**Appendix Table F58. Minimum clinically important differences in the tools to assess outcomes in patients with knee OA**

| **Author, year; Test; Worst to best; Method** | **Reference** | **Definition of minimum clinically important differences** |
| --- | --- | --- |
| Tanner, 2007349 Test: American Academy of Orthopedic Surgeons (AAOS)Sports Knee Rating Scale Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 11 questions had a mean importance ranking of 1 or less |
| Tanner, 2007349 Test: American Academy of Orthopedic Surgeons Sports Knee Rating Scale Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 3 of the AAOS questions had a mean importance ranking of at least 3 (score on a Likert scale of 0 to 5, with 0 being not experienced and 5 being experienced and very important) |
| Tanner, 2007349 Test: American Academy of Orthopedic Surgeons Sports Knee Rating Scale Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 3 questions had a top-20 FIP scores (FIP=frequency\*mean importance; the greater the FIP, the more important a symptom or disability is to patients. A high FIP indicates that a symptom or disability is both frequently experienced and most important to patients) |
| Tanner, 2007349 Test: American Academy of Orthopedic Surgeons Sports Knee Rating Scale Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | At least 51% of the patients with mild to moderate OA endorsed 11 (55%) of the AAOS questions. |
| Tanner, 2007349 Test: The Activities of Daily Living Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 5/17 of the ADL questions had a mean importance ranking of at least 3 (score on a Likert scale of 0 to 5, with 0 being not experienced and 5 being experienced and very important) |
| Tanner, 2007349 Test: The Activities of Daily Living Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 8 questions had a top-20 FIP scores (FIP=frequency\*mean importance; the greater the FIP, the more important a symptom or disability is to patients. A high FIP indicates that a symptom or disability is both frequently experienced and most important to patients) |
| Tanner, 2007349 Test: The Activities of Daily Living Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | At least 51% of the patients with mild to moderate OA endorsed 16/17 (94%) of the ADL questions |
| Tanner, 2007349 Test: The Activities of Daily Living Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | No question had a mean importance ranking of 1 or less |
| Bellamy, 2007341 Test: Bellamy et al. Low Intensity Symptom State-attainment Index  Worse to best: 100 to 0  Method: Anchor | WOMAC Likert 3.0: pain subscale | 5 threshold levels of BLISS response were performed using the WOMAC Pain Scale (WOMAC-P), from a very low level of pain to higher levels of pain. The WOMAC-P varies from 0 to 20 and in the analysis data were transformed to normalized units (NU) on a 0-100 scale. The threshold levels included: WOMAC pain score <=5, <=10, <=15, <=20, and <=25 (0=no pain, 100=extreme pain). The minimal pain intensity requirement at baseline for inclusion in this study was 35 NU, a value just above the MCAS (Maximally Clinically Acceptable Status) and PASS (Patient Acceptable Symptom State) thresholds for patient acceptable pain intensity of 33 and 32mm, respectively, for patient acceptable pain intensity.. A WOMAC-based BLISS-10, is a potentially symptom intensity state because it is a cut point at which a clinically important between-group difference (24 percentage points) is discernible. |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale (NRS) | At 18 month follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the baseline risk score group (approach proposed by Von Korff and Miglioretti) intermediate risk (5-11) was 0.341 (range: 0.218, 0.499) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 18 month follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the baseline risk score group (approach proposed by Von Korff and Miglioretti) low risk (0-4) was 0.153 (range: 0.106, 0.184) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 18 month follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the baseline risk score group (approach proposed by Von Korff and Miglioretti) possible chronic pain (12-17) was 0.656 (range: 0.552-0.781) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 18 month follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the baseline risk score group (approach proposed by Von Korff and Miglioretti) probable chronic pain (18+) was 0.888 (range: 0.815-0.974) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 3-year follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the 18-month risk -score group (approach proposed by Von Korff and Miglioretti) intermediate risk (5-10) was 0.341 (range: 0.229, 0.479) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 3-year follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the 18-month risk- score group (approach proposed by Von Korff and Miglioretti) low risk (0-4) was 0.156 (range: 0.108, 0.192) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 3-year follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the 18-month risk- score group (approach proposed by Von Korff and Miglioretti) possible chronic pain (12-17) was 0.662 (range: 0.536, 0.781) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 3-year follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the 18-month risk- score group (approach proposed by Von Korff and Miglioretti) probable chronic pain (18+) was 0.891 (range: 0.817, 0.982) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 3-year follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the baseline risk score group (approach proposed by Von Korff and Miglioretti) intermediate risk (4-10) was 0.329 (range: 0.214, 0.466) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 3-year follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the baseline risk score group (approach proposed by Von Korff and Miglioretti) low risk (0-3) was 0.165 (range: 0.132, 0.182) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 3-year follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the baseline risk score group (approach proposed by Von Korff and Miglioretti) possible chronic pain (11-17) was 0.637 (range: 0.514, 0.773) |
