Making results of patient-reported outcomes interpretable

Gordon Guyatt, MD, MSc

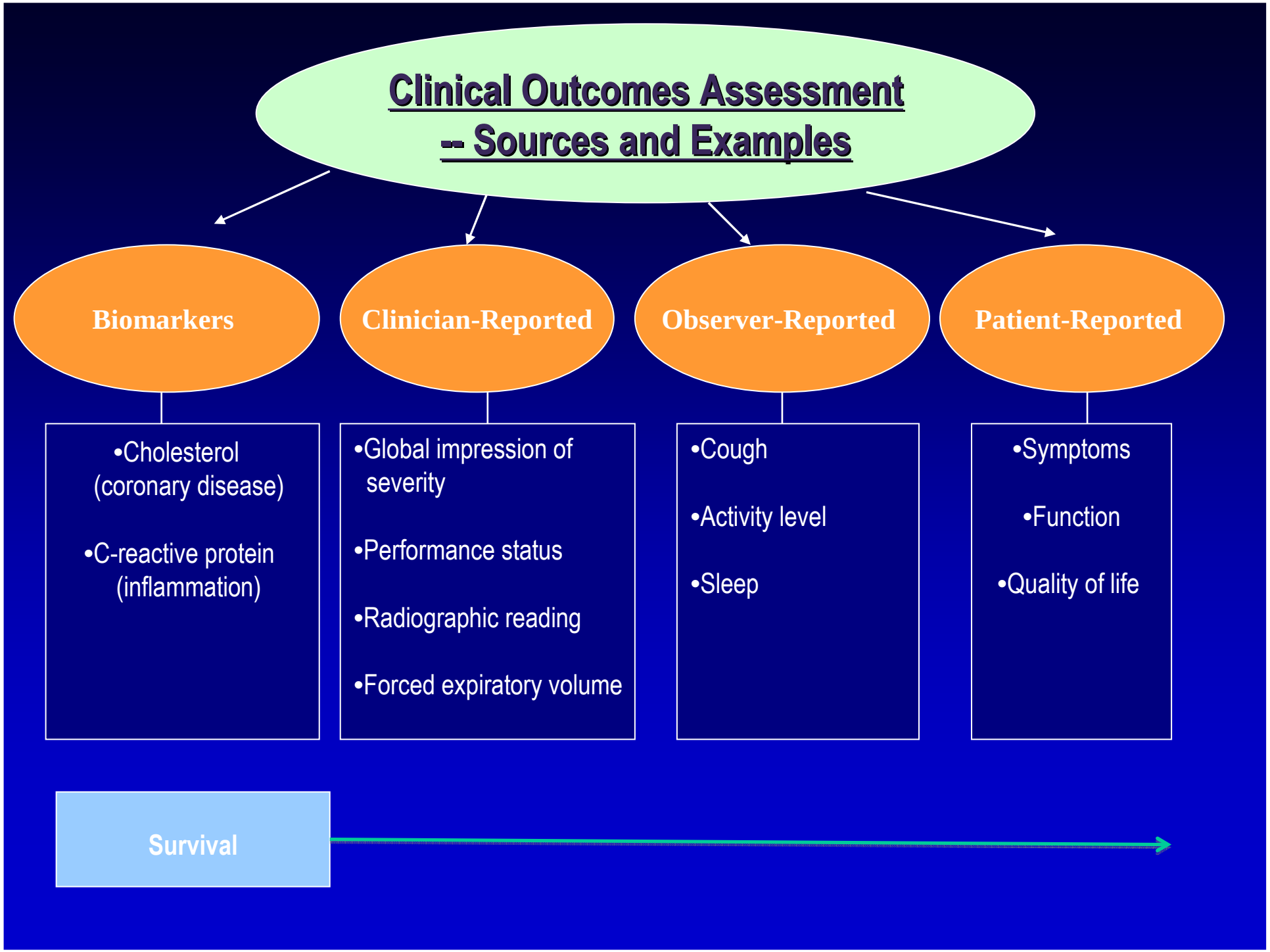
Slides available: guyatt@mcmaster.ca

Plan

- PROs in Cochrane reviews
- the problem of interpretability
- strategies for making results interpretable
  - effect sizes
  - minimal important differences
- systematic reviews and meta-analyses
  - options for summarizing effects

