Making results of patient-reported outcomes interpretable

Gordon Guyatt, MD, MSc

Slides available: guyatt@mcmaster.ca

Plan

- What is a PRO
- The problem of interpretability
- Making results interpretable individual studies
- Systematic reviews and meta-analyses
  - When studies use same or similar outcome
    - MID, range, or dichotomize
  - When studies use different outcomes
    - standardized mean difference
    - natural units
    - dichotomize – relative and absolute effects
    - Ratio of means
    - MID units

Patient-Reported Outcomes (PRO)

- **PRO**: Any report directly from patients, without interpretation by physicians or anyone else, about how they function or feel in relation to a health condition and its therapy (from diaries, questionnaires, interviews, etc.)
- Very often health-related quality of life

Plan

- What is a PROs
- *The problem of interpretability*

Interpretability

- 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?
- What other information would you like to aid interpretability?

Br J Dermatology, 2004

- Effect of alefacept on quality of life in 553 patients with psoriasis
- Alefacept significantly reduced (improved) mean Dermatology Quality of Life Scale scores compared with placebo: 4.4 vs. 1.8 at 2 weeks after the last dose ($P<0.0001$) and 3.4 vs. 1.4 at 12 weeks after the last dose ($P<0.001$).
- Magnitude of Effect?
  - trivial, small but important, large?

Plan

- PROs in Cochrane reviews
- The problem of interpretability
- *Strategies for making results interpretable in individual studies*

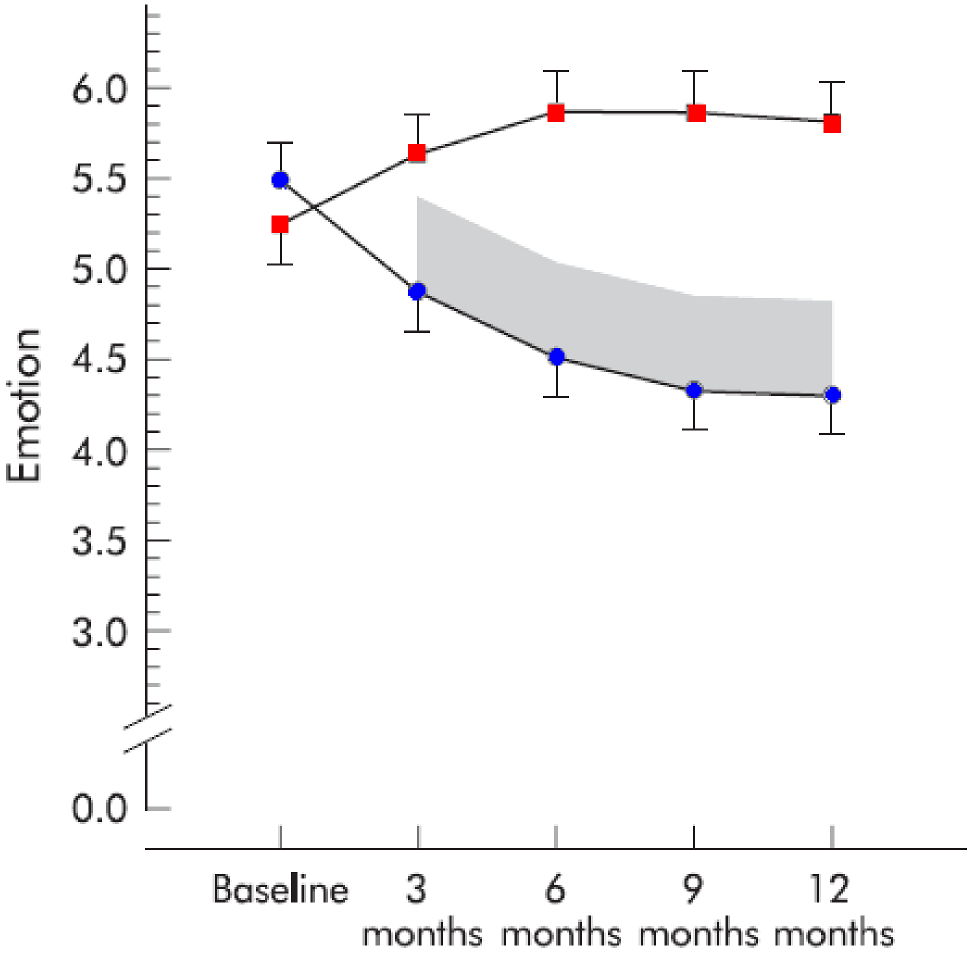
Minimally important difference

- Smallest change that patients would consider important
- Global ratings of change
  - are you the same, a little better, a lot better
- Instruments on 1 to 7 scale 0.5 often represents MID

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, emotional function

Effect of Surgery and Medical Control Treatment

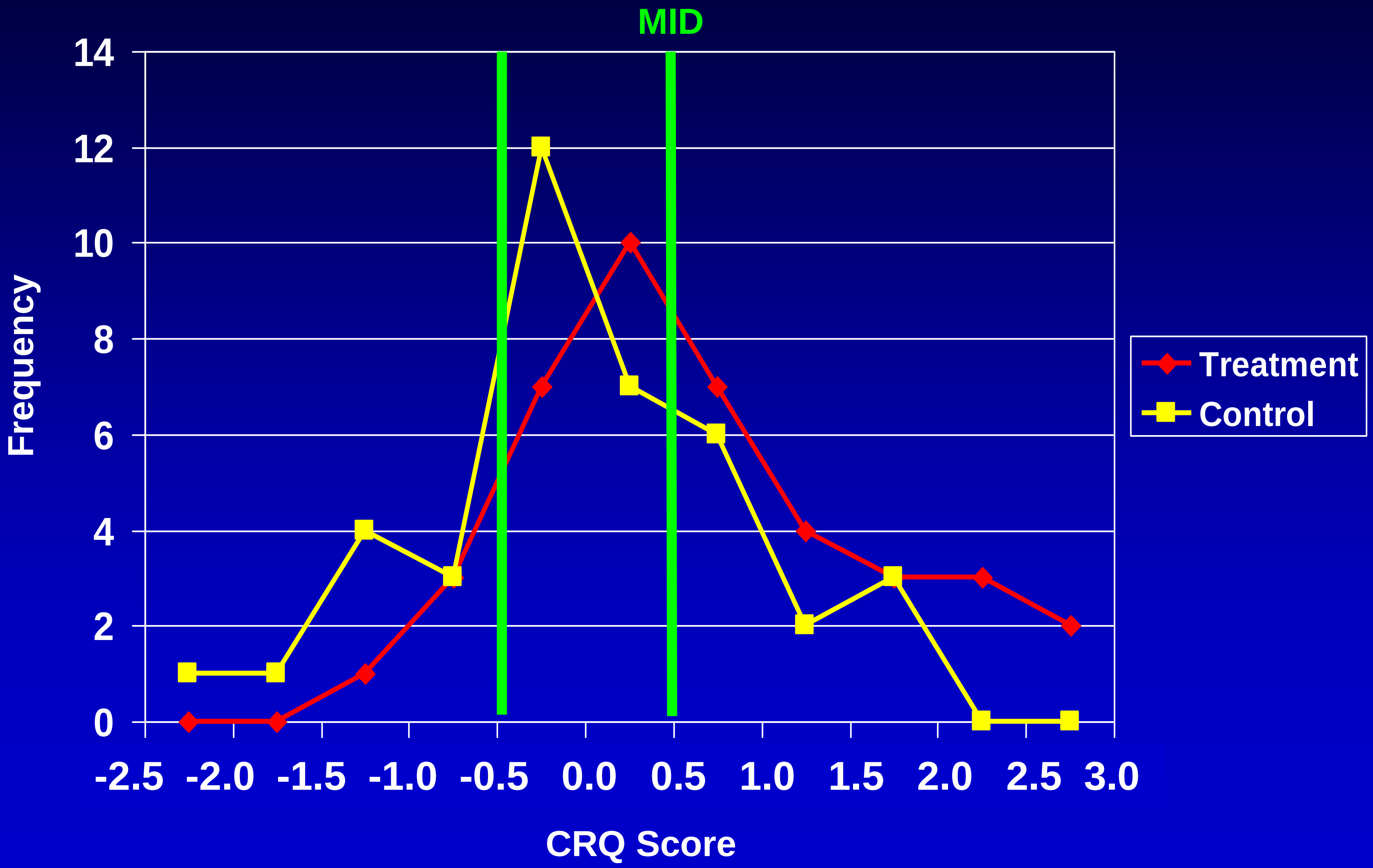


Would you recommend surgery to your patients on the basis of these results?

