Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

Minimal important changes in standard deviation units are highly variable and no universally applicable value can be determined

Yasushi et al. Journal of Clinical Epidemiology 2022;145:92-100

Prapaporn Pornsuriyasak, M.D.

Background

- Health status (HS) questionnaires become popular for measuring the effects of treatments for chronic diseases
- Defining the extent to which participants exposed to the intervention have experienced important change in HS
- Minimal important change (MIC): smallest difference in score of the domain of interest patients perceive as beneficial/important, absence of troublesome side-effects and excessive cost
    - MIC: longitudinal within-person changes in scores
    - MID: cross-sectional between-person differences

Anchor-based approach

PROM → Independent standard

- Assess which changes (improve or deteriorate) on the measurement instrument (PROM) correspond with minimal important change defined on the anchor
- Based on retrospective judgment of change
- Anchor should substantially correlates with health status instrument

Example of anchor

Transition questions

- require patients to remember prior health state and compare it to how they are feeling currently

- if correlation between global transition rating of dyspnea (overall, how much better or worse is your dyspnea in daily activities) and the change in score on a dyspnea questionnaire is over 0.5, it supports the validity of both the transition rating and the target questionnaire

Guyatt GH. Ann Intern Med 1993;70:225–30

Anchor-based method

- **‘within-patients’ score change**

  standard dyspnea score before and after treatment

- **‘between-patients’ score change**

  quality of life scores between treatment and control group

- **Sensitivity and specificity based approach**

  score that best separates patients who reported improvement and those who did not (Receiver Operating Characteristic curve)

MIC anchor-based method

- **Mean change method**
  - takes the mean change score on the measurement instrument for subcategory of patients who are minimally importantly changed according to the anchor

- **ROC method**
  - The MIC value corresponds to the optimal ROC cutoff point
  - Point on the ROC curve closest to the upper left corner of figure = MIC

    "least amount of misclassification"

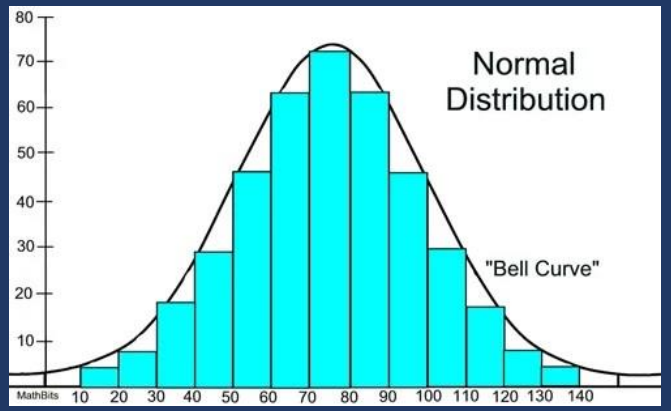
    : smallest sum of % false positives and false negatives

Disadvantage of anchor-based method

- The response susceptible to recall bias (response shift) due to retrospective judgement of change
- Patients' ratings on an anchor are more highly correlated with the follow-up score than with the baseline score
  - Unable to recall well their baseline score if long follow-up
- "Large confidence interval around the $MIC_{anchor\text{-}based}$"
- Selection of the cut-off for little better improvement on the anchor
- Standardization problem
  - Studies use different anchors with different wordings and different response options
- Anchor should correlate with PROM

  (correlation coefficient of at least 0.5)

Distribution-based approach



- Uses statistical parameters in estimating MIC
- Based on distributional characteristics of the sample
- Express the observed change to some forms of variation to obtain a standardized metric
  1) effect size (ES)
  2) standardized response means
  3) standard error of measurement (SEM)

