Making results of PROs interpretable

Gordon Guyatt, MD, MSc
Donald Patrick Ph.D

Pooling health-related quality of life outcomes in meta-analysis—a tutorial and review of methods for enhancing interpretability

Original Article

Research Synthesis Methods

Received 25 May 2011, Revised 9 September 2011, Accepted 17 October 2011 Published online 14 December 2011 in Wiley Online Library

Making results of patient-reported outcomes interpretable

Gordon Guyatt, MD, MSc
Donald Patrick Ph.D

GRADE guidelines: 13. Preparing Summary of Findings tables and evidence profiles—continuous outcomes

Journal of Clinical Epidemiology 66 (2013) 173−183

Making results of patient-reported outcomes interpretable

Gordon Guyatt, MD, MSc
Donald Patrick Ph.D

GRADE guidelines: 13. Preparing Summary of Findings tables and evidence profiles—continuous outcomes

Journal of Clinical Epidemiology 66 (2013) 173−183

PROs Interpretability

- any patient-reported outcome
  - often health-related quality of life
  - continuous variable
- problems
  - scores non-intuitive
  - CRQ mean difference 1.06
  - often different measures same construct
- how to present results of meta-analysis
  - effect trivial, small, moderate, large

Studies all use same outcome

- mean difference in natural units
  - rehab in COPD, CRQ dyspnea 1.06 on 7 point scale
- minimal important difference
  - smallest difference patients consider important
- for CRQ 0.5 on 1 to 7 scale

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?

- Yes
- No
- Not sure

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

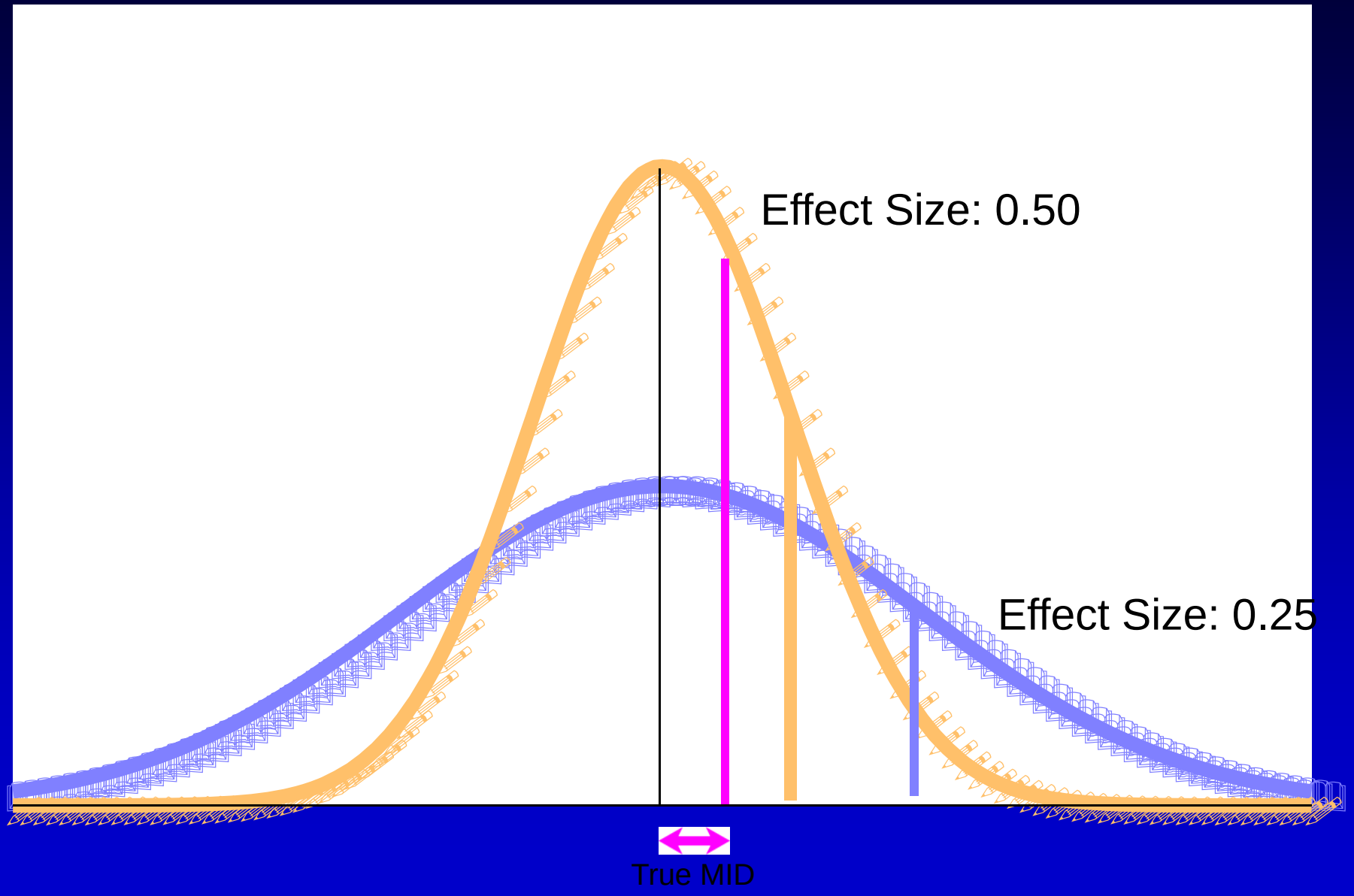
- use Rankin threshold 2 to 3
  - 2 minor handicap
  - 3 moderate handicap
  - proportion "dead or disabled"
- "death or dependency"
  - odds ratio 0.84 (95% CI 0.75 to 0.95)
  - 4% absolute risk reduction
  - NNT 25