Interpreting MID Results

- RCT respiratory rehabilitation in COPD
- Assume MID is 0.50 and patients mean improvement vs control is 0.25
- 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

CRQ Emotion Change Scores



Differences between rehabilitation and conventional care in CAL

CRQ domain	Difference between groups		Estimated proportion better on rehabilitation	Estimated proportion better on conventional care	Proportion benefiting from rehabilitation	NNT for a single patient to benefit
	Mean	P value				
Dyspnoea	0.60	0.0003	0.47	0.28	0.19	5.2
Fatigue	0.45	0.06	0.45	0.23	0.23	4.4
Emotional function	0.40	0.001	0.47	0.17	0.30	3.3

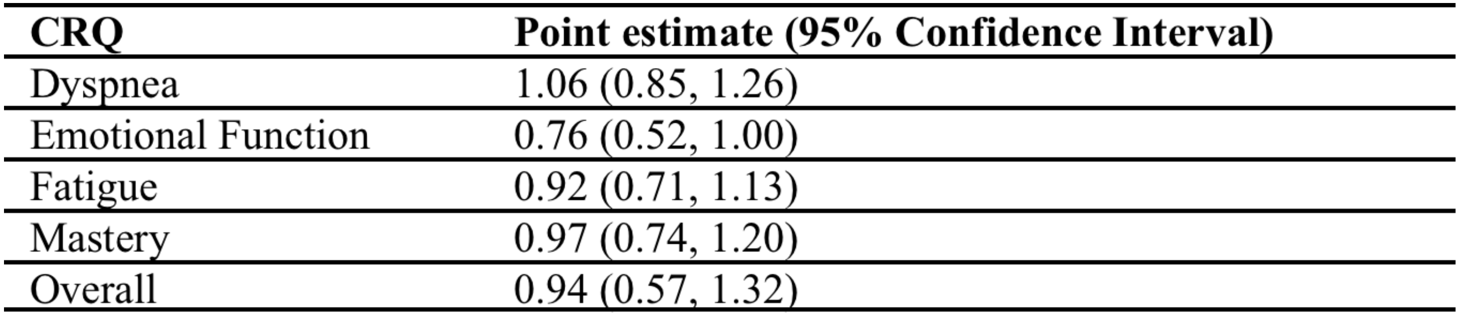
Plan

- PROs in Cochrane reviews
- The problem of interpretability
- Strategies for making results interpretable in individual studies
- *Systematic reviews and meta-analyses*
  - *When studies use same or similar outcome*
    - *MID, range, or dichotomize*

Meta-analysis

- Studies all use same or similar outcome
- Could give weighted mean difference in natural units
- Not intuitively interpretable to the audience
  - challenges in interpretation
- Solution
  - MID if available
  - Range of possible results if not

Systematic review respiratory rehabilitation



CRQ	Point estimate (95% Confidence Interval)
Dyspnea	1.06 (0.85, 1.26)
Emotional Function	0.76 (0.52, 1.00)
Fatigue	0.92 (0.71, 1.13)
Mastery	0.97 (0.74, 1.20)
Overall	0.94 (0.57, 1.32)

MID 0.5

Would you recommend respiratory rehabilitation to your patients?

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
  - mechanism unclear, 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)

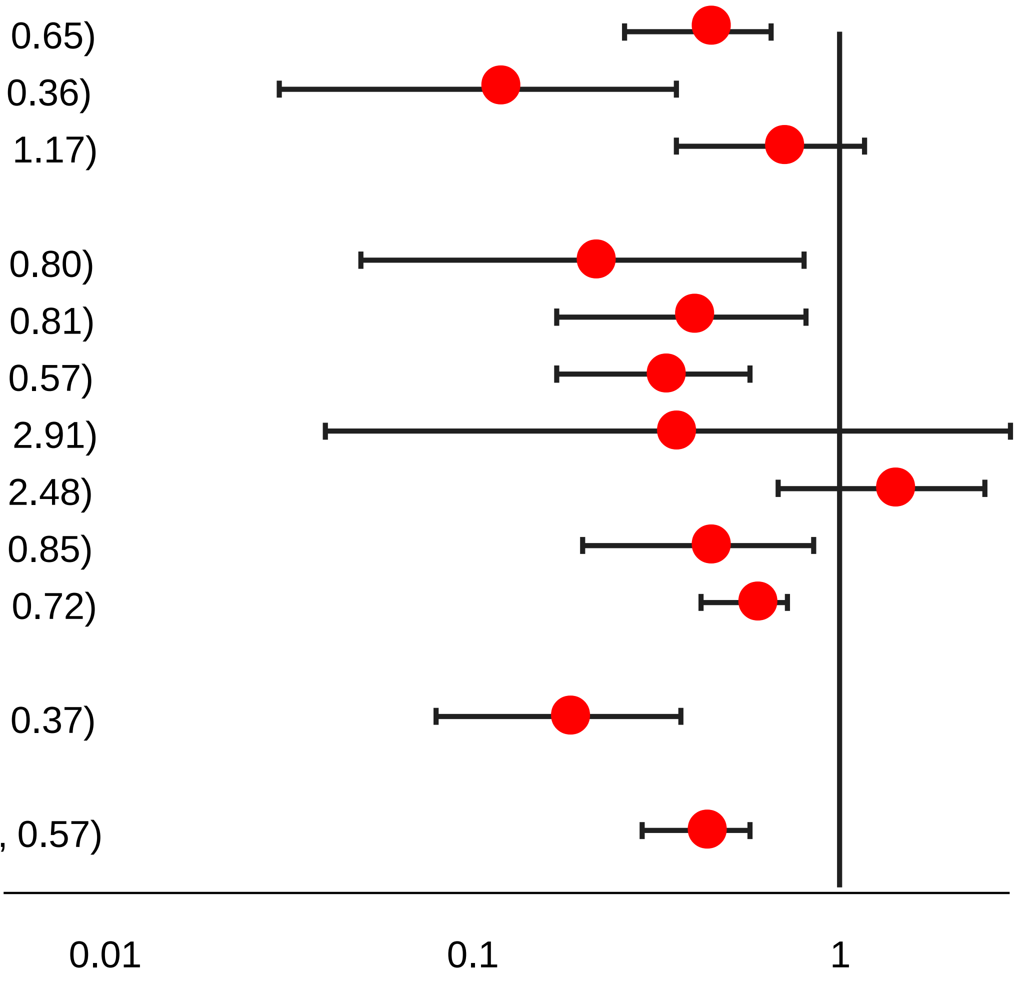
Thorp 1.30 (0.68, 2.48)

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

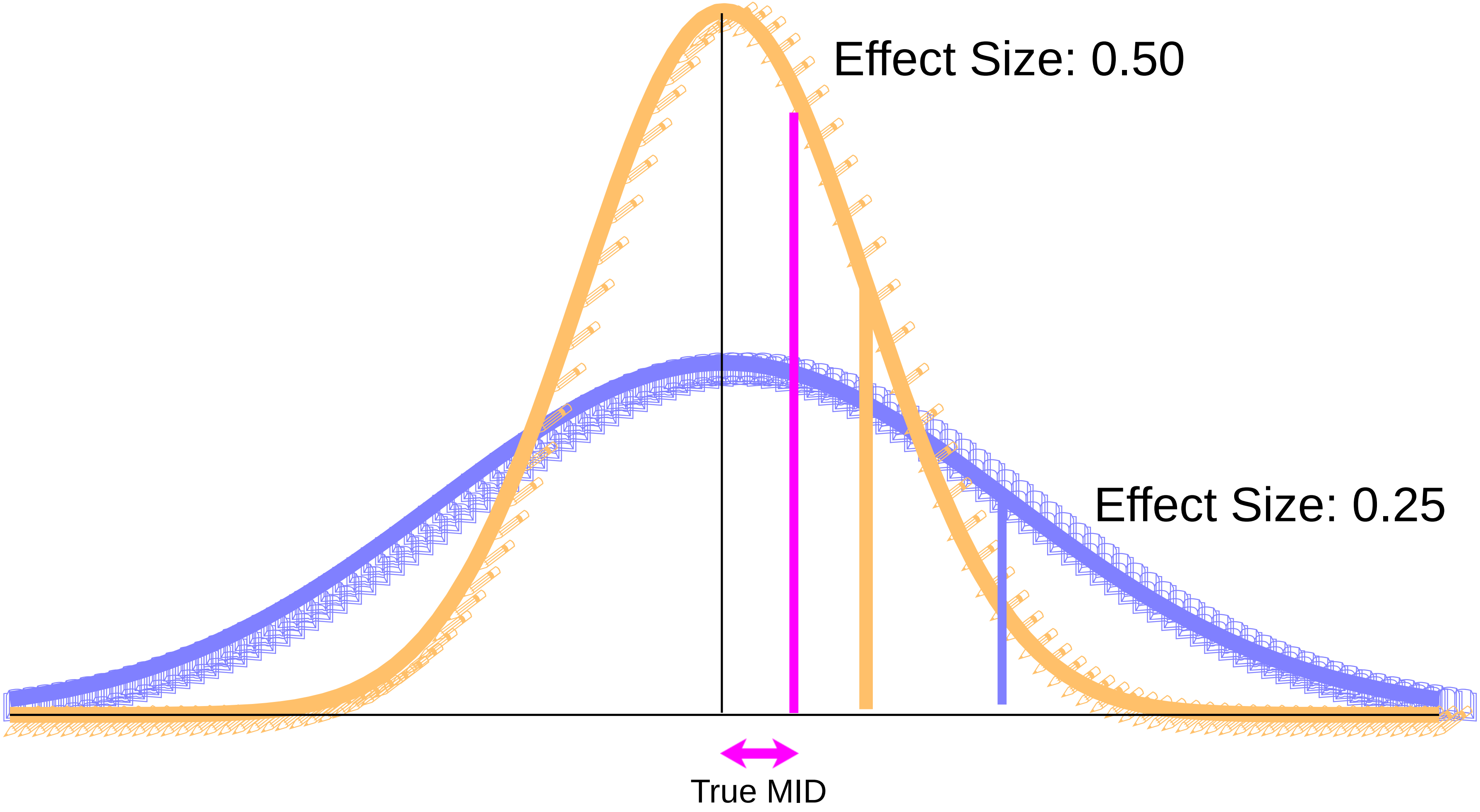
- PROs in Cochrane reviews
- The problem of interpretability
- Strategies for making results interpretable in individual studies
- Systematic reviews and meta-analyses
  - When studies use same or similar outcome
    - MID, range, or dichotomize
  - *When studies use different outcomes*
    - *Standardized mean difference*

Effect size

- Divide each effect by standard deviation
- Ultimate result in SD units
- “Effect size” or SMD

