

# Making results of patient-reported outcomes interpretable

Gordon Guyatt, MD, MSc  
Donald L. Patrick, PhD, MSPH

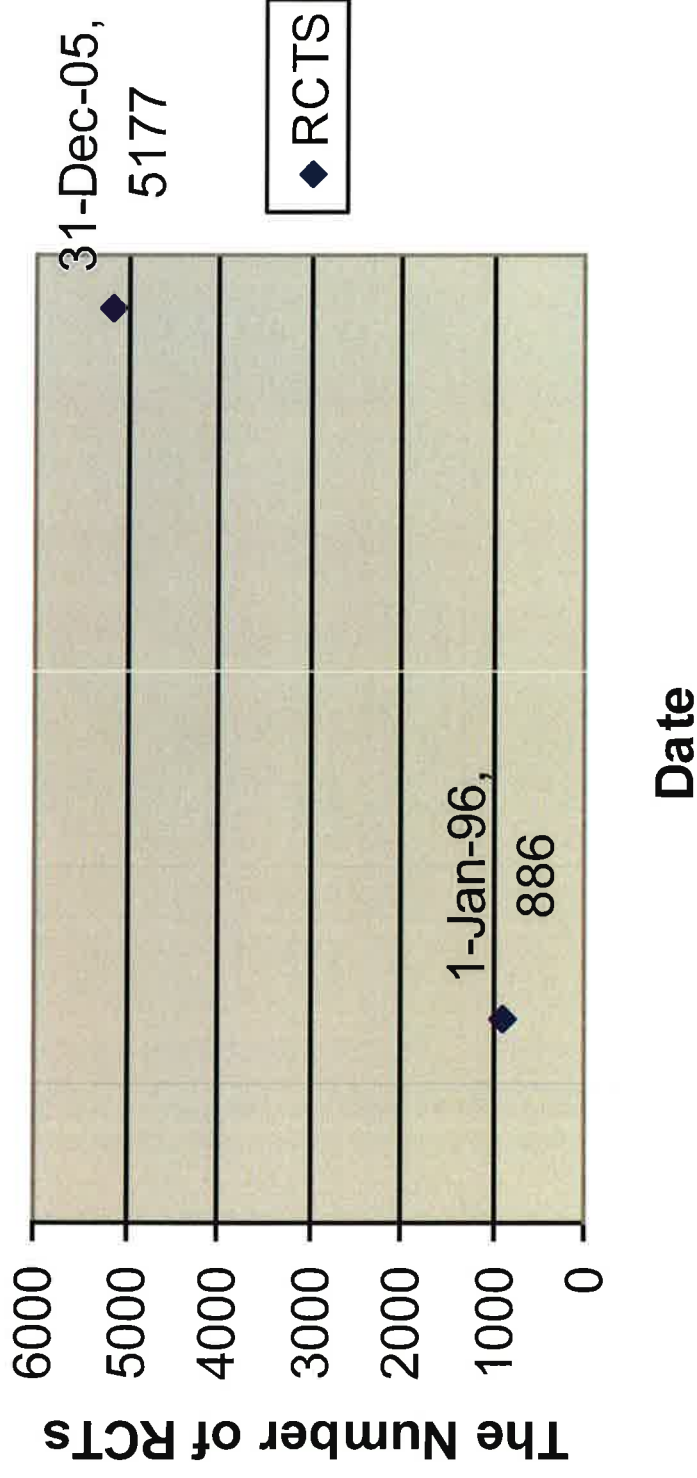
## Plan

- what are patient-reported outcomes?
- the problem of interpretability
- strategies for making results interpretable
  - effect sizes
  - minimal important differences
- systematic reviews and meta-analyses
  - options for summarizing effects

# What is a Patient-Reported Outcome (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 developed with patient input using qualitative methods a guiding principle**
- **PRO term** requires concept purported to be measured be specified
- **PRO ≠ QoL ≠ HrQoL**

## The Number of RCTs Including an Evaluation from the Patient's Perspective



# Why PROs?

- Some treatment effects known **only** to the patient, i.e. pain, symptoms, feelings
- Small changes in survival further informed by symptoms, function, and feelings
- Survival not only outcome of interest for many interventions
- Physiologic measures often do reflect how patient functions or feels
- Well-developed assessment by patients is as reliable if not more reliable than ratings of patient's condition by clinicians

# Patient-important Outcomes Sources and Examples

Survival

Morbid events

Clinician-reported

Caregiver-reported

Patient-reported

## For example

Survival curve

Myocardial infarction

Stroke

Disease exacerbation

Global Impression

Observation and test of functions

Morbid events (stroke, MI, hip fracture, etc)

Clinical events (hospitalization, coronary artery bypass surgery, etc)