Disadvantage of distribution-based method

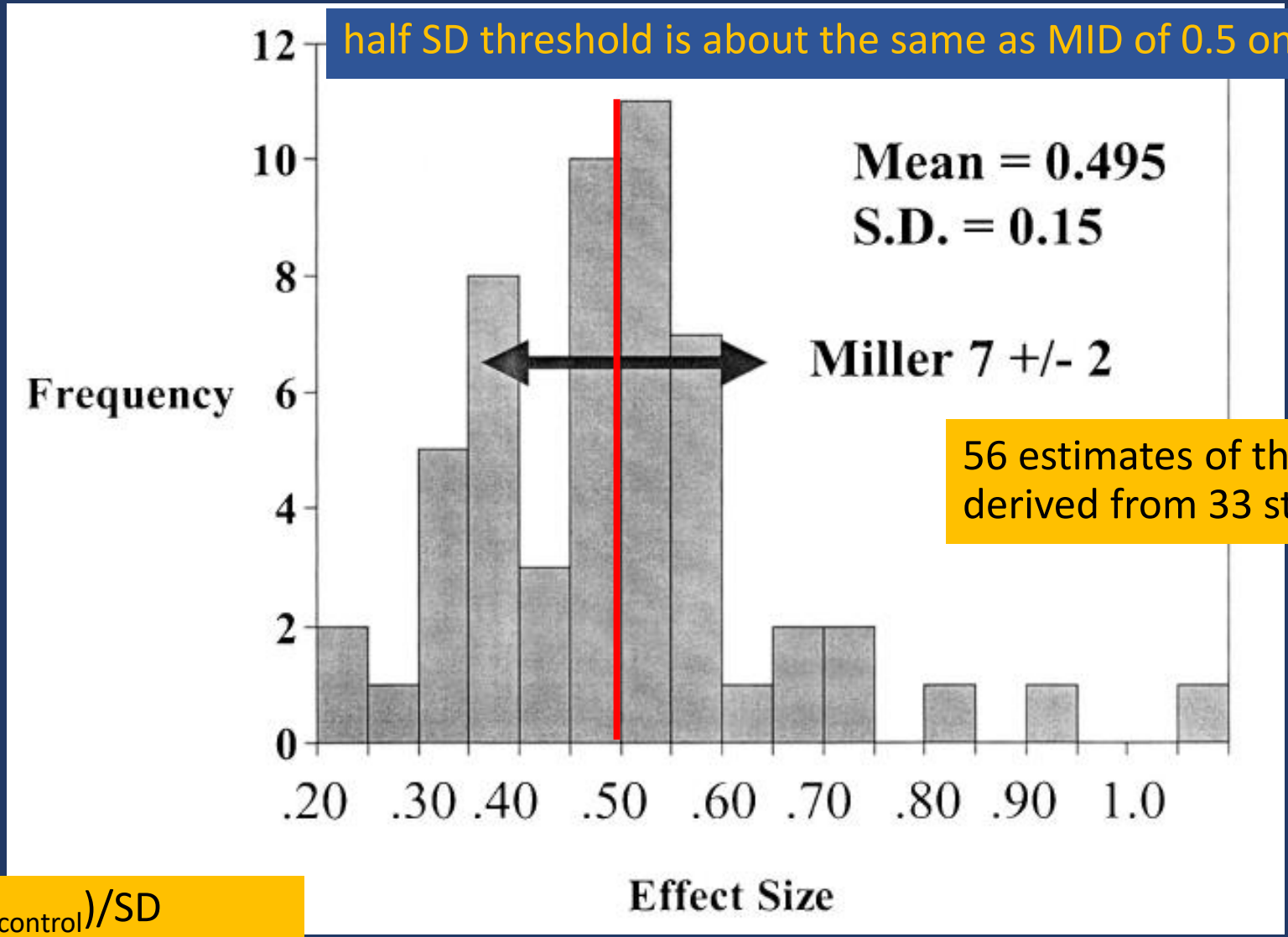
- Lack information whether the observed changes are minimally important

Distribution-based method

- MCID based on standard error of measurement (SEM)
  - how much measured test scores are spread around a “true” score
  - 1 **x SEM** is used as a benchmark for a “true” change
- MCID based on SD
  - measure of the amount of variation or dispersion of a set of values
  - 0.5 **x SD** of the baseline score often used to define MCID for patient-reported outcomes
- MCID based on effect size (ES)
  - ratio of change from baseline and SD of the baseline values
  - standardized measure of change

Why is it always 0.5 of SD at baseline?

