



# Clinically relevant differences in COPD health status: systematic review and triangulation

Harma Alma<sup>1,2</sup>, Corina de Jong<sup>1,2</sup>, Ioanna Tsiligianni<sup>1,2,3</sup>,  
Robbert Sanderma<sup>4,5</sup>, Janwillem Kocks<sup>1,2</sup> and Thys van der Molen<sup>1,2</sup>

**Affiliations:** <sup>1</sup>Dept of General Practice and Elderly Care Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. <sup>2</sup>Groningen Research Institute for Asthma and COPD (GRIAC), University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. <sup>3</sup>Clinic of Social and Family Medicine, Medical School, University of Crete, Heraklion, Greece. <sup>4</sup>Dept of Health Psychology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. <sup>5</sup>Dept of Psychology, Health and Technology, University of Twente, Enschede, The Netherlands.

**Correspondence:** Harma Alma, University Medical Center Groningen, Dept of General Practice and Elderly Care Medicine, University of Groningen, HPC FA21; Postbox 196, NL-9700 AD Groningen, The Netherlands. E-mail: h.j.alma@umcg.nl

 @ERSpublications

**Reviewed and triangulated thresholds for minimal clinically relevant improvements are  $-2.54$  (CAT),  $-0.43$  (CCQ) and  $-7.43$  (SGRQ)** <http://ow.ly/1y4v30l1yVI>

**Cite this article as:** Alma H, de Jong C, Tsiligianni I, *et al.* Clinically relevant differences in COPD health status: systematic review and triangulation. *Eur Respir J* 2018; 52: 1800412 [<https://doi.org/10.1183/13993003.00412-2018>].

**ABSTRACT** The minimal clinically important difference (MCID) quantifies when measured differences can be considered clinically relevant. This study aims to review and triangulate MCIDs of chronic obstructive pulmonary disease (COPD) health status tools.

A systematic search in PubMed, EMBASE and Cochrane Library was conducted (Prospero #CRD42015023221). Study details, patient characteristics, MCID methodology and estimates were assessed and extracted by two authors. A triangulated mean was obtained for each tool's MCID, with two-thirds weighting for anchor-based and one-third for distribution-based results. This was then multiplied by a weighted factor based upon the study size and quality rating.

Overall, 785 records were reviewed of which 21 studies were included for analysis. MCIDs of 12 tools were presented. General quality and risk of bias were average to good. Triangulated MCIDs for the COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ) and St. George's Respiratory Questionnaire (SGRQ) were  $-2.54$ ,  $-0.43$  and  $-7.43$  for improvement. Too few and/or too diverse studies were present to triangulate MCIDs of other tools.

Evidence for the MCID of the CAT and CCQ was strong and triangulation was valid. Currently used MCIDs in clinical practice for the SGRQ (4) and Chronic Respiratory Questionnaire (0.5) did not match the reviewed content, for which the MCIDs were much higher. Using too low MCIDs may lead to an overestimation of the interpretation of treatment effects. MCIDs for deterioration were scarce, which highlights the need for more research.

---

This article has supplementary material available from [erj.ersjournals.com](http://erj.ersjournals.com)

Received: Feb 27 2018 | Accepted after revision: July 18 2018

Copyright ©ERS 2018

## Introduction

Health status measurements and thresholds for clinically important change are frequently used as obligatory endpoints in medical trials, scientific research and clinical practice to evaluate the effects of an intervention [1–6]. The minimal clinically important difference (MCID) is a pivotal parameter that quantifies this threshold for clinically relevant change. It has been defined by JAESCHKE *et al.* [7] as “*the smallest difference in score in the domain of interest, which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive costs, a change in the patient’s management*”. The MCID is often used as a cut-off value to compare the percentage of patients achieving the level of clinically relevant change after intervention or pharmacotherapy in comparison to a control group [5, 6, 8]. It is also used to define sample size and to evaluate change in clinical practice. Setting an MCID too high could lead to underestimation of the interpretation of treatment effects; defining an MCID too low could result in overestimation of this interpretation.

The measurement of health status and its MCIDs is of particular interest in chronic obstructive pulmonary disease (COPD). Physiological measures like spirometry are often of interest to the physician but of limited importance to patients because these outcomes do not correlate well with their quality of life (QoL) [3, 9–16]. Patients can have similar spirometry or blood tests, but may experience very different levels of QoL and health status. QoL is “*the degree of satisfaction or dissatisfaction with various aspects of life that may be important to the individual*” [17]. Health status is considered “*the impact of health on a person’s ability to perform and derive fulfillment from the activities of daily life*” [10, 18]. It is a standardised way of measuring the concepts of health-related quality of life (HRQoL), functional status and mental well-being [2, 10]. HRQoL and health status questionnaires, often patient-reported outcomes, have received much attention in the last few years, resulting in their inclusion in Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines for the classification of patient risk groups to guide treatment [12].

Many general and disease-specific HRQoL and health status tools exist, with varying designs ranging from single items to complex multi-domain questionnaires [1, 2, 9–11, 19–22]. It is important that an instrument has strong measurement properties, including responsiveness, interpretability and good signal-to-noise ratio [1, 9, 10, 22, 23]. The MCID is an important parameter within these categories. Many authors have discussed the theory and methods to determine an instrument’s MCID [2, 4, 24–45]. These are generally divided into anchor-, distribution- and opinion-based approaches. Each method has its pros and cons. To date, there is no gold standard in defining an instrument’s MCID [2, 44]. Hence, many different practices occur, some better than others. It is recommended that both anchor- and distribution-based methods are used, combined with evidence from clinical trial data and qualitative approaches, with a systematic review or expert panel to aim for triangulation [33, 45].

Given the importance of MCIDs for research and clinical practice, there has been an increase in studies investigating the MCID of HRQoL and health status tools in COPD. Various authors have provided overviews of frequently used instruments in COPD, incidentally reporting the value or a short description of its MCID [3, 8, 10, 16, 28, 29, 38, 46–52]. In addition, some publications have presented a description of the evidence for the MCID of specific COPD outcomes, including the St. George’s Respiratory Questionnaire (SGRQ) [53], exacerbations [39], the transition dyspnoea index (TDI) [54], forced expiratory volume in 1 s (FEV<sub>1</sub>) [55], Clinical COPD Questionnaire (CCQ) [56] and COPD Assessment Test (CAT) [57]. Most studies have not recently been updated, and none of these studies attempted to evaluate the quality of the MCID methodology or aim for triangulation. Outside the field of COPD, systematic reviews have emerged that summarise, quantify and make a quality assessment of the MCIDs of patient-reported outcomes and functional status tests [31, 58–64]. Our study is the first to do so within the field of COPD. We aimed to systematically review the available evidence for the MCID of various HRQoL and health status tools used in COPD practice, to assess the quality of their methodology and to attempt to triangulate the results as a form of meta-analysis.

## Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement served as a guideline for this systematic review [65]. The study protocol was prepared and published *via* PROSPERO (#CRD42015023221) [66]. PubMed, EMBASE and the Cochrane Library were searched to identify relevant, original full-text articles on the measurement of MCIDs of HRQoL and health status tools in a COPD population. The detailed search strategy and study methods are available in the supplementary material and supplementary table S1. HRQoL and health status instruments were considered eligible when they captured more than one domain of the concept’s physical, psychological and social functioning [1, 2]. In COPD patients, this would include concepts such as breathlessness, fatigue, cough, sputum production,

physical functioning, social functioning, mental well-being and exacerbations [16]. The term health status will be used for future reference in this review.

Studies were screened, selected and analysed by two authors (H.A. and C.D.J.) independently according to the pre-defined structured protocol. Disagreement between the authors was solved by consensus or involvement of a third author (I.T., T.V.D.M. or R.S.). The quality and risk of bias of the included studies was assessed by means of a self-composed evaluation tool with 31 selected items obtained from various sources, because there was no specific tool available for evaluating studies that measure an instrument's MCID (supplementary table S2). Included concepts that were selected were from the Cochrane Risk of Bias tool [67], the COSMIN checklist [68], systematic reviews by BOHANNON and colleagues [58, 59] and recommendations in the MCID literature [2, 33, 45, 69]. Individual items were scored and an overall total score with a maximum of 62 could be obtained. Five categories were defined for the overall quality stratification, which was required for triangulation procedures. The overall risk of bias and quality assessment was given a star rating between one and five, calculated from the summed scores as follows: 0–12, one star; 13–25, two stars; 26–37, three stars; 38–49, four stars; and 50–62, five stars.