Studies use different measures

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

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

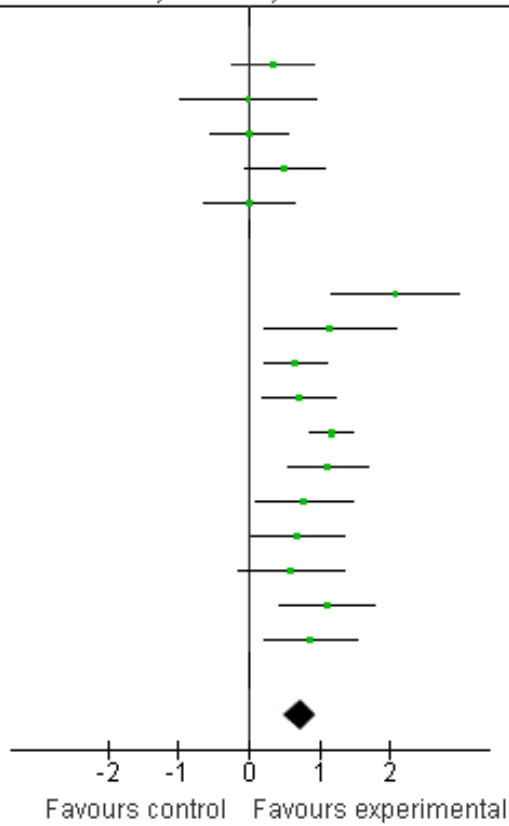


Table 5: Application of approaches to chronic respiratory rehabilitation for health-related quality of life impairment in patients with chronic airflow limitation

