

ORIGINAL ARTICLES

Measuring Patient and Clinician Perspectives to Evaluate Change in Health-Related Quality of Life Among Patients with Chronic Obstructive Pulmonary Disease

Kathleen W. Wyrwich, PhD^{1,2}, Stacie M. Metz, MPH³, Kurt Kroenke, MD^{4,5},
William M. Tierney, MD^{4,5}, Ajit N. Babu, MBBS, MPH^{6,7}, and Fredric D. Wolinsky, PhD^{2,8,9}

¹Research Methodology and Health Services Research, Saint Louis University, 3750 Lindell Blvd., McGannon Hall, Room 230, St. Louis, MO 63108, USA; ²Saint Louis University School of Public Health, St. Louis, MO, USA; ³Department of Health, West Chester University, West Chester, PA, USA; ⁴Regenstrief Institute for Health Care, Indianapolis, IN, USA; ⁵Indiana University School of Medicine, Indianapolis, IN, USA; ⁶Veterans Administration Medical Center, St. Louis, MO, USA; ⁷Amrita Institute of Medical Sciences, Kochi, India; ⁸Health Management and Policy, University of Iowa, Iowa City, IA, USA; ⁹Health Services Research, Veterans Affairs Medical Center, Iowa City, IA, USA.

CONTEXT: Many treatments aim to improve patients' health-related quality of life (HRQoL), and many care guidelines suggest assessing symptoms and their impact on HRQoL. However, there is a lack of consensus regarding which HRQoL outcome measures are appropriate to assess, and how much change on those measures depict significant HRQoL improvement.

OBJECTIVE: We used triangulation methods to identify and understand clinically important differences (CIDs) for the amount of change in HRQoL that reflects both health professionals and patients' values, among patients with chronic obstructive pulmonary disease (COPD).

DESIGN, SETTING, AND PARTICIPANTS: We incorporated three perspectives: (1) an expert panel of physicians familiar with the measurement of HRQoL in COPD patients; (2) 610 primary care COPD outpatients who completed baseline and bimonthly follow-up HRQoL interviews over the 12-month study; and (3) the primary care physicians (PCPs; $n=43$) of these outpatients who assessed their patients' disease at baseline and at subsequent PCP visits during the year long study.

MEASUREMENTS: The Chronic Respiratory Disease Questionnaire (CRQ), the Medical Outcomes Study Short Form 36-item survey (SF-36, version 2.0), and global assessments of change from each of the three perspectives for all HRQoL domains.

RESULTS: With few exceptions, the CRQ was able to detect small changes at levels reported by the patients (1–2 points) and their PCPs (1–5 points). These results confirm minimal important difference standards developed in 1989 by Jaeschke et al. anchored on patient-

perceived changes in HRQoL. In general, the expert panel and PCP CIDs were larger than the patient CIDs.

CONCLUSION: This triangulation methodology yielded improved interpretation, understanding, and insights on stakeholder perspectives of CIDs for patient-reported outcomes.

KEY WORDS: health-related quality of life; minimal important difference; chronic obstructive pulmonary disease.

DOI: 10.1007/s11606-006-0063-6

© 2007 Society of General Internal Medicine 2007;22:161–170

INTRODUCTION

The Food and Drug Administration (FDA) recently released a report titled "Innovation or Stagnation: Challenges and Opportunities on the Critical Path to New Medical Products." This report focused on the current deceleration of innovative and pioneering medical therapies capable of reaching the patients who might benefit,¹ and states that:

For many therapeutics, effectiveness criteria are best defined by the practitioners and patients who use the products. Much work needs to be done on clinical trial design and patient-driven outcome measures to ensure that endpoints in new therapeutic areas accurately reflect patient needs and values. Community (health professional and patient) consensus on appropriate outcome measures and therapeutic claims can lay a clear development path for new therapeutics, especially when there is international regulatory harmonization. (p. 24)

Although a sound understanding of the biomechanisms of potentially beneficial medical therapies will always be of great importance, most medical treatments aim to improve patients' symptoms and/or functioning and their effects on health-related quality of life (HRQoL). Whereas it is generally recognized that patient-reported outcomes are relevant, there is no

Received November 25, 2005

Revised June 21, 2006

Accepted September 18, 2006

Published online January 23, 2007

consensus regarding which HRQoL outcome measures are appropriate to assess HRQoL and how much change on those measures depicts significant improvement in HRQoL. The clinically important difference (CID) reflects the amount of change (either improvement or decline) in HRQoL that is meaningful to patients and/or their health care providers. Although some researchers and practitioners will argue that *only* the patient's perspective² should inform what constitutes a meaningful change, it is clear that the FDA has determined that decisions on which measures to use and what CID thresholds constitute small, moderate, and large changes in HRQoL should reflect both health professionals and patients' needs and values.¹

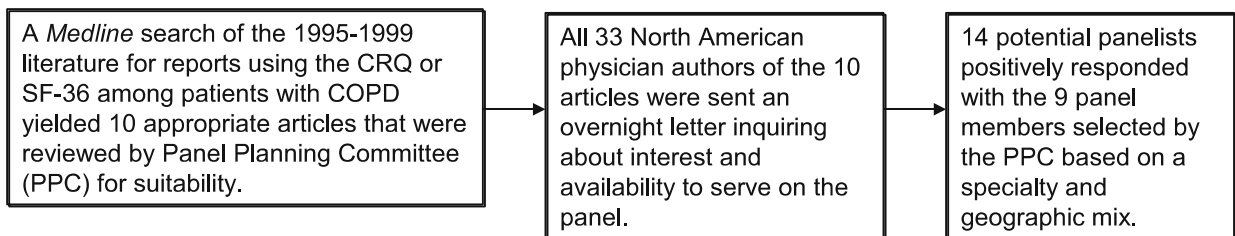
In 1989 Jaeschke et al.³ described a *minimal clinically important difference* (MCID) in HRQoL as "the smallest difference in a score of a domain of interest that patients perceive to be beneficial and that would mandate, in the absence of troublesome side effects and excessive costs, a change in the patient's management." We view CIDs as an expansion of the 1989 term because the CIDs recognize that within a disease condition, unique treatment options can have different risks. Therefore, a patient and clinician may seek a moderate or large CID in HRQoL if the investment and/or risk of a treatment is greater than minimal. Hence, CIDs examine the amount of change associated with a small, moderate, and large improvement or decline in HRQoL. Another popular term for interpret-