Study properties and MCID data results were extracted by two authors (H.A. and C.D.J.) using a standardised form. Results were categorised per health status tool. Primary outcome measures were the quality assessment of the MCIDs for health status tools in COPD; an overview of the study's MCID methods and estimates; and triangulation of the MCIDs where multiple studies per instrument existed. Triangulation was executed by first determining an anchor-based and distribution-based mean MCID per included study. The anchor-based result received a weight of two-thirds, and the distribution-based method received a weight of one-third. The results were multiplied by a weighted factor depending on the study size (n) and the quality star rating (1–5 stars). An overall triangulated mean MCID was calculated per health status tool.

## Results

The initial search in PubMed, EMBASE and the Cochrane Library resulted in 668 unique studies (figure 1). Screening the references provided an additional 117 titles, resulting in a total of 785 unique records. After screening all titles and abstracts, 78 papers were assessed for eligibility. A full-text analysis led to the removal of 57 papers (supplementary table S3), leaving 21 records for inclusion (table 1) [7, 70–89]. The initial level of agreement between the authors (H.A. and C.D.J.) was 89.2% for study selection. Cohen's kappa for the quality and risk of bias assessment was 0.42. The included studies discussed the tools CAT, CCQ, Short-Form Chronic Respiratory Questionnaire (SF-CRQ), eDiary, EuroQol Five Dimensions index (EQ-5D), Feeling Thermometer (FT), SGRQ, Short-Form Six-Dimension index (SF-6D), Short-Form 36 survey (SF-36), Quality of Life for Respiratory Illness Questionnaire (QoLRIQ) and Visual Simplified Respiratory Questionnaire (VSRQ). A description of the included tools is available in the supplementary material. Full patient characteristics, inclusion and exclusion criteria, and health status scores are available in supplementary tables S4 and S5.

Most studies scored well on the overall quality of the general study methodology with the exception of the items on systematic enrolment, similar test conditions pre- and post-measurement, the description of floor and ceiling effects, and unclear/high selective outcome reporting bias (table 2). Four studies scored lower on the description of general study methodology [7, 72, 80, 87]. Regarding the quality of the MCID methodology, various studies scored poorly on the use of more than one anchor and/or lacked (sufficient) correlations with the chosen anchor [7, 71–78, 80, 85–88]. A limited number of studies used receiver operating characteristics (ROC) curves in the anchor-based analysis, of which most failed to define sufficient area under the curve (AUC) [7, 71–75, 77–88]. A minority of studies used a global rating of change (GRC) scale with sufficient answering categories [7, 70, 74, 79, 88]. Two studies used criterion referencing [70, 84]. Some of the included studies used more than one distribution-based method [70, 71, 75, 76, 82, 84–86]. A limited number of studies measured the MCID in more than one population [7, 72, 73, 75, 87]. Certain studies determined the MCID for both improvement and deterioration [7, 71, 75, 80, 84–86, 88]. In most studies, there was a risk of ownership bias.

## CAT

Six papers discussed the MCID for the CAT [90] in pulmonary rehabilitation (PR) [70, 72, 75, 82], in patients with acute COPD exacerbation [73, 75], and in regular primary and secondary care [75, 83], with follow-up periods of 2 weeks to 12 months (table 1). The included studies received overall quality assessments of two [83], three [72, 73], four [75, 82] or five [70] stars (table 2). Anchor-based methods resulted in an MCID range for improvement on the CAT of –3.50 to –1.00 including the use of various GRCs by patients and physicians; exacerbations as a criterion; and CCQ, CRQ and SGRQ as anchors in ROC curves, linear regression analysis and mean change score calculations (figure 2, supplementary table S6)

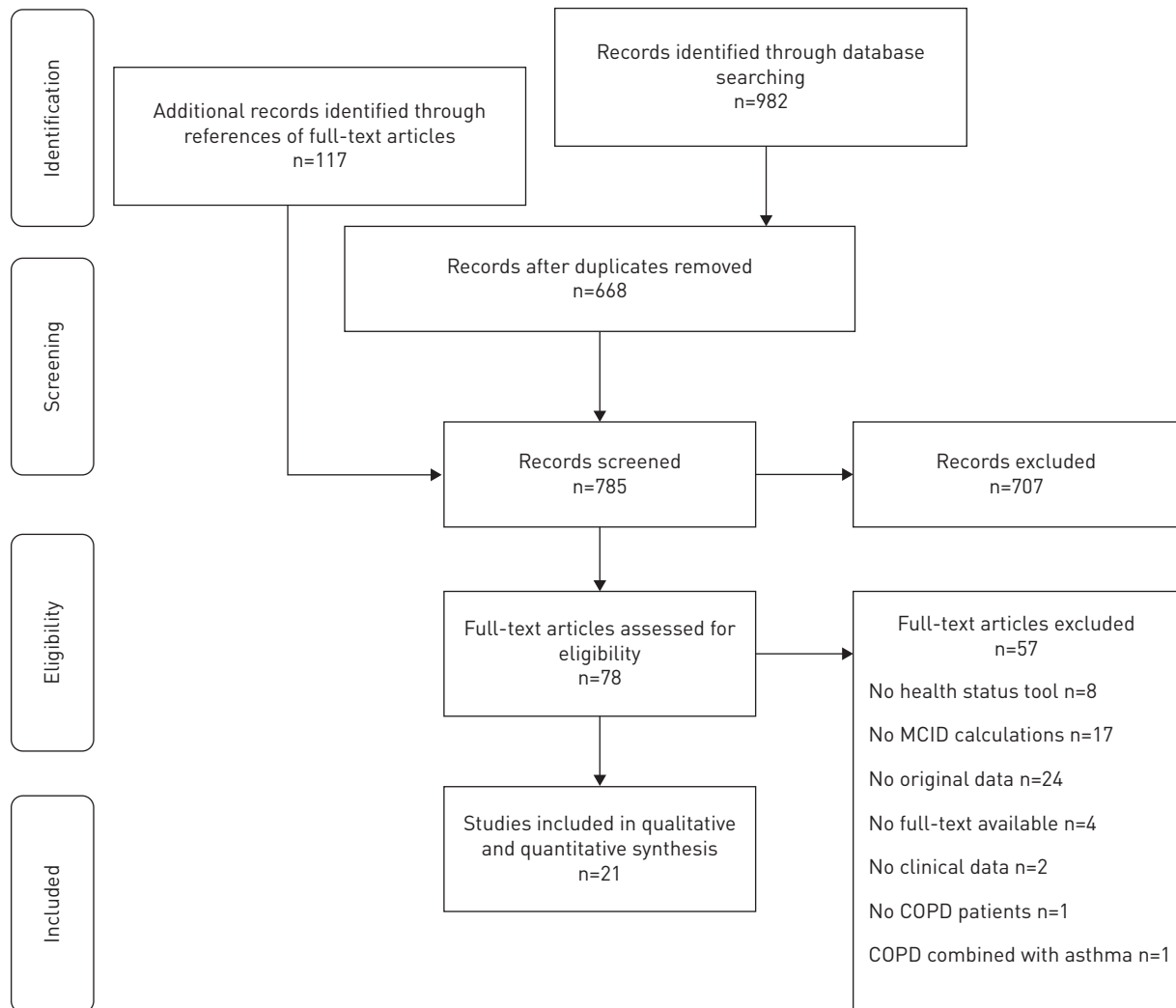


FIGURE 1 PRISMA flow-chart of the search results. MCID: minimal clinically important difference; COPD: chronic obstructive pulmonary disease.

[70, 72, 73, 75, 82]. The anchor-based MCID for deterioration ranged from +1 (n=51) to +2 (n=3) [72, 75]. Distribution-based approaches including 0.5 SD, SEM and 1.96 SEM ranged from 1.92 to 3.80 [70, 75, 82, 83], excluding the 1.96 SEM outlier of 6.43 [70]. The triangulated MCID for improvement was  $-2.54$ , excluding this 1.96 SEM outlier. No structural differences were observed between different settings or follow-up periods. However, the anchor-based MCIDs by DODD *et al.* [72] during 8 weeks' PR and by KON *et al.* [75] during 12 months' regular care were smaller (figure 2).

### CCQ

Five papers discussed the MCID for the CCQ [91] in PR [70, 76, 82], in patients with acute COPD exacerbation [74] and in regular primary and secondary care [83], with follow-up periods of 2 days to 12 months (table 1). The overall quality of the included studies was rated as two [83], four [74, 76, 82] or five [70] stars (table 2). The MCID for improvement for the CCQ total score ranged from  $-0.62$  to  $-0.34$ , including various anchor-based methods with a 15-point GRC; criterion referencing; and CAT, SGRQ and CRQ as anchors in linear regression, mean change and ROC curves (figure 3, supplementary table S6) [70, 74, 76, 82]. The MCID for deterioration for the CCQ has not been determined. Results on the domain scores are available in supplementary table S6. Distribution-based methods with 0.5 SD, SEM and 1.96 SEM ranged from 0.21 to 0.56 [70, 74, 76, 82, 83], excluding the outlier of the 0.80 estimate of the minimal detectable change (95% confidence interval) [76]. The triangulated MCID for improvement for the CCQ was  $-0.43$ . Estimates from PR, patients with acute exacerbation and regular care with various follow-up durations were similar, except for the distribution-based estimate by KOCKS *et al.* [74] in patients with exacerbation (figure 3).