Global Impression

Morbid events

Clinical events

Dependency

Functional status

Symptoms

Global Impression

Dependency

Functional status

Well Being

HRQoL

Satisfaction with treatment

Treatment Adherence

Others

# 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?



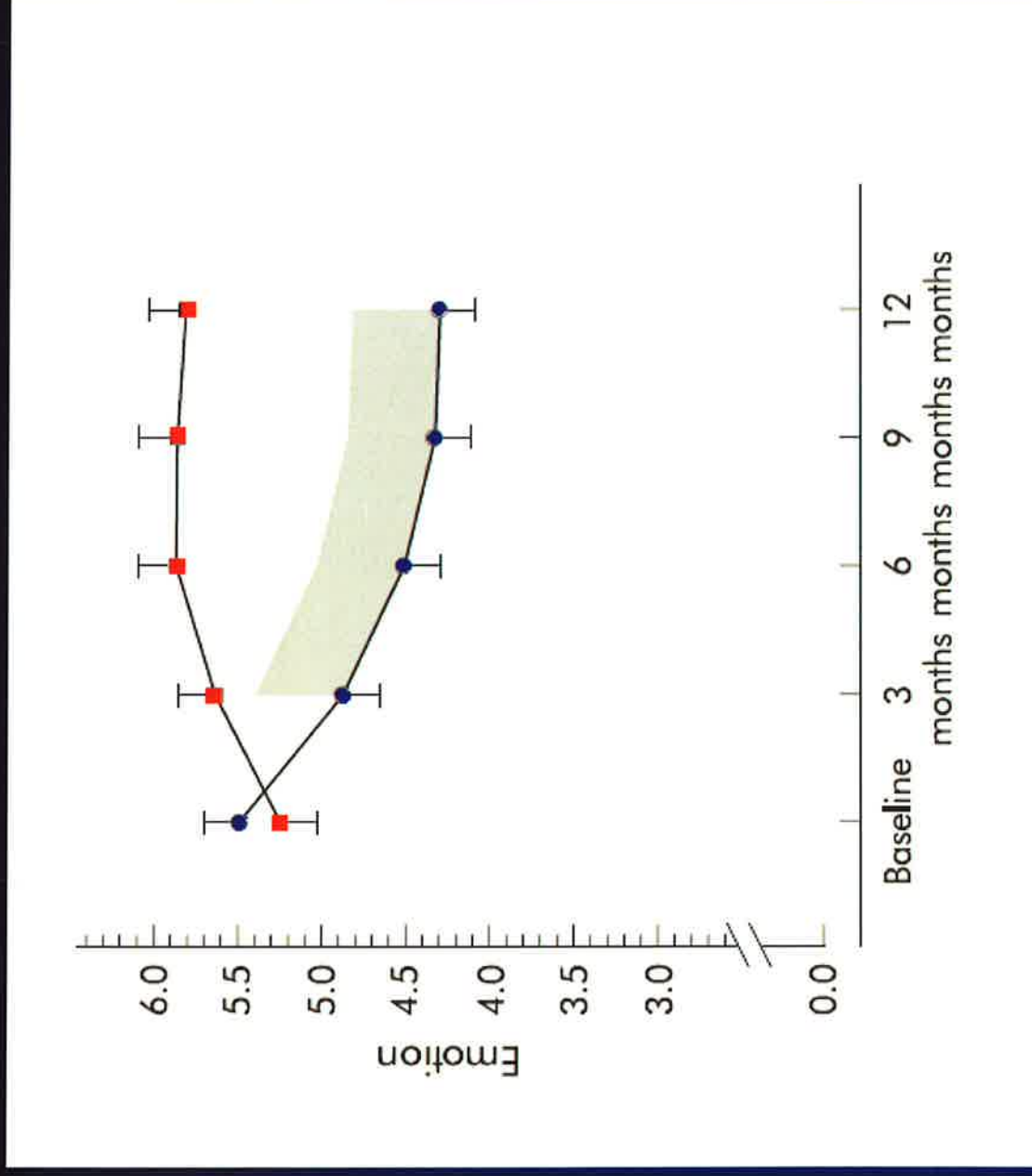
# Minimally important difference

- 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

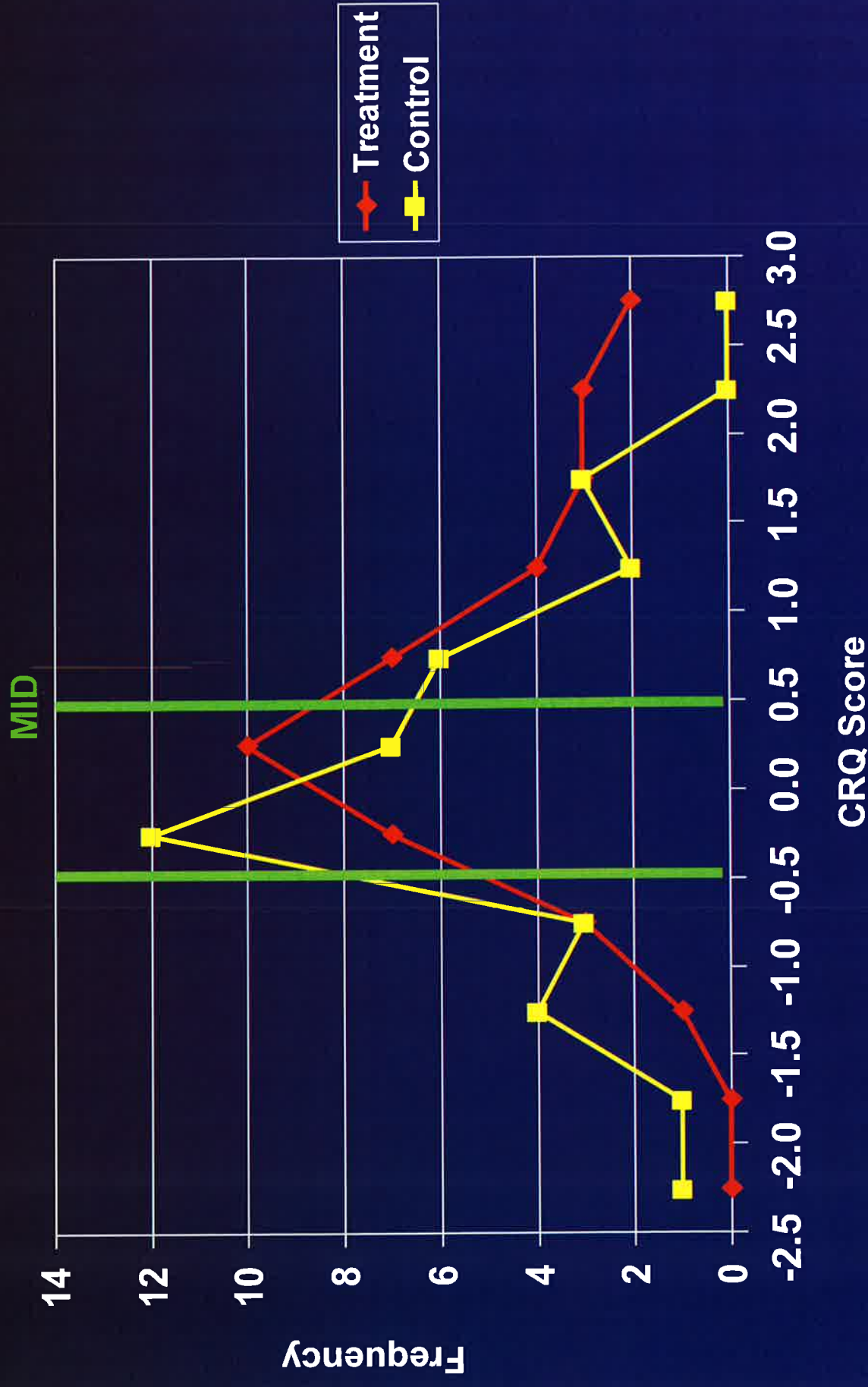


Would you recommend surgery to your patients on the basis of these results?

## What if effect smaller

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

# 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 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?



# Systematic review

- CRQ overall pooled
  - mean difference 0.7, CI 0.2 to 1.2
- St. George's MID 4
  - mean difference 6, CI 2 to 8
- what is your conclusion on size/importance of effect?