ing the magnitude of HRQoL change scores is *minimal important difference*, and is generally linked to the smallest change in HRQoL that is important to patients.²

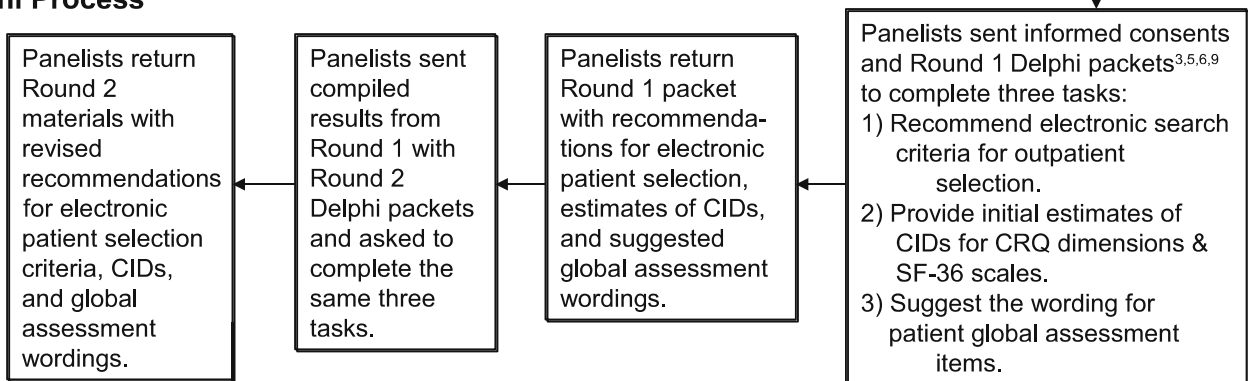
With the support of the Agency for Healthcare Research and Quality, we conducted a 4-year clinical investigation to develop criteria for assessing HRQoL in patients with chronic obstructive pulmonary disease (COPD) that could be used to evaluate therapies. Treatment goals for COPD patients focus on enhancing functioning, relieving dyspnea and other symptoms, improving emotional difficulties, and mastering situations that may irritate this condition.⁴ Therefore, we determined CID thresholds for both a well-known disease-specific HRQoL instrument, the Chronic Respiratory Disease Questionnaire (CRQ)⁵ and the most widely used generic measure of HRQoL, the Medical Outcomes Study Short Form 36-item survey (SF-36, version 2.0).⁶

We selected these two instruments for specific reasons. The CRQ was developed in 1987 by Guyatt et al.⁵ to measure the 4 important HRQoL domains associated with COPD: dyspnea, fatigue, emotional functioning, and mastery (feelings of control over the disease). The items for each of these domains use a 7-point scale, and there are 5, 4, 7, and 4 items for the respective domains. Hence, scores range from 5 to 35 in the dyspnea domain, 4 to 28 in the fatigue, 7 to 49 in emotional functioning, and 4 to 28 in mastery. A unique feature of the dyspnea domain is that each patient identifies 5 activities that are affected by

Selection



Delphi Process



Consensus Process



Figure 1. Expert panel selection, Delphi, and consensus process.

Selection

Electronic selection criteria (determined by expert panel):

- 1) Age \geq 50 years, and
- 2) Either of the following conditions:
 - a) diagnosis of: COPD or obstructive lung disease or emphysema or chronic bronchitis
 - b) medications list containing more than one prescription for oral or inhaled bronchodilators.

List of selected outpatients (n = 3128) and criteria that selected each patient sent to his/her primary care physician (PCP) for confirmation that the patient has COPD (n = 1795).

Enrollment

If the outpatient affirms all screening criteria, they are asked to participate and sign an informed consent. Enrollment data, including patient-specific dyspnea activities is solicited.

Confirmed COPD outpatients with scheduled visits with their PCP are screened for eligibility in the clinic waiting room based on affirmation to having:

- 1) COPD
- 2) experienced breathing problems in the last 4 weeks
- 3) a home telephone and
- 4) adequate hearing.

Participation

Outpatients are telephoned within 48 hours of enrollment to complete baseline interviews on the CRQ and SF-36. Additional baseline measures include: social support, sense of control, religiosity and spirituality, education, employment, income, and a current memory marker.

Every two months across the next year, participants receive follow-up interviews on the CRQ and SF-36, as well as global assessment of change items for each domain or scale of these 2 measures. A memory marker was also recorded to assist recall during the next interview.

If a participant had an exacerbation visit or a scheduled visit at least one month after a follow-up interview, but before their 2 month anniversary, the clinic coordinated an early follow-up interview within 72 hours of the visit using the same format as the regular 2 month follow-up interviews.

