 0.40 (0.29, 0.57)
Plan

- What is a patient-important outcome?
- The problem of interpretability
- How to interpret results in individual studies
  - The minimal important difference
- *How to interpret results in systematic reviews and meta-analyses*
  - *When studies use different PROMs for the same construct*
Plan

- When studies use different PROMS
  - Standardized mean difference
  - Natural units
  - Dichotomize
  - Ratio of means
  - MID units
Do clinicians understand treatment effects?

- Cross-sectional, paper-based survey
  - Academic centers in 8 countries,
  - Internal and family medicine, 531/610 (87%)
- Summary estimates hypothetical interventions vs placebo chronic pain
Clinicians’ understanding

In units of the pooled standard deviation of all the pain scores in the opioid and control groups, expressed as a standardized mean difference, a meta-analysis finds the effect of intervention A vs placebo control for patient-reported pain is 0.20 standard deviation units in favor of intervention A. The magnitude of this difference is:

- trivial difference, probably not important
- small difference, but probably important
- moderate difference, surely important
- large difference, very important
In units of the pooled standard deviation of all the pain scores in the opioid and control groups, expressed as a standardized mean difference, a meta-analysis finds the effect of intervention A vs placebo control for patient-reported pain is 0.80 standard deviation units in favor of intervention A. The magnitude of this difference is:

☐ trivial difference, probably not important
☐ small difference, but probably important
☐ moderate difference, surely important
☐ large difference, very important
In units of the pain scale, where 0 represents no pain and 10 represents the worst pain ever on numeric rating scale, a meta-analysis finds the effect of intervention C vs placebo control for patient-reported pain to be 2.0 in favor of intervention C. The magnitude of difference is:

- trivial difference, probably not important
- small difference, but probably important
- moderate difference, surely important
- large difference, very important
In units of the pain scale, where 0 represents no pain and 10 represents the worst pain ever on numeric rating scale, a meta-analysis finds the effect of intervention C vs placebo control for patient-reported pain to be 0.60 in favor of intervention C. The magnitude of difference is:

- trivial difference, probably not important
- small difference, but probably important
- moderate difference, surely important
- large difference, very important
As a relative risk (or risk ratio), where the ratio of the risk of the event occurring in the treatment group is divided by the risk in the placebo control group, a meta-analysis finds the effect of intervention D vs placebo control for patient-reported pain to be 0.80 (80%) in favor of intervention D, representing a relative risk reduction of 20% (relative to the control group, 20% fewer patients experience moderate to severe pain). The magnitude of difference is:

- trivial difference, probably not important
- small difference, but probably important
- moderate difference, surely important
- large difference, very important
As a relative risk (or risk ratio), where the ratio of the risk of the event occurring in the treatment group is divided by the risk in the placebo control group, a meta-analysis finds the effect of intervention D vs placebo control for patient-reported pain to be 0.50 (50%) in favor of intervention D, representing a relative risk reduction of 50% (relative to the control group, 50% fewer patients experience moderate to severe pain). The magnitude of difference is:

- trivial difference, probably not important
- small difference, but probably important
- moderate difference, surely important
- large difference, very important
As a risk difference, where the risk of in the treatment group is subtracted from the risk in the placebo control group, a meta-analysis finds the effect of intervention E vs placebo control for patient-reported pain to be 0.20 (20%) in favor of intervention E, representing 20 fewer patients per 100 experiencing moderate to severe pain. The magnitude of difference is:

- trivial difference, probably not important
- small difference, but probably important
- moderate difference, surely important
- large difference, very important
Plan

- When studies use different outcomes
  - *Standardized mean difference*
  - Natural units,
  - Dichotomize,
  - ratio of means,
  - MID units
Studies different instruments

- CRQ is one QoL measure for CAL
- St. George’s respiratory questionnaire another
- Some studies use one and some other?
  - What now?