What this is an is not

- not an introduction to PROs
- focus on systematic reviews
  - GRADE approach, Summary of Findings
- not for statistical beginners
  - statistically heavy


Clinical Outcomes Assessment
-- Sources and Examples
Biomarkers
Clinician-Reported
Observer-Reported
Patient-Reported
•Cholesterol (coronary disease)
•C-reactive protein (inflammation)
•Global impression of severity
•Performance status
•Radiographic reading
•Forced expiratory volume
•Cough
•Activity level
•Sleep
•Symptoms
•Function
•Quality of life
Survival

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.)
- **PROs are not concepts in and of themselves but a class of outcomes**
  - **requires concept purported to be measured be specified, i.e., respiratory symptoms, physical function, reduction in pain severity**

PROs in Cochrane Reviews

- Appear in Summary of Findings Table
- Information from published meta-analyses
- Dichotomous (yes\no) and continuous outcomes
- Focus here on continuous outcomes and how to interpret them

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?

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 ($P<0.001$).
- effect size?
  - trivial, small but important, large?

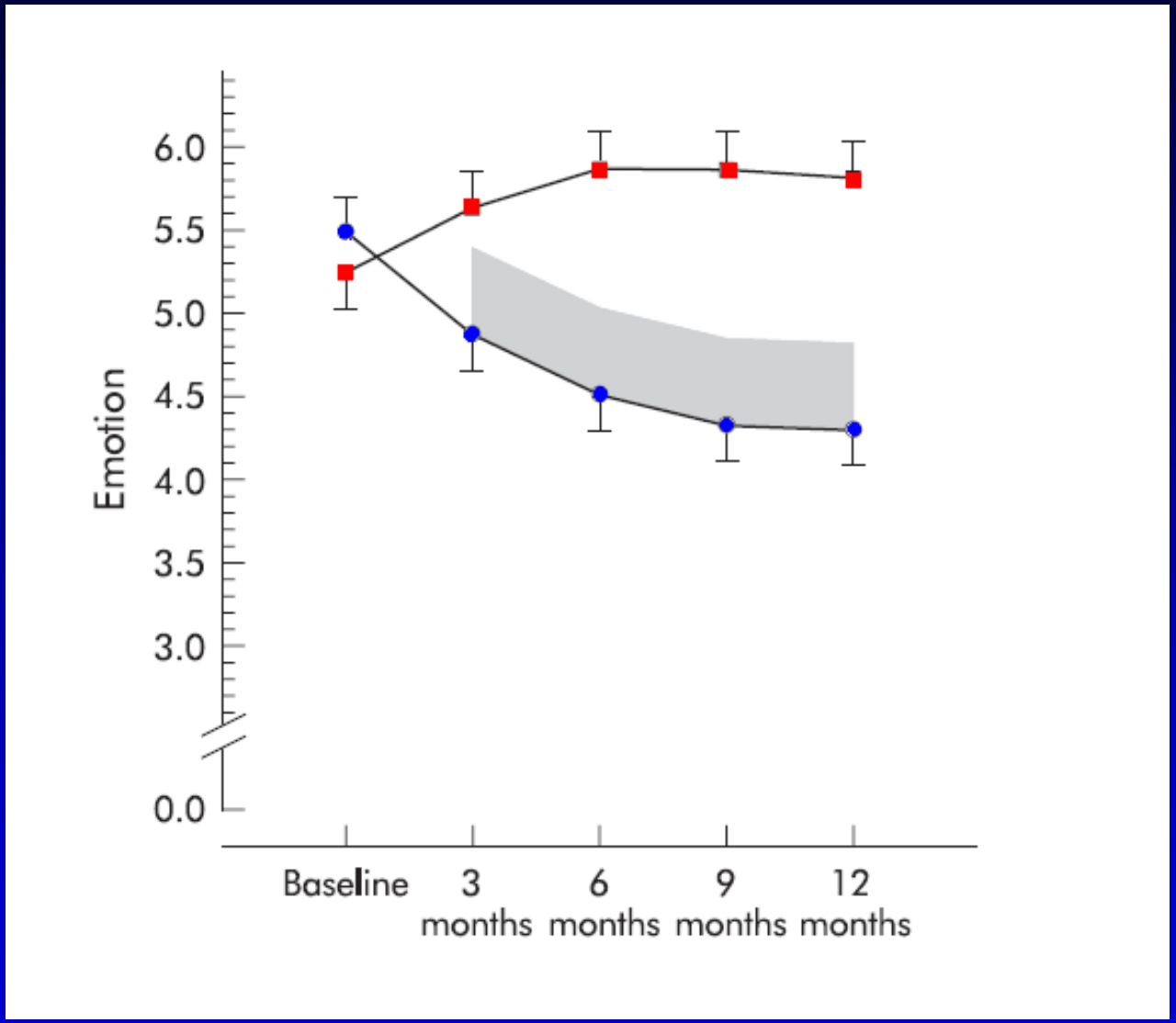
Minimally important difference

- 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?

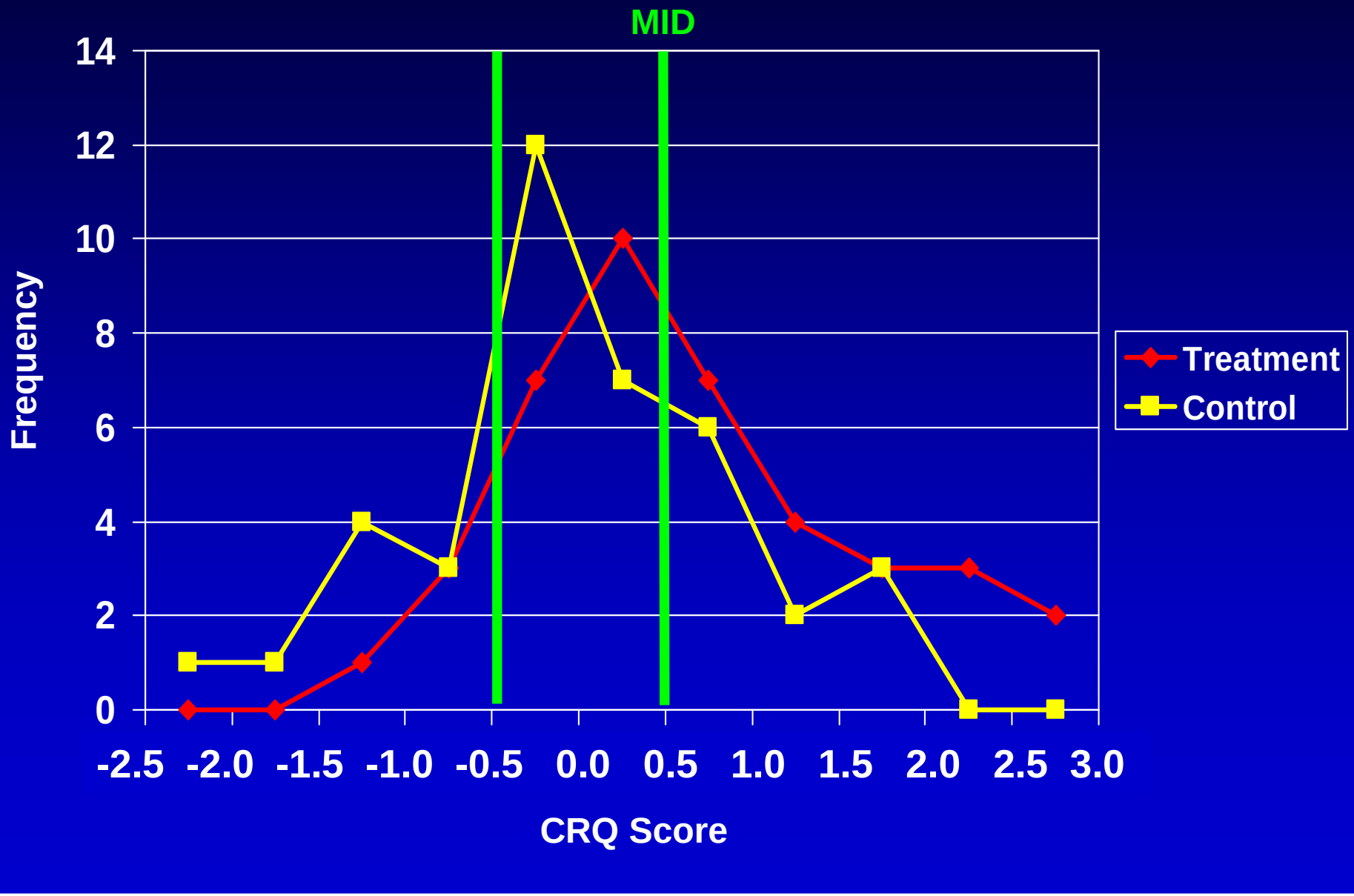
What if effect smaller

- randomized trial respiratory rehabilitation in COPD
- effect on emotional function 0.4
- important? how important?