Figure 2. Chronic obstructive pulmonary disease (COPD) outpatient selection, enrollment, and participation.

her/his shortness of breath, and then rates their limitations on each of these patient-specific activities across the time of their enrollment. We chose the CRQ instrument because of (1) the repeated demonstration of its validity among patients with chronic airflow difficulties,^{7,8} and (2) its documented acceptability among these patients in longitudinal studies, which is likely attributable to these 5 activities that each patient selects.^{8,9} Similarly, we had 3 reasons for selecting version 2.0 of the SF-36, which has eight scales, each having a 0–100 scoring range: physical functioning, roles affected by physical health, bodily pain, general health, vitality, social functioning, roles affected my emotional health, and mental health. First, the SF-36 is the most widely used HRQoL measure in the world.¹⁰ Second, version 2.0 provides greater item comprehension and improved measurement.⁶ And third, use of this generic instrument provides the potential for comparability of CIDs across different diseases or conditions.¹¹

Our goal was to report and explicate the triangulated results from evidence-based explorations to identify thresholds for the MCID for HRQoL measures and for moderate and large CIDs. Our triangulation investigation examined the perspectives of three different groups with a stake in the management: a 9-member expert panel of North American physicians familiar with both the treatment of COPD patients and the measurement of generic and COPD-specific HRQoL; outpatients with COPD attending 2 Midwestern primary care clinics who participated in bimonthly HRQoL evaluations over 1 year;

and the primary care physicians (PCPs) of these outpatients as they followed their patients clinically over time.

Although several methods were previously proposed for this purpose on interpreting HRQoL change in scores among patient with COPD, none have adequately incorporated clinicians' insight. Indeed, the seminal MCID investigation of Jaeschke et al. reported on the use of a clinical consensus panel to formulate hypothesized values of the MCID for the dyspnea, fatigue, and emotional function domains of the Chronic Heart Disease Questionnaire (CHQ) and the CRQ.³ However, the only credential for panel membership was that panelists had "extensive experience with their [CHQ and CRQ] administration." (p. 409) Therefore, our study was designed to improve on that process by assembling an expert panel of physicians who treat patients with COPD and are accomplished in their research on these patients' HRQoL.¹²

Moreover, the subsequent anchor-based MCID derivations in the Jaeschke et al.³ study did not ask the attending physician if he/she concurred with the patient's assessment of HRQoL change. To move toward a community consensus, our triangulation study collected both patient-reported changes and evaluations of clinically significant change by the patients' PCPs. Furthermore, we wanted to explore what changes in care were associated with clinician-identified changes.

In addition, Jaeschke and colleagues averaged the absolute values of change scores from patients declaring a minimal improvement or decline in HRQoL to find the MCID. However, a

Enrollment

The principal investigator (PI) for each site approached potentially eligible Primary Care Physicians (PCPs) with a substantial number of COPD patients.

Using standard academic detailing methods, the site PI's explained the study and those PCPs choosing to participate signed informed consent forms.

Confirmations

Outpatients selected using the electronic criterion were printed along with the corresponding selection standard(s).

PCPs review each of their outpatients' reports and, if needed, the entire medical chart to confirm the diagnosis of COPD for the outpatient's enrollment eligibility.

Baseline Assessments

For each of their confirmed outpatients enrolled, PCPs completed a baseline assessment indicating:

- 1) The seriousness of the patient's COPD compared to other COPD patients
- 2) The likelihood the patient would be:
 - a) hospitalized for COPD in the next 2 years
 - b) die, directly or indirectly from COPD in the next year
- 3) Whether the patient:
 - a) is on medication(s)
 - b) had laboratory tests or procedures
 - c) has seen a specialist for COPD

Follow-Up Assessments

If an enrolled outpatient had an exacerbation or scheduled office visit with the PCP at least one month, but not more than 2 months after the outpatient's interview anniversary date, the PCP completed a follow-up assessment measuring:

- 1) Has there been a change in the patient's COPD since the last time you completed a study assessment?
- 2) If yes, was the change an improvement or decline?
- 3) Did the patient's COPD change by a minimal, moderate, or large amount?
- 4) Did the clinically significant change in the patient's COPD result in:
 - a) changing medications today?
 - b) ordering laboratory tests/procedures today?
 - c) referring to a specialist today?

Figure 3. Chronic obstructive pulmonary disease (COPD) primary care physician enrollment, confirmations, and baseline and follow-up assessments.

more recent research by Cella et al.¹³ reported that these HRQoL changes are not symmetrical, and patients who reported getting worse had generally larger change scores than those with similar improvements. As a result, we hoped to improve on prior CRQ MCID studies by considering small, moderate, and large improvements in separate analyses from small, moderate, and large declines.

Finally, with the exception of expert panel reports,^{12,14} no known published studies have directly investigated CID anchors and change scores for individuals on the SF-36 in this disease group.¹⁵ Therefore, our current study report is the first known investigation that directly incorporates patients and clinicians' perceptions of important change to the SF-36 scales among patients with COPD.

METHODS

Expert Panel Processes

Figure 1 depicts our procedures for selecting and facilitating the Delphi and consensus panel processes: a more detailed description can be found elsewhere.¹² In addition to working toward consensus on the CID levels that represent small, moderate, and large improvements and declines on the CRQ and SF-36 measures, this expert panel also successfully reached agreement on the criteria for selection of appropriate outpatients with

COPD for our prospective study and advised us on the best content wordings for items used to elicit these outpatients' global assessments of change on each CRQ and SF-36 domain.