half SD threshold is about the same as MID of 0.5 on a 7-point scale



56 estimates of the minimal difference derived from 33 studies

ES= ($Mean_{Rx}$- $Mean_{control}$)/SD

Norman GR. Med Care 2003

Half SD criterion

- When patients with chronic diseases are asked to identify minimal change, the estimates fall very close to half SD
- Not an arbitrary statistical criterion but is empirically derived based on psychological theory
- We should attempt to identify factors that will result in systematic and substantial departure from 0.5 SD norm
- It would be appropriate to consider this as an approximate rule of thumb in the absence of more specific information

Norman GR. Med Care 2003

Introduction of this study

- Using distribution-based MICs when anchor-based MICs are not always available
- The relationships between the 2 methods are debated.
- Anchor-based MICs agree well with half SD of PROM
- Previous studies
  - not take into account the methodological quality of primary studies or characteristics of PROMs
  - Unclear which SD of PROM should be used in the distribution-based method:
    - Baseline SD
    - Endpoint SD
    - Change SD

Objective of this study

- To describe the anchor-based MIC estimates in SD units

- To examine if the robustness of distribution-bases MIC estimates depends on
  - the SD to be used OR
  - the methodological quality of the anchor-based estimates

Methods

- The Minimal Important Difference Inventory Dataset
- Includes 5,324 MIC estimates from 585 studies
- Using MEDLINE, EMBASE, CINAHL and PsycINFO for studies published between 1989 and October 2018, and additional relevant citations from the PROQOLID internal library
- Dataset:
  - anchor-based MICs for PROMs in adolescents (≥13 to 17) or adults (≥18)
    - health-related quality of life
    - functional ability
    - symptom severity
    - psychological distress and well- being

irrespective of participants' condition or disease, type of intervention, or nature of anchors

Methods

- Excludes
  - systematic reviews
  - conference abstracts
  - studies that authors explicitly targeted moderate or large important difference of MIC
  - combined anchor- and distribution-based approach

Eligible criteria

- Mean change method
- MIC: absolute mean change in PROM scores over time within the subgroup of participants reporting slightly improved or deteriorated using global rating of change as the anchor
- Excluded
  - studies or MIC estimates but unavailable SD of PROMs
  - did not report standard error, confidence interval or interquartile range with the number of participants

Data extraction

- study country
- population demographics
- types of PROM
- interventions administered
- anchor details (type, constructs, range of options, values and threshold selected to represent “small but important change”)
- associated measure of variability and direction
- number of patients informing MIC estimate
- credibility ratings of MIC estimates
- means and SDs of PROMs

Paired reviewers independently conducted data extraction, resolving disagreements by discussion with input from a third reviewer

Types of PROM

1) Generic (health profiles and utility measures)

2) Specific (disease/condition-specific, symptom-specific, function-specific, population-specific)

Classification and calculation of SD

- SD of baseline scores (baseline SD)
- SD of endpoint scores (endpoint SD)
- SD of change from baseline scores (change SD)

SE, CI, IQR, or range

$$\mathrm{SD} = \mathrm{SE} \times \sqrt{N}.$$

$$\mathrm{SD} = \sqrt{N} \times (\text{upper limit of CI} - \text{lower limit of CI})/3.92.$$

$$\mathrm{SD} = \mathrm{IQR}/1.35[11].$$

$$\mathrm{SD} = \mathrm{Range}/2[11].$$

Calculation of SD

- If authors reported SDs separately in subgroups of the participants (e.g., improved, no change, or deteriorated)
- Overall SD

$$SD = \sqrt{\frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2 + \frac{N_1 N_2}{(N_1 + N_2)\left(M_1^2 + M_2^2 - 2M_1 M_2\right)}}{N_1 + N_2 - 1}}$$

Cochrane Handbook for Systematic Reviews of Interventions version 6.0: Cochrane(editors); 2019

Outcomes

- Primary outcome: distribution of MICs in SD units

$$MIC\ in\ SD\ units = \frac{MIC}{SD}$$

Sensitivity analysis by excluding MIC estimates with low credibility

Credibility instrument for judging the trustworthiness of MID

Signalling question	Response options	
	High credibility	Low credibility
Core criteria		
Is the patient or necessary proxy responding directly to both the PROM and the anchor?	Yes	No/impossible to tell
Is the anchor easily understandable and relevant for patients or necessary proxy?		
Has the anchor shown good correlation with the PROM?		
Is the MID precise?	Definitely yes/to a great extent	Definitely no/not so much/impossible to tell
Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference?		

