



“From Where I Stand”: using multiple anchors yields different benchmarks for meaningful improvement and worsening in the rheumatoid arthritis flare questionnaire (RA-FQ)

Susan J. Bartlett^{1,2,3} · Vivian P. Bykerk⁴ · Orit Schieir⁵ · Marie-France Valois¹ · Janet E. Pope⁶ · Gilles Boire⁷ · Carol Hitchon⁸ · Glen Hazlewood^{3,9} · Louis Bessette¹⁰ · Edward Keystone¹¹ · Carter Thorne¹² · Diane Tin¹² · Clifton O. Bingham III¹³ On behalf of CATCH Investigators

Accepted: 2 August 2022 / Published online: 8 September 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Purpose The Rheumatoid Arthritis Flare Questionnaire (RA-FQ) is a patient-reported measure of disease activity in RA. We estimated minimal and meaningful change from the perspective of RA patients, physicians, and using a disease activity index.

Methods Data were from 3- to 6-month visits of adults with early RA enrolled in the Canadian Early Arthritis Cohort. Participants completed the RA-FQ, the Patient Global Assessment of RA, and the Patient Global Change Impression at consecutive visits. Rheumatologists recorded joint counts and MD Global. Clinical Disease Activity Index (CDAI) scores were computed. We compared mean RA-FQ change across categories using patients, physicians, and CDAI anchors.

Results The 808 adults were mostly white (84%) women (71%) with a mean age of 55 and moderate-high disease activity (85%) at enrollment. At V2, 79% of patients classified their RA as changed; 59% were better and 20% were worse. Patients reporting they were a lot worse had a mean RA-FQ increase of 8.9 points, whereas those who were a lot better had a -6.0 decrease. Minimal worsening and improvement were associated with a mean 4.7 and -1.8 change in RA-FQ, respectively, while patients rating their RA unchanged had stable scores. Physician and CDAI classified more patients as worse than patients, and minimal and meaningful RA-FQ thresholds differed by group.

Conclusion Thresholds to identify meaningful change vary by anchor used. These data offer new evidence demonstrating robust psychometric properties of the RA-FQ and offer guidance about improvement or worsening, supporting its use in RA care, research, and decision-making.

Keywords Rheumatoid arthritis · Patient-reported outcomes · Flare · Minimally important differences

✉ Susan J. Bartlett
Susan.bartlett@mcgill.ca

¹ Centre for Outcomes Research and Evaluation, McGill University, 5252 de Maisonneuve, #3D.57, Montreal, QC H4A 3S5, Canada

² Research Institute, McGill University Health Center, Montreal, QC, Canada

³ Arthritis Research Canada, Vancouver, Canada

⁴ Hospital for Special Surgery, Weill Cornell Medical College, New York, NY, USA

⁵ Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

⁶ St. Joseph's Health Care London, University of Western Ontario, London, ON, Canada

⁷ University of Sherbrooke, Sherbrooke, QC, Canada

⁸ University of Manitoba, Winnipeg, MB, Canada

⁹ University of Calgary, Calgary, AB, Canada

¹⁰ University of Laval, Quebec City, QC, Canada

¹¹ Mount Sinai Hospital, Toronto, ON, Canada

¹² The Arthritis Center, Newmarket, ON, Canada

¹³ Johns Hopkins Medicine, Baltimore, MD, USA

Introduction

Rheumatoid arthritis (RA) affects up to 1% of adults and is three times more common in women than men [1, 2]. The joint pain, swelling, and damage associated with RA greatly affect physical, emotional, and social health and significantly impair health-related quality of life (HRQL) [3–5]. Up to half of people with RA are unable to work due to disability after 10 years, a trend that remains unchanged despite the introduction of biologics 20 years ago [6]. One reason for this may be that people with RA experience both unexpected temporary and sustained increases (flares) in RA activity which vary in frequency, severity, and impact. Disease flares are periods of increased inflammatory disease and have been defined as a cluster of symptoms of sufficient duration and intensity to require a review and possible change of existing treatment [7]. RA flares are important to identify and manage as they contribute to joint damage, disability, and increased cardiovascular disease that greatly impact health-related quality of life (HRQL) in people with RA [8].

Current treatment guidelines emphasize tight disease control and a “treat-to-target approach” in RA with remission or low disease activity as the treatment goal to reduce disability and improve HRQL [9]. To achieve this, RA is closely monitored, a patient’s level of disease activity is calculated using a validated algorithm, and therapies adjusted as needed until this target is reached. However, reliable monitoring of disease activity in real time is hampered by the lack of a gold standard and use of different approaches and tools of varying complexity that can yield different results.

Patient-reported outcomes (PROs) have been a cornerstone of RA disease monitoring since the late 1970s when the Patient Global Assessment (PGA), an 11-point numeric rating scale (NRS) was introduced [10]. In clinical practice, physicians monitor the number of swollen and tender joints and ask patients about their pain and disability to form an overall impression of disease activity (MD Global Assessment [MDGA]). Joint counts and MDGA constitute clinician-reported outcomes (CLIN-ROs). In the early 1990s, researchers created a composite disease activity index by combining the PGA with MD joint counts and a biomarker (first ESR and later CRP) in a weighted algorithm to create the Disease Activity Score (DAS) [11]. Thresholds were identified to classify RA patients as being in remission, low, moderate, or high disease activity states and better match treatment to current disease activity. The inclusion of biomarkers largely relegated the DAS to research given the need for laboratory results and complex calculations. More recently, the PGA and CLIN-ROs (i.e., joint counts and MDGA) were

combined into a simple summative score known as the Clinical Disease Activity Index (CDAI). CDAI can be easily calculated during clinical encounters to inform treatment decision-making at the visit [12]. However, growing evidence suggests that CLIN-ROs overestimate improvement and underestimate progression of disease activity [13]. This prompted international members of the Outcome Measures in Rheumatology (OMERACT) RA Flare Working Group to co-create with patients a new composite PRO of RA disease activity—the RA Flare Questionnaire (RA-FQ) [14]. In the RA-FQ, patients rate RA symptoms and function over the past week [15]. International patient focus groups were conducted to identify relevant domains [16] and items were refined through a Delphi exercise that included several hundred patients, clinicians, researchers, and other stakeholders [17]. The RA-FQ has been shown to be valid, reliable, and responsive to change in international clinical trials and observational studies [17, 18]. To increase utility, evidence supporting score interpretation is needed.

The minimally important difference (MID) has been defined as the average change score of patients reporting a “small” or “minimal” difference [19, 20]. Meaningful change is average score change where RA patients consider themselves to be *a lot better* or *a lot worse*) [21, 22]. In RA, meaningful change helps to establish if treatment is sufficiently controlling RA inflammation to allow patients to feel and function better in everyday life and as such, for patient-reported outcomes, patients should define these values [23]. For regulatory activities, the U.S. FDA also recommends identifying meaningful within-patient score change from the patient perspective [24]; however, in clinical practice, physician assessment and CDAI level often drive treatment decisions given the emphasis on treat-to-target for RA management.