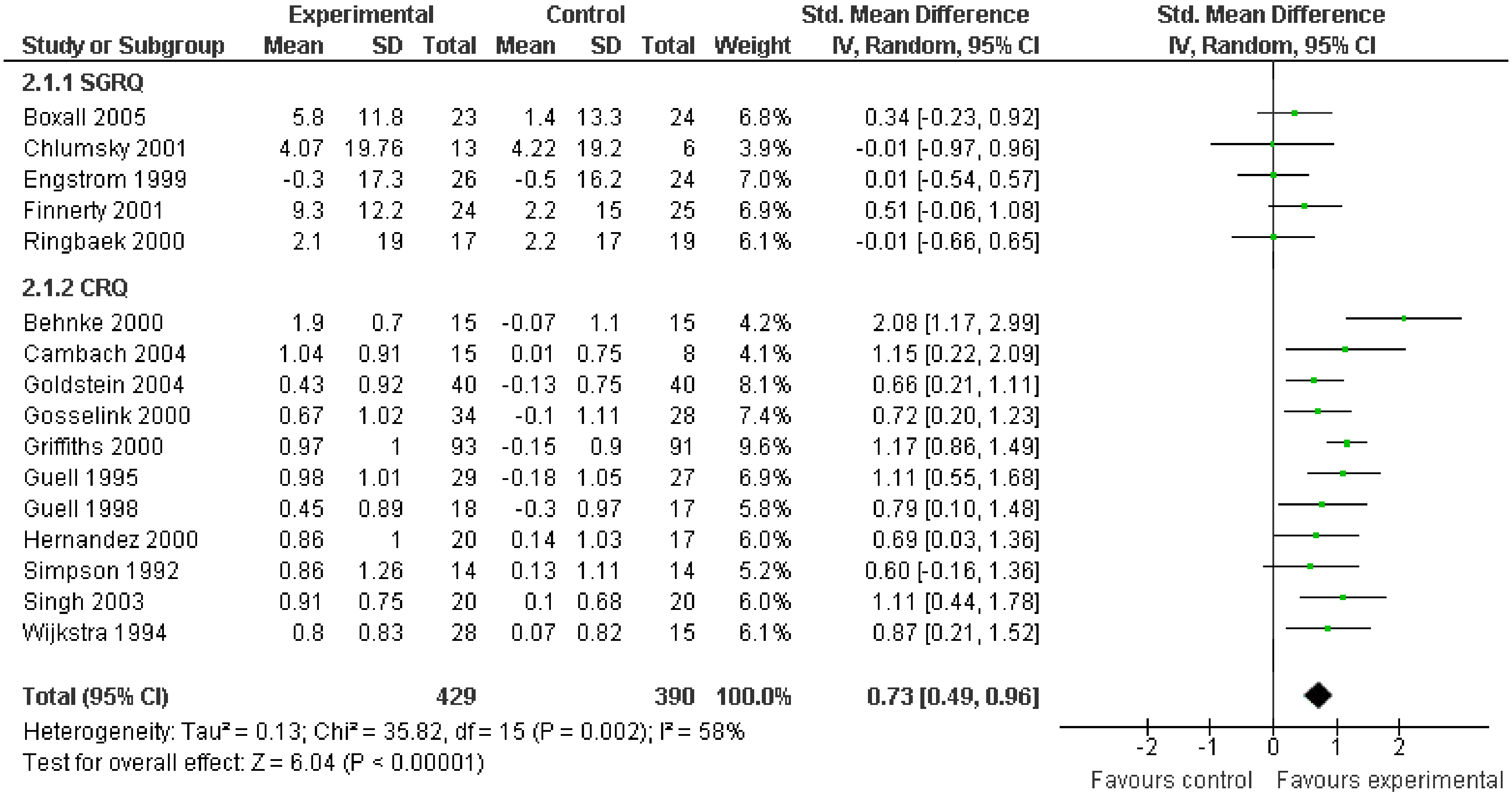
Cohen:
Small effect 0.2 SD units
Moderate effect 0.5
Large effect 0.8

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



Effect Size: 0.50
Effect Size: 0.25
True MID

Results – SD Units



	Experimental			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
2.1.1 SGRQ								
Boxall 2005	5.8	11.8	23	1.4	13.3	24	6.8%	0.34 [-0.23, 0.92]
Chlumsky 2001	4.07	19.76	13	4.22	19.2	6	3.9%	-0.01 [-0.97, 0.96]
Engstrom 1999	-0.3	17.3	26	-0.5	16.2	24	7.0%	0.01 [-0.54, 0.57]
Finnerty 2001	9.3	12.2	24	2.2	15	25	6.9%	0.51 [-0.06, 1.08]
Ringbaek 2000	2.1	19	17	2.2	17	19	6.1%	-0.01 [-0.66, 0.65]
2.1.2 CRQ								
Behnke 2000	1.9	0.7	15	-0.07	1.1	15	4.2%	2.08 [1.17, 2.99]
Cambach 2004	1.04	0.91	15	0.01	0.75	8	4.1%	1.15 [0.22, 2.09]
Goldstein 2004	0.43	0.92	40	-0.13	0.75	40	8.1%	0.66 [0.21, 1.11]
Gosselink 2000	0.67	1.02	34	-0.1	1.11	28	7.4%	0.72 [0.20, 1.23]
Griffiths 2000	0.97	1	93	-0.15	0.9	91	9.6%	1.17 [0.86, 1.49]
Guell 1995	0.98	1.01	29	-0.18	1.05	27	6.9%	1.11 [0.55, 1.68]
Guell 1998	0.45	0.89	18	-0.3	0.97	17	5.8%	0.79 [0.10, 1.48]
Hernandez 2000	0.86	1	20	0.14	1.03	17	6.0%	0.69 [0.03, 1.36]
Simpson 1992	0.86	1.26	14	0.13	1.11	14	5.2%	0.60 [-0.16, 1.36]
Singh 2003	0.91	0.75	20	0.1	0.68	20	6.0%	1.11 [0.44, 1.78]
Wijkstra 1994	0.8	0.83	28	0.07	0.82	15	6.1%	0.87 [0.21, 1.52]
Total (95% CI)			**429**			**390**	**100.0%**	**0.73 [0.49, 0.96]**

Heterogeneity: Tau^2 = 0.13; Chi^2 = 35.82, df = 15 (P = 0.002); I^2 = 58%

Test for overall effect: Z = 6.04 (P < 0.00001)

Table 5: Application of approaches to chronic respiratory rehabilitation for health-related quality of life impairment in patients with chronic airflow limitation

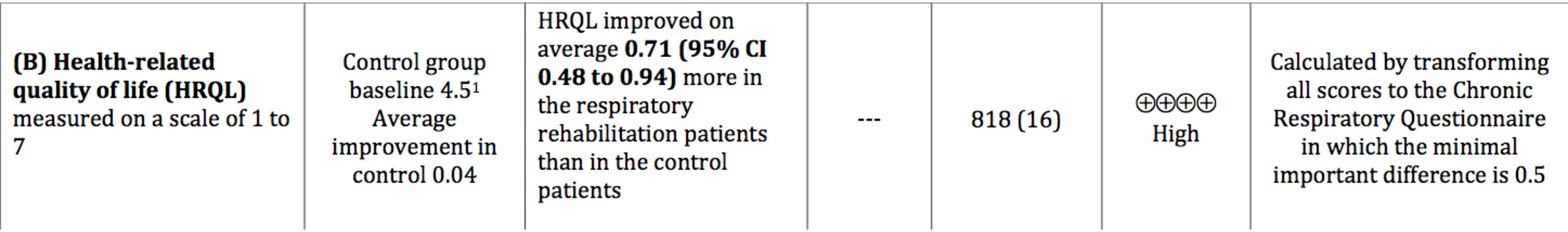
Outcomes	Estimated baseline score/proportion improving in control patients	Absolute increase in proportion improving in patients receiving respiratory rehabilitation	Relative Effect (95% CI)	Number of Participants (studies)	Confidence in effect estimate[1]	Comments
(A) Health-related quality of life (HRQL) Investigators measured HRQL using different instruments. Higher scores mean better HRQL.	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		---	818 (16)	⊕⊕⊕⊕ High	As a rule of thumb, 0.2 SD represents a small difference, 0.5 moderate, and 0.8 large

Plan

- The problem of interpretability
- Strategies for making results interpretable in individual studies
- Systematic reviews and meta-analyses
  - When studies use same or similar outcome
    - MID, range, or dichotomize
  - ***When studies use different outcomes***
    - Standardized mean difference
    - *Natural units*

Conversion to familiar units

- All instruments into most familiar
  - Two statistical approaches
- Multiply SD units X SD of most familiar
  - May be challenging to decide which SD
  - Vulnerable to heterogenity
- Rescale to units of most familiar
  - St. George's 0 to 100
  - Multiply by 7/100 to go to CRQ units
    - Statistical approach to get variance



(B) Health-related quality of life (HRQL) measured on a scale of 1 to 7	Control group baseline 4.5[1] Average improvement in control 0.04	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	Calculated by transforming all scores to the Chronic Respiratory Questionnaire in which the minimal important difference is 0.5

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

What if mean difference 0.4

Limitations to presentation?