TABLE 1 Study characteristics of the included studies

Study	Subjects n	Study setting	Study period	Follow-up period
<b>COPD Assessment Test (CAT)</b>				
ALMA <i>et al.</i> , 2016 [70]	449	Inpatient PR; Klinik Bad Reichenhall, Germany	February 2013–July 2014	3 weeks
DODD <i>et al.</i> , 2011 [72]	297	Multidisciplinary PR in primary and secondary care; London, UK	January–August 2010	8 weeks
JONES <i>et al.</i> , 2012 [73]	65	1) Regular care treatment for acute COPD exacerbation in primary and secondary care; USA	1) February 2009–April 2009	2 weeks
KON <i>et al.</i> , 2014 [75]	59	2) PR; Canada and USA	2) July 2009–December 2009	6 weeks
	565	1) Outpatient PR; Harefield Hospital, London, UK	1) April 2010–December 2012	8 weeks
	147	2) Hospital discharge after admission from acute wards; Hillingdon Hospital, London, UK	2) November 2011–December 2012	3 months
SMID <i>et al.</i> , 2017 [82]	164	3) Regular care outpatient clinics; Harefield Hospital, London, UK	3) January 2012–August 2012	12 months
	419	In- or outpatient PR; CIRO Rehabilitation Network, Horn, The Netherlands	April 2012–September 2014	8 or 16 weeks
TSILIGIANNI <i>et al.</i> , 2012 [83]	90	Primary and secondary regular care; Crete, Greece	July 2010–June 2011	2 and 6 weeks
<b>Clinical COPD Questionnaire (CCQ)</b>				
ALMA <i>et al.</i> , 2016 [70]	449	Inpatient PR; Klinik Bad Reichenhall, Germany	February 2013–July 2014	3 weeks
KOCKS <i>et al.</i> , 2006 [74]	168	Oral <i>versus</i> intravenous prednisolone for acute COPD exacerbation; Isala Klinieken, Zwolle, The Netherlands	June 2001–May 2003	Days 1–7, 6 weeks, 12 months
KON <i>et al.</i> , 2014 [76]	261	Outpatient PR; Harefield Hospital, London, UK	November 2011–January 2013	8 weeks
SMID <i>et al.</i> , 2017 [82]	419	In- or outpatient PR; CIRO Rehabilitation Network, Horn, The Netherlands	April 2012–September 2014	8 or 16 weeks
TSILIGIANNI <i>et al.</i> , 2012 [83]	90	Primary and secondary regular care; Crete, Greece	July 2010–June 2011	2 and 6 weeks
<b>Short-Form Chronic Respiratory Questionnaire (SF-CRQ)</b>				
TSAI <i>et al.</i> , 2008 [71]	301	Regular care for acute COPD exacerbation in 29 emergency departments; USA and Canada	2000–2001	2 weeks
JAESCHKE <i>et al.</i> , 1989 [7]	28/23/20/10/21	1) Inpatient PR for 4–6 weeks	1) September 1983–July 1984	Discharge and 2, 6, 12 and 24 weeks
	24	2) Clinical trial on inhaled salbutamol and oral theophylline for four 2-week periods <i>versus</i> placebo	2) NR	4×2-week periods
REDELMEIER <i>et al.</i> , 1996 [80]	112	Small groups of patients in supervised PR	February 1992–February 1994	Not applicable
WYRWICH <i>et al.</i> , 2007 [88]	9	1) Expert panel discussion	1) 2000–2001	Every 2 months
	554/504/484/462/462	2) Regular primary care; Wishard Health Services, Indianapolis, IN, USA, and the St. Louis Veteran Affairs Medical Center, St. Louis, MO, USA	2) August 2000–November 2001	Every 2 months
	43	3) Regular primary care; Wishard Health Services, Indianapolis, IN, USA, and the St. Louis Veteran Affairs Medical Center, St. Louis, MO, USA	3) August 2000–November 2001	Not applicable
<b>EQ-5D Utility Index (UI) and VAS</b>				
NOLAN <i>et al.</i> , 2016 [78]	616	1) Regular care in respiratory clinics; Harefield Hospital, London, UK	1) April 2012–October 2014	Baseline
WALTERS and BRAZIER, 2005 [86]	324	2) Outpatient PR clinics; Harefield Hospital, London, UK	2) August 2013–October 2014	8 weeks
	97/81	Regular care in chest clinic of a city teaching hospital	NR	6 and 12 months
ZANINI <i>et al.</i> , 2015 [89]	439	Inpatient PR tertiary healthcare center; Italy	January 2009–December 2012	3 weeks

Continued

TABLE 1 Continued

Study	Subjects n	Study setting	Study period	Follow-up period
<b>St. George's Respiratory Questionnaire (SGRQ)</b>				
ALMA <i>et al.</i> , 2016 [70]	449	Inpatient PR; Klinik Bad Reichenhall, Germany	February 2013–July 2014	3 weeks
SCHÜNEMANN <i>et al.</i> , 2003 [81]	84	PR patients; University of Toronto and McMaster University, Hamilton, ON, Canada	NR	3 months
TSILIGIANNI <i>et al.</i> , 2012 [83]	90	Primary and secondary regular care; Crete, Greece	July 2010–June 2011	2 and 6 weeks
WELLING <i>et al.</i> , 2015 [87]	110/86	Bronchoscopic lung volume reduction; University Medical Center Groningen, The Netherlands	NR	1 and 6 months
<b>SF-6D and SF-36</b>				
WALTERS and BRAZIER, 2003 [85]	60	Regular care in chest clinic of a city teaching hospital	NR	1 year
WALTERS and BRAZIER, 2005 [86]	97/81	Regular care in chest clinic of a city teaching hospital	NR	6 and 12 months
WYRWICH <i>et al.</i> , 2007 [88]	9	1) Expert panel discussion	1) 2000–2001	Every 2 months
	554/504/484/462/462	2) Regular primary care; Wishard Health Services, Indianapolis, IN, USA, and the St. Louis Veteran Affairs Medical Center, St. Louis, MO, USA	2) August 2000–November 2001	Every 2 months
	43	3) Regular primary care; Wishard Health Services, Indianapolis, IN, USA, and the St. Louis Veteran Affairs Medical Center, St. Louis, MO, USA	3) August 2000–November 2001	Not applicable
<b>Other tools</b>				
KULICH <i>et al.</i> , 2015 [77] – eDiary	177	Phase III multicentre clinical trial (SHINE study) on QVA149 (dual bronchodilator)	September 2010–February 2012	26 weeks
PEREZ <i>et al.</i> , 2009 [79] – VSRQ	373	RCT 18 µg tiotropium once daily <i>versus</i> placebo in 123 centres; France	May 2002–April 2004	2 weeks and 3 months
SCHÜNEMANN <i>et al.</i> , 2003 [81] – FT	84	PR patients; University of Toronto and McMaster University, Hamilton, ON, Canada	NR	3 months
VAN STEL <i>et al.</i> , 2003 [84] – QoLRIQ	108	Individualised inpatient PR; Asthma Center Heideheugel, The Netherlands	January 1996–December 1997	3–6 months

COPD: chronic obstructive pulmonary disease; EQ-5D: EuroQol Five Dimensions; VAS: visual analogue scale; SF-6D: Short-Form Six-Dimension index; SF-36: Short-Form 36 survey; VSRQ: Visual Simplified Respiratory Questionnaire; FT: Feeling Thermometer; QoLRIQ: Quality of Life for Respiratory Illness Questionnaire; PR: pulmonary rehabilitation; RCT: randomised controlled trial; NR: not reported.