Our goal was to identify the mean minimal and meaningful score change for the RA-FQ between groups of patients who were improving, worsening, or experienced no change in disease activity. Scores were obtained at consecutive visits 3 months apart from patients, treating physicians, and in relation to changes in CDAI levels. We also explored score changes in response to both improving and worsening RA as others have noted these may vary based on the direction of change [25].

Methods

Design

We used data from two consecutive visits (the 3- and 6-month follow-ups) of RA patients enrolled in the Canadian Early Arthritis Cohort (CATCH). CATCH is a prospective

observational inception cohort of adults with early RA who are enrolled around the time of diagnosis and beginning treatment and followed at pre-determined intervals [26]. These time points were chosen as 3 months which is a typical time frame used to judge the effectiveness of initial RA treatments, and it also allowed for sufficient variation in patients experiencing improvement and worsening at the second visit. Ethics approval was previously obtained at each of the 16 participating CATCH sites across Canada. Written informed consent was obtained from participants at enrollment, and the study was conducted in accordance with the Declaration of Helsinki.

Participants

Participants were adults 18+ years of age who were enrolled in the Canadian Early Arthritis Cohort (CATCH) from 2011 when the RA-FQ was implemented to March 2017 (when the Patient Global Impression of Change (PGIC) was discontinued) and who had RA-FQ and PGIC scores available at the 3- and 6-month visits.

Outcomes

Rheumatoid arthritis flare questionnaire (RA-FQ)

The RA-FQ contains five items that ask respondents to rate their pain, physical function, stiffness, fatigue, and participation over the past week using 11-point NRS (0 = none to 10 = severe). The RA-FQ was co-created with patients and in accordance with best practice methods [27–29]. The conceptual framework evolved from international focus groups with RA patients [16] and Delphi exercises with RA patients, clinicians, researchers, and other stakeholders [3]. Psychometric performance of the RA-FQ was examined in multiple international clinical trials and longitudinal observational studies, including CATCH [18, 30, 31]. Rasch analysis showed acceptable fit to the Rasch model, with items and people covering a broad measurement continuum with appropriate targeting of items to people, ordered thresholds, minimal differential item functioning by language, sex, or age [17]. A summative score across items is defensible, yielding an interval score (0–50) where higher scores reflect worsening disease activity.

Patient-reported outcomes (PROs)

The Patient Global Assessment is an 11-point NRS that asks “Considering all the ways arthritis affects you, how well are you doing today” with responses ranging from 0 = very well to 10 = very poorly [10]. At the second visit, they also completed a 5-point RA Global Impression of Change rating (“Compared to your last visit would you say that your

arthritis is *a lot better, a little better, the same, a little worse, and a lot worse*”).

Physician-reported outcomes (CLIN-ROs)

At each visit, physicians counted the number of swollen and tender joints (from a total of 28). Swollen joint counts represent the most widely used “objective” indicator of inflammatory activity reflecting inflamed synovial tissue, while tender joint counts are thought to indicate the patient’s level of pain [32]. Among the seven RA Core Data set measures in the American College of Rheumatology for clinical trials, joint counts are weighted heavily compared to the other five Core Data Set measures [33]. The MD Global Assessment is a 0–10 rating scale with higher scores representing higher RA disease activity.

Clinical disease activity

The Clinical Disease Activity Index (CDAI), an index that sums tender and swollen joint counts and patient and MD Global Assessments, was calculated for each visit [34]. CDAI is widely used to classify RA disease activity level (remission ≤ 2.8 , low $2.8 < \leq 10$, moderate $10 < \leq 22$, and high > 22) and help guide treatment decisions. We used CDAI and not indices that include biomarkers as DAS-28-CRP and SDAI as scores on all of measures are moderately-highly correlated [35] and CDAI is widely used in care as part of shared decision-making [36].

Anchors

Patients were asked “In the past month, is your arthritis (RA)...” and given seven categories from which to choose from much worse to much better. Meaningful within-person differences reflected ratings of *much worse* and *worse* for deterioration and *much better* and *better* for improvement; minimal within-person differences reflected ratings of *a little better* or *a little worse*. To identify meaningful and minimal score changes from the physician perspective, we calculated the difference in the MD global, a static global impression of (RA) severity (0–10) obtained at the two visits. This method was recently recommended by the U.S. FDA as it may be less likely to recall error [29]. We are aware of only one study by an OMERACT committee that has identified clinically important changes in physician global assessments of RA activity using a consensus process among rheumatologists, methodologists, and other stakeholders [37]. They reported a change of 2.3 in an 11-point NRS represented clinically important changes in RA patients and 1.4 in clinical trials of new therapeutic agents [37]. As 1- and 2-point change in patient global assessments have been associated with being *slightly better* and *much better* in patients in

studies of musculoskeletal pain [38], including RA [39], we used these thresholds to represent minimal and meaningful change from the physician perspective. This approach is also consistent with the widely used ACR20 criteria (i.e., 20% improvement of core outcomes including physician global) to define improvement. We also estimated score changes associated with a 1 or 2 category change in CDAI levels, as this is a common secondary outcome in RA research.

Statistical methods

Descriptive statistics were calculated to summarize patient characteristics and outcomes. Our overall approach was to calculate the mean within-patient difference between the two visits from three perspectives—patients, physicians, and between CDAI categories. We calculated Pearson correlation coefficients between the change in RA-FQ total score and components and with other disease activity indicators. Mean score change was also calculated for traditional clinical indicators of RA disease activity (CDAI, Patient Global, swollen, and tender joints) across change categories. To visually examine the separation between changes among groups across all levels of the RA-FQ change scores and enhance interpretation, we generated empirical cumulative distribution function (eCDF) curves by patient, MD, and CDAI change categories. Analyses were completed using SAS (v. 9.4) and plots were constructed using R (v 3.4.3).

Results

Participants

Participants were 808 middle-aged adults who were mostly white (84%) and female (71%). At enrollment, 85% were classified as having moderate or high RA disease activity (Table 1). At 3 months, about half (43%) were in moderate-high disease activity, 39% in Low Disease Activity, and 18% in Remission; by 6 months, 65% were in Remission or Low Disease Activity. Mean RA-FQ scores at enrollment through the 3- and 6-month visits were 25.5, 15.6, and 13.9, respectively. Correlations between changes in the RA-FQ total and component scores and disease activity indicators ranged from 0.33 for swollen joint counts to 0.89 for pain and physical function supporting the appropriateness of the data to evaluate change scores [40].

Patient perspective: minimal and meaningful change

At the second visit, 59% of patients rated their RA disease activity as improved, 20% worse, and 21% the same as the previous visit. Mean changes in RA-FQ total score

Table 1 Characteristics of participants at enrollment ($n = 808$)

	Mean (SD) or n (%)
Age (years)	55 (15)
Women (%)	575 (71%)
White race	679 (84%)
Education ^a > High school (%)	508 (63%)
Patient Global Disease Activity (0–10)	5.8 (2.9)
Swollen Joints (0–28)	6.9 (5.9)
Tender Joints (0–28)	7.8 (6.4)
MD Global Assessment (0–10)	4.8 (2.5)
Clinical Disease Activity Index (0–76)	25.3 (13.9)
Remission	10 (1%)
Low disease activity	98 (12%)
Moderate disease activity	263 (33%)
High disease activity	421 (52%)
Missing	16 (2%)

^aMissing 20 (2%)

and component scores by patient change categories are shown in Table 2. RA-FQ scores changed in the expected directions, and the magnitude of change was similar within components. Worsening disease activity was associated with larger mean changes in RA-FQ scores than improvement. Similar patterns were evident in changes in traditional disease activity indicators (CDAI, patient and MD global assessments, and tender and swollen joint counts). Cumulative distribution function (CDF) curves by patient RA change categories, shown in Fig. 1, suggest that throughout most of the range of change scores, there was clear separation of curves supporting discrimination among categories by patients, with larger spread across scores related to worsening.