# Solution for RCT interpretation

- 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

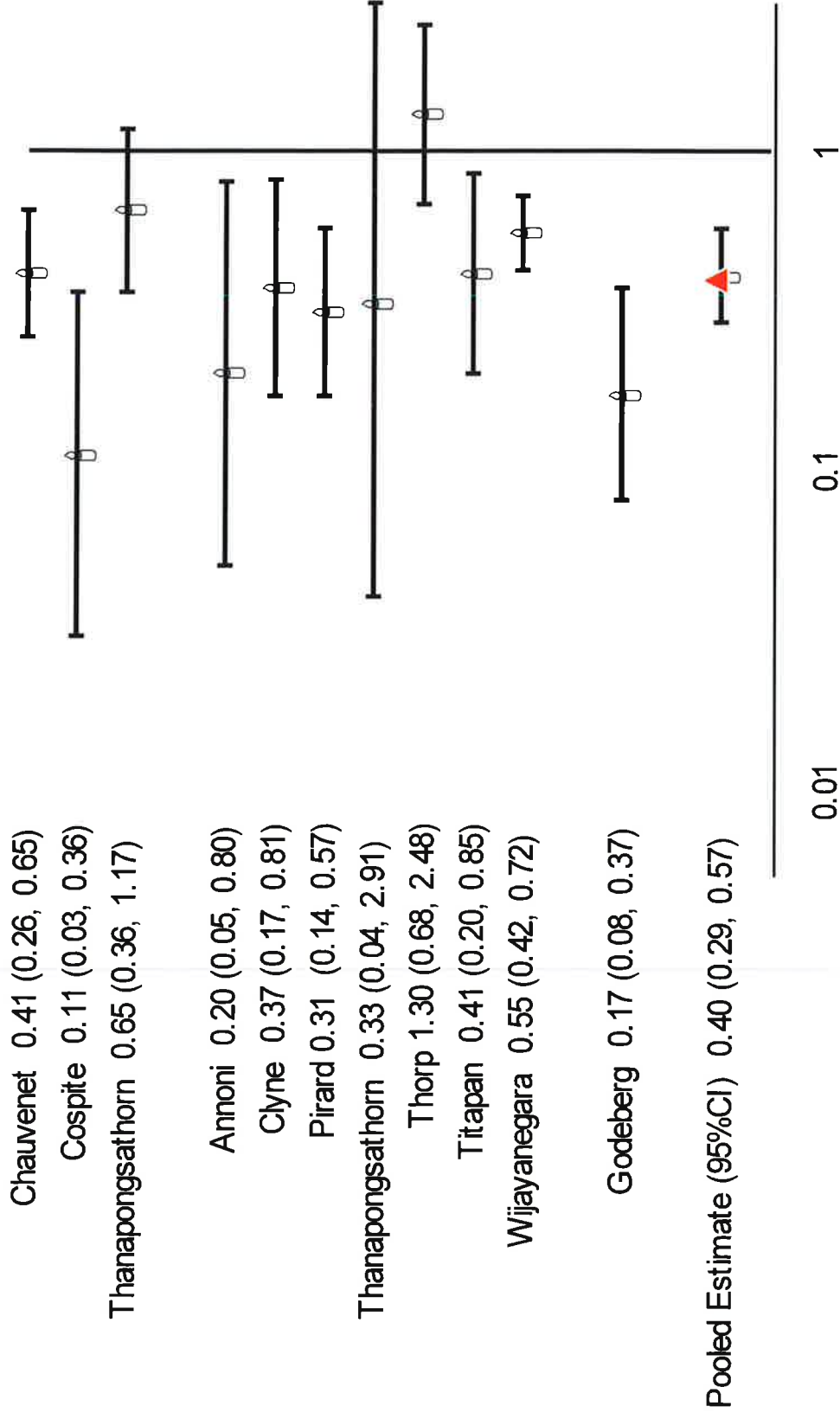
# Systematic review of RCTs of thrombolysis in acute stroke

- use Rankin threshold 2 to 3
  - 2 minor handicap
  - 3 moderate handicap
  - proportion "dead or disabled"
- "death or dependency"
  - 55.2% in thrombolysis, 68.3% in control
  - 42% odds reduction
  - 13.1% absolute risk reduction
  - NNT 7 to 8