- Divide each effect by standard deviation
- Ultimate result in SD units
- “Effect size” or SMD
- Study shows effect size of 0.4
  - Trivial, small but important, medium or large effect?
Cohen:
small effect 0.2 SD units
moderate effect 0.5
large effect 0.8

More recent suggestions in terms of MID
across all instruments 0.5 or 0.35

Rules of thumb likely to be limited
# Results – SD Units

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Experimental Mean</th>
<th>SD</th>
<th>Total</th>
<th>Control Mean</th>
<th>SD</th>
<th>Total</th>
<th>Weight</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.1 SGRQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boxall 2005</td>
<td>5.8</td>
<td>11.8</td>
<td>23</td>
<td>1.4</td>
<td>13.3</td>
<td>24</td>
<td>6.8%</td>
<td>0.34 [-0.23, 0.92]</td>
<td></td>
</tr>
<tr>
<td>Chlumsky 2001</td>
<td>4.07</td>
<td>19.76</td>
<td>13</td>
<td>4.22</td>
<td>19.2</td>
<td>6</td>
<td>3.9%</td>
<td>-0.01 [-0.97, 0.96]</td>
<td></td>
</tr>
<tr>
<td>Engstrom 1999</td>
<td>-0.3</td>
<td>17.3</td>
<td>26</td>
<td>-0.5</td>
<td>16.2</td>
<td>24</td>
<td>7.0%</td>
<td>0.01 [-0.54, 0.57]</td>
<td></td>
</tr>
<tr>
<td>Finnerly 2001</td>
<td>9.3</td>
<td>12.2</td>
<td>24</td>
<td>2.2</td>
<td>15</td>
<td>25</td>
<td>6.9%</td>
<td>0.51 [-0.06, 1.08]</td>
<td></td>
</tr>
<tr>
<td>Ringbaek 2000</td>
<td>2.1</td>
<td>19</td>
<td>17</td>
<td>2.2</td>
<td>17</td>
<td>19</td>
<td>6.1%</td>
<td>-0.01 [-0.66, 0.65]</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>429</strong></td>
<td><strong>390</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>429</strong></td>
<td><strong>390</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>0.73 [0.49, 0.96]</strong></td>
<td><strong>0.73 [0.49, 0.96]</strong></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.13; Chi² = 35.82, df = 15 (P = 0.002); I² = 58%
Test for overall effect: Z = 6.04 (P < 0.000001)
### Table 5: Application of approaches to chronic respiratory rehabilitation for health-related quality of life impairment in patients with chronic airflow limitation

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Estimated baseline score/proportion improving in control patients</th>
<th>Absolute increase in proportion improving in patients receiving respiratory rehabilitation</th>
<th>Relative Effect (95% CI)</th>
<th>Number of Participants (studies)</th>
<th>Confidence in effect estimate</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Health-related quality of life (HRQL)</td>
<td>The HRQL score in the respiratory rehabilitation group improved on average 0.72 (95% CI 0.48 to 0.96) SDs more in the respiratory rehabilitation patients than in control patients</td>
<td>---</td>
<td>818 (16)</td>
<td>High</td>
<td>As a rule of thumb, 0.2 SD represents a small difference, 0.5 moderate, and 0.8 large</td>
<td></td>
</tr>
</tbody>
</table>
Plan

- When studies use different outcomes
  - Standardized mean difference
  - *Natural units*
  - Dichotomize
  - Ratio of means
  - MID units
Conversion to familiar units

- All instruments into most familiar
  - Two statistical approaches

- Rescale to units of most familiar
  - St. George’s 0 to 100
  - Divide by 7 to go to CRQ units
(B) Health-related quality of life (HRQL) measured on a scale of 1 to 7

| Control group baseline 4.5 | HRQL improved on average 0.71 (95% CI 0.48 to 0.94) more in the respiratory rehabilitation patients than in the control patients | --- | 818 (16) | High |

- Confident encourage
- Possibly encourage
- Probably discourage
- Certainly discourage

What if mean difference 0.4

Vulnerable to no one benefits/everyone benefits
Plan

- When studies use different outcomes
  - Standardized mean difference
  - Natural units
  - *Dichotomize*
  - Ratio of means
  - MID units
Dichotomize

Assume standard symmetrical distribution
Assume equal variance in intervention and control groups

Effect Size

P(Imp|T) - P(Imp|C)
Yields relative and absolute effects

50% RRR in number of patients severe pain
  - Big or small effect?