Physician ratings: minimal and meaningful change

At the second visit, 44% of physicians rated RA disease activity as improved, 24% had worsened, and 32% rated it the same as the previous visit. As compared with patients, physicians were more likely to classify patients as the same or worse. Mean changes in RA-FQ total score and component scores by physician change categories 0, |1|, and ≥ 2 are shown in Table 3. RA-FQ scores changed in the expected directions, and the magnitude of change was similar within components. In contrast to patients, improvement was associated with larger mean changes in RA-FQ scores than worsening disease activity. CDF curves by physician RA change categories shown in Fig. 2 suggest that throughout much of the range of scores, physicians had difficulty discriminating between patients who were *the same or a little worse* at the second visit.

Table 2 Change in RA-FQ scores at second visit by patient global impression of change categories

Domain	A lot better (N=346; 43%)			A little better (N=132; 16%)			The same (N=174; 21%)			A little worse (N=94; 12%)			A lot worse (N=62; 8%)		
	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD
RA-FQ total (0–50)	-6.0	(-7.1, -4.9)	10.3	-1.8	(-3.2, -0.3)	8.4	-0.1	(-1.3, 1.1)	8.1	4.7	(2.9, 6.6)	9.1	8.9	(5.1, 12.7)	15.0
RA-FQ components (0–10)															
Pain	-1.2	(-1.4, -0.9)	2.4	-0.4	(-0.8, 0.0)	2.3	0.0	(-0.2, 0.3)	1.8	1.3	(0.8, 1.7)	2.2	2.0	(1.2, 2.9)	3.3
Physical function	-1.3	(-1.6, -1.1)	2.4	-0.3	(-0.6, 0.1)	2.1	0.0	(-0.3, 0.3)	2.1	0.9	(0.4, 1.4)	2.4	1.8	(0.8, 2.7)	3.7
Fatigue	-1.1	(-1.4, -0.8)	2.6	-0.4	(-0.7, 0.0)	1.9	0.0	(-0.3, 0.3)	2.1	0.7	(0.3, 1.1)	2.1	1.3	(0.5, 2.1)	3.2
Stiffness	-1.1	(-1.4, -0.9)	2.4	-0.4	(-0.7, 0.0)	2.0	-0.1	(-0.4, 0.2)	2.0	1.1	(0.6, 1.5)	2.2	1.8	(1.0, 2.7)	3.3
Participation	-1.2	(-1.5, -1.0)	2.5	-0.1	(-0.5, 0.3)	2.1	-0.1	(-0.4, 0.2)	2.2	0.8	(0.4, 1.3)	2.2	2.0	(1.1, 2.8)	3.4
Disease activity indicators															
CDAI ^a	-5.3	(-6.3, -4.3)	9.1	-3.3	(-5.4, -1.3)	11.5	-0.8	(-2.0, 0.5)	8.1	1.7	(-0.1, 3.5)	8.8	6.8	(3.7, 9.8)	12.0
Patient global (0–10)	-1.3	(-1.5, -1.0)	2.7	-0.5	(-0.9, -0.1)	2.1	-0.1	(-0.4, 0.2)	2.1	1.3	(0.8, 1.8)	2.4	2.9	(2.1, 3.6)	3.1
MD global (0–10)	-1.2	(-1.4, -1.0)	1.9	-0.7	(-1.1, -0.3)	-0.1	-0.1	(-0.4, 0.2)	1.9	0.1	(-0.3, 0.5)	2.8	0.7	(0.0, 1.5)	2.8
Swollen joints (28)	-1.4	(-1.7, 1.0)	3.2	-1.0	(-1.8, -0.2)	4.6	-0.4	(-0.9, 0.0)	3.0	0.0	(-0.7, 0.7)	3.4	1.3	(0.2, 2.5)	4.6
Tender joints (28) [®]	-1.5	(-1.9, -1.1)	3.9	-1.3	(-2.2, -0.3)	5.5	0.0	(-0.7, 0.6)	4.3	0.3	(-0.7, 1.2)	4.5	2.2	(0.8, 3.5)	5.4

^aClinical Disease Activity Index

CDAI categories: minimal and meaningful change

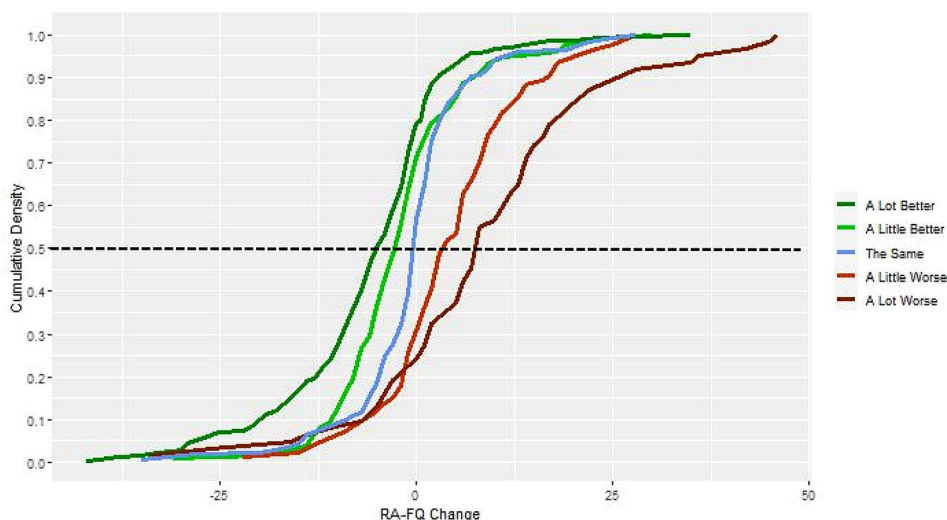
Scores changes in relation to CDAI change by |1| or ≥ 2 categories are shown in Table 4. At the second visit, 56% remained within the same, 29% had improved, and 15% had worsened by one or more CDAI categories. The thresholds for a 1 category change were similar to patient and physician thresholds representing meaningful change. Improvement was associated with a larger mean RA-FQ score change for patients starting with moderate-high disease activity where a two-category change (i.e., to remission) was nearly double that for 1 category (to low disease activity). Similarly, for worsening disease activity, patients starting in low disease activity had numerically larger mean changes than individuals who started in remission at the first visit. Across disease activity levels, patients whose CDAI category was unchanged had similar scores at both visits. CDF curves illustrate considerable variability within and wide separation of curves between change categories among patients within the same CDAI category at both visits (see Fig. 3).