# Flavonoids 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)

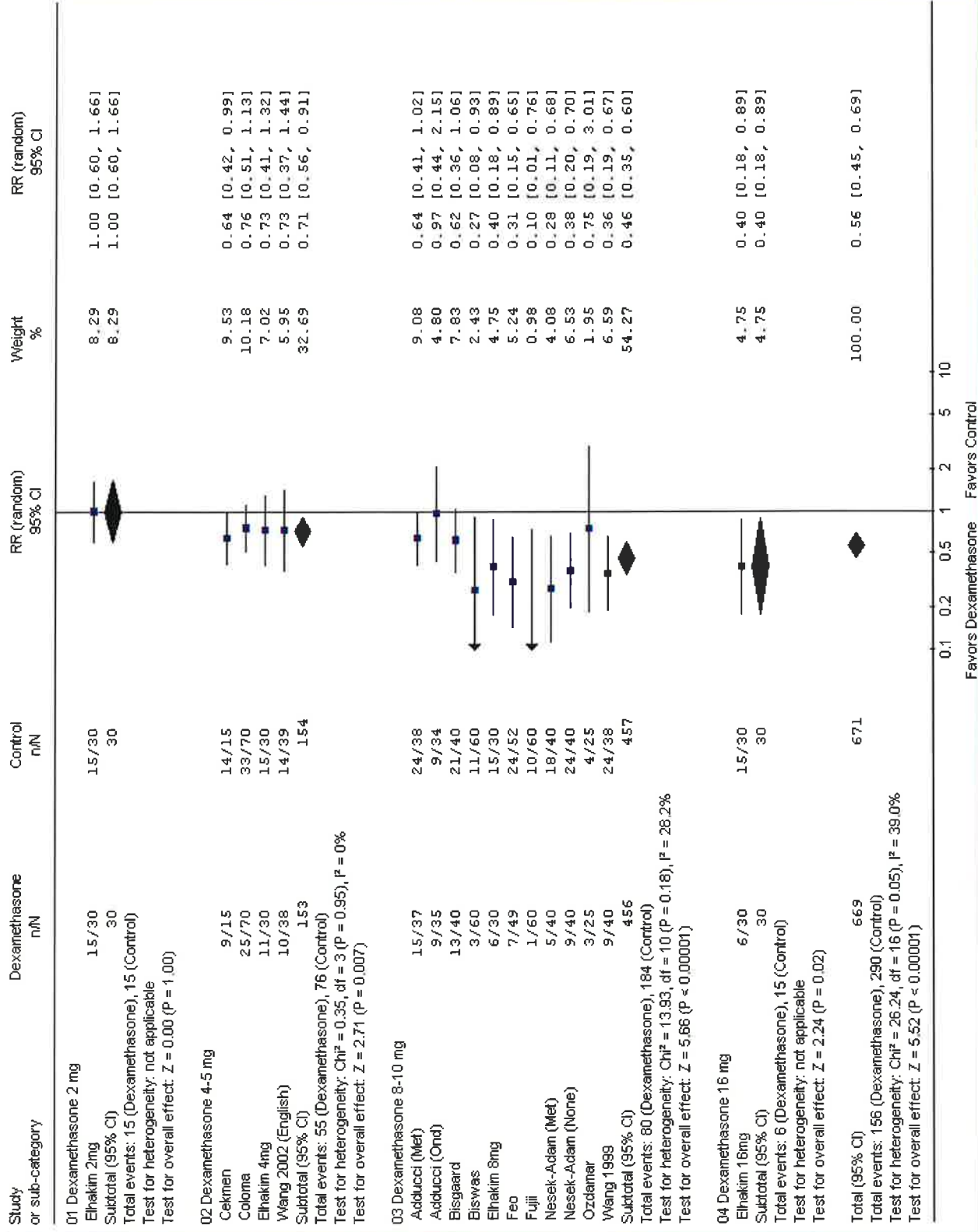


# Systematic reviews, meta-analysis

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

# Steroids for laparoscopic Cholecystectomy

- systematic review
- nausea and vomiting
  - 16 RCTs
- pain
  - 5 RCTs



Pain - no dichotomies, multiple measures - what to do?



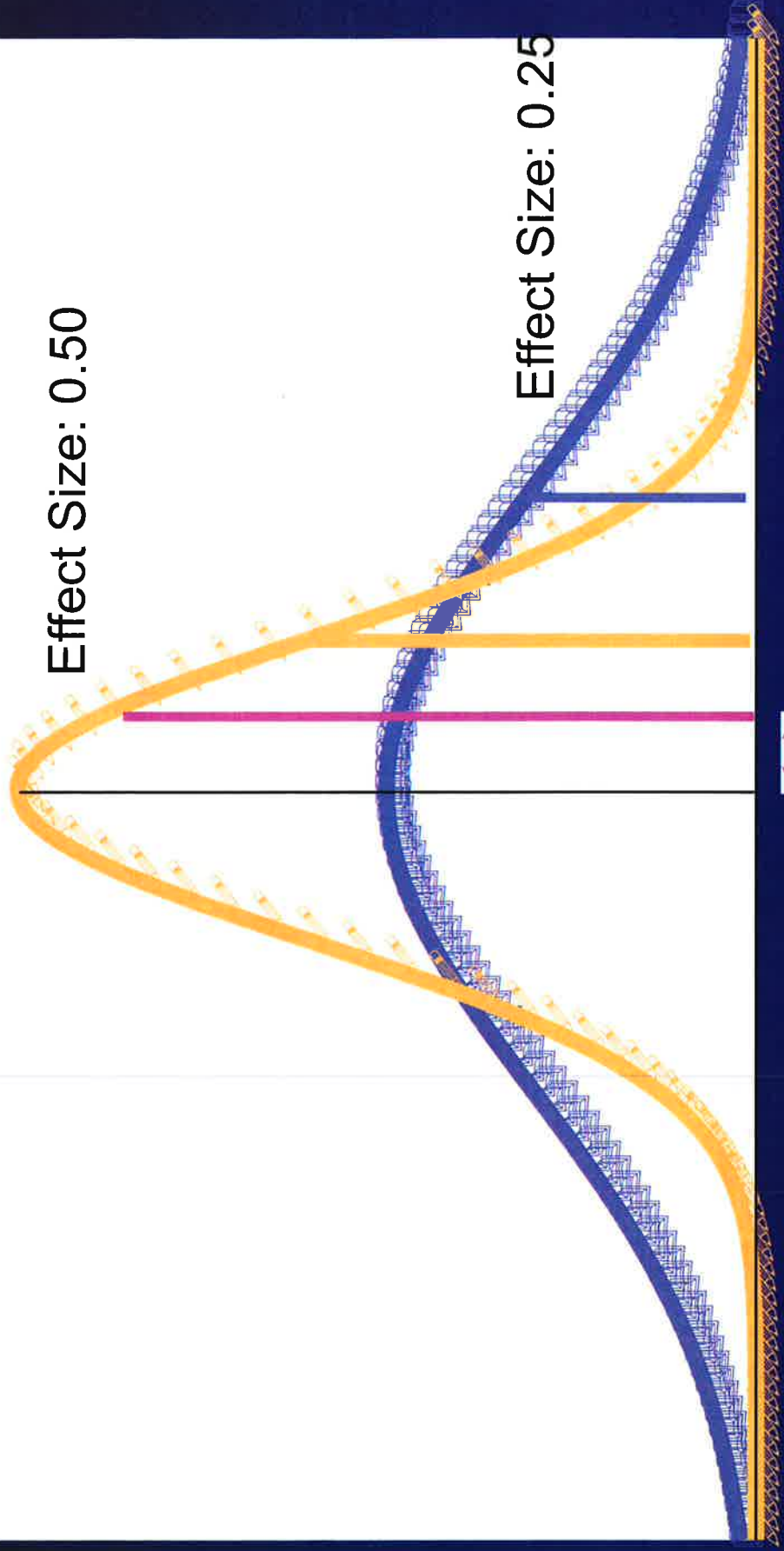
Methods of reporting	Pooled estimate and 95% CI	
	COPD meta-analysis	Dexamethasone meta-analysis
<i>Category 1: Methods derived from standard deviation units</i>		
i) SMD	SMD=0.73 (95% CI 0.49 to 0.96)	SMD=-0.79(95%CI -0.17 to -1.41)
ii) Conversion of SMD to dichotomies (to OR) (Suissa)	<sup>1</sup> OR=3.22 (95% CI 2.24 to 4.74)	<sup>2</sup> OR=0.21 (95% CI 0.04 to 0.76)
iii) Conversion of SMD to OR (Hasselbad/Hedges)	OR=3.74 (95% CI 2.42 to 5.68)	OR=0.23 (95% CI 0.08 to 0.74)
iv) Conversion of SMD to dichotomies (to NNT) (Suissa)	<sup>1</sup> NNT=3.6 (95% CI 2.7 to 5.4)	<sup>2</sup> NNT=6.7 (95% CI 5.3 to 22.7)
<i>Category 2: Not derived from SMD, not relying on MID</i>		
i) Conversion of all instruments to the most popular	<sup>3</sup> MD=0.76 (95% CI 0.64 to 0.88)	<sup>4</sup> MD=4 (95%CI 3.2 to 4.7)
iv) Ratio of means (ROM)	<sup>5</sup> Not applicable	ROM=0.87 (95% CI 0.78 to 0.98)
<i>Category 3: Not derived from SMD, depends on knowledge of MID</i>		
i) Minimally important difference (MID) units	<sup>6</sup> MD=1.75 (95% CI 1.37 to 2.13)	<sup>7</sup> Not applicable
ii) Conversion of MDs to dichotomies (to OR)	OR=3.52 (95% CI 2.60 to 4.76)	<sup>7</sup> Not applicable
iii) Conversion of MDs to dichotomies (to NNT)	NNT=4.3 (95% CI 3.34 to 6.0)	<sup>7</sup> Not applicable