Applying the MID

- assume MID is 0.50 and patients mean improvement 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



Number Needed to Treat

- Number needed to treat (NNT) for 1 person to achieve a specified change in a PRO (responder criteria)
- *NNT = 100/(pT – pC)*
- pT is the percentage of patients who improved in the treatment group, and
- pC is the proportion of patients who improved in the control group

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	No 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

Systematic reviews, meta-analysis

- seldom have original data from individual studies to apply thresholds
- individual studies my use different PROs to measure same concepts

Meta-analysis

- studies all use same or similar outcome
  - not intuitively interpretable to the audience
- could give weighted mean difference in natural units
  - 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)

Would you recommend respiratory rehabilitation to your patients?

Outcomes	Absolute risks (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)
	Estimated risk **Without stockings**	Corresponding risk **With stockings (95% CI)**			
Symptomatic deep vein thrombosis – inferred from surrogate, symptomless deep vein thrombosis	**Low risk population**		**RR 0.10** (0.04 to 0.25)	2637 (9 studies)	⊕⊕⊕◯ **Moderate** due to indirectness [4]
	5 per 10,000	**0.5 per 10,000** (0 to 1.25)			
	High risk population				
	18 per 10,000	**1.8 per 10,000** (1 to 8)			
Oedema Post-flight values measured on a scale from 0, no oedema, to 10, maximum oedema.	The mean oedema score ranged across control groups from **6.4 to 8.9**.	The mean oedema score in the intervention groups was on average **-4.72 lower** (95% CI −4.91 to -4.52).		1246 (6 studies)	⊕⊕◯◯ **Low**[4] Due to risk of bias (unblinded, unvalidated measure)

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

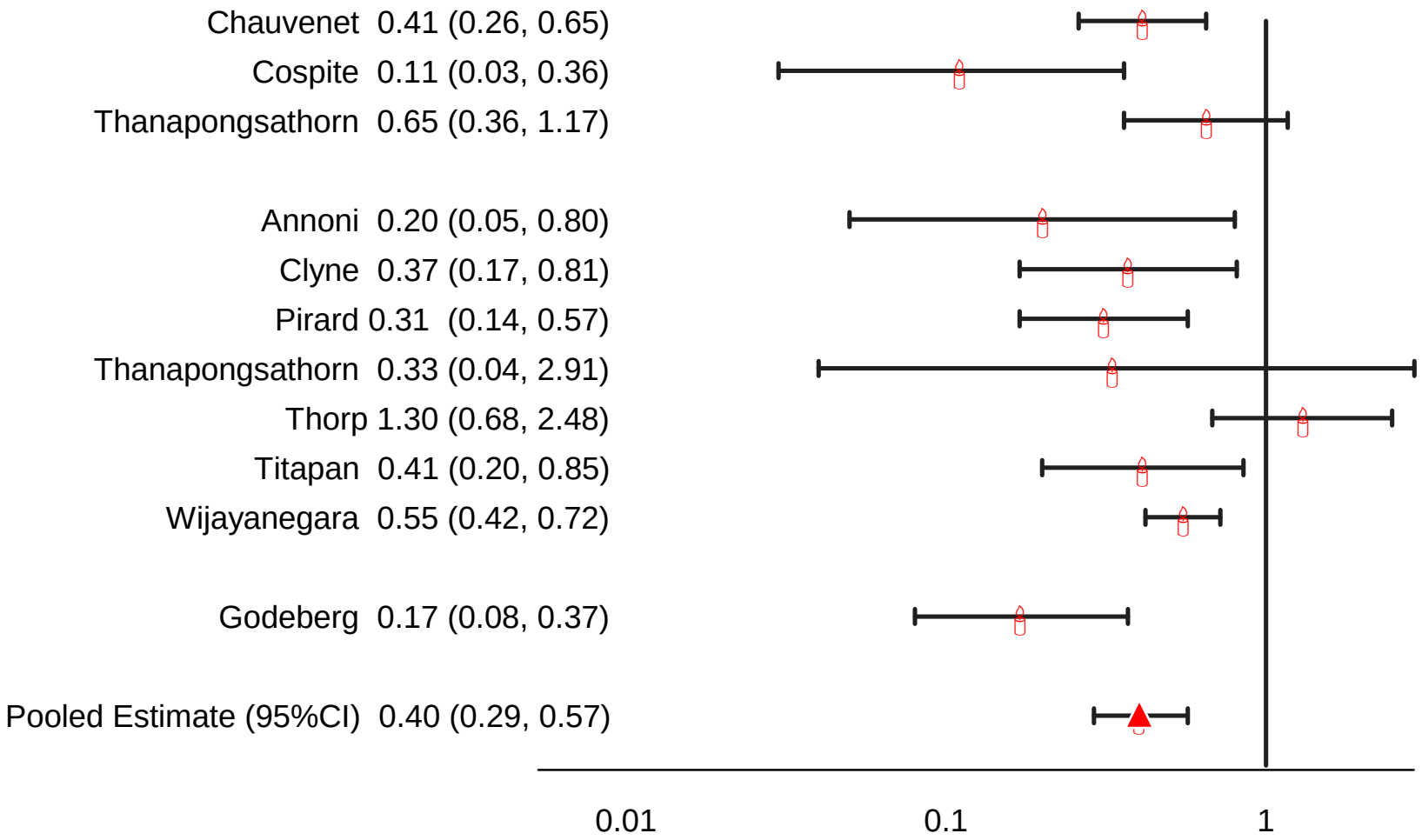
Systematic review of RCTs of thrombolysis in acute stroke