Chronic Obstructive Pulmonary Disease Outpatients

Details of patient enrollment and participation are outlined in Figure 2. First, the expert panel-derived criteria for potential enrollment were applied to data in the electronic medical records systems at Wishard Health Services in Indianapolis and the St. Louis Veteran Affairs Medical Center. These criteria are listed in the first box of Figure 2. The generality of these potentially eligible criteria led to many patients being initially selected ($n=3,128$). This electronic search cast the widest administrative net for potential cases. When these patients' PCPs reviewed these selected charts to confirm the potential COPD diagnoses in their individual patients, far fewer COPD cases ($n=1,795$) were identified. The patients thus identified were interviewed for eligibility and interest in this study at their next scheduled clinic visit, but many either refused to participate or were deemed ineligible because they did not know they had lung disease, did not have a phone in their home, or could not hear well enough to use the telephone. Interested and eligible participants signed informed consents. Next, they provided enrollment data by answering several

CRQ**Dyspnea**

Overall, has there been a change in your shortness of breath during your daily activities since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say your shortness of breath is worse, about the same, or better?

Fatigue

Overall, has there been a change in your level of fatigue since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say your fatigue is worse, about the same, or better?

Emotional Function

Overall, has there been a change in how you are feeling emotionally since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say that emotionally you are feeling worse, about the same, or better?

Mastery

Overall, has there been a change in your ability to cope with your breathing problems since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say that your coping is worse, about the same, or better?

SF-36**Physical Functioning**

Overall, has there been a change in your ability to perform physical activities since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say your ability to perform physical activities is worse, about the same, or better?

Role-Physical

Overall, has there been a change in your physical ability to get things done since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say your physical ability to get things done is worse, about the same, or better?

Bodily Pain

Overall, has there been a change in your pain since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say your pain is worse, about the same, or better?

General Health Perception

Overall, has there been a change in how you would rate your general health since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you rate your general health as being worse, about the same, or better?

Mental Health

Overall, has there been a change in how you would rate your mental health since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you rate your mental health as being worse, about the same, or better?

Vitality

Overall, has there been a change in your energy level since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say your energy level is worse, about the same, or better?

Social Functioning

Overall, has there been a change in your social activities because of your health since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say your participation in social activities because of your health is worse, about the same, or better?

Role-Emotional

Overall, has there been a change in your emotional ability to get things done since your last telephone interview on [PRIOR INTERVIEW DATE]? Would you say your emotional ability to get things done is worse, about the same, or better?

Figure 4. Global change assessment items for the Chronic Respiratory Disease Questionnaire (CRQ) domains and Medical Outcomes Study Short Form 36-item survey (SF-36) scales.

demographic and comorbidity items, as well as selected the five patient-specific activity items for the CRQ's dyspnea domain. Within 48 hours of their enrollment in the clinic, they were contacted by telephone for their baseline HRQoL interviews.

The baseline interviews lasted on average of 35 minutes and included the CRQ, the SF-36, the American Board of Internal Medicine Patient Satisfaction Questionnaire,¹⁶ the personal stress scale from the National Opinion Research Council National Health Survey,¹⁷ a 5-item subset of the Medical Outcomes Study social support scale,¹⁸ Mirowsky and Ross' sense of control measure,¹⁹ the 2-summary religiosity and spirituality items from the Fetzer instrument,²⁰ educational attainment, employment status, perceived income adequacy, and income. To assist our participants with recall at future telephone interviews, we employed the use of memory markers at the end of each interview that allowed participants to record a significant or memorable aspect about their day ("I had lunch with my sister Frances today"). This memory marker was then read back to the participant at their next interview to assist with the recall.

Subsequently, Louis Harris and Associates conducted follow-up telephone interviews every 2 months for a 1-year period during which they readministered the CRQ and the SF-36, as well as global assessments of change for each relevant dimension. However, if during the year an outpatient visited the PCP at least 1 month—but not more than 2 months—after a telephone interview, this triggered an early follow-up telephone interview within 72 hours after the PCP office visit. These *visit-triggered* interviews were intended to capture the HRQoL of some outpatients at a juncture when PCP assessments of disease-related changes could also be obtained. Moreover, we hoped that many of these early follow-up interviews would reflect extreme fluctuations in HRQoL because of exacerbations that prompted such PCP visits.

Primary Care Physicians

Forty-three PCPs at the Indianapolis²⁹ and the St. Louis¹⁴ sites volunteered to participate in this study as depicted in Figure 3. By confirming disease eligibility; completing baseline evaluations of disease severity; estimating the potential for hospitalization and mortality; and indicating referrals, tests, and medication use, these PCPs provided necessary and direct input from health care professionals both before and during patient enrollments. If a patient participant returned for a PCP office visit during the participant's year of enrollment, the PCP also assessed disease-related change over time so that PCP-rated change could be directly compared (linked) to the patients' reports of change in their HRQoL. If the PCP rated a participant to have a clinically significant change in their COPD, the PCP was also prompted to evaluate the change as an improvement or a decline; the magnitude of that change (small, moderate, or large); and whether the change resulted in a medication change, ordering laboratory test or procedures, or a referral to a specialist. We provided no specific definitions for the change assessments (no change; or small, moderate, or large improvement or decline), and each party (patient and PCP) was blinded to the other's assessment.

Statistical Procedures

We calculated each CRQ domain or SF-36 scale change scores (time 2–time 1, where time 2=current follow up assessment

and time 1=immediately previous bimonthly assessment) for each adjacent pair of completed interviews. The domain or scale change scores were then categorized based on the patient's transition rating index (TRI) for the dimension using the global assessments of change in HRQoL that were included in each follow-up interview. A complete list of the global assessment items used for each of the 4 CRQ domains and 8 SF-36 scales are provided in Figure 4. If the participant responded that he/she was better, we then asked him/her to describe by how much better using a TRI of 1 (almost the same, hardly any better) to 7 (a very great deal better). Likewise, if the participant was worse, he/she provided a –1 (almost the same, hardly any worse) to –7 (a very great deal worse) TRI. The TRI equaled 0 for those participants who reported being "about the same." The TRIs were classified using the traditional approach described by Juniper et al.²¹ and endorsed by Jaeschke et al. (Fig. 5).