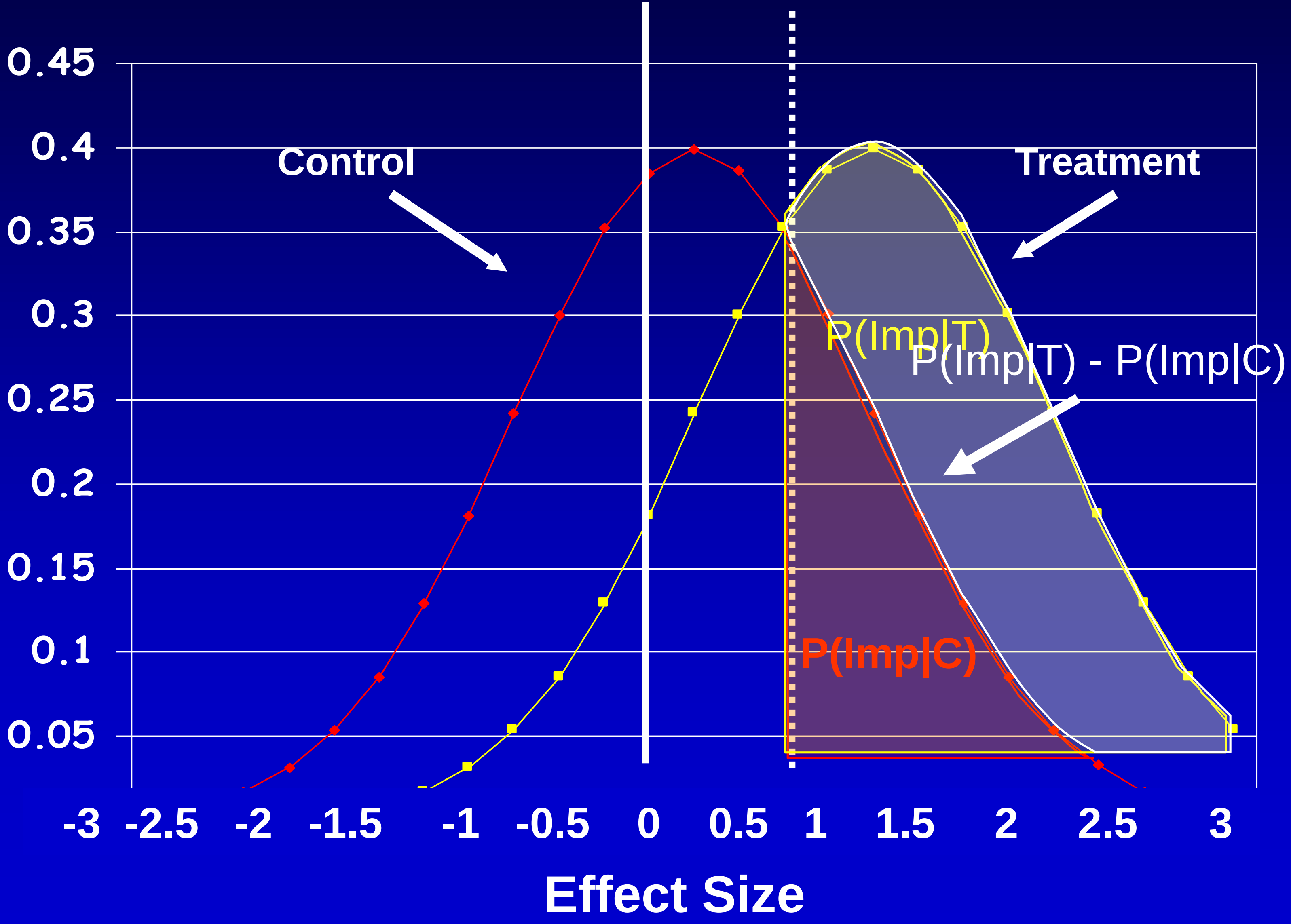
Vulnerable to no one benefits/everyone benefits

Plan

- PROs in Cochrane reviews
- The problem of interpretability
- Strategies for making results interpretable in individual studies
- Systematic reviews and meta-analyses
  - When studies use same or similar outcome
    - MID, range, or dichotomize
  - When studies use different outcomes
    - Standardized mean difference
    - Natural units
    - *Dichotomize – relative and absolute effects*

Dichotomize

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



Dichotomize

- Relative and absolute effects
- Number of statistical approaches relying on SMD
- Normal distribution/equal variance
    - Furukawa

6A, for situations in which the event is undesirable, reduction in adverse events with the intervention

Control group response rate	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
SMD = -0.2	-0.03	-0.05	-0.07	-0.08	-0.08	-0.08	-0.07	-0.06	-0.040
SMD = -0.5	-0.06	-0.11	-0.15	-0.17	-0.19	-0.20	-0.20	-0.17	-0.12
SMD = -0.8	-0.08	-0.15	-0.21	-0.25	-0.29	-0.31	-0.31	-0.28	-0.22
SMD = -1.0	-0.09	-0.17	-0.24	-0.23	-0.34	-0.37	-0.38	-0.36	-0.29

6B for situations in which the event is desirable, increase in positive responses to the intervention

Control group response rate	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
SMD = 0.2	0.04	0.61	0.07	0.08	0.08	0.08	0.07	0.05	0.03
SMD = 0.5	0.12	0.17	0.19	0.20	0.19	0.17	0.15	0.11	0.06
SMD = 0.8	0.22	0.28	0.31	0.31	0.29	0.25	0.21	0.15	0.08
SMD = 1.0	0.29	0.36	0.38	0.38	0.34	0.30	0.24	0.17	0.09

Limitations

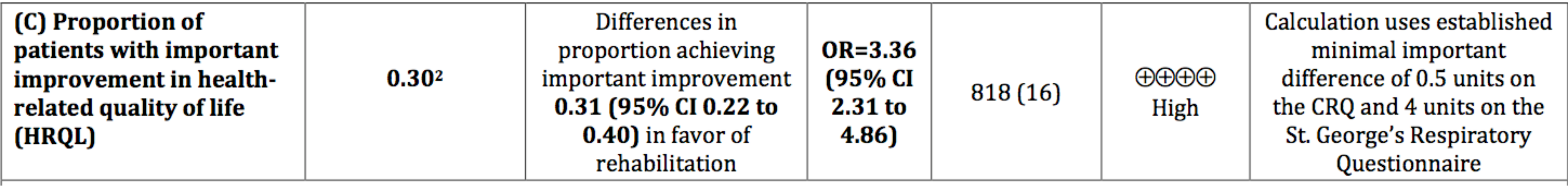
- Dichotomous outcome may not be clear
  - pain continuous outcome
  - threshold severe, moderate, mild?
- Control proportion may not be clear
  - Differs a lot only at extremes
- Based on SMD
  - Vulnerable to population heterogeneity

Other statistical approaches

- Relying on SMD
  - Cox/Snell; Hasselbad/Hedges
- Similar assumptions
- Doesn’t require specifying control group rate

Alternative

- If know MID for all instruments can go to individual studies
- Calculate proportion benefiting in each individual study
- Combine proportions across studies
- Alternative convert to same units and WMD to risk difference
- Doesn’t depend on SMD



(C) Proportion of patients with important improvement in health-related quality of life (HRQL)	**0.30**[2]	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

Furukawa RD 0.28

Plan

- PROs in Cochrane reviews
- The problem of interpretability
- Strategies for making results interpretable in individual studies
- Systematic reviews and meta-analyses
  - When studies use same or similar outcome
    - MID, range, or dichotomize
  - When studies use different outcomes
    - Standardized mean difference
    - Natural units
    - Dichotomize – relative and absolute effects
    - ***Ratio of means***

Ratio of Means (RoM)

$$\mathrm{RoM} = \frac{\mathrm{mean}_{\mathrm{exp}}}{\mathrm{mean}_{\mathrm{control}}}$$

- Requires estimate of variance of this ratio – this can be estimated using the delta method:
  - $\mathrm{Var}_{\mathrm{ln(RoM)}} = \dfrac{\mathrm{var}_{\mathrm{exp}}}{(\mathrm{mean}_{\mathrm{exp}}{}^{2})} + \dfrac{\mathrm{var}_{\mathrm{control}}}{(\mathrm{mean}_{\mathrm{control}}{}^{2})}$

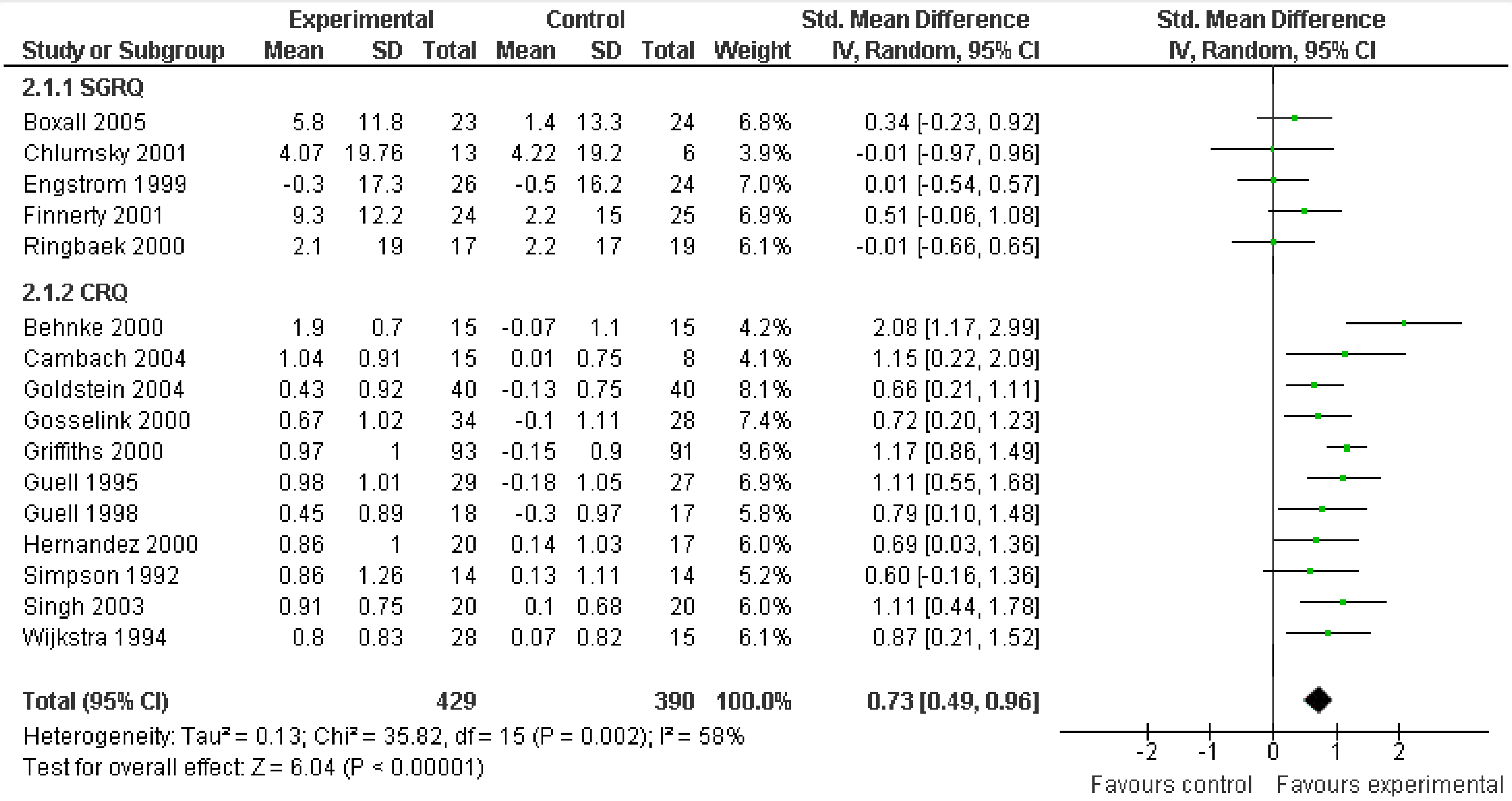
Ratio of means

- Analogous to relative risk
  - Greater absolute difference with greater control risk
- Requires natural zero
- Cannot use if results reported as change and changes go in opposite directions in the two groups