TABLE 2 Quality assessment and risk of bias of the included studies

First author	Inclusion criteria clear	Exclusion criteria clear	Systematic enrolment	Follow-up completed	Missing data reported	% lost to follow-up <25%	Pre and post measurements available	Time interval follow-up stated	Similar test conditions	Health status description	Validation health status tool	Floor effects	Ceiling effects	MCID calculation	Description anchor	Measurement properties anchor	Anchor golden standard	> 1 Anchor used	Anchor correlations calculated	Anchor correlations ≥0.50	ROC curves produced	Area under the curve ≥0.70	Global Rating of Change used	Anchor question options ≥11	Minimal criterion used	≥ 1 Distribution-based method	≥ 1 Population used	Improvement and deterioration	Selective outcome reporting bias low	Funding bias risk low	Ownership bias risk low	Overall rating (1–5 stars)	
<b>CAT</b>																																	
ALMA [70]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★★
DODD [72]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
JONES [73]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
KON [75]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
SMID [82]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
TSILIGIANNI [83]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★
<b>CCQ</b>																																	
ALMA [70]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★★
KOCKS [74]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
KON [76]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
SMID [82]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
TSILIGIANNI [83]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★
<b>SF-CRQ</b>																																	
TSAI [71]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
JAESCHKE [7]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
REDELMEIER [80]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★
WYRICH [88]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
<b>EQ-5D UI and VAS</b>																																	
NOLAN [78]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
WALTERS [86]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
ZANINI [89]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
<b>SGRQ</b>																																	
ALMA [70]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★★
SCHÜNEMANN [81]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
TSILIGIANNI [83]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★
WELLING [87]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
<b>SF-6D and SF-36</b>																																	
WALTERS [85]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
WALTERS [86]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
WYRICH [88]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
<b>Other tools</b>																																	
KULICH [77]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
PEREZ [79]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
SCHÜNEMANN [81]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★
VAN STEL [84]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★

Data overall:

- Indicates that respective item is clearly defined, poses a low risk of bias or the considered test was performed
- Indicates that respective item is unclear
- Indicates that respective item is not clearly defined, poses a high risk of bias or this test was not performed

Data on the item "improvement and deterioration":

- Indicates both improvement and deterioration MCIDs
- MCID only for improvement or deterioration calculated
- Unclear whether MCID is for improvement, deterioration or both

MCID: minimal clinically important difference; ROC: receiver operator characteristics; CAT: COPD Assessment Test; CCQ: Clinical COPD Questionnaire; SF-CRQ: Short-Form Chronic Respiratory Questionnaire; EQ-5D UI: EuroQol Five Dimensions Utility Index; VAS: visual analogue scale; SGRQ: St George’s Respiratory Questionnaire; SF-6D: Short-Form Six-Dimension index; SF-36: Short-Form 36 survey.

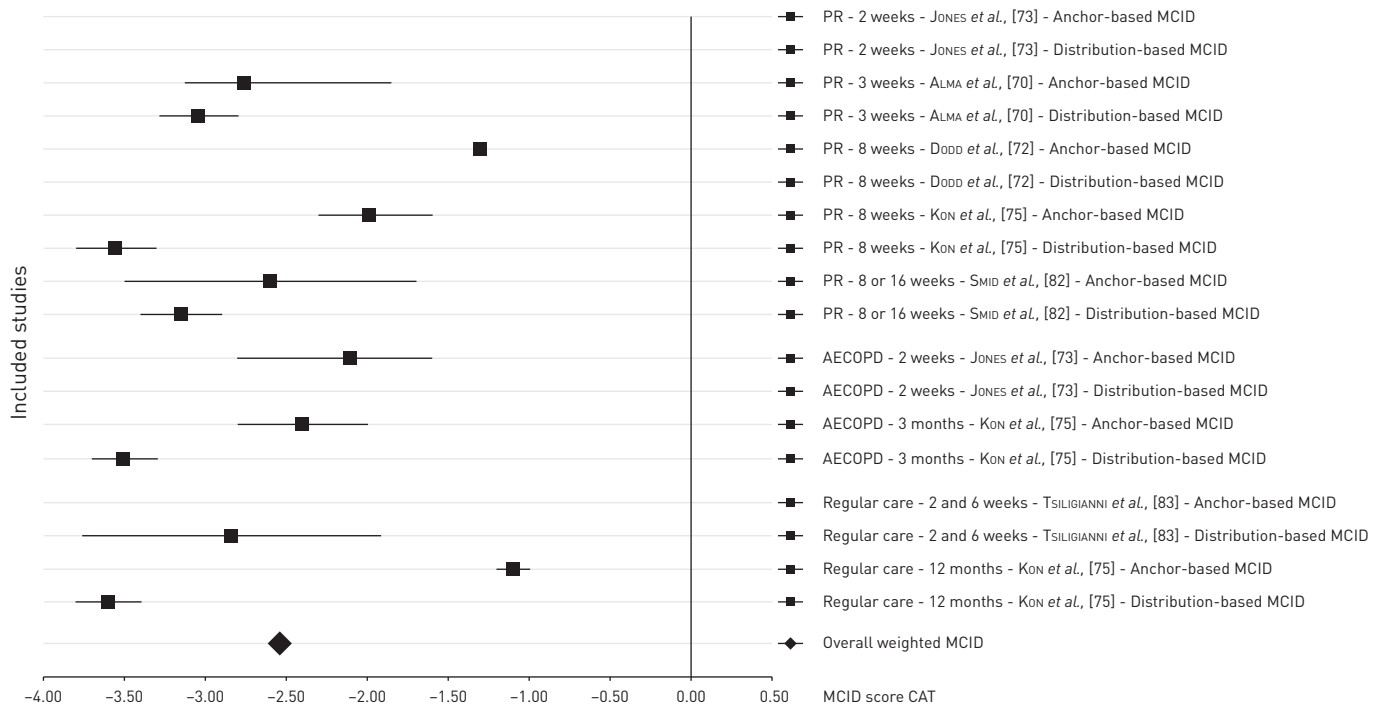


FIGURE 2 Overview of the minimal clinically important difference (MCID) study results and triangulation for the COPD Assessment Test (CAT). Data are presented as mean study MCIDs for anchor-based and distribution-based methods (squared estimates). The horizontal lines represent the range of estimates provided in the respective study. The larger diamond represents the triangulated MCID. Results are categorised per setting in correspondence with the duration of follow-up period. PR: pulmonary rehabilitation; AECOPD: acute exacerbation of chronic obstructive pulmonary disease.

### SF-CRQ

Four papers reported the MCID for the SF-CRQ [92] in PR [7, 80], in patients with acute COPD exacerbation [71], in a salbutamol trial [7], in regular primary care [88] and by means of expert opinions [88] (table 1), with follow-up periods of 2 weeks to 24 weeks. Overall quality was rated two [80], three [7, 71] or four [88] stars (table 2). MCID estimates for both improvement and deterioration were determined in all studies (supplementary table S6). The MCIDs resulted from the anchor-based method using a five-, seven- or 15-point GRC for both within- and between-subject change [7, 71, 80, 88]. The MCIDs for improvement for the SF-CRQ ranged from 0.30 to 1.60 as the average domain score change (two items per domain), and for deterioration from -0.60 to -0.06 [71]. The MCIDs for the CRQ for improvement per item score were 0.40 to 1.00 (dyspnoea), 0.25 to 0.50 (fatigue), 0.14 (emotion) and zero

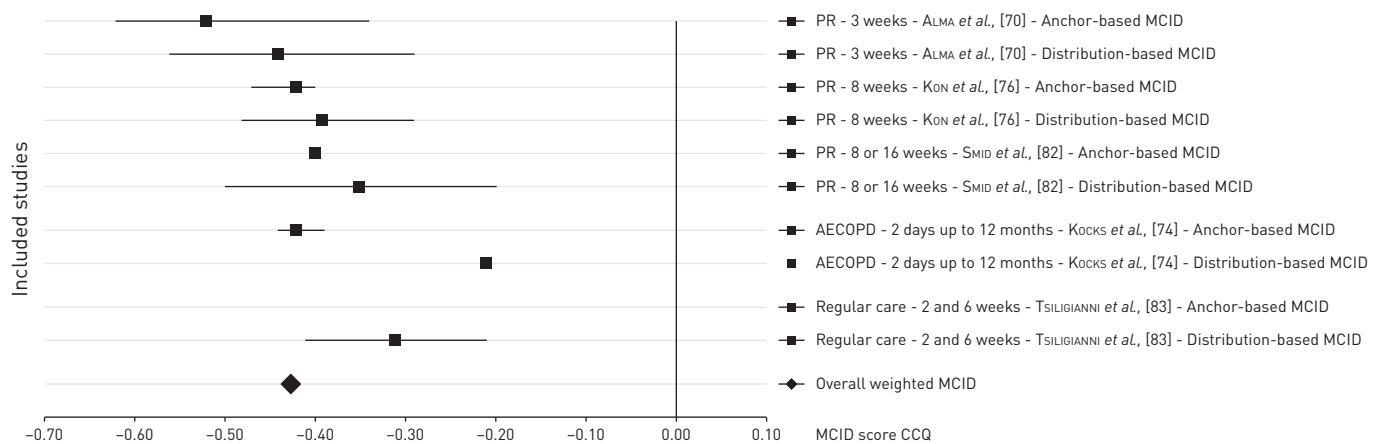


FIGURE 3 Overview of the minimal clinically important difference (MCID) study results and triangulation for the Clinical COPD Questionnaire (CCQ). Data are presented as mean study MCIDs for anchor-based and distribution-based methods (squared estimates). The horizontal lines represent the range of estimates provided in the respective study. The larger diamond represents the triangulated MCID. Results are categorised per setting in correspondence with the duration of follow-up period. PR: pulmonary rehabilitation; AECOPD: acute exacerbation of chronic obstructive pulmonary disease.



to 0.25 (mastery). The MCIDs for the CRQ for deterioration per item score were  $-0.20$  (dyspnoea),  $-0.50$  (fatigue),  $-0.14$  to zero (emotion) and  $-0.50$  to  $-0.25$  (mastery) [88]. A combined MCID for improvement and deterioration was per item 0.09 to 0.62 (dyspnoea), 0.50 to 0.68 (fatigue), 0.57 to 0.87 (emotion) and 0.23 to 0.75 (mastery) [7, 80, 88]. Owing to the limited number of studies, the diversity of domains and scoring approaches and the small number of patients in certain studies [7], no triangulation was performed.