- 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)



Studies used different instruments measure same construct

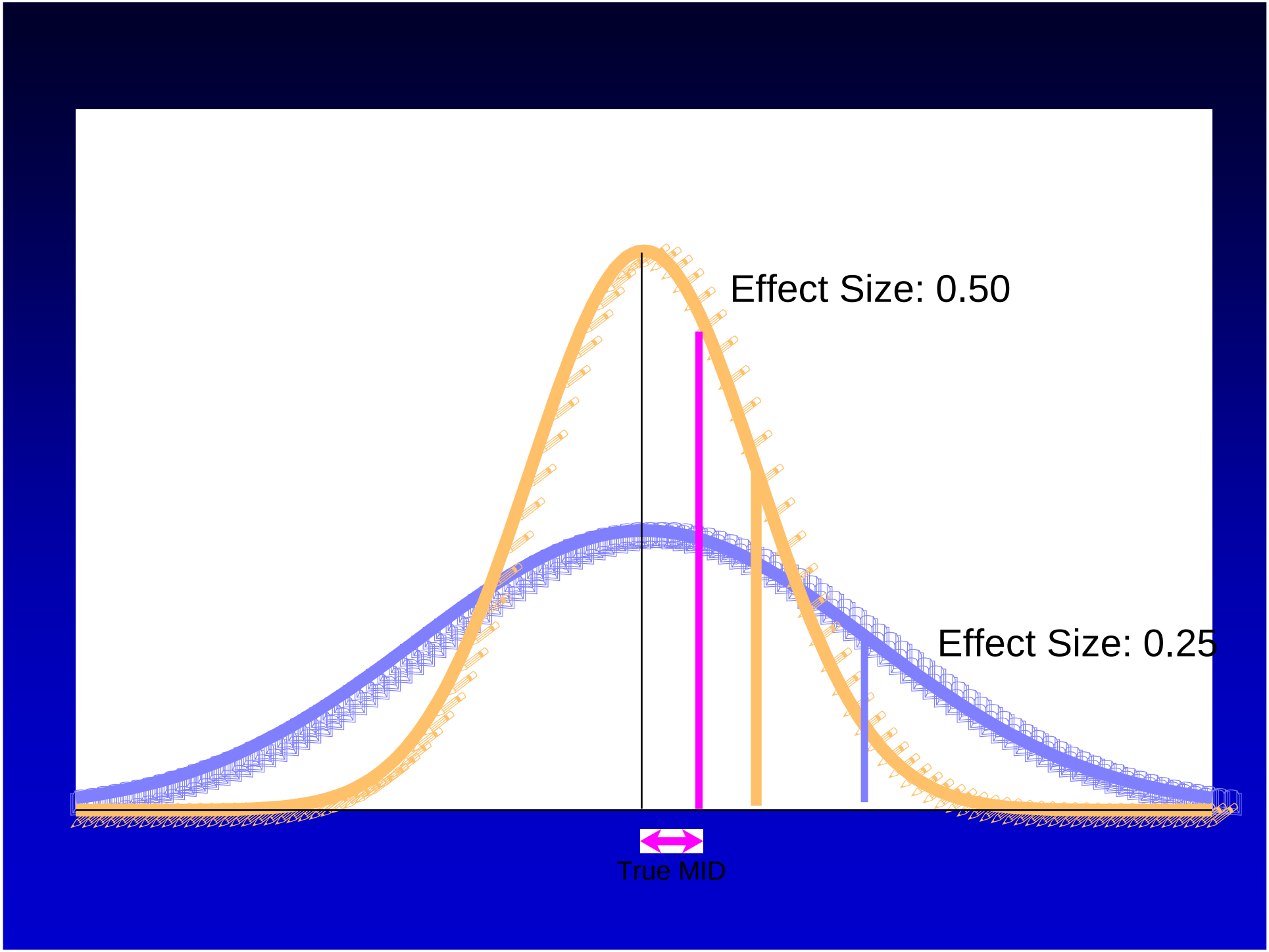
Approach	Advantages	Disadvantages
(A) Standard deviation (SD) units (standardized mean difference; effect size)	Widely used	Interpretation challenging Can be misleading depending on whether population very homogenous or heterogeneous
(B) Present as natural units	May be viewed as closer to primary data	Few instruments sufficiently used in clinical practice to make units easily interpretable
(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)
(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
(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