Once classified into the appropriate change categories, we determined the optimal thresholds for no change, and small, moderate, and large improvements and declines for each HRQoL measure by averaging all of the change scores within a patient-perceived TRI classification.³ We used the PCP's blinded assessment of change in the patient's COPD because the last linked clinical encounter in the same manner to classify linked patient change scores in all HRQoL measures before averaging (Fig. 3). All analyses were conducted using SPSS, version 12.0.²²

Large Decline	}	-7 A very great deal better
		-6 A great deal worse
Moderate Decline	}	-5 A good deal worse
		-4 Moderately worse
Small Decline	}	-3 Somewhat worse
		-2 A little worse
No Change	}	-1 Almost the same, hardly any worse at all
		0 About the same
		1 Almost the same, hardly any better at all
Small Improvement	}	2 A little better
		3 Somewhat better
Moderate Improvement	}	4 Moderately better
		5 A good deal better
Large Improvement	}	6 A great deal better
		7 A very great deal better

Figure 5. Transition rating index for global estimates of change.

Table 1. Recommendations for Small, Moderate, and Large Change Thresholds on the Chronic Respiratory Disease Questionnaire (CRQ) and Medical Outcomes Study Short Form 36-Item Survey (SF-36), Version 2.0 by the North American Physician Expert Panel

HRQoL measure	Value of one state change	Number of items	Small change	Moderate change	Large change
CRQ					
Dyspnea	1	5	3	6	9
Fatigue	1	4	2	4	6
Emotional	1	7	5	10	15
Mastery	1	4	3	6	9
SF-36					
Physical functioning	5	10	10	20	30
Role physical	6.25	4	12.5	25	37.5
Bodily pain	10	2	10	20	30
General health	5	5	10	20	30
Vitality	6.25	4	12.5	25	37.5
Social functioning	12.5	2	12.5	25	37.5
Role emotional	8.33	3	8.33	16.67	25
Mental health	5	5	10	20	30

The panel reached consensus that, in the absence of evidence to the contrary, all of the thresholds were symmetrical (i.e., improvement and decline magnitudes were equivalent).

RESULTS

After consideration of the panel’s premeeting Delphi activities, each member of this expert panel was asked to articulate their personal understandings of a clinically important change on each CRQ domain and SF-36 scale at the face-to-face consensus meeting. Panelists then agreed to calculate a unit of measurement for each domain and scale named a *state change*, which is equivalent to the amount of change in the total score on most domain or scale items that results in moving up just one response choice. Using the state change metric, this North American physician expert panel reached consensus on the CIDs for small, moderate, and large improvements and declines (Table 1). State changes for the CRQ dimensions were always 1 unit in magnitude, but the SF-36 state changes varied from 5 to 12.5 units on the 0–100 scales. The expert panel recommended that small CID thresholds were at least 2 state changes for all CRQ dimensions, and at least 1 state change for each SF-36 scale. Although the panel process was quite insightful, it is important to note that our investigation had *a priori* designated the expert panel CIDs, which would be the least important of the three perspectives sampled in this study.¹²

Table 2 presents the characteristics of our COPD outpatient sample at baseline. These data reflect the considerable socioeconomic burdens this population faced, with one quarter being self-identified minorities, over half never having completed high school, and two thirds reporting annual incomes below \$20,000. In addition, nearly all COPD enrollees had fair or poor self-rated health, were either unable to work or had retired, and had a smoking history of more than 20 pack years. Of the 610 outpatients who completed baseline interviews, 554 (91%) completed their first follow-up interview 2 months later. The 4-, 6-, 8-, 10-, and 12-month completion rates were 522 (86%), 504 (83%), 484 (79%), 462 (76%), and 462 (76%),

respectively, and demonstrated excellent cooperation from this generally disadvantaged patient sample.

During the 2,988 follow-up interviews, most participants (59–74%) reported no change or being about the same compared to their prior interview when asked to make a global assessment of change for each CRQ dimension and SF-36 scale (left-hand columns of Table 3). Of these 2,988 follow-up interviews, 410 were linked to PCP office visits. Therefore, we had both the patient’s assessment of change and the PCP’s assessment of change in the patient’s COPD for these linked interviews, which occurred mostly within 24 hours of the visit with a maximum allowance of 72 hours between the visit and the follow-up patient interview. Most PCPs (80%) reported no change in the patient’s condition. As a result, there are few or no patient/PCP-linked encounters identifying either improve-

Table 2. Baseline Demographics and Health Characteristics among Patients with Chronic Obstructive Pulmonary Disease (COPD) Attending Midwest Medical Centers (August 2000–November 2001)

	COPD outpatients, N (%) or mean
Age	
50–54	70 (11.5)
55–64	195 (32.0)
65–74	228 (37.4)
>75	117 (19.2)
Sex	
Female	217 (35.6)
Male	393 (64.4)
Race	
White	467 (76.6)
Black	128 (21.0)
Other	14 (2.3)
Unknown	1 (0.2)
Education	
<High school	341 (55.9)
High school	150 (24.6)
>High school	118 (19.3)
Unknown	1 (0.2)
Annual household income	
<10,000	150 (24.6)
10,000–14,999	138 (22.6)
15,000–19,999	127 (20.8)
20,000–24,999	84 (13.8)
≥\$25,000	81 (13.3)
Unknown	30 (4.9)
Employment status	
Employed for wages	63 (10.3)
Unable to work	246 (40.3)
Retired	267 (43.8)
Others	34 (5.6)
Patient-reported general health	
Excellent	8 (1.3)
Very good	35 (5.7)
Good	118 (19.3)
Fair	241 (39.5)
Poor	208 (34.1)
Smoking status	
Never	37 (6.1)
<20 pack y	157 (25.7)
≥20 pack y	409 (67.0)
Unknown	7 (1.1)
Sense of control*	
	18.7
Religiosity†	
	63.9
Social support†	
	64.5
Stress†	
	55.9
Patient satisfaction†	
	81.7

*Ranges from 0 (denies responsibility) to 100 (takes responsibility).