Plan

- The problem of interpretability
- Strategies for making results interpretable in individual studies
- Systematic reviews and meta-analyses
  - When studies use same or similar outcome
    - MID, range, or dichotomize
  - When studies use different outcomes
    - Standardized mean difference
    - Natural units
    - Dichotomize – relative and absolute effects
    - Ratio of means
    - ***MID units***

Results – SD Units

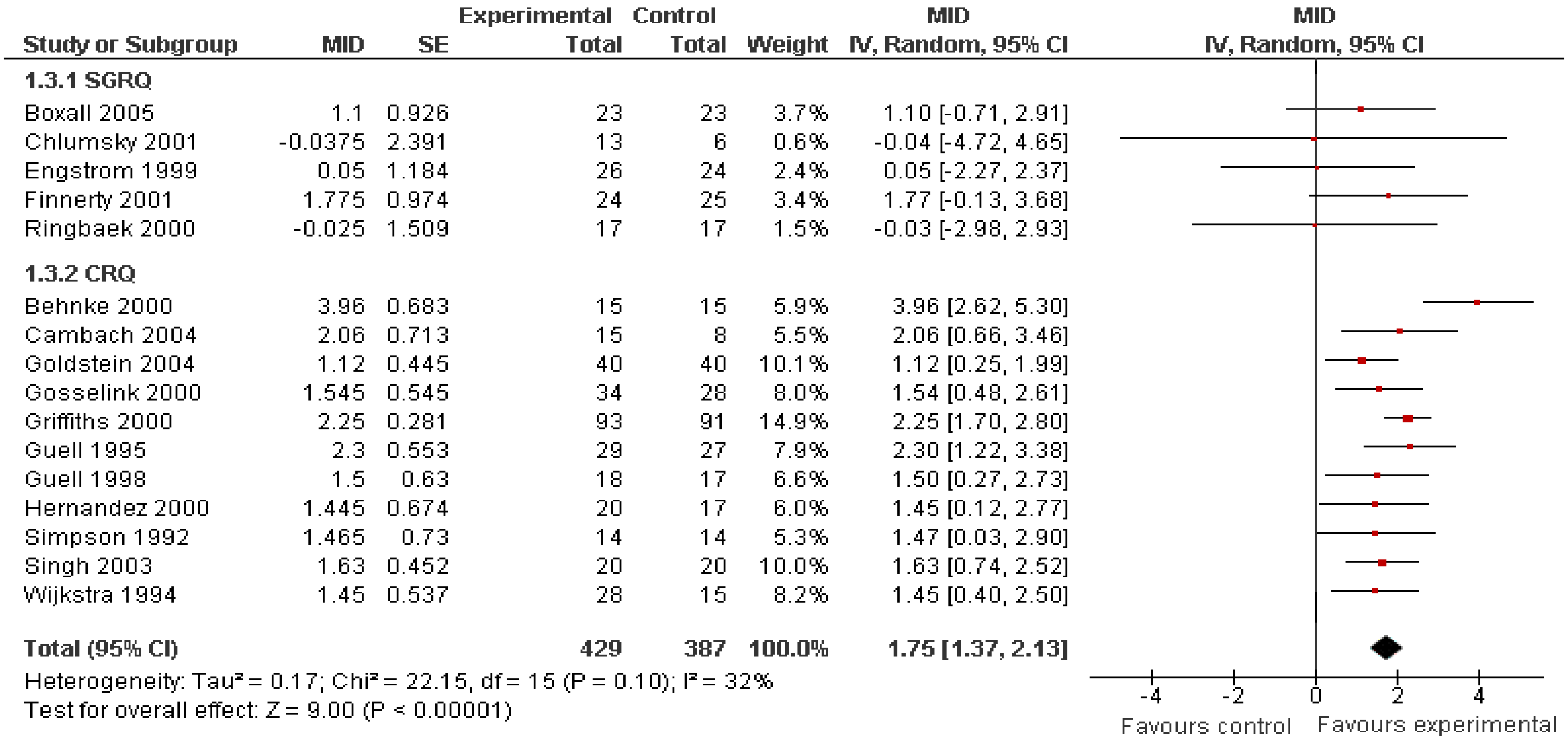


	Experimental			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
2.1.1 SGRQ								
Boxall 2005	5.8	11.8	23	1.4	13.3	24	6.8%	0.34 [-0.23, 0.92]
Chlumsky 2001	4.07	19.76	13	4.22	19.2	6	3.9%	-0.01 [-0.97, 0.96]
Engstrom 1999	-0.3	17.3	26	-0.5	16.2	24	7.0%	0.01 [-0.54, 0.57]
Finnerty 2001	9.3	12.2	24	2.2	15	25	6.9%	0.51 [-0.06, 1.08]
Ringbaek 2000	2.1	19	17	2.2	17	19	6.1%	-0.01 [-0.66, 0.65]
2.1.2 CRQ								
Behnke 2000	1.9	0.7	15	-0.07	1.1	15	4.2%	2.08 [1.17, 2.99]
Cambach 2004	1.04	0.91	15	0.01	0.75	8	4.1%	1.15 [0.22, 2.09]
Goldstein 2004	0.43	0.92	40	-0.13	0.75	40	8.1%	0.66 [0.21, 1.11]
Gosselink 2000	0.67	1.02	34	-0.1	1.11	28	7.4%	0.72 [0.20, 1.23]
Griffiths 2000	0.97	1	93	-0.15	0.9	91	9.6%	1.17 [0.86, 1.49]
Guell 1995	0.98	1.01	29	-0.18	1.05	27	6.9%	1.11 [0.55, 1.68]
Guell 1998	0.45	0.89	18	-0.3	0.97	17	5.8%	0.79 [0.10, 1.48]
Hernandez 2000	0.86	1	20	0.14	1.03	17	6.0%	0.69 [0.03, 1.36]
Simpson 1992	0.86	1.26	14	0.13	1.11	14	5.2%	0.60 [-0.16, 1.36]
Singh 2003	0.91	0.75	20	0.1	0.68	20	6.0%	1.11 [0.44, 1.78]
Wijkstra 1994	0.8	0.83	28	0.07	0.82	15	6.1%	0.87 [0.21, 1.52]
Total (95% CI)			**429**			**390**	**100.0%**	**0.73 [0.49, 0.96]**

Heterogeneity: Tau² = 0.13; Chi² = 35.82, df = 15 (P = 0.002); I² = 58%

Test for overall effect: Z = 6.04 (P < 0.00001)

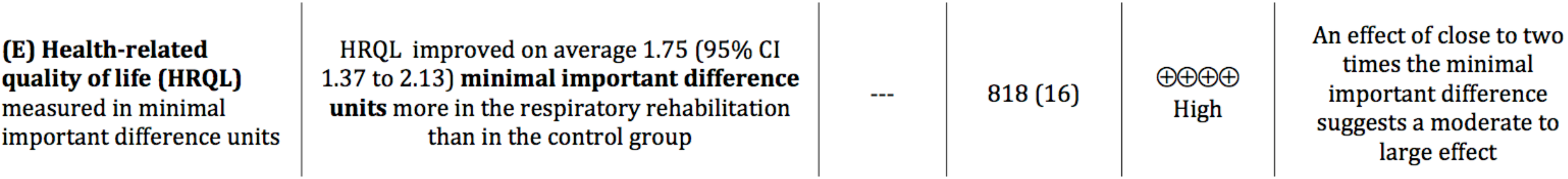
Results – MID Units



Study or Subgroup	MID	SE	Experimental Total	Control Total	Weight	MID IV, Random, 95% CI
1.3.1 SGRQ						
Boxall 2005	1.1	0.926	23	23	3.7%	1.10 [-0.71, 2.91]
Chlumsky 2001	-0.0375	2.391	13	6	0.6%	-0.04 [-4.72, 4.65]
Engstrom 1999	0.05	1.184	26	24	2.4%	0.05 [-2.27, 2.37]
Finnerty 2001	1.775	0.974	24	25	3.4%	1.77 [-0.13, 3.68]
Ringbaek 2000	-0.025	1.509	17	17	1.5%	-0.03 [-2.98, 2.93]
1.3.2 CRQ						
Behnke 2000	3.96	0.683	15	15	5.9%	3.96 [2.62, 5.30]
Cambach 2004	2.06	0.713	15	8	5.5%	2.06 [0.66, 3.46]
Goldstein 2004	1.12	0.445	40	40	10.1%	1.12 [0.25, 1.99]
Gosselink 2000	1.545	0.545	34	28	8.0%	1.54 [0.48, 2.61]
Griffiths 2000	2.25	0.281	93	91	14.9%	2.25 [1.70, 2.80]
Guell 1995	2.3	0.553	29	27	7.9%	2.30 [1.22, 3.38]
Guell 1998	1.5	0.63	18	17	6.6%	1.50 [0.27, 2.73]
Hernandez 2000	1.445	0.674	20	17	6.0%	1.45 [0.12, 2.77]
Simpson 1992	1.465	0.73	14	14	5.3%	1.47 [0.03, 2.90]
Singh 2003	1.63	0.452	20	20	10.0%	1.63 [0.74, 2.52]
Wijkstra 1994	1.45	0.537	28	15	8.2%	1.45 [0.40, 2.50]
Total (95% CI)			**429**	**387**	**100.0%**	**1.75 [1.37, 2.13]**