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

Conversion to familiar units

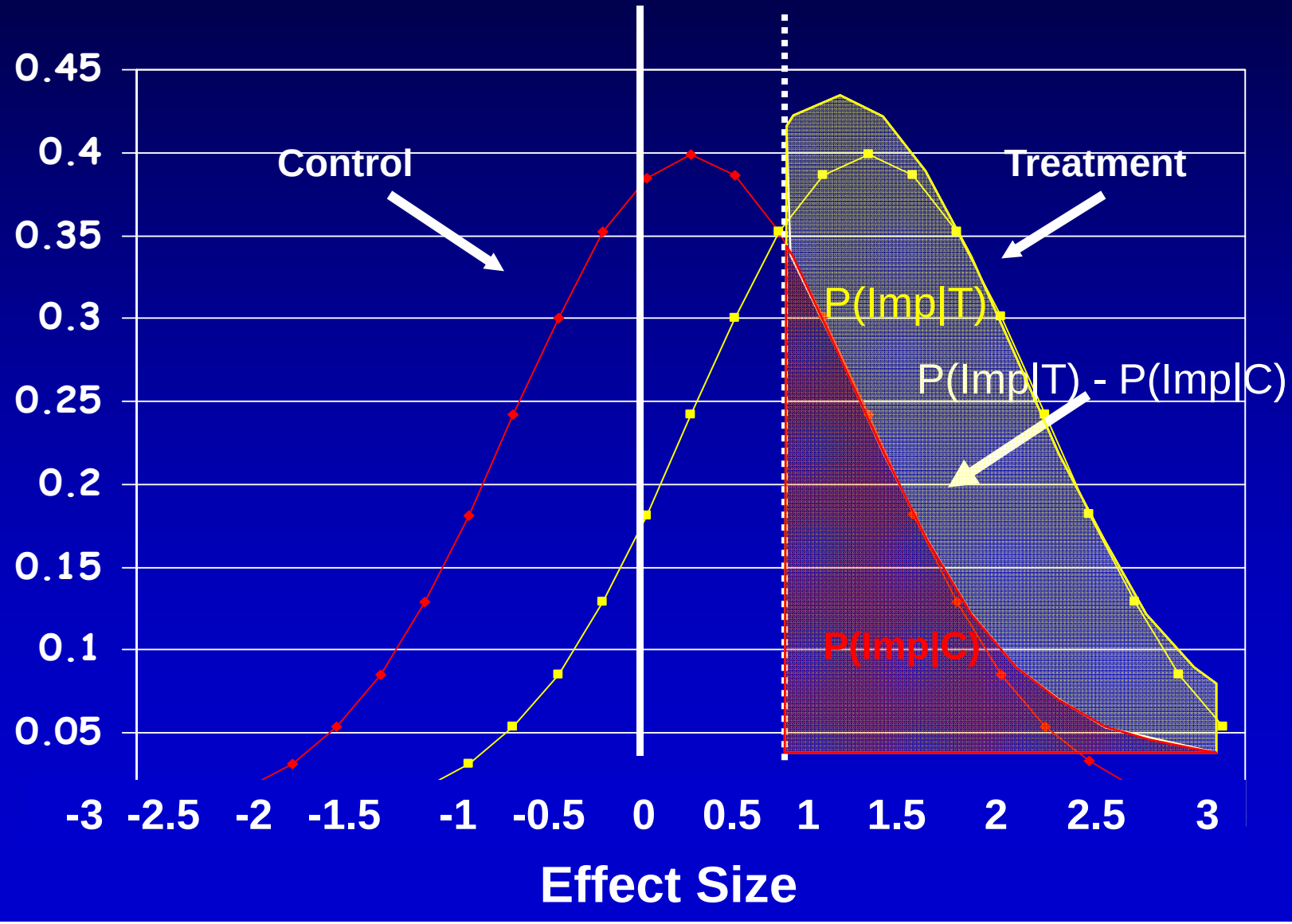
- all instruments into most familiar
  - two statistical approaches
- multiply SD units X SD of most familiar
  - may be challenging to decide which SD
  - vulnerable to heterogenity
- rescale to units of most familiar
  - St. George's 0 to 100
  - divide by 7 to go to CRQ units

(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

What if mean difference 0.3?
Vulnerable to no one benefits/everyone benefits

Dichotomize

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



Dichotomize

- number of statistical approaches relying on SMD
- normal distribution/equal variance
  - Furukawa
  - other approaches, similar assumptions

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

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

Furukawa RD 0.28

MID units

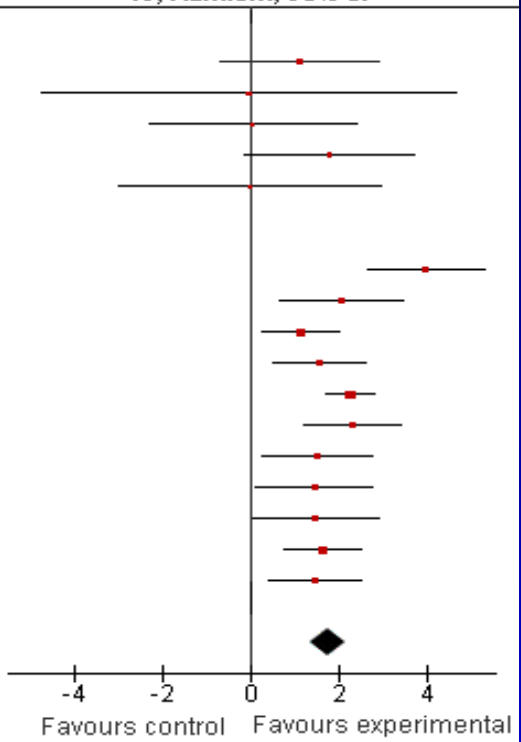
- Cochrane review of respiratory rehabilitation for COPD
- using 16 trials, we compared the existing method with the MID method
- trials employed two widely used disease-specific HRQL instruments
  - Chronic Respiratory Disease Questionnaire (CRQ)
  - St. Georges Respiratory Questionnaire (SGRQ)

Results - 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

Conclusions re interpretability

- if possible use natural dichotomies
- many approaches rely on SD units
  - suffer from problem of heterogeneity
- 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 in SoF, especially MID
- one of approaches should be dichotomy