# Effect size

- divide each effect by standard deviation
- ultimate result in SD units
- between "effect size" or SMD
- within standardized response mean

Effect Size: 0.50

Effect Size: 0.25



True MID

Methods of reporting	Pooled estimate and 95% CI	
	COPD meta-analysis	Dexamethasone meta-analysis
Category 1: <i>Methods derived from standard deviation units</i>		
i) SMD	SMD=0.73 (95% CI 0.49 to 0.96)	SMD=-0.79(95%CI -0.17 to -1.41)

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

## Making SD units interpretable

- convert back to natural units of most popular measure
- pooled effect 0.5 SD units
  - pooled SD on 100 point scale 20
  - effect in natural units 10
- vulnerability
  - effect size distortion by heterogeneity
  - SD of studies using most popular varies

# Avoiding heterogeneity problem

- convert all measures to units of most popular
  - most popular 0 - 100
  - alternative 0 to 7
- multiply all scores by 100/7
  - get weighted mean difference
  - alternative to standardized mean difference
- result in natural units

Pooled estimate and 95% CI

COPD meta-analysis      Dexamethasone meta-analysis

Category 2: Not derived from SMD, not relying on MID

i) Conversion of all instruments to the most popular

<sup>3</sup>MD=0.76 (95% CI 0.64 to 0.88)

<sup>4</sup>MD=4 (95%CI 3.2 to 4.7)

- vulnerable to differences in instrument properties
  - assumes clinicians understand natural units (MID)
  - vulnerable to interpretation problem

Outcomes	Illustrative Comparative Risks (95% CI)		Relative Effect (95% CI)	Number of participants (studies)	Quality of the Evidence <sup>1</sup> <b>(GRADE)</b>	Comments
	Assumed risk with Placebo	Corresponding risk with Dexamethasone				
Post-operative pain (B) Measured on a scale from 0, no pain, to 100, worst pain imaginable. <sup>4</sup>	The mean post-operative pain scores with placebo ranged from 43 to 54	The mean pain score in the intervention groups was on average 15 (3 to 27) lower		539 (5 studies)	Low <sup>2,3</sup>	Scores estimated based on an SMD of 0.79 (95% CI -1.41 to -0.17) <sup>4</sup>

1- Quality rated from 1 (very low quality) to 4 (high quality), 2- Evidence limited by heterogeneity between studies, 3- Evidence limited by imprecise data (small sample size or event rate), 4- A standard deviation of 0.5 represents a moderate difference between groups,

## MID units

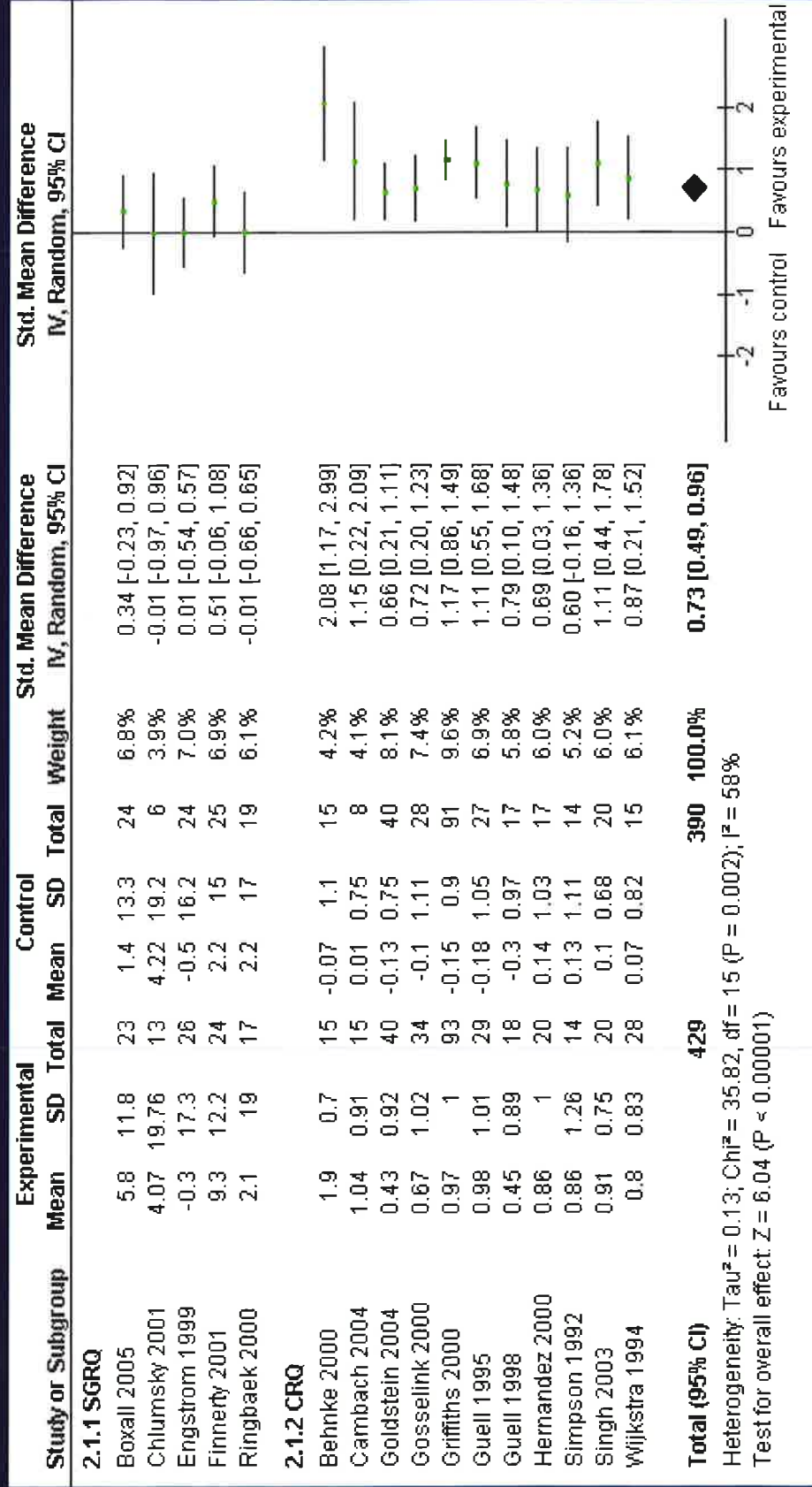
- Cochrane review of respiratory rehabilitation for COPD
- using 16 trials, we compared the existing method with the MID method
- trials employed two widely used disease-specific HRQL instruments
  - Chronic Respiratory Disease Questionnaire (CRQ)
  - St. Georges Respiratory Questionnaire (SGRQ)