Heterogeneity: Tau² = 0.17; Chi² = 22.15, df = 15 (P = 0.10); I² = 32%

Test for overall effect: Z = 9.00 (P < 0.00001)



(E) Health-related quality of life (HRQL) measured in minimal important difference units	HRQL improved on average 1.75 (95% CI 1.37 to 2.13) **minimal important difference units** more in the respiratory rehabilitation than in the control group	---	818 (16)	⊕⊕⊕⊕ High	An effect of close to two times the minimal important difference suggests a moderate to large effect

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

Steroids for laparoscopic Cholecystectomy

- Systematic review
- Nausea and vomiting
  - 16 RCTs
- Pain
  - 5 RCTs

Standardized mean difference

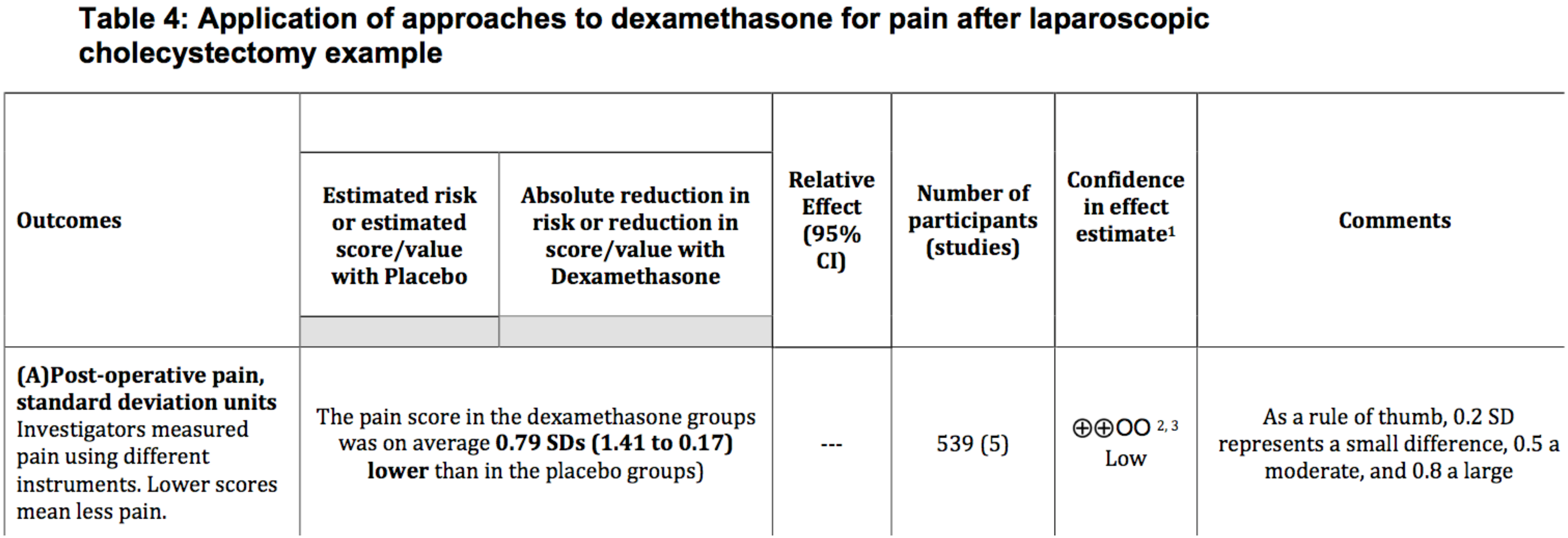
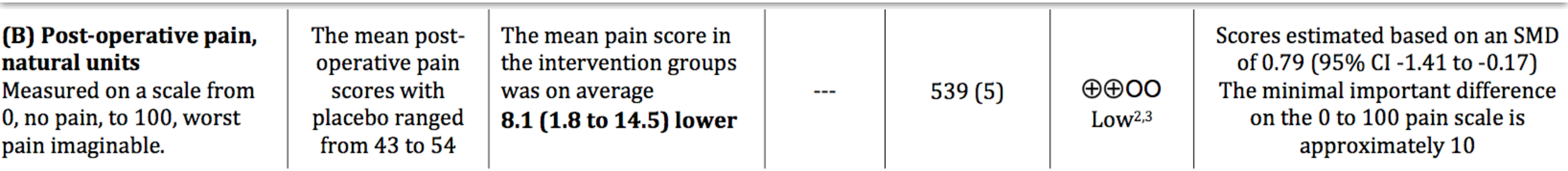


Table 4: Application of approaches to dexamethasone for pain after laparoscopic cholecystectomy example

Outcomes	Estimated risk or estimated score/value with Placebo	Absolute reduction in risk or reduction in score/value with Dexamethasone	Relative Effect (95% CI)	Number of participants (studies)	Confidence in effect estimate[1]	Comments
(A)Post-operative pain, standard deviation units Investigators measured pain using different instruments. Lower scores mean less pain.	The pain score in the dexamethasone groups was on average **0.79 SDs (1.41 to 0.17) lower** than in the placebo groups)		---	539 (5)	⊕⊕OO [2, 3] Low	As a rule of thumb, 0.2 SD represents a small difference, 0.5 a moderate, and 0.8 a large

- Large effect
- Moderate effect
- Small effect
- Trivial or no effect

Natural Units

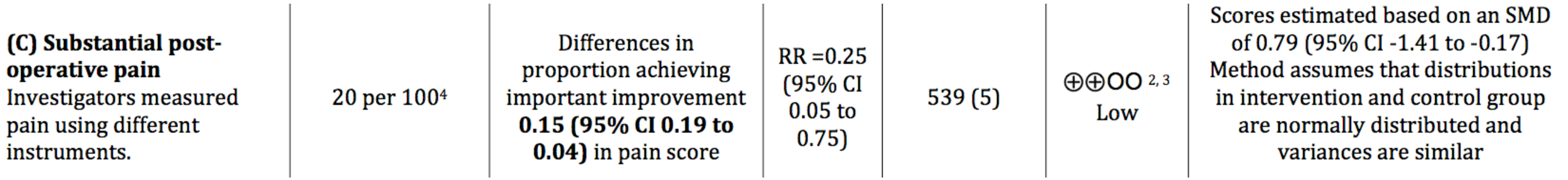


(B) Post-operative pain, natural units Measured on a scale from 0, no pain, to 100, worst pain imaginable.	The mean post-operative pain scores with placebo ranged from 43 to 54	The mean pain score in the intervention groups was on average **8.1 (1.8 to 14.5) lower**	---	539 (5)	⊕⊕OO Low[2,3]	Scores estimated based on an SMD of 0.79 (95% CI -1.41 to -0.17) The minimal important difference on the 0 to 100 pain scale is approximately 10

- Large effect
- Moderate effect
- Small effect
- Trivial or no effect

Using direct conversion method
3.5 (0.5 to 6.5) lower

Risk difference

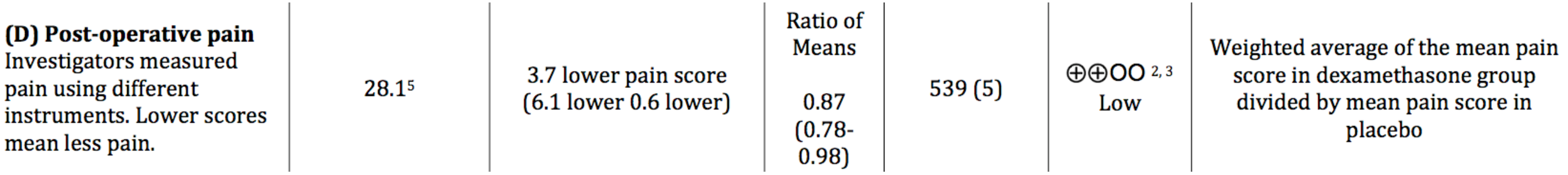


(C) Substantial post-operative pain Investigators measured pain using different instruments.	20 per 100[4]	Differences in proportion achieving important improvement **0.15 (95% CI 0.19 to 0.04)** in pain score	RR =0.25 (95% CI 0.05 to 0.75)	539 (5)	⊕⊕OO [2, 3] Low	Scores estimated based on an SMD of 0.79 (95% CI -1.41 to -0.17) Method assumes that distributions in intervention and control group are normally distributed and variances are similar

- Large effect
- Moderate effect
- Small effect
- Trivial or no effect

Using MID 0.03 (0.01 less to 0.07 more)

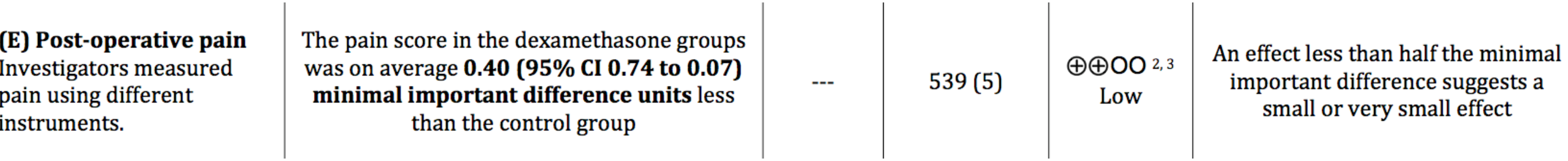
Ratio of Means



(D) Post-operative pain Investigators measured pain using different instruments. Lower scores mean less pain.	28.1[5]	3.7 lower pain score (6.1 lower 0.6 lower)	Ratio of Means 0.87 (0.78-0.98)	539 (5)	⊕⊕○○ [2, 3] Low	Weighted average of the mean pain score in dexamethasone group divided by mean pain score in placebo