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

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference 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)

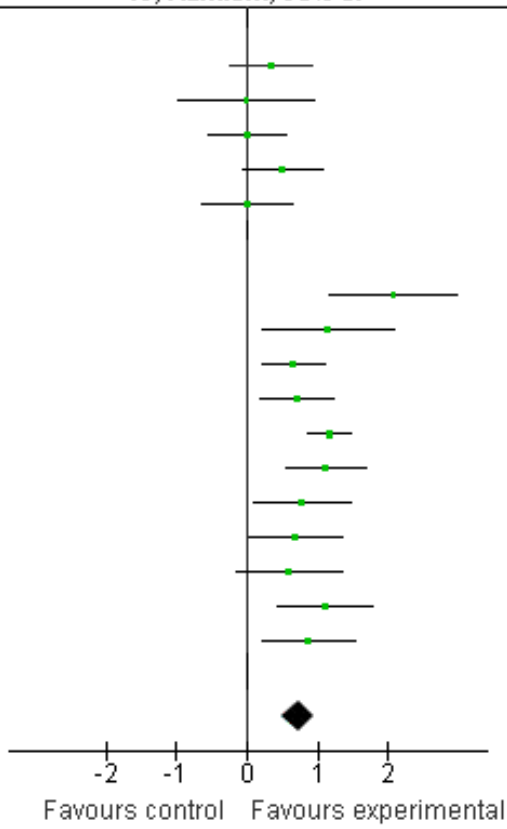


Table 5: Application of approaches to chronic respiratory rehabilitation for health-related quality of life impairment in patients with chronic airflow limitation

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

- confident encourage
- possibly encourage
- probably discourage
- certainly discourage

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
  - divide by 7 to go to CRQ units