*Threshold of correlation coefficient of at least 0.5

Results

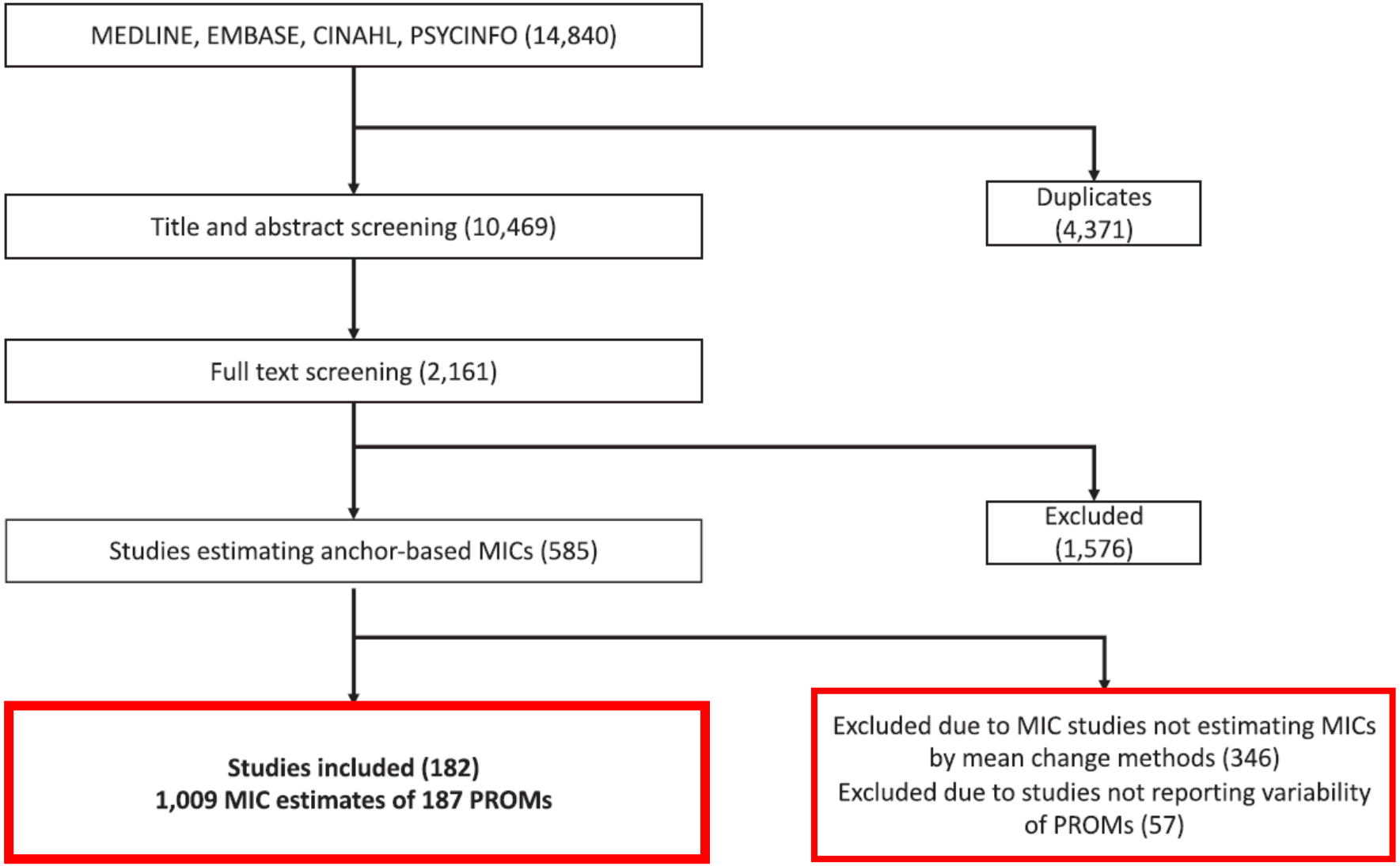


Fig. 1. Flow diagram of study eligibility. PROM, patient-reported outcome measure; MIC, minimal important change.