Discussion

The RA-FQ is one of only a few RA measures that can summarize the extent and complexity of the broad impact of RA symptoms on how people feel and function as a single number. We have previously shown that the RA-FQ is sensitive to changes in RA symptoms and function [17]. This is the first study to identify group-based within-person minimal and meaningful change scores for the RA-FQ. Notably, thresholds for change were influenced by the anchor, direction, and baseline values. Values derived from patients are contrasted with those of treating physicians and in relation to change in CDAI levels, the keystone of the treat-to-target approach.

Our results hold implications for researchers, methodologists, and clinicians. First, the thresholds for minimal and meaningful change varied depending on the anchor used. This in turn impacted the proportion of patients who were classified as having improved or worsened. For example, among patients, 59% classified themselves as *better* or *much better*, whereas physicians classified 44% as improved; 29% had improved at least one CDAI category. Patients were least likely to be classified as worse using a change in CDAI level (15%) and most likely to be seen as having worsened disease activity by physicians (24%), with patients landing in between (20%). Second, patients had larger thresholds for defining worsening RA whereas physicians had larger thresholds for defining improvement. In effect, physicians were looking for more resolution of symptoms/impacts than patients to be confident RA was much better (score change - 7.3 vs. - 6.0, respectively), but had a lower threshold

Fig. 1 Cumulative distribution function curves of RA-FQ change between by patient change categories



to judge symptoms and RA activity as much worse (score change +5.7 vs. +8.9). The lack of separation between CDF curves of physician RA change categories suggested they can detect substantial changes in disease activity but had difficulty recognizing when patients were *a little better* or *a little worse* as compared to the previous visit. Establishing the overlap among thresholds that both patients and physicians consider meaningful and worthwhile holds important implications for clinical trials, comparative effectiveness research, and optimal management of RA in care settings. There are likely to be scenarios where more or less stringent definitions of change are appropriate which will influence the choice of anchor (e.g., change that is associated with a greater likelihood increased joint damage may better reflected by physician or CDAI categories). In addition, in RA, duration of change may be as important as the magnitude of change; for example, in RA clinical trials, it has been proposed that increased disease activity should persist for at least 7 days before being classified as an inflammatory flare [17]. Discrepancies in thresholds are also important to identify as these may contribute to patient dissatisfaction with care, treatment non-adherence, poorer disease outcomes, and increase in health care utilization and costs [41].

Two recent systematic reviews concluded that patient and physician perspectives regarding RA disease status often significantly differ [33, 34]; when discordance exists, up to 79% of patients generally perceived their disease as being more active than their treating physician. As global assessments of physicians and patients generally have good reliability [42], this discrepancy can be explained, in part, because pain, disability, and for fatigue may feature prominently in patient assessments, whereas joint counts and acute phase reactants influence physician ratings [43]. Our data

suggest assessments of change also differ where patients are more likely to perceive improvement than providers. Further, among patients the RA-FQ score change associated with meaningful improvement vs. worsening differed. Meaningful improvement for patients was associated with 6-point decrease in RA-FQ, whereas meaningful worsening was associated with a 9-point increase; similar patterns were seen for minimal worsening or improvement. This suggests that either patients may be more vigilant to identify improvement or perhaps that the threshold for disease to be perceived as worse is higher than that for improvement. At the same time, the graphical displays suggest patients appear better able to identify worsening RA at a finer level than improvement.

The pattern observed with physicians differed from patients. First, clinicians classified fewer patients as improved between visits with more classified as the *same* or *a lot worse*. In contrast to patients, on average, a numerically larger RA-FQ score change was associated with improvement versus worsening. It is interesting to note that the change in joint count (and CDAI) also was higher for patients rated as *a lot worse* or *a lot better* as compared with patient-based assessments of change. This is likely because physicians view joint counts as a reliable indicator of inflammatory disease activity. Physicians also appeared less able to differentiate patients who were *a little worse* from those who had *the same* level of disease activity at the previous visit. Changes in RA-FQ scores in relation to changes in CDAI levels were robust and similar in the direction of worsening or improvement; scores were also stable between visits in participants whose CDAI level was unchanged. Overall, triangulating meaningful RA-FQ change scores among patient and clinician ratings with RA disease activity classifications

Table 3 Change in RA – FQ by physician change categories* (N = 787)

	A lot better (N=198; 25%)			A little better (N=149; 19%)			The same (N=253; 32%)			A little worse (N=93; 12%)			A lot worse (N=94; 12%)		
	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD
RA-FQ total (range 0–50)	- 7.3	(- 8.8, - 5.7)	10.8	- 3.3	(- 5.0, - 1.7)	10.2	0.2	(- 0.9, 1.2)	8.6	0.8	(- 1.4, 3.0)	10.7	5.7	(3.4, 8.1)	11.3
RA-FQ Components (0–10)															
Pain	- 1.5	(- 1.9, - 1.2)	2.5	- 0.6	(- 1.0, - 0.2)	2.4	0.1	(- 0.1, 0.4)	2.1	0.4	(- 0.1, 0.9)	2.7	1.2	(0.7, 1.8)	2.7
Physical function	- 1.6	(- 1.9, - 1.2)	2.4	- 0.9	(- 1.2, - 0.5)	2.5	0.1	(- 0.2, 0.4)	2.1	0.1	(- 0.5, 0.6)	2.7	1.3	(0.8, 1.9)	2.6
Fatigue	- 1.0	(- 1.4, - 0.6)	2.7	- 0.6	(- 1.0, - 0.2)	2.4	- 0.2	(- 0.5, 0.1)	2.3	- 0.1	(- 0.6, 0.4)	2.3	0.8	(0.2, 1.3)	2.6
Stiffness	- 1.7	(- 2.1, - 1.4)	2.5	- 0.6	(- 1.0, - 0.2)	2.4	0.2	(- 0.1, 0.4)	2.1	0.2	(- 0.3, 0.7)	2.2	1.5	(1.0, 2.0)	2.4
Participation	- 1.4	(- 1.8, - 1.1)	2.6	- 0.7	(- 1.1, - 0.3)	2.4	0.0	(- 0.3, 0.3)	2.1	0.2	(- 0.4, 0.8)	2.8	1.0	(0.4, 1.5)	2.7
Disease activity indicators															
Clinical disease activity index	- 12.4	(- 13.9, - 11.0)	10.3	- 3.9	(- 4.8, - 3.1)	5.2	0.0	(- 0.5, 0.5)	4.1	3.0	(1.8, 4.3)	6.2	10.6	(8.6, 12.6)	9.8
Physician global	- 3.8	(- 4.2, - 3.4)	1.4	- 1.0	-	-	0.0	-	-	1.0	-	-	2.4	(1.7, 3.2)	0.8
Patient global (0–10)	- 1.7	(- 2.0, - 1.3)	2.7	- 0.6	(- 1.0, - 0.1)	2.8	0.2	(- 0.1, 0.5)	2.2	0.5	(- 0.1, 1.1)	2.8	1.2	(0.7, 1.8)	2.7
Swollen joints (28)	- 3.6	(- 4.1, - 3.0)	4.0	- 1.2	(- 1.6, - 0.8)	2.3	- 0.2	(- 0.4, 0.1)	2.0	0.7	(0.2, 1.2)	2.4	3.0	(2.1, 3.9)	4.4
Tender joints (28)	- 3.9	(- 4.7, - 3.1)	5.7	- 1.1	(- 1.6, - 0.7)	2.8	0.0	(- 0.3, 0.3)	2.3	0.8	(0.1, 1.6)	3.7	3.5	(2.4, 4.5)	5.0