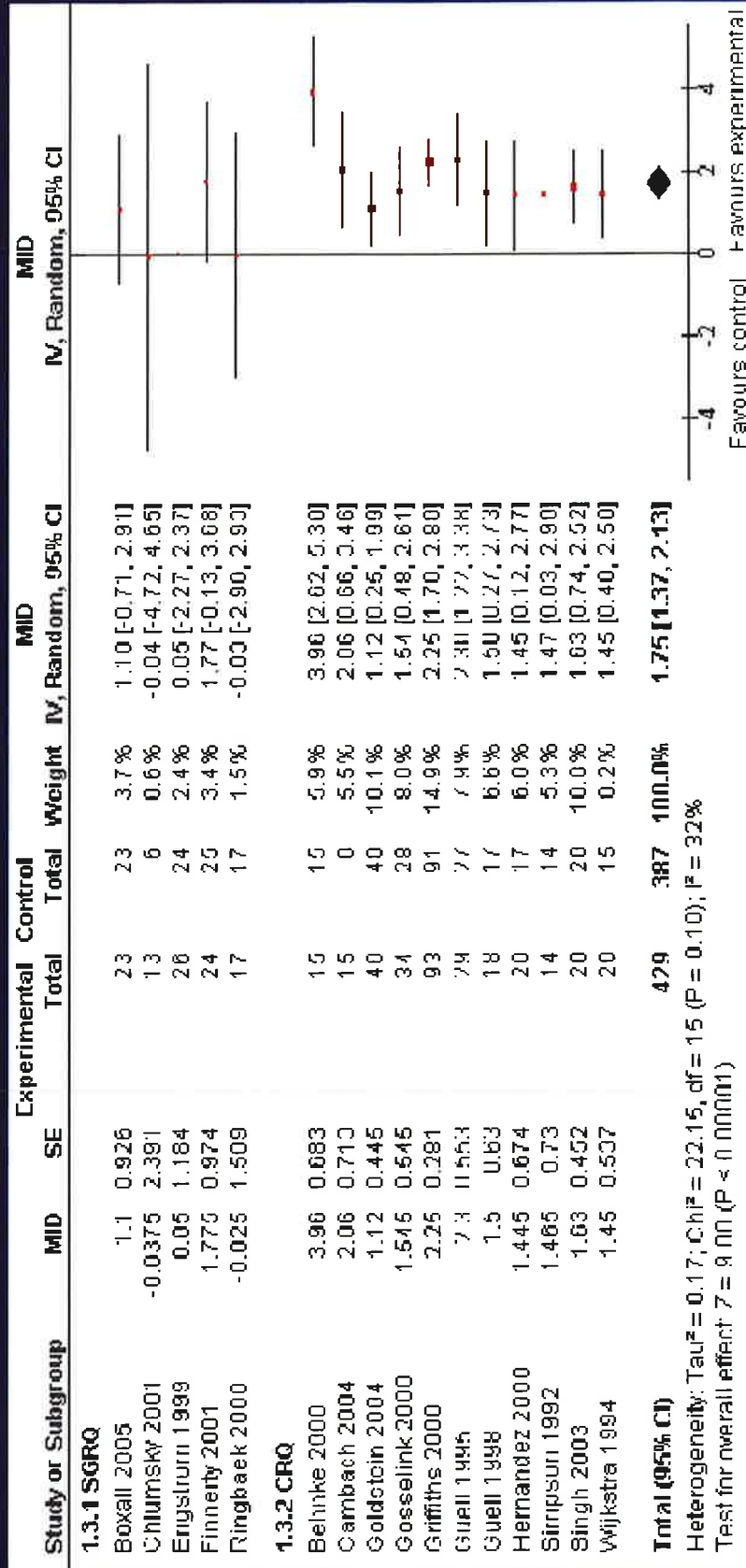
# Results

<b>CRQ</b>	<b>Mean Difference (95% CI)</b>
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)
<b>SGRQ</b>	
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



# Results - MID Units



# MID Units

- suggests a large effect:
  - the pooled estimate twice the smallest difference patients perceive as important
- MID approach
  - prevents introducing inconsistency depending on the SD
  - intuitive interpretation
  - vulnerable to all-or-nothing misinterpretation