Characteristics of included studies and PROMs

Characteristics	
Study-level ($n = 182$)	
Year of publication	2013 (2009 to 2016)
Number of total participants	191 (100 to 306)
Age groups	
Children or adolescents	5 (3)
Adults or elderly	169 (93)
Not reported	8 (4)
Interventions	
Pharmacological intervention	31 (17)
Surgical or invasive intervention	40 (22)
Rehabilitation	17 (9)
Mixture	50 (27)
Others	20 (11)
Not available	24 (13)
PROM-level ($n = 187$)	
Types of PROMs[a]	
Generic, Health profile	7 (4)
Generic, Utility measure	7 (4)
Specific, Disease/condition	140 (74)
Specific, Symptom	26 (14)
Specific, Function	6 (3)
Specific, Population	1 (1)
Number of MIC estimates per PROM	2 (1 to 4)

Characteristics of MIC estimates

Credibility assessment of MIC estimates[a]	
Is the patient or necessary proxy responding directly to both the PROM and the anchor?	
Yes	984 (98)
No / impossible to tell	25 (2)
Is the anchor easily understandable and relevant for patients or necessary proxy?	
Definitely yes / to a great extent	974 (97)
Definitely no / not so much / impossible to tell	35 (3)
Has the anchor shown good correlation with the PROM?	
Definitely yes / to a great extent	86 (9)
Definitely no / not so much / impossible to tell	332 (33)
Not reported	591 (59)
Is the MIC precise?	
Definitely yes / to a great extent	100 (10)
Definitely no / not so much / impossible to tell	909 (90)
Does the threshold or difference between groups on the anchor used to estimate the MIC reflect a small but important differe	
Definitely yes / to a great extent	671 (67)
Definitely no / not so much / impossible to tell	3385 (33)