#### EQ-5D Utilities Index and visual analogue scale

Three papers discussed the MCID for the EQ-5D-5L Utility Index (UI), visual analogue scale (VAS) and/or EQ-5D-3L-VAS [93, 94] in PR [78, 89], and for the EQ-5D-3L in regular secondary care [86], with follow-up periods of 3 weeks to 12 months (table 1). The quality assessment differed from three [86] to four [78, 89] stars (table 2). The MCID for improvement for the EQ-5D VAS scale ranged from 6.50 to 10.10 [78, 89]. The anchor-based MCID for improvement for the EQ-5D-3L and EQ-5D-5L UI ranged from  $-0.128$  to 0.063 [78, 86]. Estimates for deterioration ranged from  $-0.007$  to 0.039 [86]. A combined MCID for improvement and deterioration ranged from  $-0.011$  to zero [86]. Distribution-based results ranged from  $-0.050$  to 0.150 (supplementary table S6) [78, 86]. Owing to the limited number of included studies and the diversity of the results, no triangulation was executed.

#### SF-6D and SF-36

Two papers discussed the MCID for the SF-6D [85] in regular care [85, 86], and one of the SF-36 [95] in regular care and by means of expert opinions [88] (table 1). Both studies on the SF-6D [85, 86] were of average quality (three stars); the study on the SF-36 by WYRWICH *et al.* [88] was of good quality (four stars) (table 2). The MCID for improvement for the various SF-36 domains ranged from 2 to 11 using a GRC scored by the patient or the physician [88]. The range for deterioration was  $-6$  to  $+4$ . The expert-based panel determined that values of 8.33–12.50 represented minimal changes [88]. The MCID for improvement for the SF-6D using a five-point GRC ranged from  $-0.004$  to 0.054; for deterioration ranged from 0.012 to 0.028; and combined ranged from 0.010 to 0.036 [85, 86]. The distribution-based estimates for the SF-6D ranged from 0.044 to 0.410 using the standardised response mean, effect size and 0.5 SD (supplementary table S6) [85, 86]. Owing to the limited number of included studies and diversity of the results, no triangulation was performed.

#### SGRQ

Four studies analysed the MCID for the SGRQ [96] during PR [70, 81], bronchoscopic lung volume reduction (BLVR) [87] and regular primary and secondary care [83], with follow-up periods of 2 weeks to 6 months (table 1). The included studies scored two [83], three [81, 87] or five [70] stars. The MCID was determined for improvement only. Anchor-based approaches resulted in an MCID for the SGRQ between  $-9.28$  and  $-6.30$  using a 15-point GRC; criterion referencing; and CAT, CCQ, FEV<sub>1</sub>, 6-min walking distance and residual volume as anchors in linear regression, mean change and ROC curves (figure 4, supplementary table S6) [70, 87]. An outlying anchor-based result in SCHÜNEMANN *et al.* [81] was  $-3.05$  using the CRQ dyspnoea domain in linear regression analysis. Distribution-based results ranged from 2.40 to 10.19 using 0.2–0.8 SD, SEM and 1.96 SEM [70, 81, 83, 87]. The triangulated MCID for improvement was  $-7.43$ . Estimates from 3 weeks' PR [70], and 1 and 6 months' BLVR [87] were similar (figure 4). However,

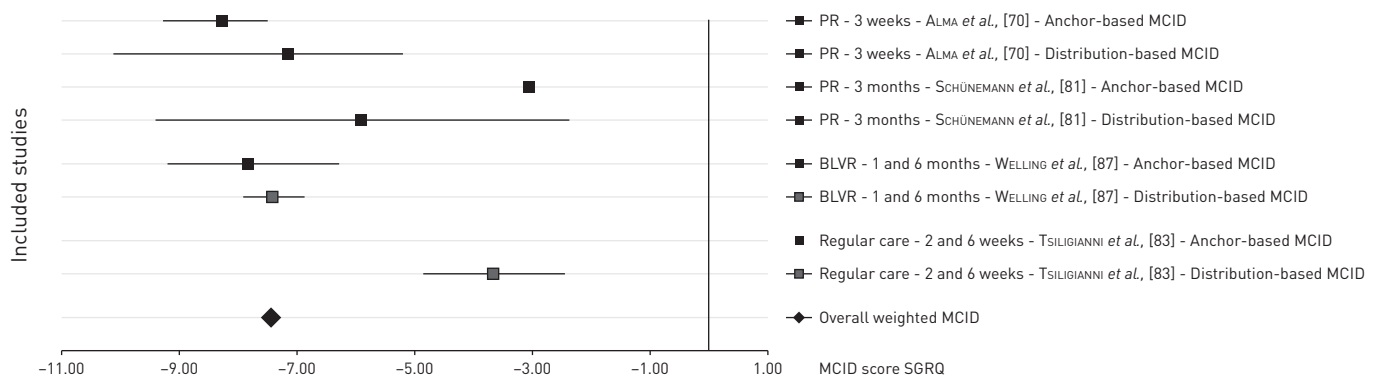


FIGURE 4 Overview of the minimal clinically important difference (MCID) study results and triangulation for the St George's Respiratory Questionnaire (SGRQ). Data are presented as mean study MCIDs for anchor-based and distribution-based methods (squared estimates). The horizontal lines represent the range of estimates provided in the respective study. The larger diamond represents the triangulated MCID. Results are categorised per setting in correspondence with the duration of follow-up period. PR: pulmonary rehabilitation; BLVR: bronchoscopic lung volume reduction.

the 3 months' PR anchor-based estimate by SCHÜNEMANN *et al.* [81] was much smaller using the CRQ dyspnoea domain as anchor. The distribution-based result by TSILIGIANNI *et al.* [83] was also much smaller when measured in regular care.

### Other tools

The other tools discussed included the eDiary [77], VSRQ [79], FT [81] and the QoLRIQ [84] (table 1). These studies were of average (three stars) [77, 81] to good (four stars) [79, 84] quality (table 2). All studies measured the MCID using anchor-based methods with multiple anchors using reasonable methodology. The MCID for improvement for the eDiary was  $-0.64$  to  $-0.52$  using a seven-point GRC and the TDI as the anchor (supplementary table S6) [77]. The MCID for improvement for the VSRQ was 3.20 to 3.50 using a 15-point GRC and the SGRQ as the anchor in linear regression [79]. The MCID for the FT was 4.10 to 16.30 using the CRQ fatigue domain and the SGRQ total and domain scores as anchors in linear regression, as well as the distribution-based methods  $0.2$ – $0.8$  SD [79]. The MCID for improvement for the QoLRIQ was 0.51 to 0.64, and for deterioration 0.37 to 0.49 using a five-point GRC [84]. The distribution-based results ranged from 0.18 to 0.45 [84]. No triangulation was executed owing to the low number of studies per instrument.

## Discussion

### Summary of main results

The current systematic review provides a unique overview and triangulation of 21 papers including 12 different COPD HRQoL and health status tools and their MCID methodology, quality and estimates. The tools included are the CAT, CCQ, SF-CRQ, eDiary, EQ-5D, FT, QoLRIQ, SF-6D, SF-36, SGRQ and VSRQ. The overall quality of the methodology and MCID calculation was average to good, with one study scoring excellent and two studies scoring poor. Triangulated MCIDs for CAT, CCQ and SGRQ were  $-2.54$ ,  $-0.43$  and  $-7.43$ , respectively, for improvement, without structural differences between various settings and/or follow-up duration. The other instruments had too few or too heterogeneous studies to attempt triangulation; however, ranges have been presented. Studies on MCIDs for deterioration were scarce or non-existent for all tools.