†Transformed from 0=worst (least religious, lowest support, most stress, and least satisfaction) to 100=best.

Table 3. Sample Size (Percentages) within Change Categorizations by Patient and Primary Care Physician (PCP) Reports of Change on the Chronic Respiratory Disease Questionnaire (CRQ) Domains and Medical Outcomes Study Short Form 36-Item Survey (SF-36) Scales

HRQoL measure	Patient-perceived changes						PCP-perceived changes							
	Large decline	Moderate decline	Small decline	No change	Small improve	Moderate improve	Large improve	Large decline	Moderate decline	Small decline	No change	Small improve	Moderate improve	Large improve
CRQ														
Dyspnea	121 (4)	198 (7)	313 (11)	1,807 (63)	212 (7)	162 (6)	68 (2)	3 (1)	23 (6)	28 (7)	311 (80)	15 (4)	8 (2)	0 (0)
Emotional	70 (2)	133 (5)	208 (7)	1,850 (64)	266 (9)	240 (8)	115 (4)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
Fatigue	93 (3)	171 (6)	349 (12)	1,867 (65)	197 (7)	152 (5)	54 (2)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
Mastery	52 (2)	82 (3)	158 (5)	2,053 (71)	247 (9)	196 (7)	96 (3)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
SF-36														
Physical functioning	109 (4)	216 (8)	242 (8)	1,944 (68)	188 (7)	132 (5)	49 (2)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
Role physical	111 (4)	190 (7)	284 (10)	2,007 (70)	139 (5)	110 (4)	42 (1)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
Bodily pain	135 (5)	274 (10)	314 (11)	1,790 (62)	144 (5)	146 (5)	75 (3)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
General health	112 (4)	214 (7)	406 (14)	1,698 (59)	218 (8)	154 (5)	70 (2)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
Vitality	113 (4)	219 (8)	368 (13)	1,778 (62)	199 (7)	150 (5)	57 (2)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
Social functioning	115 (4)	152 (5)	200 (7)	2,110 (74)	122 (4)	117 (4)	46 (2)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
Role emotional	70 (2)	133 (5)	208 (7)	2,132 (74)	150 (5)	126 (4)	59 (2)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)
Mental health	64 (2)	118 (4)	195 (7)	2,031 (71)	181 (6)	185 (6)	98 (3)	3 (1)	23 (6)	30 (7)	330 (80)	15 (4)	9 (2)	0 (0)

ments or declines (right hand columns of Table 3). Quadratic weighted kappa statistics for agreement²³ between patient-reported changes among those with linked PCP assessment and the associated PCP assessments of change ranged from 0.29 (dyspnea domain of the CRQ) to 0.07 (role emotional scale of the SF-36), demonstrating poor agreement among these linked raters.

For the 80 clinical encounters when a PCP reported a clinically significant change, half (n=40) were coupled with a change in medication. Of these, most often (n=37) the PCP stated that there had been a decline in the participant's COPD. Few linked PCP encounters with clinically significant changes resulted in ordering laboratory tests or procedures (n=16) or referral to a specialist (n=5), but nearly all of the patients requiring these changes in care had a PCP-rated decline in their condition.

Table 4 reports the mean thresholds, respectively, for the change scores (time 2–time 1) of the patients classified into each cell of Table 3. With few exceptions, the CRQ measures were able to detect small changes at levels reported by the patients and their PCPs. This is evidenced by the patient-perceived small decline thresholds being 1–2 points (equivalent to 1 or 2 state changes), and the optimal small improvement thresholds for both patients and PCPs being 1–5 points. An exception involves the CRQ mastery dimension and small improvements rated by PCPs, and there is prior corroboration that longitudinal validity is problematic for this domain.⁵ Patient-perceived moderate and large improvements and declines also seem credible, again with the exception of the mastery domain. The PCP-perceived average change scores for those with clinically significant declines do not display trends seen in the patient-perceived results, and PCP-related moderate improvement thresholds for dyspnea and mastery are somewhat uninterrupted given the no change and small improvement results.

In contrast, the SF-36 results are more problematic because small changes rated by patients or their PCPs yielded average change thresholds that were generally smaller than one state change value (which range from 5 to 12.5, see Table 1) for the SF-36 scales. It is counterintuitive to recommend that an important change has happened when a threshold is at a level less than the resulting scale scores change that occurs when an individual shifts to an adjacent response category on only one item in a scale.^{12,14} Our SF-36 results demonstrate that among the small improvement and decline thresholds, only the patient-perceived changes in mental health (small decline) and role physical (small improvement), and the PCP-classified changes for physical functioning, role physical, and general health (small improvement) meet this “at least one state change” criterion. Although some moderate and large change category thresholds do meet the state change criterion, many do not and it is difficult to interpret trends in many of the SF-36 scale results.