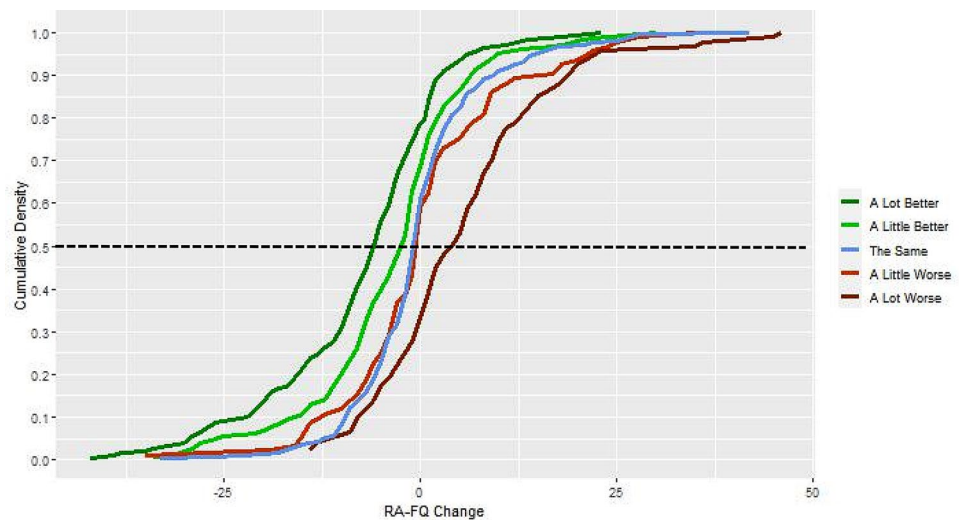
*1- and 2-point change in physician global assessments were used to identify a little or a lot of change, respectively

yielded robust results adding further evidence that disease activity as captured by the RA-FQ represents a well-defined concept that can be reliably measured. Evaluation of within-patient change establishes both the responsiveness of the RA-FQ and patient-relevant thresholds of change. The use of multiple anchors also demonstrates that benefits of a treatment as judged by physicians and using a treat-to-target approach are also perceived as clinically meaningful by patients. Conversely, both physicians and CDAI change may suggest patients are deteriorating before patients perceive their RA has changed.

While thresholds identified at the group level can inform policy and comparisons between different treatments, individual-level thresholds are necessary to inform clinical treatment decisions [44]. Examining within-patient change visually also can offer new information. The CDF curves presenting a continuous view of the proportion of patients within each category experiencing scores changes. The physical separation between curves suggest that most patients who said they were *a lot better* or *a lot worse* were relatively distinct from those reporting they were *a little better or worse* or *the same throughout most of the continuum of score changes*. We also evaluated changes in traditional RA clinical indicators and observed that changes in other disease activity indicators (e.g., mean CDAI and joint counts) were largest when using the CDAI anchors followed by physician change categories. This is not surprising since joint counts and physician global impression of disease activity constitute 3 of 4 components of the CDAI. When physicians rated patients as *a lot better* CDAI decreased by 12 points; similarly, when patients were rated as *a lot worse*, CDAI increased 11 points. The change in CDAI scores we observed were notably larger than the minimal clinically important differences that have been identified for CDAI (i.e., -6 points when starting with moderate disease activity and 2 points for worsening when starting in remission/low disease activity) in one study [45]. The change in CDAI when using patient reports of feeling *a lot better* was -5 points, whereas *a lot worse* was 7 points.

These findings add to the growing body of evidence supporting strong measurement properties (reliability, construct, content and criterion validity, responsiveness) of the RA-FQ. One reason may be that the RA-FQ was developed in accordance with best practice methods [27–29]. The conceptual framework of the RA-FQ evolved from international focus groups with RA patients [16] and Delphi exercise with RA patients, clinicians, researchers, and other stakeholders [3]. Performance of the RA-FQ was examined in multiple international clinical trials and longitudinal observational studies, including CATCH [18, 30, 31]. Rasch analysis showed acceptable fit to the Rasch model, with items and people covering a broad measurement continuum with

Fig. 2 Cumulative distribution function curves by physician change categories (|1| or |2| point change on MD Global for minimal and meaningful change, respectively)



appropriate targeting of items to people, ordered thresholds, minimal differential item functioning by language, sex, or age [17].

The strengths of this study include the use of multiple anchors to identify meaningful change in a large well characterized and diverse real-world sample of people with RA. Patients had been recently diagnosed and had started disease-modifying treatments which often take several months to reach maximum therapeutic effectiveness; only 21% rated their RA as the same at both visits. Values identified may also generalize to situations where patients are starting a new RA treatment as a result of flaring. We used patient and clinician impressions and CDAI, a disease activity index that combines these perspectives with joint counts is viewed as “the most specific quantitative clinical measure” in RA research and care [33]. In this early RA cohort, many patients reported a change in their RA status at the second visit, and these reports were supported by changes in the standardized indicators of RA disease activity recommended in international clinical practice guidelines [46, 47] and used in clinical trials as part of a treat-to-target approach. We also visualized scores associated with patient change categories with CDFs, a technique recently recommended by the US Food and Drug Administration (FDA) [24]. There are limitations. A recent systematic review showed how different methods including predictive modeling have been applied to identify change thresholds [40]. While some, such as distribution-based methods, have fallen out favor, others such as the use of bookmarking are being applied more widely [40, 48]. We used the mean change method which is the most widely used method [40], is currently recommended by the FDA [24], and is used in RA trials identifying change scores

for NIH-PROMIS™ measures [21]; however, this approach may yield values higher than actual minimal thresholds [40]. We evaluated mean score changes in relation to disease activity levels at the first visit for CDAI only. In patients with established RA, different values may be obtained for improvement and worsening as patients gain more experience with disease flares. We also observed that mean change values for worsening and improving RA varied according to the baseline value. Values derived from our cohort of recently diagnosed patients may be higher than those estimated in established patients with less active disease at baseline or patients with painful comorbidities, such as osteoarthritis. As baseline values influenced the magnitude of change, additional work is needed to determine change values for patients who begin in states of remission or low disease activity or participate in tapering trials.