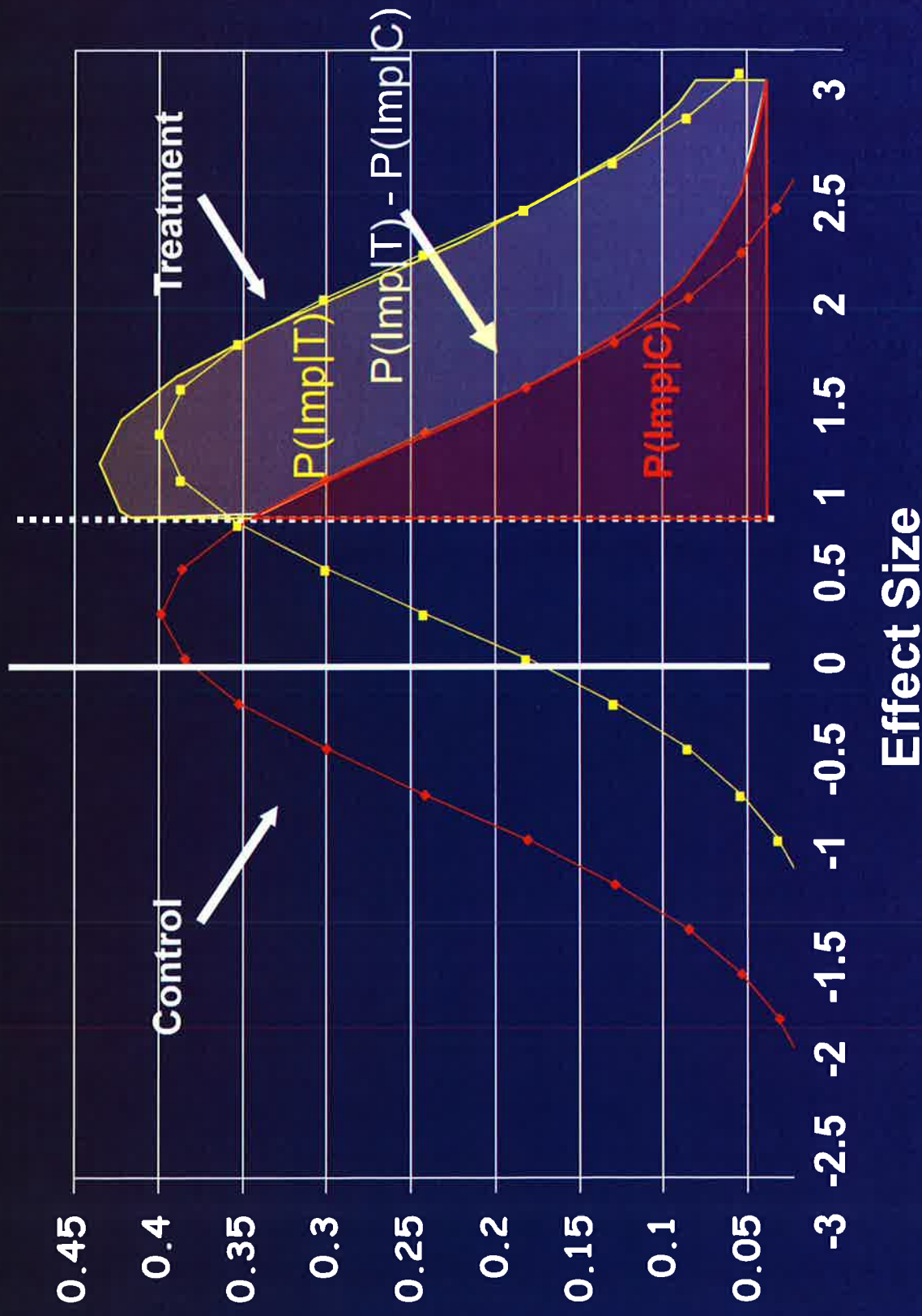
# Applying the MID

- 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

# No dichotomies in primary studies

Assume standard symmetrical distribution

Assume equal variance in intervention and control groups



# Approaches to dichotomizing

- Suissa: output risk in control and intervention group
- assumes normal distribution
  - but not equal variance
  - requires specification of control group risk
- specify control group risk and generate intervention group risk
- from event rates generate odds ratio

# Approaches to dichotomizing

- Hasselbad and Hedges
- assume logistic distribution
  - doesn't require control event risk
  - assumes normality, equal variance
- Cox and Snell
- Kramer - ROC, AUC
- all can generate OR, RD, NNT
- all vulnerable to heterogeneity, normality



# Furukawa Suissa approach to generate NNTs

Control group response rate	10	20	30	40	50	60	70	80	90
ES = 0.2	25	17	14	13	13	13	15	20	33
ES = 0.5	9	6	5	5	5	5	7	9	16
ES = 0.8	5	4	3	3	4	4	5	7	12
ES = 1.0	4	3	3	3	3	4	4	6	11

**Pooled estimate and 95% CI**

<b>Methods of reporting</b>	<b>COPD meta-analysis</b>	<b>Dexamethasone meta-analysis</b>
<i>Category 1: Methods derived from standard deviation units</i>		
i) SMD	SMD=0.73 (95% CI 0.49 to 0.96)	SMD=-0.79(95%CI -0.17 to -1.41)
ii) Conversion of SMD to dichotomies (to OR) (Suissa)	<sup>1</sup> OR=3.22 (95% CI 2.24 to 4.74)	<sup>2</sup> OR=0.21 (95% CI 0.04 to 0.76)
iii) Conversion of SMD to OR (Hasselbad/Hedges)	OR=3.74 (95% CI 2.42 to 5.68)	OR=0.23 (95% CI 0.08 to 0.74)
iv) Conversion of SMD to dichotomies (to NNT) (Suissa)	<sup>1</sup> NNT=3.6 (95% CI 2.7 to 5.4)	<sup>2</sup> NNT=6.7 (95% CI 5.3 to 22.7)

# Ratio of Means (RoM)

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

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

$$\bullet \text{Var}_{\ln(\text{RoM})} = \frac{\text{var}_{\text{exp}}}{(\text{mean}_{\text{exp}})^2} + \frac{\text{var}_{\text{control}}}{(\text{mean}_{\text{control}})^2}$$