### Interpretation of findings

COPD assessment and management should include health status instruments combined with the number of exacerbations to decide on patients' classification and therapy [12]. These tools are difficult to use in daily practice and scientific research without guidelines on what change may be considered clinically relevant [1, 3, 5, 9, 10, 22, 23]. The MCID parameter aims to quantify this threshold at the group level [2, 6, 8]. It is an obligatory endpoint in clinical trials in which the percentage of patients achieving clinically relevant change is compared to the percentage of control patients achieving this change [2, 6, 8]. The MCID may also indicate to what extent an individual patient experiences relevant change over time. It is therefore of pivotal concern that the MCID is well established, else this may result in over- or underestimation of the interpretation of treatment effects. This review has provided insight into the quality and quantity of the MCID for various HRQoL and health status instruments in COPD, with CAT, CCQ, CRQ and SGRQ as the most important tools. The short CAT and CCQ are recommended especially for use in clinical practice; the lengthier CRQ and SGRQ are more applicable for scientific research [12, 52, 97].

The MCID for improvement for the CAT was between  $-4$  and  $-1$  with the majority of estimates between  $-3$  and  $-2$ , resulting in a triangulated MCID of  $-2.54$  [70, 72, 73, 75, 82, 83]. This estimate was valid from multiple studies performed at the group level, demonstrating consistency of results for different settings with various follow-up periods. Stability of MCIDs during various follow-up periods was also demonstrated by ALMA *et al.* [98]. Because CAT only allows for integer scores at the individual level [90], a change of  $-3$  could be considered a clinically relevant improvement for use in daily clinical practice. The MCID for deterioration for the CAT was between 1 and 2; however, this resulted from two studies with a limited number of patients [72, 75]. The MCID for the CCQ was valid for improvement only, ranging from  $-0.60$  to  $-0.20$ , leading to a triangulated estimate of  $-0.43$  valid from PR, patients with acute exacerbation, and regular care [70, 74, 76, 82, 83]. Because the quality of the included studies for CAT and CCQ was average to excellent, and the quality and size of the study was integrated in the triangulation, these estimates are valid for use in clinical practice and scientific research. The triangulated estimates for CAT and CCQ are close to the accepted MCIDs currently used in practice of  $-2$  and  $-0.40$ , respectively [12]. MCIDs for deterioration were not readily available.

The MCID for the SGRQ in the current review ranged from  $-11$  to  $-2$  for improvement, with most estimates between  $-10$  and  $-6$ , resulting in a triangulated MCID of  $-7.43$ . However, the MCID for the SGRQ extensively used in clinical practice is 4. This estimate was based on analyses by JONES and colleagues in 1991 and 2005 [53, 95]. The evidence found in this review suggests the MCID to be double

that, and the studies that formed the basis for the currently accepted MCID for the SGRQ [53, 95] did not meet the study criteria for this review. It is therefore questionable how grounded the currently accepted MCID of 4 is. Our evidence for the MCID for the SGRQ was of average to excellent quality, which validates the triangulated value of 7.43. The MCID for the SGRQ of 4 for improvement has also been used as an anchor for the MCID for the CAT [70, 73, 75, 82], CCQ [70, 76, 82], VSRQ [79] and FT [81]. It may in fact have led to lower MCIDs for these tools. However, these tools have used other anchors and techniques in addition to the SGRQ to determine the MCID, validating their currently estimated (triangulated) values. Still, careful selection of anchors should be advocated.

CRQ item MCIDs of 0.5 points have been regularly used for both improvement and deterioration based upon JAESCHKE *et al.* [7] and REDELMEIER *et al.* [80]. However, the ranges of item MCIDs for both the CRQ and SF-CRQ were wider and more inconsistent based on the current review. The assessed quality of both studies was poor to average [7, 80]. The MCID for the CRQ of 0.5 might therefore be too simple an interpretation of results from methods of questionable quality. Owing to the inconsistent results, variety of scoring techniques and limited size of the studies, no triangulation was executed. The item MCID of 0.50 points has been used as an anchor for the MCIDs of the CAT [75], CCQ [76], SGRQ [81] and EQ-5D [78], which may have affected their MCIDs. The use of the CRQ as anchor did not result in structurally different results for the MCID for the CAT or CCQ. However, it did result in an outlying, possibly (too) low, MCID for the SRGQ using the CRQ dyspnoea domain as reference [81].

The general health status instruments EQ-5D, SF-6D and SF-36 had less evidence for their MCIDs in COPD. The UI estimates for the EQ-5D and SF-6D varied and inconsistently ranged from minus to plus scores, including the zero estimate. The MCID for improvement for the VAS of the EQ-5D was between six and 10 [89]. General instruments may be applicable when comparing HRQoL between patient disease groups; however, it is not valid to use their MCIDs to evaluate therapy outcome within patient groups. To compare within patient groups, disease-specific health status tools such as CAT and CCQ are more valid, with well-established MCIDs. The other tools in our review, including the eDiary, FT, QoLRIQ and VSRQ, each had only one study available regarding its MCID. More research is required for these instruments for them to be used in clinical practice.

In general, determining an MCID for an HRQoL and health status instrument requires a combination of anchor- and distribution-based methods, preferably measured in multiple settings over various follow-up periods [33, 45]. All included studies in this review, except for one [83], used anchor-based methods in measuring the MCID. Most studies, except for five papers [7, 72, 73, 80, 83], combined anchor-, distribution- and/or opinion-based methods as recommended. Studies regarding the CRQ, SF-6D and SF-36 in general did not use multiple anchors to determine the MCID. Furthermore, in most studies, the presentation of anchor correlations or correlations  $\geq 0.50$ , the use of ROC curves with AUC  $\geq 0.70$ , and the use of a GRC with  $\geq 11$  answering options was poor. Most studies did not use multiple distribution-based methods either. These would be points of attention for future MCID determination processes.

#### ***Strength and limitations of the current study***

This study is the first to systematically address the MCIDs of COPD HRQoL and health status tools. Although other papers have provided an overview of instruments, none of them has addressed the methodology, values and triangulation of MCIDs [3, 8, 10, 16, 28, 29, 38, 39, 46–57]. This study had a structurally defined protocol that was thoroughly executed by two independent reviewers. This review could be a starting point for further discussion. There are currently no fixed guidelines on how to judge studies measuring an instrument's MCID nor guidelines for triangulation. Triangulation should involve a combination of anchor- and distribution-based methods combined with evidence from clinical trials data and qualitative approaches [33, 45]; however, this does not provide clear guidelines. The current authors have selected elements from existing assessment tools for their own risk of bias and quality assessment to evaluate the studies, their MCID methodologies and quality. This combined tool has not been validated. However, because it contains elements of established checklists and the MCID literature, the authors feel that there has been a sound evaluation of the overall quality and risk of bias of the included studies. The triangulation process as performed in the current review has also not been used before, but takes into account the MCID methodology, study quality and study size.

#### ***Implications for practice and future research***

This systematic review highlights gaps in the current MCID evidence and a need for further research. First, evidence is limited or lacking for the MCID for deterioration for all COPD HRQoL and health status tools. This is relevant because MCIDs for improvement are not necessarily similar to those for deterioration [2]. COPD is a progressive disease causing deterioration of HRQoL over time, which makes MCIDs for deterioration clinically important [99, 100]. Second, the MCIDs for improvement for the CAT

and CCQ were well established in different settings for various follow-up durations. The triangulated MCIDs could be used in diverse clinical practice. However, the frequently used MCIDs for SGRQ and CRQ lacked evidence for the currently used clinical values of, respectively, 4 and 0.5 points. These estimates have been used as obligatory endpoints in clinical trials for many years. This might have resulted in an overestimation of treatment effects. MCIDs for SGRQ and CRQ should thus be reconsidered; however, this has major clinical consequences for the currently approved evidence-based therapies in COPD. It would be impossible to re-evaluate all clinical trials to date that have used the SGRQ and/or CRQ as outcome parameters. However, it may be worthwhile reviewing the existing evidence to observe which level of evidence remains valid with alternative MCIDs for the SGRQ and CRQ. Current knowledge and guidelines of health status measurement and its MCIDs have evolved over time. Older instruments such as the SGRQ and CRQ and their MCIDs may not have evolved with current guidelines. Last, evidence for the MCID for the eDiary, EQ5D, FT, QoLRIQ, SF-6D, SF-36 and VSRQ was limited in terms of consistency and number of studies available, highlighting the need for more research before they can reliably be used in clinical practice.

### Conclusion

This study provides a first comprehensive and systematic assessment of MCIDs for COPD HRQoL and health status instruments. It highlights pros and cons in the used methodology, as well as gaps in the evidence. Triangulated MCIDs for CAT, CCQ and SGRQ were, respectively,  $-2.54$ ,  $-0.43$  and  $-7.43$  for improvement. These values may be integrated in future GOLD guidelines. This is an important step for clinicians and patients, who could easily use MCIDs in their scientific research and clinical practice. The MCIDs of the CAT and CCQ are well established; however, the reviewed MCID poorly matched with currently used values for the SGRQ (4 points) and CRQ (0.5 points). It is recommended that CAT or CCQ are used as outcome parameters for health status in COPD and that MCIDs for SGRQ and CRQ are recalculated. Evidence for the MCID for the other tools included was inconsistent, too heterogeneous or too limited. The fact that MCIDs for deterioration are scarce highlights a clear need for more well-designed studies.