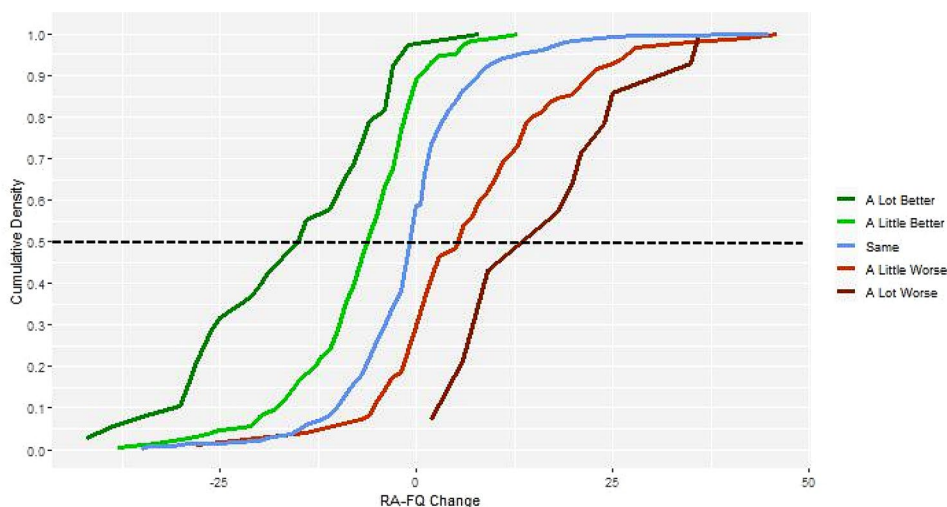
In summary, in a large diverse cohort of real-world patients recently diagnosed with RA, we compared multiple anchors to derive meaningful and minimal changes from the perspective of patients, clinicians, and using CDAI, a disease activity index used widely as part of treat-to-target approaches. We found similar patterns overall, but some important differences in the actual value of thresholds. Further, our results suggest that the benchmarks used to classify RA treatments as a success or failure differ depending on the anchor used. These findings contribute new information that can be used to interpret RA-FQ scores in research and patient care. They also demonstrate that proportion of patients classified as “responders” to a new treatment could vary considerably depending on the anchor used to define meaningful change.

Table 4 Change in RA-FQ by change in Clinical Disease Activity levels between visits

	Better						Worse											
	2 categories MOD-high to REM (N=38; 5%)			1 category MOD-high to LDA (N=97; 12%)			1 Category LDA to REM (N=94; 12%)			1 Category REM to LDA (N=35; 5%)			1 category LDA to MOD-high (N=61; 8%)			2 categories REM to MOD-high (N=14; 2%)		
	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD	Δ	95% CI	SD
RA-FQ Total	-15.8	(-19.7, -11.8)	5.4	-8.4	(-10.3, -6.5)	9.6	-5.4	(-6.7, -4.1)	6.3	6.8	(2.4, 11.2)	12.9	7.4	(4.5, 10.3)	11.4	16.9	(10.6, 23.1)	10.8
Components																		
Pain	-3.6	(-4.5, -2.8)	-1.1	-1.8	(-2.3, -1.4)	2.1	-1.1	(-1.4, -0.8)	1.5	1.8	(0.9, 2.7)	2.6	2.0	(1.3, 2.6)	2.7	4.3	(2.8, 5.8)	2.6
Function	-3.1	(-4.0, -2.3)	-1.2	-1.7	(-2.2, -1.2)	2.4	-1.2	(-1.6, -0.9)	1.5	1.4	(0.4, 2.4)	2.9	1.6	(0.9, 2.4)	2.8	3.5	(1.8, 5.2)	3.0
Fatigue	-2.3	(-3.4, -1.2)	-1.0	-1.4	(-1.8, -0.9)	2.3	-1.0	(-1.5, -0.6)	2.2	1.3	(0.3, 2.2)	2.9	0.7	(0.0, 1.4)	2.6	2.4	(0.3, 4.4)	3.5
Stiffness	-3.7	(-4.5, -2.9)	-1.0	-1.8	(-2.2, -1.3)	2.3	-1.0	(-1.3, -0.6)	1.6	1.4	(0.4, 2.4)	3.0	1.8	(1.2, 2.5)	2.6	3.3	(2.0, 4.6)	2.3
Participation	-3.0	(-3.9, -2.1)	-1.1	-1.7	(-2.3, -1.2)	2.6	-1.1	(-1.4, -0.8)	1.6	1.0	(0.0, 2.0)	2.9	1.3	(0.6, 2.0)	2.6	3.4	(1.7, 5.2)	3.1
Disease Activity																		
CDAP ^a	-2.0	(-2.3, -1.7)	-0.5	-1.2	(-1.4, -1.0)	0.9	-0.5	(-0.5, -0.4)	2.5	0.4	(0.4, 0.5)	0.2	1.0	(0.8, 1.2)	0.8	1.9	(1.2, 2.5)	1.1
MD Global	-4.2	(-5.0, -3.4)	-1.4	-1.8	(-2.3, -1.3)	2.5	-1.4	(-1.7, -1.1)	1.5	2.4	(1.4, 3.3)	2.7	2.5	(1.8, 3.2)	2.7	5.1	(3.8, 6.5)	2.3
Swollen Joints	-6.0	(-7.5, -4.5)	-1.0	-3.4	(-4.1, -2.6)	3.6	-1.0	(-1.2, -0.7)	1.2	0.4	(0.0, 0.8)	1.1	2.4	(1.4, 3.4)	4.1	4.0	(0.9, 7.1)	5.4
Tender Joints	-5.9	(-7.8, -4.0)	-1.0	-4.6	(-5.6, -3.6)	4.9	-1.0	(-1.2, -0.8)	1.2	0.9	(0.4, 1.4)	1.5	3.1	(2.2, 4.1)	3.6	6.9	(3.7, 10.1)	5.5

^aClinical Disease Activity Index

Fig. 3 Cumulative distribution function curves by CDAI change categories



Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s1136-022-03227-7>.

Acknowledgements The authors wish to acknowledge and thank CATCH Principal Investigators (M Baron, I Colmegna, S Fallavollita, D Haaland, B Haraoui, S Jamal, R Joshi, B Nair, P Panopoulos, L Rubin, E Villeneuve, M Zummer) and CATCH participants for their contributions.

M Baron, I Colmegna, S Fallavollita, D Haaland, B Haraoui, S Jamal, R Joshi, B Nair, P Panopoulos, L Rubin, E Villeneuve, and M Zummer

Author contributions The CATCH study was conceived by VB. VB, SB, JP, GB, CH, GH, LB, EK, CT, and DT serve on the CATCH Steering committee; OS serves as Scientific Director. SB, VB, and CO conceived the study design, and MFV completed the statistical analyses under the supervision of OS and SB. SB and CO wrote the first draft of the manuscript. All critically revised the manuscript and approved the final published version.

Funding The Canadian Early Arthritis Cohort (CATCH) study was independently designed and was implemented by the investigators. It has been financially supported through unrestricted research grants from Amgen, Pfizer Canada, AbbVie, Medexus, Inc., Eli Lilly, Merck Canada, Sandoz Canada Biopharmaceuticals, Hoffmann-La Roche, Janssen Biotech, UCB Canada, Bristol Myers Squibb Canada, and Sanofi Genzyme. The funders played no part in planning or conducting the study.

Data availability Available upon request via email to corresponding author.

Code availability Not applicable.

Declarations

Conflicts of interest All authors declare no conflicts of interest relevant to this article.

Ethical approval Ethics approval was previously obtained from each of the 16 participating CATCH sites across Canada. The study was conducted in accordance with the Declaration of Helsinki.

Consent to participate Written informed consent was obtained from participants at study enrollment.