- Large effect
- Moderate effect
- Small effect
- Trivial or no effect

MID Units



(E) Post-operative pain Investigators measured pain using different instruments.	The pain score in the dexamethasone groups was on average **0.40 (95% CI 0.74 to 0.07) minimal important difference units** less than the control group	---	539 (5)	⊕⊕◯◯ [2, 3] Low	An effect less than half the minimal important difference suggests a small or very small effect

- Large effect
- Moderate effect
- Small effect
- Trivial or no effect

Summary of results

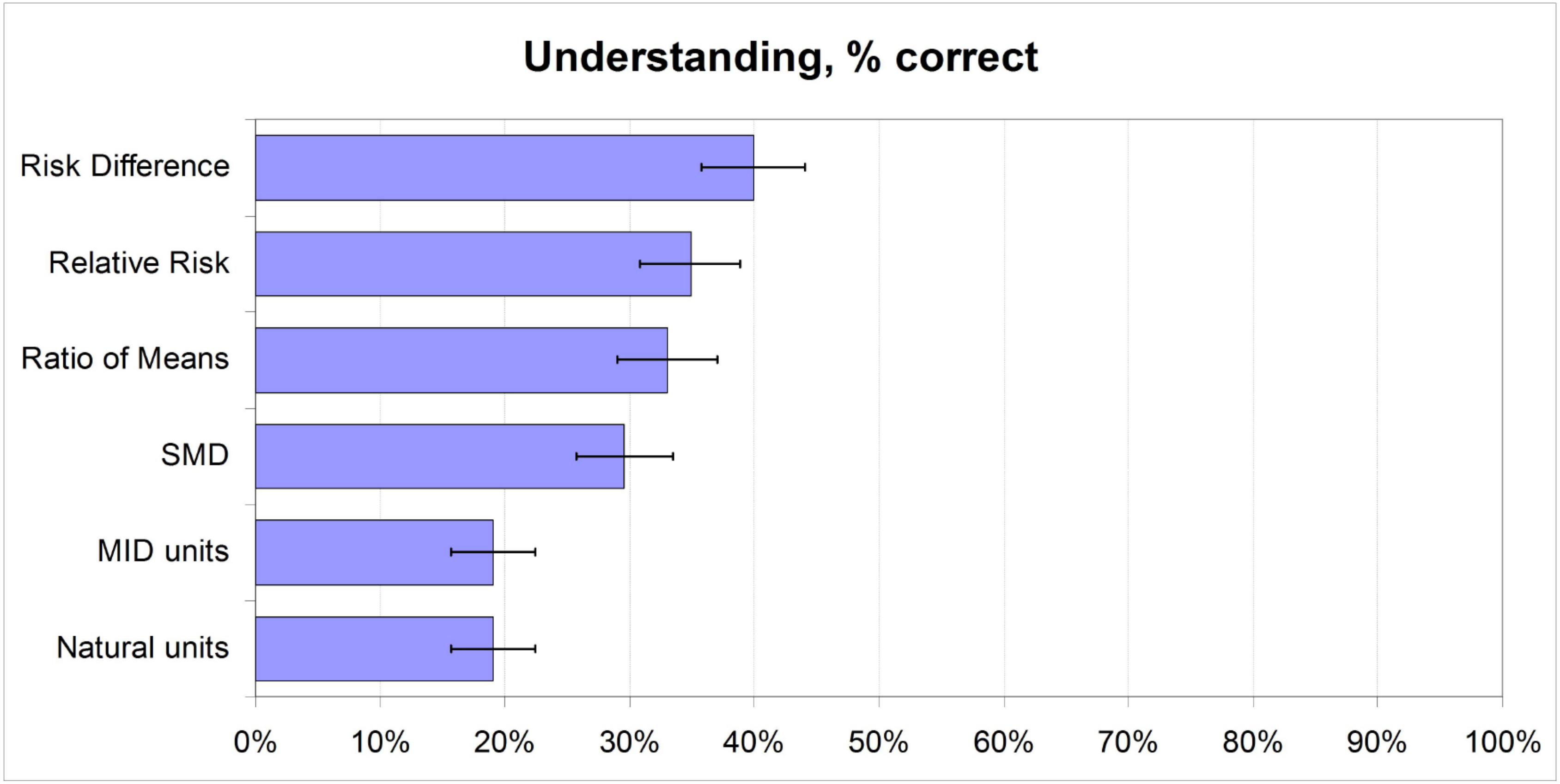
- SMD 0.79
- Natural units 3.5 to 8.1 on 100 pt scale
- Dichotomy
  - based on SMD risk difference 0.15
  - based on MID 0.03
- Ratio of means 0.87
- 0.40 MID units
- Discrepancy? Explanation

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
- Results depicted as small or large effect for 6 statistical presentation approaches
- Response options
  - *trivial difference, probably not important*
  - *small difference, but probably important*
  - *moderate difference, surely important*
  - *large difference, very important*

Results: Correct answers

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

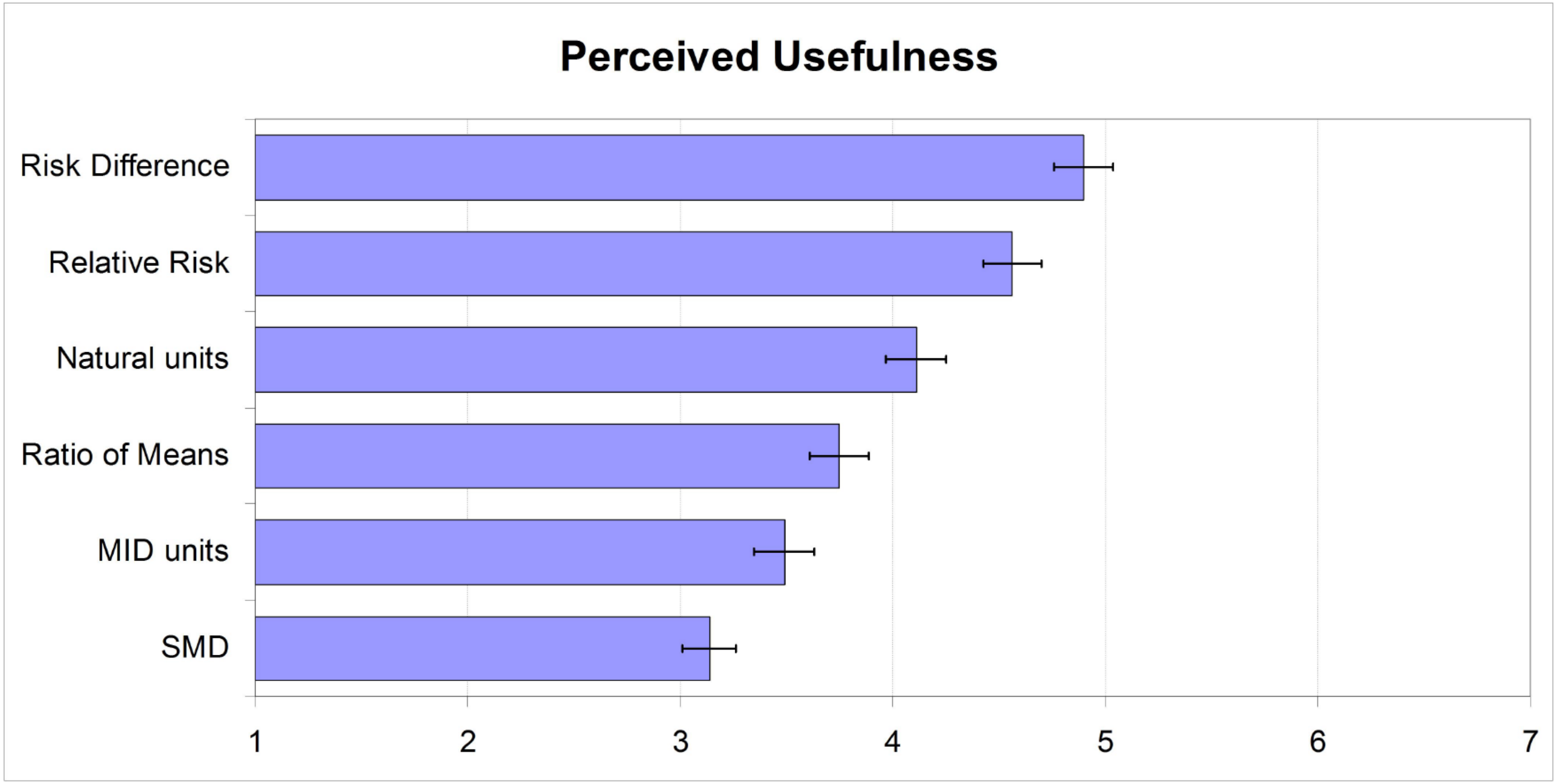


In pooled standard deviation units of all pain scores in the treatment and control groups, a meta-analysis finds the effect of intervention A versus placebo control for patient-reported pain to be 0.20 standard deviation units in favour of intervention A. Please clearly indicate whether this presentation approach is useful:

1	*2*	*3*	*4*	*5*	*6*	*7*
Not useful in understanding size and importance of the effect						*Extremely useful in understanding the size and importance of the effect*

Results: Usefulness

Figure 4: Perceived Usefulness, *n* = 531



Higher scores represent higher perceived usefulness

Informing a practice guideline

- Patients with knee pain
  - Degenerative knee disease
  - Impact of arthroscopy, lavage, debridement, menisectomy