Can’t tell
  - Could be reduction from 2% to 1%
  - Or reduction from 40% to 20%
| (C) Proportion of patients with important improvement in health-related quality of life (HRQL) | 0.30² | Differences in proportion achieving important improvement 0.31 (95% CI 0.22 to 0.40) in favor of rehabilitation | OR=3.36 (95% CI 2.31 to 4.86) | 818 (16) | High | Calculation uses established minimal important difference of 0.5 units on the CRQ and 4 units on the St. George's Respiratory Questionnaire |

- Confident encourage
- Possibly encourage
- Probably discourage
- Certainly discourage
When studies use different outcomes

- Standardized mean difference
- Natural units
- Dichotomize
- *Ratio of means*
- MID units
Ratio of means

RoM = \frac{\text{mean}_{\text{exp}}}{\text{mean}_{\text{control}}}

- Can tell us for instance:
  - Treatment had 30% less pain than control

- Analogous to relative risk
  - Greater absolute difference with greater control risk
When studies use different outcomes

- Standardized mean difference
- Natural units
- Dichotomize
- Ratio of means
- MID units
## Results – MID Units

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>MID</th>
<th>SE</th>
<th>Control Total</th>
<th>Weight</th>
<th>MID IV, Random, 95% CI</th>
<th>MID IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.3.1 SGRQ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boxall 2005</td>
<td>1.1</td>
<td>0.926</td>
<td>23</td>
<td>3.7%</td>
<td>1.10 [-0.71, 2.91]</td>
<td></td>
</tr>
<tr>
<td>Chlumsky 2001</td>
<td>-0.0375</td>
<td>2.391</td>
<td>13</td>
<td>6.0%</td>
<td>-0.04 [-4.72, 4.65]</td>
<td></td>
</tr>
<tr>
<td>Engstrom 1999</td>
<td>0.05</td>
<td>1.184</td>
<td>26</td>
<td>2.4%</td>
<td>0.05 [-2.27, 2.37]</td>
<td></td>
</tr>
<tr>
<td>Finnerty 2001</td>
<td>1.775</td>
<td>0.974</td>
<td>24</td>
<td>3.4%</td>
<td>1.77 [-0.13, 3.68]</td>
<td></td>
</tr>
<tr>
<td>Ringbaek 2000</td>
<td>-0.025</td>
<td>1.509</td>
<td>17</td>
<td>1.5%</td>
<td>-0.03 [2.98, 2.93]</td>
<td></td>
</tr>
<tr>
<td><strong>1.3.2 CRQ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behnke 2000</td>
<td>3.96</td>
<td>0.683</td>
<td>15</td>
<td>5.9%</td>
<td>3.96 [2.62, 5.30]</td>
<td></td>
</tr>
<tr>
<td>Cambach 2004</td>
<td>2.06</td>
<td>0.713</td>
<td>15</td>
<td>5.5%</td>
<td>2.06 [0.66, 3.46]</td>
<td></td>
</tr>
<tr>
<td>Goldstein 2004</td>
<td>1.12</td>
<td>0.445</td>
<td>40</td>
<td>10.1%</td>
<td>1.12 [0.25, 1.99]</td>
<td></td>
</tr>
<tr>
<td>Gosselink 2000</td>
<td>1.545</td>
<td>0.545</td>
<td>34</td>
<td>8.0%</td>
<td>1.54 [0.48, 2.61]</td>
<td></td>
</tr>
<tr>
<td>Griffiths 2000</td>
<td>2.25</td>
<td>0.281</td>
<td>93</td>
<td>14.9%</td>
<td>2.25 [1.70, 2.80]</td>
<td></td>
</tr>
<tr>
<td>Guell 1995</td>
<td>2.3</td>
<td>0.553</td>
<td>29</td>
<td>7.9%</td>
<td>2.30 [1.22, 3.38]</td>
<td></td>
</tr>
<tr>
<td>Guell 1998</td>
<td>1.5</td>
<td>0.63</td>
<td>18</td>
<td>6.6%</td>
<td>1.50 [0.27, 2.73]</td>
<td></td>
</tr>
<tr>
<td>Hernandez 2000</td>
<td>1.445</td>
<td>0.674</td>
<td>20</td>
<td>6.0%</td>
<td>1.45 [0.12, 2.77]</td>
<td></td>
</tr>
<tr>
<td>Simpson 1992</td>
<td>1.465</td>
<td>0.73</td>
<td>14</td>
<td>5.3%</td>
<td>1.47 [0.03, 2.90]</td>
<td></td>
</tr>
<tr>
<td>Singh 2003</td>
<td>1.63</td>
<td>0.452</td>
<td>20</td>
<td>10.0%</td>
<td>1.63 [0.74, 2.52]</td>
<td></td>
</tr>
<tr>
<td>Wijkstra 1994</td>
<td>1.45</td>
<td>0.537</td>
<td>28</td>
<td>8.2%</td>
<td>1.45 [0.40, 2.50]</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>429</td>
<td>387</td>
<td>100.0%</td>
<td>1.75</td>
<td>[1.37, 2.13]</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 0.17$; $\chi^2 = 22.15$, df = 15 ($P = 0.10$); $I^2 = 32$