Consent to publication Not applicable

References

1. Myasoedova, E., Crowson, C. S., Kremers, H. M., Therneau, T. M., & Gabriel, S. E. (2010). Is the incidence of rheumatoid arthritis rising?: Results from Olmsted County, Minnesota, 1955–2007. *Arthritis and Rheumatism*, *62*, 1576–1582.
2. Myasoedova, E., Davis, J. M., 3rd., Crowson, C. S., & Gabriel, S. E. (2010). Epidemiology of rheumatoid arthritis: Rheumatoid arthritis and mortality. *Current Rheumatology Reports*, *12*, 379–385.
3. Bartlett, S. J., Hewlett, S., Bingham, C. O., 3rd., Woodworth, T. G., Alten, R., Pohl, C., Choy, E. H., Sanderson, T., Boonen, A., Bykerk, V., et al. (2012). Identifying core domains to assess flare in rheumatoid arthritis: An OMERACT international patient and provider combined Delphi consensus. *Annals of the Rheumatic Diseases*, *71*, 1855–1860.
4. Sanderson, T., Morris, M., Calnan, M., Richards, P., & Hewlett, S. (2010). Patient perspective of measuring treatment efficacy: The rheumatoid arthritis patient priorities for pharmacologic interventions outcomes. *Arthritis Care Res (Hoboken)*, *62*, 647–656.
5. Gossec, L., Dougados, M., Rincheval, N., Balanescu, A., Boumpas, D. T., Canadello, S., Carmona, L., Daures, J. P., de Wit, M., Dijkmans, B. A., et al. (2009). Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: A EULAR initiative. *Annals of the Rheumatic Diseases*, *68*, 1680–1685.
6. Ward, M. M. (2022). Trends in permanent work disability associated with rheumatoid arthritis in the United States, 1999–2015. *Arthritis Care Res (Hoboken)*, *74*, 1321–1324.
7. Bingham, C. O., III., Pohl, C., Woodworth, T. G., Hewlett, S. E., May, J. E., Rahman, M. U., Witter, J. P., Furst, D. E., Strand, C. V., Boers, M., & Alten, R. E. (2009). Developing a standardized definition for disease “flare” in rheumatoid arthritis (OMERACT 9 Special Interest Group). *Journal of Rheumatology*, *36*, 2335–2341.

8. Myasoedova, E., Chandran, A., Ilhan, B., Major, B. T., Michet, C. J., Matteson, E. L., & Crowson, C. S. (2016). The role of rheumatoid arthritis (RA) flare and cumulative burden of RA severity in the risk of cardiovascular disease. *Annals of the Rheumatic Diseases*, *75*, 560–565.
9. Smolen, J. S., Breedveld, F. C., Burmester, G. R., Bykerk, V., Dougados, M., Emery, P., Kvien, T. K., Navarro-Compan, M. V., Oliver, S., Schoels, M., et al. (2016). Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Annals of the Rheumatic Diseases*, *75*, 3–15.
10. Nikiphorou, E., Radner, H., Chatzidionysiou, K., Desthieux, C., Zabalán, C., van Eijk-Hustings, Y., Dixon, W. G., Hyrich, K. L., Askling, J., & Gossec, L. (2016). Patient global assessment in measuring disease activity in rheumatoid arthritis: A review of the literature. *Arthritis Research & Therapy*, *18*, 251.
11. Prevoo, M. L., t'Hof, M. A., Kuper, H. H., van Leeuwen, M. A., van De Putte, L. B., & van Riel, P. L. (1995). Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis and Rheumatism*, *38*, 44–48.
12. Aletaha, D., & Smolen, J. (2005). The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis. *Clinical and Experimental Rheumatology*, *23*, S100–S108.
13. Oderda, G. M., & Balfe, L. M. (2011). Comparative effectiveness research (CER): A summary of AHRQ's CER on therapies for rheumatoid arthritis. *Journal of Managed Care Pharmacy*, *17*, S19–S24.
14. Bingham, C. O., 3rd., Alten, R., Bartlett, S. J., Bykerk, V. P., Brooks, P. M., Choy, E., Christensen, R., Furst, D. E., Hewlett, S. E., Leong, A., et al. (2011). Identifying preliminary domains to detect and measure rheumatoid arthritis flares: Report of the OMERACT 10 RA Flare Workshop. *Journal of Rheumatology*, *38*, 1751–1758.
15. Bykerk, V. P., Lie, E., Bartlett, S. J., Alten, R., Boonen, A., Christensen, R., Furst, D. E., Hewlett, S., Leong, A. L., Lyddiatt, A., et al. (2014). Establishing a core domain set to measure rheumatoid arthritis flares: Report of the OMERACT 11 RA flare Workshop. *Journal of Rheumatology*, *41*, 799–809.
16. Hewlett, S., Sanderson, T., May, J., Alten, R., Bingham, C. O., 3rd., Cross, M., March, L., Pohl, C., Woodworth, T., & Bartlett, S. J. (2012). "I'm hurting, I want to kill myself": Rheumatoid arthritis flare is more than a high joint count—an international patient perspective on flare where medical help is sought. *Rheumatology (Oxford)*, *51*, 69–76.
17. Bartlett, S. J., Barbic, S. P., Bykerk, V. P., Choy, E. H., Alten, R., Christensen, R., den Broeder, A., Fautrel, B., Furst, D. E., Guillemin, F., et al. (2017). Content and construct validity, reliability, and responsiveness of the rheumatoid arthritis flare questionnaire: OMERACT 2016 workshop report. *Journal of Rheumatology*, *44*, 1536–1543.
18. Bartlett, S. J., Bykerk, V. P., Cooksey, R., Choy, E. H., Alten, R., Christensen, R., Furst, D. E., Guillemin, F., Halls, S., Hewlett, S., et al. (2015). Feasibility and domain validation of rheumatoid arthritis (RA) flare core domain set: Report of the OMERACT 2014 RA flare group plenary. *Journal of Rheumatology*, *42*, 2185–2189.
19. Wyrwich, K. W., Norquist, J. M., Lenderking, W. R., & Acaster, S. (2013). Methods for interpreting change over time in patient-reported outcome measures. *Quality of Life Research*, *22*, 475–483.
20. Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, *10*, 407–415.
21. Bartlett, S. J., Gutierrez, A. K., Andersen, K. M., Bykerk, V. P., Curtis, J. R., Haque, U. J., Orbai, A. M., Jones, M. R., & Bingham, C. O., 3rd. (2022). Identifying minimal and meaningful change in a patient-reported outcomes measurement information system for rheumatoid arthritis: Use of multiple methods and perspectives. *Arthritis Care Res (Hoboken)*, *74*, 588–597.
22. Wells, G., Beaton, D., Shea, B., Boers, M., Simon, L., Strand, V., Brooks, P., & Tugwell, P. (2001). Minimal clinically important differences: Review of methods. *Journal of Rheumatology*, *28*, 406–412.
23. Beaton, D., Maxwell, L., Grosskleg, S., Shea, B., Tugwell, P. S., Bingham, C. O., III., Conaghan, P., D'Agostino, M. A., Hofstetter, C., March, L., et al. (2021). *The OMERACT Handbook 2.1*. Ottawa: OMERACT.
24. U.S. Food and Drug Administration. Discussion Document for Patient-Focused Drug Development Public Workshop on Guidance 3: Select, Develop, or Modify Fit-for-purpose Clinical Outcome Assessments. In *Methods to Identify What is Important to Patients & Select, Develop or Modify Fit-for-Purpose Clinical Outcomes Assessments* Rockville, MD; Workshop Date October 15–16, 2018.
25. Strand, V., & Singh, J. A. (2010). Newer biological agents in rheumatoid arthritis: Impact on health-related quality of life and productivity. *Drugs*, *70*, 121–145.
26. Bykerk, V. P., Jamal, S., Boire, G., Hitchon, C. A., Haraoui, B., Pope, J. E., Thorne, J. C., Sun, Y., & Keystone, E. C. (2012). The Canadian Early Arthritis Cohort (CATCH): Patients with new-onset synovitis meeting the 2010 ACR/EULAR classification criteria but not the 1987 ACR classification criteria present with less severe disease activity. *Journal of Rheumatology*, *39*, 2071–2080.
27. Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: Part 2—assessing respondent understanding. *Value Health*, *14*, 978–988.
28. Reeve, B. B., Wyrwich, K. W., Wu, A. W., Velikova, G., Terwee, C. B., Snyder, C. F., Schwartz, C., Revicki, D. A., Moinpour, C. M., McLeod, L. D., et al. (2013). ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research*, *22*, 1889–1905.
29. US Food and Drug Administration. (2020). Principles for selecting, developing, modifying, and adapting patient-reported outcome instruments for use in medical device evaluation; Draft Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders, vol. FDA-2020-D-1564. pp. 53820–53822. Food and Drug Administration, Health and Human Services, Docket No. FDA-2020-D-1564; Document Number:2022-01377. <https://www.fda.gov/media/141565/download>.
30. Bartlett, S. J., Bykerk, V. P., Fautrel, B., Guillemin, F., den Broeder, A., Alten, R., Christensen, R., Choy, E., Furst, D., Hewlett, S., et al. (2017). The RA Flare Questionnaire (RA-FQ) is responsive to change in ra symptoms and impacts in clinical and observational trials. *Annals of the Rheumatic Diseases*, *76*, 470–470.
31. Bykerk, V. P., Bingham, C. O., Choy, E. H., Lin, D., Alten, R., Christensen, R., Furst, D. E., Hewlett, S., Leong, A., March, L., et al. (2016). Identifying flares in rheumatoid arthritis: Reliability and construct validation of the OMERACT RA Flare Core Domain Set. *RMD Open*, *2*, e000225.
32. Scott, I. C., & Scott, D. L. (2014). Joint counts in inflammatory arthritis. *Clinical and Experimental Rheumatology*, *32*, 7–12.
33. Sokka, T., & Pincus, T. (2009). Joint counts to assess rheumatoid arthritis for clinical research and usual clinical care: advantages and limitations. *Rheumatic Disease Clinics North America*, *35*, 713–722.