- Outcome: Pain and function
  - Variety of instruments

Our Approach

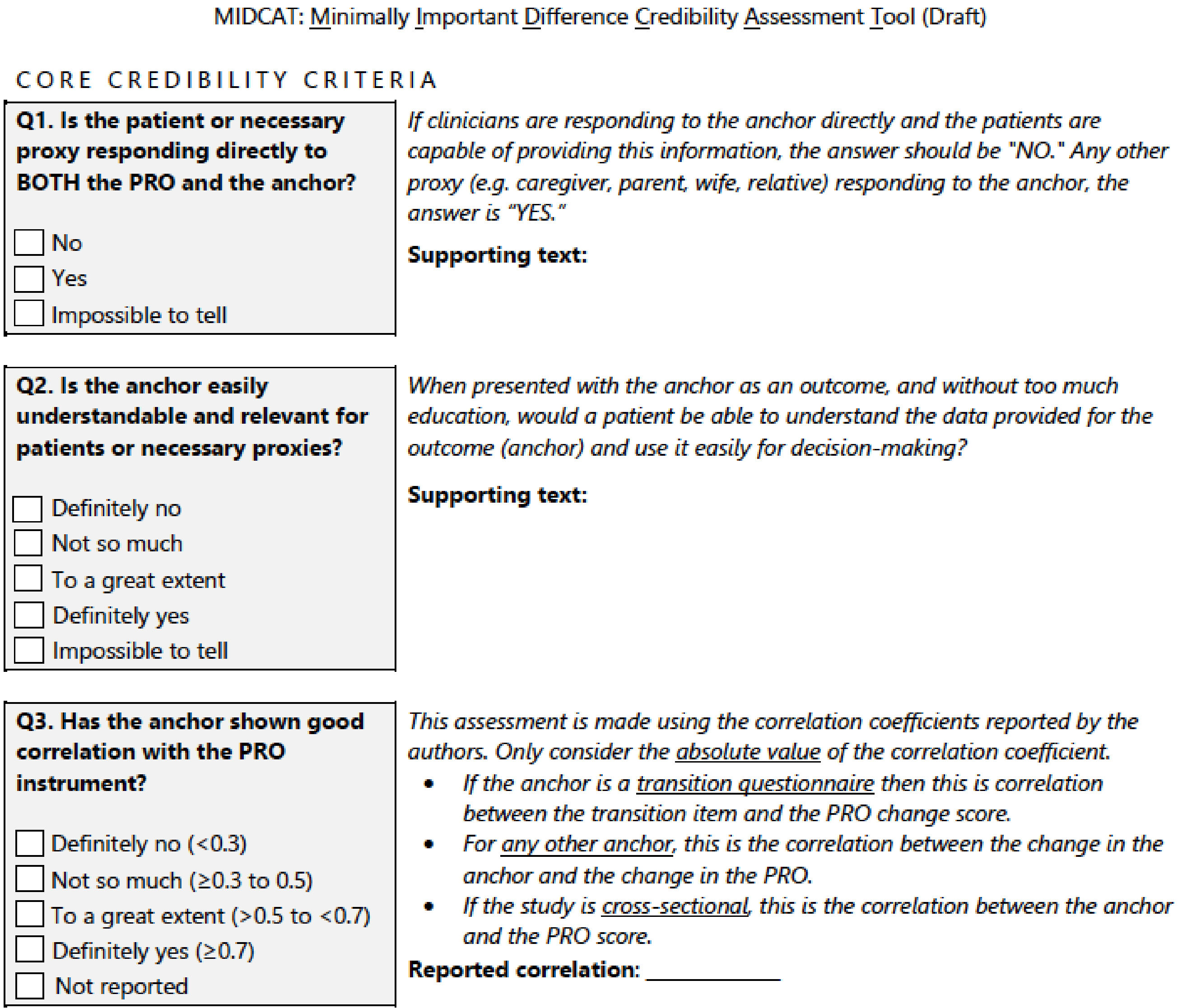
1. **Systematic review**
    - What amount of change on a given instrument's scale is important to patients?
        - minimal important difference (MID)
    - Systematically searched for empirical studies estimating anchor-based MIDs for instruments included in meta-analysis for benefit
    - Assessed credibility of identified MIDs by applying a single criterion: correlation between change in PRO and the transition item anchor ≥0.4
2. **Identified a range of credible MIDs for each key outcome measure and used the median**

Our Approach

3. **Meta-analysis – results presented in two ways:**
   - Mean difference
     - Scores transformed to the scale of an index instrument (the highest in the hierarchy)
   - Risk difference

Outcome Timeframe	Study results and measurements	Absolute effect estimates		Certainty in effect estimates (Quality of evidence)
		Conservative management	Arthroscopy	
Short term (3 months)				
Pain (difference in change from baseline)	index instrument (KOOS pain sub scale) Scale: 0-100 High better, **MID 12** Data from 1231 patients in 10 studies	15.0 points (Mean)	20.0 points (Mean)	High
		Mean Difference **5.4 more** (CI 95% 1.9 more - 8.8 more)		
Pain (difference in patients who achieve a change higher than the MID)	Data from 1102 patients in 9 studies	669 per 1000	793 per 1000	High
		Difference: **124 more per 1000**		

Credibility assessment of MIDs



MIDCAT: Minimally Important Difference Credibility Assessment Tool (Draft)

CORE CREDIBILITY CRITERIA

Q1. Is the patient or necessary proxy responding directly to BOTH the PRO and the anchor?

□ No
□ Yes
□ Impossible to tell

If clinicians are responding to the anchor directly and the patients are capable of providing this information, the answer should be "NO." Any other proxy (e.g. caregiver, parent, wife, relative) responding to the anchor, the answer is "YES."

Supporting text:

Q2. Is the anchor easily understandable and relevant for patients or necessary proxies?

□ Definitely no
□ Not so much
□ To a great extent
□ Definitely yes
□ Impossible to tell

When presented with the anchor as an outcome, and without too much education, would a patient be able to understand the data provided for the outcome (anchor) and use it easily for decision-making?

Supporting text:

Q3. Has the anchor shown good correlation with the PRO instrument?

□ Definitely no (<0.3)
□ Not so much (≥0.3 to 0.5)
□ To a great extent (>0.5 to <0.7)
□ Definitely yes (≥0.7)
□ Not reported

This assessment is made using the correlation coefficients reported by the authors. Only consider the absolute value of the correlation coefficient.

- *If the anchor is a transition questionnaire then this is correlation between the transition item and the PRO change score.*
- *For any other anchor, this is the correlation between the change in the anchor and the change in the PRO.*
- *If the study is cross-sectional, this is the correlation between the anchor and the PRO score.*

Reported correlation: ___________

Approach	Advantages	Disadvantages	Recommendation
(A) Standard deviation (SD) units (standardized mean difference; effect size)	Widely used	Interpretation challenging Can be misleading depending on whether population very homogenous or heterogeneous	Do not use as the only approach
(B) Present as natural units	May be viewed as closer to primary data	Few instruments sufficiently used in clinical practice to make units easily interpretable	Approaches to conversion to natural units include those based on SD units and re-scaling approaches. We suggest the latter. In rare situations when instrument very familiar to front line clinicians seriously consider this presentation.
(C) Relative and absolute effects	Very familiar to clinical audiences and thus facilitate understanding Can apply GRADE guidance for large and very large effects	Involve assumptions that may be questionable (particularly methods based on SD units)	If the minimal important difference is known use this strategy in preference to relying on SD units Always seriously consider this option
(D) Ratio of means	May be easily interpretable to clinical audiences Involves fewer questionable assumptions than some other approaches Can apply GRADE guidance for large and very large effects	Cannot be applied when measure is change and therefore negative values possible Interpretation requires knowledge and interpretation of control group mean	Consider as complementing other approaches, particularly the presentation of relative and absolute effects
(E) Minimal important difference units	May be easily interpretable to audiences Not vulnerable to population heterogeneity	Only applicable when minimal important difference is known To the extent that MID is uncertain, this approach will be less attractive	Consider as complementing other approaches, particularly the presentation of relative and absolute effects

Conclusions re interpretability

- If possible use natural dichotomies
- Many approaches rely on SD units
  - suffer from problem of heterogeneity
  - important limitation
- Approaches not relying on SD units preferable
  - ideally know (more or less) MID
  - can present in MID units and proportions
  - approaches complementary

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

TWEETS 1,873 FOLLOWING 487 FOLLOWERS 753 FAVORITES 324

EBCPMcMaster
@EBCPMcMaster

McMaster Centre for Evidence-Based Clinical Practice, birthplace of Evidence-Based Medicine, promoting the use of the best evidence to inform clinical decisions

Hamilton, On. Canada
ebcp.mcmaster.ca

24 Photos and videos

Pinned Tweet
EBCPMcMaster @EBCPMcMaster · Sep 9
Forbidding Dr. Google - Moving forward to inform our patients better and evidence-based. We need the tools ow.ly/S0mri #EBM #EBCP
3 1

EBCPMcMaster @EBCPMcMaster · 53s
Systematic review findings on post stroke pain management are inconsistent with current clinical practice guidelines ow.ly/SfuBF

EBCPMcMaster retweeted
Kari Tikkinen @KariTikkinen · Sep 13
DTCA and lax consensus guidelines led up to 40-fold increase in testosterone prescriptions: onlinelibrary.wiley.com/store/10.1111/

Who to follow · Refresh · View all

PublicHealth Library @ham...
Followed by Sam Craigie an...
Follow

Julie Wood @woodjulieuk
Followed by Jani Ruotsalain...
Follow

Jonathan Craig @Jonathan...
Followed by Joerg Meerpohl...
Follow

Find friends

Trends · Change
NHS England
Exclusive: NHS England in talks over indemnity costs amid threat to...
355 Tweets about this trend

Ebcp.mcmaster.ca

Evidence-Based Clinical Practice Workshop

@EBCPMcMaster

Follow us on Twitter