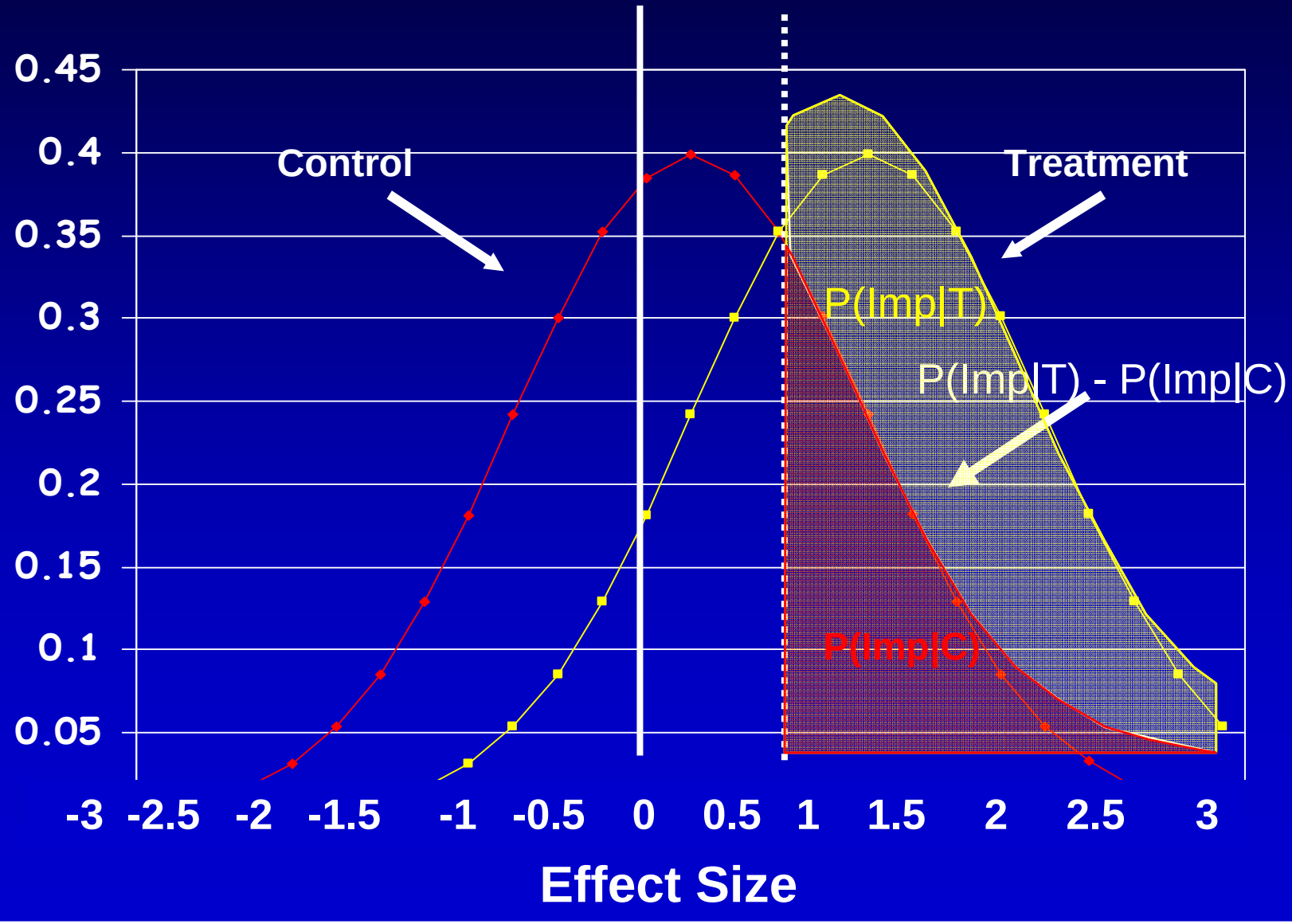
(B) Health-related quality of life (HRQL) measured on a scale of 1 to 7	Control group baseline 4.5[1] Average improvement in control 0.04	HRQL improved on average **0.71 (95% CI 0.48 to 0.94)** more in the respiratory rehabilitation patients than in the control patients	---	818 (16)	⊕⊕⊕⊕ High	Calculated by transforming all scores to the Chronic Respiratory Questionnaire in which the minimal important difference is 0.5

vulnerable to no one benefits/everyone benefits

- confident encourage
- possibly encourage
- probably discourage
- certainly discourage

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
  - Suissa/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

- not necessarily clear what is the outcome that is increasing/decreasing
- specify the control group proportion
  - 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
- don't require specifying control group rate
- Kraemer ROC AUC

Alternative

- if know MID for all instruments can go to individual studies
- calculate proportion benefiting in each individual study
- combine proportions across studies
- 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

Suissa/Furukawa RD 0.28
Kraemer RD 0.40

Ratio of Means (RoM)

$$RoM = \frac{mean_{exp}}{mean_{control}}.$$

- Requires estimate of variance of this ratio - this can be estimated using the delta method:

  - $Var_{ln(RoM)} = \frac{var_{exp}}{(mean_{exp}^2)} + \frac{var_{control}}{(mean_{control}^2)}$