**Acknowledgements:** We wish to acknowledge D.G. van Ittersum at the University of Groningen for helping to define this review's search strategy.

**Author contributions:** H. Alma, C. de Jong, R. Sanderman and T. van der Molen designed the current systematic review. H. Alma and C. de Jong performed the search, article selection and data extraction procedures. H. Alma wrote the first draft, while C. de Jong, I. Tsiligianni, J. Kocks, R. Sanderman and T. van der Molen actively participated in the review process. R. Sanderman and T. van der Molen supervised and participated in different steps of the study, as well as in writing. All authors participated in various steps in the study, edited the manuscript and gave approval for submission.

**Conflict of interest:** I. Tsiligianni received personal fees from Boehringer Ingelheim, Novartis, AstraZeneca and GlaxoSmithKline, all outside the submitted work. J. Kocks reports personal fees from Novartis, IPCRG and Springer Media; research grants and personal fees from Boehringer Ingelheim and GlaxoSmithKline; research grants from Stichting Zorgdraad; and travel arrangements from Chiesi BV, GlaxoSmithKline BV and IPCRG, all outside the submitted work. T. van der Molen reports personal reimbursements from GlaxoSmithKline, TEVA, AstraZeneca and Boehringer Ingelheim; and study grants from AstraZeneca and GlaxoSmithKline, all outside the submitted work. After this study was terminated, he became an employee of GlaxoSmithKline. T. van der Molen developed the CCQ and holds the copyright.

**Support statement:** This review received financial support from the Junior Scientific Masterclass, part of the University of Groningen, to enable the PhD research position of H. Alma. Funding information for this article has been deposited with the Crossref Funder Registry.

### References

- 1 U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: Patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006; 4: 79.
- 2 Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; 56: 395–407.
- 3 Jones PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 2001; 56: 880–887.
- 4 Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002; 14: 109–114.
- 5 Jones PW. Health status: what does it mean for payers and patients? *Proc Am Thorac Soc* 2006; 3: 222–226.
- 6 Jones PW, Rennard S, Tabberer M, et al. Interpreting patient-reported outcomes from clinical trials in COPD: a discussion. *Int J COPD* 2016; 11: 3069–3078.
- 7 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–415.

- 8 Jones PW, Beeh KM, Chapman KR, *et al.* Minimal clinically important differences in pharmacological trials. *Am J Respir Crit Care Med* 2014; 189: 250–255.
- 9 Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993; 118: 622–629.
- 10 Curtis JR, Patrick DL. The assessment of health status among patients with COPD. *Eur Respir J* 2003; 21: Suppl. 41, 36–45.
- 11 Jones PW. Issues concerning health-related quality of life in COPD. *Chest* 1995; 107: Suppl., 187S–193S.
- 12 Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD (2017 Report). 2017. [www.goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/](http://www.goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/) Date last accessed: 17 November 2017.
- 13 Westwood M, Bourbeau J, Jones PW, *et al.* Relationship between FEV1 change and patient-reported outcomes in randomised trials of inhaled bronchodilators for stable COPD: a systematic review. *Respir Res* 2011; 12: 40-9921-12-40.
- 14 Tsiligianni I, Kocks J, Tzanakis N, *et al.* Factors that influence disease-specific quality of life or health status in patients with COPD: a review and meta-analysis of Pearson correlations. *Prim Care Respir J* 2011; 20: 257–268.
- 15 Martin AL, Marvel J, Fahrback K, *et al.* The association of lung function and St. George's Respiratory Questionnaire with exacerbations in COPD: a systematic literature review and regression analysis. *Respir Res* 2016; 17: 40.
- 16 Jones P, Miravittles M, Van der Molen T, *et al.* Beyond FEV1 in COPD: a review of patient-reported outcomes and their measurement. *Int J COPD* 2012; 7: 697–670.
- 17 Farquhar M. Definitions of quality of life: a taxonomy. *J Adv Nurs* 1995; 22: 502–508.
- 18 Reardon JZ, Lareau SC, ZuWallack R. Functional status and quality of life in chronic obstructive pulmonary disease. *Am J Med* 2006; 119: Suppl. 1, 32–37.
- 19 Bergner M, Rothman ML. Health status measures: an overview and guide for selection. *Annu Rev Public Health* 1987; 8: 191–210.
- 20 Bergner M. Quality of life, health status, and clinical research. *Med Care* 1989; 27: Suppl., S148–S156.
- 21 Jones PW, Lareau S, Mahler DA. Measuring the effects of COPD on the patient. *Respir Med* 2005; 99: Suppl. B, S11–S18.
- 22 Fitzpatrick R, Davey C, Buxton MJ, *et al.* Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998; 2: i-iv,1–74.
- 23 Davidson M, Keating J. Patient-reported outcome measures (PROMs): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. *Br J Sports Med* 2014; 48: 792–796.
- 24 Angst F, Aeschlimann A, Angst J. The minimal clinically important difference (MCID) raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 2017; 82: 128–136.
- 25 Copay AG, Subach BR, Glassman SD, *et al.* Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007; 7: 541–546.
- 26 Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD* 2005; 2: 63–67.
- 27 Dworkin RH, Turk DC, Wyrwich KW, *et al.* Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008; 9: 105–121.
- 28 Jones PW. Interpreting thresholds for a clinically significant change in health status in asthma and COPD. *Eur Respir J* 2002; 19: 398–404.
- 29 Make B. How can we assess outcomes of clinical trials: the MCID approach. *COPD* 2007; 4: 191–194.
- 30 Norman GR. The relation between the minimally important difference and patient benefit. *COPD* 2005; 2: 69–73.
- 31 Ostelo RW, Deyo RA, Stratford P, *et al.* Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 2008; 33: 90–94.
- 32 Revicki DA, Cella D, Hays RD, *et al.* Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes* 2006; 4: 70.
- 33 Revicki D, Hays RD, Cella D, *et al.* Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; 61: 102–109.
- 34 Sloan JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. *COPD* 2005; 2: 57–62.
- 35 Turner D, Schünemann HJ, Griffith LE, *et al.* The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol* 2010; 63: 28–36.
- 36 Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *J Eval Clin Pract* 2000; 6: 39–49.
- 37 Wells G, Beaton D, Shea B, *et al.* Minimal clinically important differences: review of methods. *J Rheumatol* 2001; 28: 406–412.
- 38 Make B, Casaburi R, Leidy NK. Interpreting results from clinical trials: understanding minimal clinically important differences in COPD outcomes. *COPD* 2005; 2: 1–5.
- 39 Chapman KR, Bergeron C, Bhutani M, *et al.* Do we know the minimal clinically important difference (MCID) for COPD exacerbations? *COPD* 2013; 10: 243–249.
- 40 Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics* 2000; 18: 419–423.
- 41 Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *Int J Integr Care* 2002; 2: e15.
- 42 Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; 41: 582–592.
- 43 Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991; 59: 12–19.
- 44 Cook CE. Clinimetrics corner: the minimal clinically important change score (MCID): a necessary pretense. *J Man Manip Ther* 2008; 16: E82–E83.