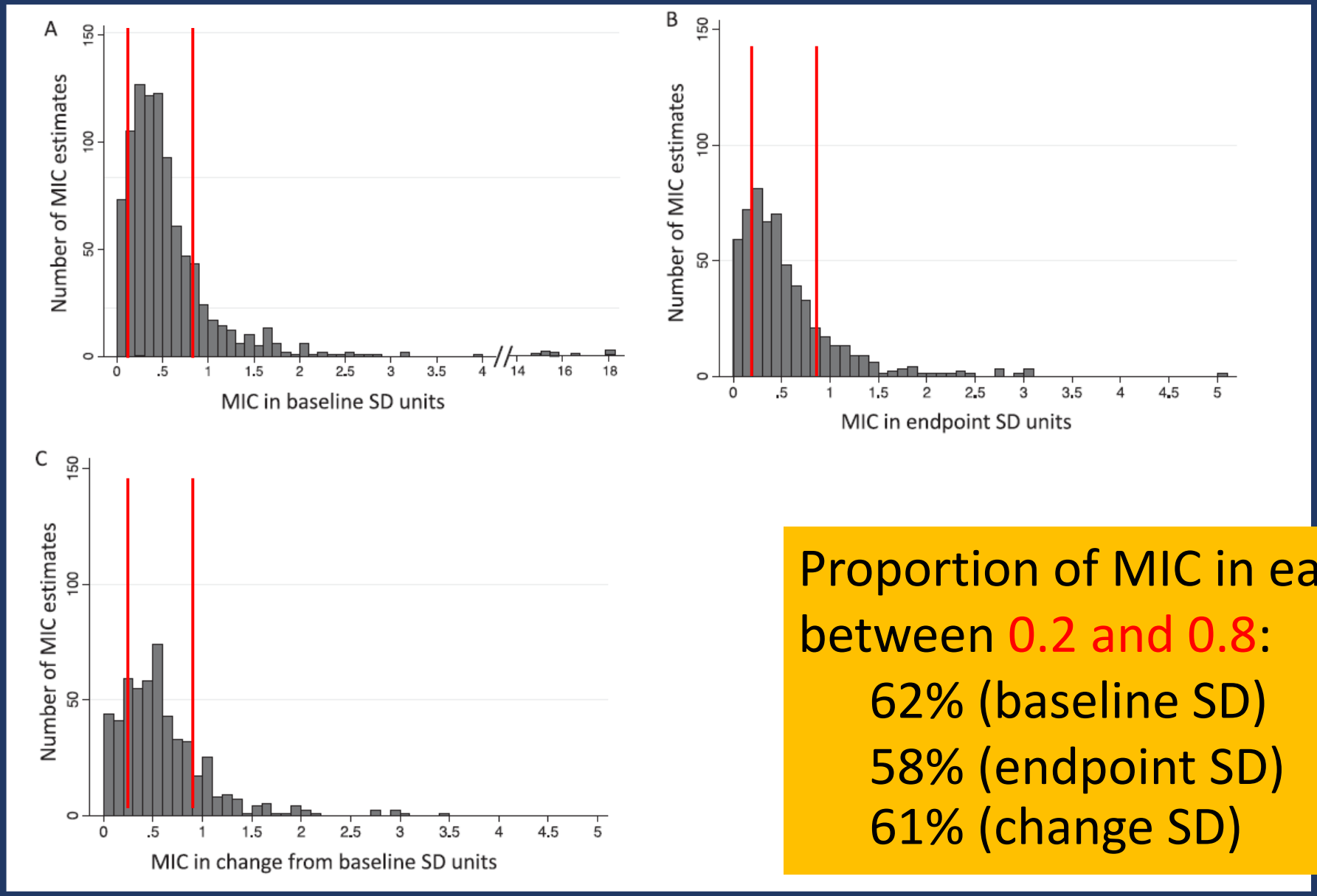
Sample size < 100

Summary statistics of MIC estimates in each SD unit

Type sf SD units	Mean (SD)	95%CI	Median	IQR	Range
Primary analysis					
93% Baseline ($n = 931$)	0.67 (1.52)	0.58 to 0.77	0.43	0.25 to 0.69	0 to 18.11
58% Endpoint ($n = 582$)	0.55 (0.52)	0.50 to 0.59	0.42	0.22 to 0.70	0 to 5.02
53% Change ($n = 530$)	0.59 (0.47)	0.55 to 0.63	0.51	0.28 to 0.78	0 to 3.42
Sensitivity analysis[a]					
Baseline ($n = 320$)	0.42 (0.37)	0.38 to 0.46	0.35	0.18 to 0.52	0 to 2.77
Endpoint ($n = 190$)	0.43 (0.32)	0.38 to 0.47	0.37	0.19 to 0.60	0 to 2.32
Change ($n = 155$)	0.50 (0.32)	0.45 to 0.55	0.46	0.26 to 0.69	0 to 1.72

[a] Sensitivity analysis excluding minimal important change estimates with less credible methodology

Distribution of MIC estimates



Proportion of MIC in each SD unit between 0.2 and 0.8:

62% (baseline SD)

58% (endpoint SD)

61% (change SD)

Discussion

- Neither SD (baseline score, endpoint score, change from baseline score) yielded universally applicable and widely generalizable value of MIC in SD unit
- Broad distribution of MIC in SD unit was not narrowed down by sensitivity analysis that excluded MICs estimates with less credible methodology
- No constant or consistent value for MIC in SD unit universally used for PROMs (contrary to 0.5 SD for MIC reported by Norman et al)

Discussion

- Variation depended on
  - Participants' characteristics at baseline (disease severity)
    - heterogeneity of sample in different studies yielded different SDs and different MIC in SD unit despite the same absolute MICs
    - large variation in MIC estimates by the same method across studies and across different methods within studies
  - Different types of SD
- Not clear whether the variation was due to differences between populations or to methodological problems of MIC approach

Discussion

- Approximately 60% of MICs in SD units were within the range from 0.2 to 0.8
- If MIC in SD unit did not fall in this range, we may rely on distribution-based approach to interpret the change in PROM
- Estimates of MICs are substantially larger for subgroups of patients with high baseline values
- If anchor and PROM are not correlated well, MIC can be zero or grossly erroneous