DISCUSSION

The results of our triangulation investigation provide several “off the shelf” CIDs for the CRQ domains when used to benchmark important changes in HRQoL domains in similar patient populations. In so doing, they also serve as a procedural outline for future multiple stakeholder studies to determine CIDs in other HRQoL measures and/or other patient populations, as well as an insight to the complications that occur when

Table 4. Mean Change Scores Using Patient and Primary Care Physician (PCP) Reports of Change on the Chronic Respiratory Disease Questionnaire (CRQ) Domains and Medical Outcomes Study Short Form 36-Item Survey (SF-36) Scales (Scores Truncated to Integer Values Because of Space Constraints)

HRQoL measure	Patient-perceived changes					PCP-perceived changes								
	Large decline	Moderate decline	Small decline	No change	Small improve	Moderate improve	Large improve	Large decline	Moderate decline	Small decline	No change	Small improve	Moderate improve	Large improve
CRQ														
Dyspnea	-3	-2	-1	0	2	2	5	-4	0	1	2	5	1	-
Emotional	-4	-2	-2	0	1	2	3	5	1	2	1	2	4	-
Fatigue	-3	-2	-1	0	1	2	4	0	0	0	1	1	2	-
Mastery	-1	-2	-2	0	1	1	1	-1	-2	-1	1	0	1	-
SF-36														
Physical functioning	-6	-4	-2	0	4	4	6	-17	-2	-1	-1	5	3	-
Role physical	-12	-8	-3	1	8	12	11	8	2	3	3	9	17	-
Bodily pain	-9	-7	-3	1	6	8	6	7	-7	-2	3	2	8	-
General health	-6	-3	-3	0	3	1	11	15	-4	-1	0	11	8	-
Vitality	-7	-7	-5	1	5	9	10	2	0	0	0	2	3	-
Social functioning	-15	-9	-6	1	6	10	6	4	-1	-5	2	6	14	-
Role emotional	-11	-7	-5	0	7	8	9	-8	3	1	1	2	16	-
Mental health	-9	-11	-6	0	4	6	8	17	-1	4	1	4	7	-

three streams of data to not converge. Beyond these products, our results also demonstrate 2 previously documented HRQoL measurement issues: (1) the merits of disease-specific measures for monitoring individual HRQoL change;^{24,25} and (2) the gap between expert panelists, patients, and their PCPs' perceptions of change^{26,27} in individual patients.

Replicating our findings in COPD populations that differ from our sample is important to determine the generalizability of our CID results to this disease group. Our expert panel's results matched the 1989 CHQ/CRQ clinical consensus panel recommendations.⁵ Our patient-perceived thresholds are similar, and in some domains, a bit lower than those ascertained by Jaeschke and colleagues' pioneering HRQoL interpretation study.⁵

Despite our substantial sample size (n=610 at baseline), there was a limited number of PCP-reported improvements (n=24) or declines (n=56) in HRQoL among these outpatients over a year of enrollment, with no PCP reporting a large improvement. The scarcity of change scores available for the PCP-classified CID estimates not only questions their stability, but also underscores the need for very large (if not prohibitive) sample sizes to observe a substantial number of improvements or declines within small time intervals. In addition, there was poor agreement between the patient and PCP assessments of change, especially in these problematic cells.

These results provide investigators with benchmarks for evaluating patient-reported changes in HRQoL using these measures with COPD patients. Hence, they are a valuable resource for future interventions with primary and secondary HRQoL outcomes, demonstrating not just a statistically significant difference but also a clinically significant difference in patient change, which is currently recommended by the FDA for patient-reported outcome measures. Nonetheless, there are important limitations to these results. First, because of busy clinic schedules, we did not ask PCPs to rate their perceived changes in each HRQoL domain and scale at all linked patient encounters; instead, we asked PCPs to evaluate whether the participant had a clinically significant change in their COPD. Therefore, it is possible that patients and PCPs were rating different constructs. Yet, even the most COPD-related CRQ dyspnea domain, which assessed change in activities affected by shortness of breath, yielded a dismal weighted kappa value ($\kappa=.29$) when compared to the PCPs' COPD change rating among linked patients. The poor agreement in statistics are numerically affected by the overwhelming number of patients in the "no change" category, but nonetheless demonstrate a need for increased dialogue between patients and physicians regarding when important changes occur.

We hoped that our PCP ratings of clinically significant changes in COPD would also be documented through changes in care. Of the 80 encounters where PCPs did note a change, 47 of these were associated with a positive response to at least one change in care item. Moreover, most of the patients who had a clinically significant PCP change rating but no reported change in care action were rated as improved. Although we recognize instances where an important change in condition may not require a change in care, we encourage other researchers interested in documenting clinicians' assessments of change to also explore treatment responses to these evaluations.²⁸

Another limitation in this field of inquiry is the retrospective global change assessments that anchor the patients' change classifications. Several studies have demonstrated that patients have great difficulty in accurately remembering their

prior health state (time 1) and then comparing it to their current health (time 2).^{29–31} As a result, the change assessments are often highly correlated with current health, and uncorrelated with the time 1 assessment. To assist with this recall, we employed the use of memory markers to potentially improve the accuracy of the anchor-based methodology. Nonetheless, the use of these single-item global assessments to classify change in a domain measured through multiple items at each time point may be intuitively flawed, despite the fact that it does reflect the patient's overall perceptions of HRQoL changes. This ambiguity deserves further investigation that may necessitate improved patient-reported anchors. In addition, there is a potential problem of response shift³² from changes in patients or PCPs understanding of conceptualization of (1) the domain or scale being assessed over their enrollment span; or (2) the response options used with the HRQoL instruments, the TRIs, or the PCP assessments.

Our findings have implications for clinical research, as suggested by the FDA's call for community consensus, and for clinical practice. In these data, patients tend to report smaller changes as more clinically meaningful than do physicians. Conversely, physicians may look for changes of a magnitude that warrant clinical actions or show unquestionable improvements. The demonstrated disconnection between these two perspectives in achieving agreement regarding *who has changed* and the *level of important change* is needed. Integrating both patient and physician perspectives is important to improve research applications and interpretation of these important outcomes,³³ and patient–physician communications regarding health status, treatment options, and shared decision making for living with chronic disease like COPD.