Avoiding heterogeneity problem: 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

MID units

- Cochrane review of respiratory rehabilitation for COPD
- using 16 trials, we compared the existing method with the MID method
- trials employed two widely used disease-specific HRQL instruments
  - Chronic Respiratory Disease Questionnaire (CRQ)
  - St. Georges Respiratory Questionnaire (SGRQ)

Results

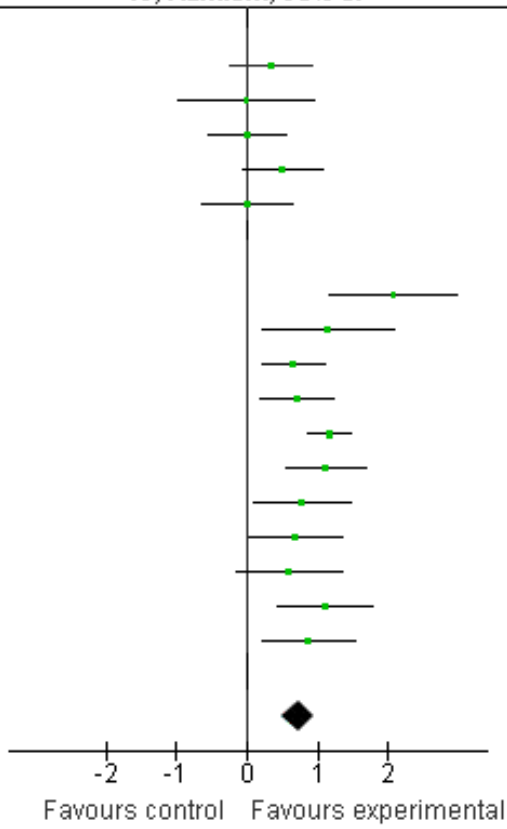
CRQ	**Mean Difference (95% CI)**
Dyspnea	1.06 (0.85, 1.26)
Emotional Function	0.76 (0.52, 1.00)
Fatigue	0.92 (0.71, 1.13)
Mastery	0.97 (0.74, 1.20)
Overall	0.94 (0.57, 1.32)
SGRQ	
Activities	4.78 (1.72, 7.83)
Impacts	6.27 (2.47, 10.08)
Symptoms	4.68 (0.25, 9.61)
Overall	6.11 (3.24, 8.98)

Results - SD Units

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference 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)

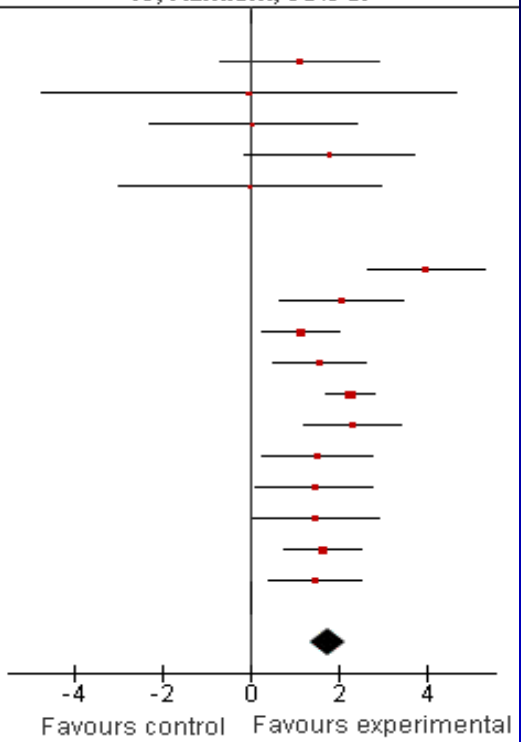


Results - MID Units

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: $Tau^2 = 0.17$; $Chi^2 = 22.15$, df = 15 (P = 0.10); $I^2 = 32\%$

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



(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
- 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
- no effect

Using MID method 3.5 (0.5 to 6.5) lower

Dichotomy

(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
- 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)	⊕⊕OO [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
- no effect

MID units

E) Post-operative pain nvestigators measured ain using different nstruments.	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)	⊕⊕OO [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
- no effect

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 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, especially MID
- one of approaches should be dichotomy

For copies of the slides

- Contact

guyatt@mcmaster.ca