Test for overall effect: $Z = 9.00$ ($P < 0.00001$)
Do clinicians understand treatment effects?

- Cross-sectional, paper-based survey
  - Academic centers in 8 countries,
  - Internal and family medicine, 531/610 (87%)
- Summary estimates hypothetical interventions vs placebo chronic pain
Do clinicians understand treatment effects?

- Objective: determine clinicians understanding and perspective of 6 approaches to presenting continuous outcomes
  - Standardized Mean Difference
  - Natural units
  - Ratio of Means
  - Relative Risk
  - Absolute Risk
  - Ratio of Means
  - MID Units

- Random assign 1 of 4 versions, differing magnitude of effect and presentation order
<table>
<thead>
<tr>
<th>Presentation method</th>
<th>Small effect</th>
<th>Large effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMD</td>
<td>0.20</td>
<td>0.80</td>
</tr>
<tr>
<td>MD in natural units</td>
<td>0.60</td>
<td>2.00</td>
</tr>
<tr>
<td>RR</td>
<td>0.80</td>
<td>0.50</td>
</tr>
<tr>
<td>RD</td>
<td>0.04</td>
<td>0.20</td>
</tr>
<tr>
<td>RoM</td>
<td>0.92 (8% less)</td>
<td>0.63 (37% less)</td>
</tr>
<tr>
<td>MID units</td>
<td>0.60</td>
<td>2.00</td>
</tr>
</tbody>
</table>
Results: Correct answers

Figure 3: Understanding of the presentation approaches, $n = 531$

Understanding, % correct

- Risk Difference
- Relative Risk
- Ratio of Means
- SMD
- MID units
- Natural units
18. In pooled standard deviation units of all pain scores in the treatment and control groups, a meta-analysis finds the effect of intervention A vs placebo control for patient-reported pain to be 0.20 standard deviation units in favoring of intervention A. Please clearly indicate whether this presentation approach is useful:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not useful in understanding size and importance of the effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Extremely useful in understanding the size and importance of the effect</td>
</tr>
</tbody>
</table>

Results: Usefulness

Figure 4: Perceived Usefulness, $n = 531$

Perceived Usefulness

- Risk Difference
- Relative Risk
- Natural units
- Ratio of Means
- MID units
- SMD

Higher scores represent higher perceived usefulness
Conclusions

- Patient-reported outcomes often critical
  - Almost always patient-important
  - Symptoms, function, quality of life
- Interpretations can be challenging
- Single study, SR all same instrument
  - Fine if everyone familiar with units
  - If not, need MID or dichotomize
- SR different instruments
  - SMD most used – limitations
  - Look for natural units, conversion RR and RD
More conclusions

- Use more than one method
  - Decreases selection bias
  - If similar reassuring
  - If not, need to explain, appropriate doubt
- If very familiar instrument, use as approach
- Use comments in SoF, especially MID
- One of approaches should be dichotomy
For copies of the slides

- Contact
  guyatt@mcmaster.ca