34. Aletaha, D., & Smolen, J. S. (2007). The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) to monitor patients in standard clinical care. *Best Practice & Research Clinical Rheumatology*, *21*, 663–675.
35. Dhaon, P., Das, S. K., Srivastava, R., & Dhakad, U. (2018). Performances of Clinical Disease Activity Index (CDAI) and Simplified Disease Activity Index (SDAI) appear to be better than the gold standard Disease Assessment Score (DAS-28-CRP) to assess rheumatoid arthritis patients. *International Journal of Rheumatic Diseases*, *21*, 1933–1939.
36. Yun, H., Chen, L., Xie, F., Patel, H., Boytsov, N., Zhang, X., & Curtis, J. R. (2020). Do patients with moderate or high disease activity escalate rheumatoid arthritis therapy according to treat-to-target principles? results from the rheumatology informatics system for effectiveness registry of the American College of Rheumatology. *Arthritis Care & Research*, *72*, 166–175.
37. Goldsmith, C. H., Boers, M., Bombardier, C., & Tugwell, P. (1993). Criteria for clinically important changes in outcomes: Development, scoring and evaluation of rheumatoid arthritis patient and trial profiles OMERACT Committee. *The Journal of Rheumatology*, *20*, 561–565.
38. Farrar, J. T., Young, J. P., Jr., LaMoreaux, L., Werth, J. L., & Poole, M. R. (2001). Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*, *94*, 149–158.
39. Salaffi, F., Stancati, A., Silvestri, C. A., Ciapetti, A., & Grassi, W. (2004). Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *European Journal of Pain*, *8*, 283–291.
40. Terwee, C. B., Peipert, J. D., Chapman, R., Lai, J.-S., Terluin, B., Cella, D., Griffith, P., & Mokkink, L. B. (2021). Minimal important change (MIC): A conceptual clarification and systematic review of MIC estimates of PROMIS measures. *Quality of Life Research*, *30*, 2729–2754.
41. Starfield, B., Wray, C., Hess, K., Gross, R., Birk, P. S., & D’Lugoff, B. C. (1981). The influence of patient-practitioner agreement on outcome of care. *American Journal of Public Health*, *71*, 127–131.
42. Rohekar, G., & Pope, J. (2009). Test-retest reliability of patient global assessment and physician global assessment in rheumatoid arthritis. *Journal of Rheumatology*, *36*, 2178–2182.
43. Desthieux, C., Hermet, A., Granger, B., Fautrel, B., & Gossec, L. (2016). Patient-physician discordance in global assessment in rheumatoid arthritis: A systematic literature review with meta-analysis. *Arthritis Care Res (Hoboken)*, *68*, 1767–1773.
44. Beaton, D. E., Bombardier, C., Katz, J. N., Wright, J. G., Wells, G., Boers, M., Strand, V., & Shea, B. (2001). Looking for important change/differences in studies of responsiveness. OMERACT MCID Working. Group Outcome Measures in Rheumatology. *Minimal Clinically Important Difference. J Rheumatol.*, *28*, 400–405.
45. Curtis, J. R., Yang, S., Chen, L., Pope, J. E., Keystone, E. C., Haraoui, B., Boire, G., Thorne, J. C., Tin, D., Hitchon, C. A., et al. (2015). Determining the minimally important difference in the clinical disease activity index for improvement and worsening in early rheumatoid arthritis patients. *Arthritis Care Res (Hoboken)*, *67*, 1345–1353.
46. Anderson, J., Caplan, L., Yazdany, J., Robbins, M. L., Neogi, T., Michaud, K., Saag, K. G., O’Dell, J. R., & Kazi, S. (2012). Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. *Arthritis Care Res (Hoboken)*, *64*, 640–647.
47. Bykerk, V. P., Akhavan, P., Hazlewood, G. S., Schieir, O., Dooley, A., Haraoui, B., Khraishi, M., Leclercq, S. A., Legare, J., Mosher, D. P., et al. (2012). Canadian Rheumatology Association recommendations for pharmacological management of rheumatoid arthritis with traditional and biologic disease-modifying antirheumatic drugs. *Journal of Rheumatology*, *39*, 1559–1582.
48. Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology*, *56*, 395–407.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.