Acknowledgement: This research was funded by grants from the Agency for Healthcare Research and Quality to Dr. Wolinsky (R01 HS10234) and Dr. Wyrwich (K02 HS11635).

Potential Financial Conflicts of Interest: None disclosed.

Corresponding Author: Kathleen W. Wyrwich, Research Methodology and Health Services Research, Saint Louis University, 3750 Lindell Blvd., McGannon Hall, Room 230, St. Louis, MO 63108, USA (e-mail: wyrwichk@slu.edu).

REFERENCES

- Food and Drug Administration. Innovation or Stagnation: Challenges and Opportunities on the Critical Path to New Medical Products. Washington, DC: US Department of Health and Human Services; March 2004.
- Schunemann H, Guyatt GH. Commentary—goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res.* 2005;40(2):593.
- Jaeschke R, Singer J, Guyatt G. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10:407–15.
- Global Initiative for Chronic Obstructive Lung Disease (GOLD). <http://www.goldcopd.com>; October 26, 2001.
- Guyatt G, Berman L, Townsend M, Pugsley S, Chambers L. A measure of quality of life for clinical trials in chronic lung disease. *Thorax.* 1987;42:773–8.
- Ware J, Kosinski M, Dewey J. How to Score Version Two of the SF-36 Health Survey. Lincoln, RI: QualityMetric Inc.; 2000.
- Lacasse Y, Wong E, Guyatt G. A systematic overview of the measurement properties of the chronic respiratory questionnaire. *Can Respir J.* 1997;4(3):131–9.
- Schunemann H, Griffith L, Jaeschke R, et al. A comparison of the original chronic respiratory questionnaire with a standardized version. *Chest.* 2003;124(4):1421–9.
- Wyrwich K, Tierney W, Wolinsky F. Further evidence supporting a SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol.* 1999;52(9):861–73.
- Brazier J, Harper R, Jones N. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *Br Med J.* 1992;305:160–4.
- Patrick D, Deyo R. Generic and disease-specific measures in assessing quality of life. *Med Care.* 1989;27:S217–32.
- Wyrwich K, Fihn S, Tierney W, Kroenke K, Babu A, Wolinsky F. Clinically important differences in health-related quality of life for patients with chronic obstructive pulmonary disease: an expert panel report. *J Gen Intern Med.* 2003;18(3):196–202.
- Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res.* 2002;11:207–21.
- Wyrwich KW, Tierney WM, Babu AN, Kroenke K, Wolinsky FD. A comparison of clinically important differences in health-related quality of life for patients with chronic lung disease, asthma, or heart disease. *Health Serv Res.* 2005;40(2):577–92.
- Curtis JR, Patrick DL. The assessment of health status among patients with COPD. *Eur Respir J.* 2003;21:36S–45S.
- Webster GD. Final Report of the ABIM Patient Satisfaction Project. Philadelphia: American Board of Internal Medicine; 1988.
- McHorney CA, Lerner J. The 1990 NORC National Health Survey: Documentation and Codebook. Chicago: National Opinion Research Council; 1991.
- Sherbourne CD, Stewart AL. The MOS social support survey. *Soc Sci Med.* 1991;32:705–14.
- Mirowsky J, Ross CE. Eliminating defense and agreement bias from measures of the sense of control: a 2×2 index. *Soc Psychol Q.* 1991;54:127–45.
- Fetzer Institute. Multidimensional Measurement of Religiosity/Spirituality for Use in Health Research. Kalamazoo: John E. Fetzer Institute; 1999.
- Juniper E, Guyatt G, Willan A, Griffith L. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol.* 1994;47:81–7.
- Statistical Package for the Social Sciences. SPSS for Windows, release 13.0.0. Chicago: SPSS, Inc.; 2005.
- Streiner D, Norman G. Reliability. In: *Health Measurement Scales*, 3rd edn., Chap. 8. Oxford: Oxford University Press; 2003.
- Jones G, Jenkinson C, Kennedy S. Evaluating the responsiveness of the endometriosis health profile questionnaire: the EHP-30. *Qual Life Res.* 2004;13:705–13.
- Singh S, Sodergren S, Hyland M, Williams J, Morgan M. A comparison of three disease-specific and two generic health-status measures to evaluate the outcome of pulmonary rehabilitation in COPD. *Respir Med.* 2001;95:71–7.
- Juniper E, Price D, Stampone P, Creemers J, Mol S, Fireman P. Clinically important improvements in asthma-specific quality of life, but no difference in conventional clinical indexes in patients changed from conventional beclomethasone dipropionate to approximately half the dose of extrafine beclomethasone dipropionate. *Chest.* 2002;121:1824–32.
- Chauhan C, Eppard W, Perroti J. The patient advocate perspective on assessing the clinical significance for quality-of-life measures. *J Cancer Integr Med.* 2004;2(4):155–7.
- Naylor CD, Llewellyn-Thomas HA. Can there be a more patient-centred approach to determining clinically important effect sizes for randomised trials? *J Clin Epidemiol.* 1994;47:787–95.
- Norman G, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lessons of Cronbach. *J Clin Epidemiol.* 1997;50(8):869–79.
- Guyatt G, Norman G, et al. A critical look at transition ratings. *J Clin Epidemiol.* 2002;55:900–8.
- Stratford P, Binkley J, Solomon P, et al. Defining the minimal level of detectable change for the Roland Morris Questionnaire. *Phys Ther.* 1996;76:359–65.
- Schwartz C, Sprangers M. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med.* 1999;48:1531–48.
- Pulmonary Allergy Drugs Advisory Committee. New Drug Application (NDA) 21-395, Spiriva (Tiotropium bromide). Gaithersburg, MD: Food and Drug Administration; 2002.