Discussion

Anchor-based approach to estimate precise MIC requires

- Close correlation between anchor and PROMs
- Sample size calculation
- Inferences about the magnitude of treatment effect would differ at the extremes of CI around the MIC estimate when the CI is wide

Limitations of the study

- Sensitivity analysis excluding studies with less methodological credibility to estimate the MIC actually included studies which did not report correlations or studies with small sample sizes

Has the anchor shown good correlation with the PROM?	
Definitely yes / to a great extent	86 (9)
Definitely no / not so much / impossible to tell	332 (33)
Not reported	591 (59)

- Not assess other distribution approach: standard error of the mean (SEM)

Conclusion

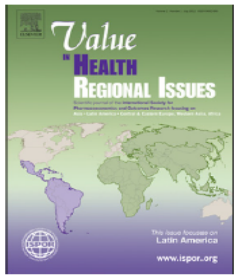
- Converting anchor-based MICs to SD units resulted in
  - highly variable estimates
  - made it difficult to determine a universal value of MICs in SD units for the distribution method

VALUE IN HEALTH REGIONAL ISSUES 12C (2017) 57–62

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/vhri



Minimal Clinically Important Differences (MCIDs) of the Thai Version of the Leicester Cough Questionnaire for Subacute and Chronic Cough

Prapaporn Pornsuriyasak, MD[1,], Poungrat Thungtitigul, MD[1], Theerasuk Kawamatawong, MD[1], Surinder S. Birring, MD[2], Tipaporn Pongmesa, PhD[3]*

[1]*Faculty of Medicine, Division of Pulmonary and Critical Care Medicine, Department of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand;* [2]*Division of Asthma, Allergy and Lung Biology, King's College London, London, UK;* [3]*Faculty of Pharmacy, Silpakorn University, Nakhon Prathom, Thailand*

- Chronic cough > 3 weeks
- PROM: Leicester Cough Questionnaire (Thai-translated): 19 items in 3 domains, 7-point response (3-21)

  the lower score = the greater impairment
- Baseline, f/u 4 weeks, 4-16 weeks, > 16 weeks after treatment
- Anchor: Global Rating of Change (GRC) scale of cough impact on QoL (7-point scale response)
  - unchanged (1/0/-1)
  - small change(3, 2, -2, -3)
  - moderate change (5, 4, -4, -5)
  - large change(7, 6, -6, -7)

minimal but clinically important changes (MCID)

Analysis

- Pearson correlation coefficients/Spearman rank correlation coefficients to determine relationship between baseline and follow-up score changes in LCQ and GRC score
- Anchor-based method
  - MCID: changes of score in LCQ score between baseline and follow-up were averaged within the categories of the GRC (+2, +3 = small improvement)
- Distribution-based method
  - MCID calculated using half of SD, third of SD

  (derived from the LCQ-T scores at baseline, at follow-up, and from the baseline to follow-up score change)

Results: MCID of the LCQ

Domain	Anchor-based method (N = 59)	Distribution-based method (N = 107)							
		SEM (Cronbach α)	Baseline		Follow-up		Change		Mean	
			1/2 SD	1/3 SD	1/2 SD	1/3 SD	1/2 SD	1/3 SD	1/2 SD	1/3 SD
Total	1.1	0.8	2.0	1.3	1.8	1.2	2.0	1.4	1.9	1.3
Physical	0.4	0.3	0.6	0.4	0.5	0.3	0.6	0.4	0.6	0.4
Psychological	0.4	0.3	0.7	0.5	0.7	0.4	0.8	0.5	0.7	0.5
Social	0.4	0.3	0.8	0.5	0.7	0.5	0.8	0.5	0.8	0.5

- GRC score significantly correlated with
  - LCQ score pre (r = 0.01 , p = NS)
  - LCQ follow-up (r = 0.5, p <0.001)
  - LCQ changes (r = 0.3, p= 0.01)
- LCQ score pre vs LCQ follow-up (r = 0.43, p<0.001)

Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

Thank you