# Avoiding heterogeneity problem:

## Ratio of means

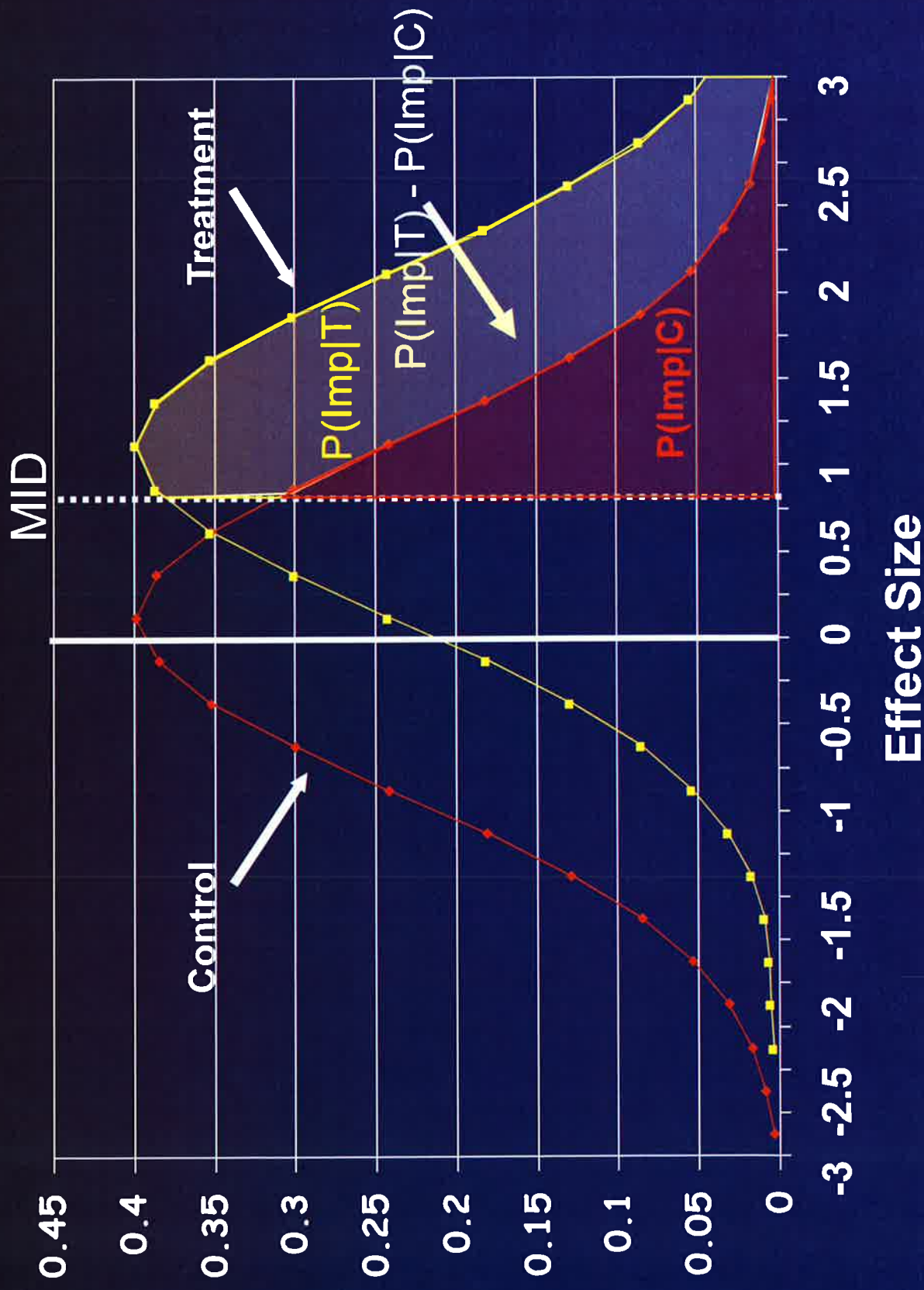
- analogous to relative risk
  - greater absolute difference with greater control risk
- requires natural zero
- cannot use if results reported as change

Methods of reporting	Pooled estimate and 95% CI	
	COPD meta-analysis	Dexamethasone meta-analysis
<i>Category 2: Not derived from SMD, not relying on MID</i>		
i) Conversion of all instruments to the most popular	<sup>3</sup> MD=0.76 (95% CI 0.64 to 0.88)	<sup>4</sup> MD=4 (95%CI 3.2 to 4.7)
iv) Ratio of means (ROM)	<sup>5</sup> Not applicable	ROM=-0.87 (95% CI 0.78 to 0.98)

## Avoiding heterogeneity problem

- back to MID
- effect in MID units
- dichotomy - risk difference

# No dichotomies in primary studies



Methods of reporting	Pooled estimate and 95% CI	
	COPD meta-analysis	Dexamethasone meta-analysis
<i>Category 3: Not derived from SMD, depends on knowledge of MID</i>		
i) Minimally important difference (MID) units	<sup>6</sup> MD=1.75 (95% CI 1.37 to 2.13)	<sup>7</sup> Not applicable
ii) Conversion of MDs to dichotomies (to OR)	OR=3.52 (95% CI 2.60 to 4.76)	<sup>7</sup> Not applicable
iii) Conversion of MDs to dichotomies (to NNT)	NNT=4.3 (95% CI 3.34 to 6.0)	<sup>7</sup> Not applicable

		Pooled estimate and 95% CI	
Methods of reporting	COPD meta-analysis	Dexamethasone meta-analysis	
<i>Category 1: Methods derived from standard deviation units</i>			
i) SMD	SMD=0.73 (95% CI 0.49 to 0.96)	SMD=-0.79(95%CI -0.17 to -1.41)	
ii) Conversion of SMD to dichotomies (to OR) (Suissa)	<sup>1</sup> OR=3.22 (95% CI 2.24 to 4.74)	<sup>2</sup> OR=0.21 (95% CI 0.04 to 0.76)	
iii) Conversion of SMD to OR (Hasselbad/Hedges)	OR=3.74 (95% CI 2.42 to 5.68)	OR=0.23 (95% CI 0.08 to 0.74)	
iv) Conversion of SMD to dichotomies (to NNT) (Suissa)	<sup>1</sup> NNT=3.6 (95% CI 2.7 to 5.4)	<sup>2</sup> NNT=6.7 (95% CI 5.3 to 22.7)	
<i>Category 2: Not derived from SMD, not relying on MID</i>			
i) Conversion of all instruments to the most popular	<sup>3</sup> MD=0.76 (95% CI 0.64 to 0.88)	<sup>4</sup> MD=4 (95%CI 3.2 to 4.7)	
iv) Ratio of means (ROM)	<sup>5</sup> Not applicable	ROM=0.87 (95% CI 0.78 to 0.98)	
<i>Category 3: Not derived from SMD, depends on knowledge of MID</i>			
i) Minimally important difference (MID) units	<sup>6</sup> MD=1.75 (95% CI 1.37 to 2.13)	<sup>7</sup> Not applicable	
ii) Conversion of MDs to dichotomies (to OR)	OR=3.52 (95% CI 2.60 to 4.76)	<sup>7</sup> Not applicable	
iii) Conversion of MDs to dichotomies (to NNT)	NNT=4.3 (95% CI 3.34 to 6.0)	<sup>7</sup> Not applicable	



# 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