- 45 Leidy NK, Wyrwich KW. Bridging the gap: using triangulation methodology to estimate minimal clinically important differences (MCIDs). *COPD* 2005; 2: 157–165.
- 46 Cazzola M, Hanania NA, MacNee W, *et al.* A review of the most common patient-reported outcomes in COPD – revisiting current knowledge and estimating future challenges. *Int J COPD* 2015; 10: 725–738.
- 47 Haughney J, Gruffydd-Jones K. Patient-centred outcomes in primary care management of COPD – what do recent clinical trial data tell us? *Prim Care Respir J* 2004; 13: 185–197.
- 48 Jayadevappa R, Cook R, Chhatre S. Important difference to infer changes in health related quality of life—a systematic review. *J Clin Epidemiol* 2017; 89: 188–198.
- 49 Kocks JWH, Asijec GM, Tsiligianni IG, *et al.* Functional status measurement in COPD: a review of available methods and their feasibility in primary care. *Prim Care Respir J* 2011; 20: 269–275.
- 50 Ekström M, Currow DC, Johnson MJ. Outcome measurement of refractory breathlessness: endpoints and important differences. *Curr Opin Support Palliat Care* 2015; 9: 238–243.
- 51 Weldam SW, Schuurmans MJ, Liu R, *et al.* Evaluation of quality of life instruments for use in COPD care and research: a systematic review. *Int J Nurs Stud* 2013; 50: 688–707.
- 52 Kocks JWH, Seys SF, van Duin TS, *et al.* Assessing patient-reported outcomes in asthma and COPD patients: which can be recommended in clinical practice? *Curr Opin Pulm Med* 2018; 24: 18–23.
- 53 Jones PW. St. George's Respiratory Questionnaire: MCID. *COPD* 2005; 2: 75–79.
- 54 Mahler DA, Witek TJ Jr. The MCID of the Transition Dyspnea Index is a total score of one unit. *COPD* 2005; 2: 99–103.
- 55 Donohue JF. Minimal clinically important differences in COPD lung function. *COPD* 2005; 2: 111–124.
- 56 Zhou Z, Zhou A, Zhao Y, *et al.* Evaluating the Clinical COPD Questionnaire: a systematic review. *Respirology* 2017; 22: 251–262.
- 57 Gupta N, Pinto LM, Morogan A, *et al.* The COPD Assessment Test: a systematic review. *Eur Respir J* 2014; 44: 873–884.
- 58 Bohannon RW, Glenney SS. Minimal clinically important difference for change in comfortable gait speed of adults with pathology: a systematic review. *J Eval Clin Pract* 2014; 20: 295–300.
- 59 Bohannon RW, Crouch PT. Minimal clinically important difference for change in 6-minute walk test distance in adults with pathology: a systematic review. *J Eval Clin Pract* 2017; 23: 377–381.
- 60 Binenbaum Y, Amit M, Billan S, *et al.* Minimal clinically important differences in quality of life scores of oral cavity and oropharynx cancer patients. *Ann Surg Oncol* 2014; 21: 2773–2781.
- 61 Keurentjes JC, Van Tol FR, Fiocco M, *et al.* Minimal clinically important differences in health-related quality of life after total hip or knee replacement: a systematic review. *Bone Joint Res* 2012; 1: 71–77.
- 62 Schrover R, Evans K, Giugliani R, *et al.* Minimal clinically important difference for the 6-min walk test: literature review and application to Morquio A syndrome. *Orphanet J Rare Dis* 2017; 12: 78.
- 63 Doganay Erdogan B, Leung YY, Pohl C, *et al.* Minimal clinically important difference as applied in rheumatology: an OMERACT Rasch Working Group systematic review and critique. *J Rheumatol* 2016; 43: 194–202.
- 64 Maltenfort M, Diaz-Ledezma C. Statistics in brief: minimum clinically important difference – availability of reliable estimates. *Clin Orthop Relat Res* 2017; 475: 933–946.
- 65 Moher D, Liberati A, Tetzlaff J, *et al.* The PRISMA Group Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *PLoS Med* 2009; 6: e1000097.
- 66 Alma H, de Jong C, Tsiligianni I, *et al.* Measuring clinically relevant change in health status for patients with COPD: a systematic review and meta-analysis. PROSPERO 2015: CRD42015023221. [www.crd.york.ac.uk/PROSPERO/display\\_record.php?ID=CRD42015023221](http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42015023221) Date last accessed: October 2017.
- 67 Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. The Cochrane Collaboration, 2011. <http://handbook-5-1.cochrane.org/> Date last updated: March 2011.
- 68 Terwee CB, Mokkink LB, Knol DL, *et al.* Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; 21: 651–657.
- 69 Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther* 2009; 17: 163–170.
- 70 Alma H, de Jong C, Jelusic D, *et al.* Health status instruments for patients with COPD in pulmonary rehabilitation: defining a minimal clinically important difference. *NPJ Prim Care Respir Med* 2016; 26: 16041.
- 71 Tsai CL, Hodder RV, Page JH, *et al.* The short-form chronic respiratory disease questionnaire was a valid, reliable, and responsive quality-of-life instrument in acute exacerbations of chronic obstructive pulmonary disease. *J Clin Epidemiol* 2008; 61: 489–497.
- 72 Dodd JW, Hogg L, Nolan J, *et al.* The COPD Assessment Test (CAT): response to pulmonary rehabilitation. A multicentre, prospective study. *Thorax* 2011; 66: 425–429.
- 73 Jones PW, Harding G, Wiklund I, *et al.* Tests of the responsiveness of the COPD Assessment Test following acute exacerbation and pulmonary rehabilitation. *Chest* 2012; 142: 134–140.
- 74 Kocks JW, Tuinenga MG, Uil SM, *et al.* Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. *Respir Res* 2006; 7: 62.
- 75 Kon SS, Canavan JL, Jones SE, *et al.* Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. *Lancet Respir Med* 2014; 2: 195–203.
- 76 Kon SS, Dilaver D, Mittal M, *et al.* The Clinical COPD Questionnaire: response to pulmonary rehabilitation and minimal clinically important difference. *Thorax* 2014; 69: 793–798.
- 77 Kulich K, Keininger DL, Tiplady B, *et al.* Symptoms and impact of COPD assessed by an electronic diary in patients with moderate-to-severe COPD: psychometric results from the SHINE study. *Int J Chron Obstruct Pulmon Dis* 2015; 10: 79–94.
- 78 Nolan CM, Longworth L, Lord J, *et al.* The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. *Thorax* 2016; 71: 493–500.
- 79 Perez T, Arnould B, Grosbois JM, *et al.* Validity, reliability, and responsiveness of a new short Visual Simplified Respiratory Questionnaire (VSRQ) for health-related quality of life assessment in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2009; 4: 9–18.
- 80 Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996; 49: 1215–1219.

- 81 Schünemann HJ, Griffith L, Jaeschke R, *et al.* Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *J Clin Epidemiol* 2003; 56: 1170–1176.
- 82 Smid DE, Franssen FM, Houben-Wilke S, *et al.* Responsiveness and MCID estimates for CAT, CCQ, and HADS in patients with COPD undergoing pulmonary rehabilitation: a prospective analysis. *J Am Med Dir Assoc* 2017; 18: 53–58.
- 83 Tsiligianni IG, van der Molen T, Moraitaki D, *et al.* Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulm Med* 2012; 12: 20.
- 84 Van Stel HF, Maillé AR, Colland VT, *et al.* Interpretation of change and longitudinal validity of the quality of life for respiratory illness questionnaire (QoLRIQ) in inpatient pulmonary rehabilitation. *Qual Life Res* 2003; 12: 133–145.
- 85 Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual Life Outcomes* 2003; 1: 4.
- 86 Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005; 14: 1523–1532.
- 87 Welling JB, Hartman JE, Ten Hacken NH, *et al.* The minimal important difference for the St George's Respiratory Questionnaire in patients with severe COPD. *Eur Respir J* 2015; 46: 1598–1604.
- 88 Wyrwich KW, Metz SM, Kroenke K, *et al.* Measuring patient and clinician perspectives to evaluate change in health related quality of life among patients with chronic obstructive pulmonary disease. *J Gen Intern Med* 2007; 22: 161–170.
- 89 Zanini A, Aiello M, Adamo D, *et al.* Estimation of minimal clinically important difference in EQ-5D visual analog scale score after pulmonary rehabilitation in subjects with COPD. *Respir Care* 2015; 60: 88–95.
- 90 Jones PW, Harding G, Berry P, *et al.* Development and first validation of the COPD Assessment Test. *Eur Respir J* 2009; 34: 648–654.
- 91 Van der Molen T, Willemse BW, Schokker S, *et al.* Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health Qual Life Outcomes* 2003; 1: 13.
- 92 Guyatt GH, Berman LB, Townsend M, *et al.* A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42: 773–778.
- 93 EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199–208.
- 94 Herdman M, Gudex C, Lloyd A, *et al.* Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011; 20: 1727–1736.
- 95 Ware J, Kosinski M, Dewey J. How to Score Version Two of the SF-36 Health Survey. Lincoln, RI, QualityMetric Inc., 2000.
- 96 Jones PW, Quirk FH, Baveystock CM. The St. George's Respiratory Questionnaire. *Respir Med* 1991; 85: 25–31.
- 97 Cave AJ, Atkinson L, Tsiligianni IG, *et al.* Assessment of COPD wellness tools for use in primary care: an IPCRG initiative. *Int J Chron Obstruct Pulmon Dis* 2012; 7: 447–456.
- 98 Alma H, de Jong C, Jelusic D, *et al.* Assessing health status over time: impact of recall period and anchor question on the minimal clinically important difference of COPD health status tools. *Health Qual Life Outcomes* 2018; 16: 130.
- 99 Yoo JY, Kim YS, Kim SS, *et al.* Factors affecting the trajectory of health-related quality of life in COPD patients. *Int J Tuberc Lung Dis* 2016; 20: 738–746.
- 100 Sundh J, Montgomery S, Hasselgren M, *et al.* Change in health status in COPD: a seven-year follow-up cohort study. *NPJ Prim Care Respir Med* 2016; 26: 16073.