| Thomas, 2008342 Test: Chronic Pain Grade Method:  Anchor | Numerical Rating Scale | At 3-year follow-up: the mean for the predicted probability of Chronic Pain Grade II-IV knee pain for patients who belonged to the baseline risk score group (approach proposed by Von Korff and Miglioretti) probable chronic pain (18+) was 0.878 (range: 0.806, 0.967) |
| Belo, 2009436 Test: American College of Rheumatology (ACR) Method:   Anchor | Baseline and 12-month follow-up values of WOMAC ; SF-36; KSS (Knee Society Score); Lysholm Knee Scoring Scale; and baseline values of Tampa Scale for Kinesiophobia | The clinical ACR classification criteria of knee OA have no prognostic value for predicting persisting knee complaints or an increase of disability at 1-year of follow-up in adult patients with non-traumatic knee complaints in GP. |
| Bellamy, 2001435 Test: Dictionary of the Rheumatic Diseases Method Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for grip strength (Dictionary of the Rheumatic Diseases Method) =37.5 |
| Bellamy, 2001435 Test:  Investigators subject opinion of patients general condition Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for investigator's opinion of patient's general condition =0.90 |
| Bellamy, 2001435 Test: Physicians' estimate of disease activity Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for physician's estimate of disease activity =0.78 |
| Bellamy, 2001435 Test: Subjective pain evaluation by patient Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for subjective pain evaluation by patient=0.78 |
| Bellamy, 2001435 Test: Patient estimate of disease activity Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for patient estimate of disease activity =1 |
| Bellamy, 2001435 Test: Patient's opinion of general condition Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for patient's opinion of general condition =0.9 |
| Redelmeier, 1993347 Test: Health Assessment Questionnaire (HAQ) Method:  Anchor | After completing the initial questionnaires, participants were asked to talk to each other in one-on-one conversations. At the end of each dialogue, participants were told to rate their disability relative to their conversational partner. The rating question was, "Compared with this person, my physical ability to function during the past week was ..." The response categories were as follows: "much better," "somewhat better," "about the same," "somewhat worse," and "much worse". | HAQ scores needed to improve by 0.17 units for respondents on an average to stop rating themselves as about the same and start rating themselves as somewhat better (0.15 to -0.02) |
| Redelmeier, 1993347 Test: Health Assessment Questionnaire Method:  Anchor | After completing the initial questionnaires, participants were asked to talk to each other in one-on-one conversations. At the end of each dialogue, participants were told to rate their disability relative to their conversational partner. The rating question was, "Compared with this person, my physical ability to function during the past week was ..." The response categories were as follows: "much better," "somewhat better," "about the same," "somewhat worse," and "much worse". | HAQ scores needed to differ by 0.22 units for average respondents to stop rating themselves as about the same and start rating themselves as somewhat worse (0.37 to 0.15) |
| Redelmeier, 1993347 Test: Health Assessment Questionnaire Method:  Anchor | After completing the initial questionnaires, participants were asked to talk to each other in one-on-one conversations. At the end of each dialogue, participants were told to rate their disability relative to their conversational partner. The rating question was, "Compared with this person, my physical ability to function during the past week was ..." The response categories were as follows: "much better," "somewhat better," "about the same," "somewhat worse," and "much worse". | The threshold of symptomatic clinical importance for OA =0.20 |
| Redelmeier, 1993347 Test: Health Assessment Questionnaire Method:  Anchor | After completing the initial questionnaires, participants were asked to talk to each other in one-on-one conversations. At the end of each dialogue, participants were told to rate their disability relative to their conversational partner. The rating question was, "Compared with this person, my physical ability to function during the past week was ..." The response categories were as follows: "much better," "somewhat better," "about the same," "somewhat worse," and "much worse". | The threshold of symptomatic clinical importance for less disabled participants was significantly lower than the threshold for more disabled participants (0.08 vs. 0.29 HAQ units; p<0.05) |
| Redelmeier, 1993347 Test: Health Assessment Questionnaire Method:  Anchor | After completing the initial questionnaires, participants were asked to talk to each other in one-on-one conversations. At the end of each dialogue, participants were told to rate their disability relative to their conversational partner. The rating question was, "Compared with this person, my physical ability to function during the past week was ..." The response categories were as follows: "much better," "somewhat better," "about the same," "somewhat worse," and "much worse". | The average patient would find an HAQ score difference on the order of 0.2 units to be an important symptomatic difference |
| Tanner, 2007349 Test: International Knee Documentation Committee Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 3 questions had a top-20 FIP scores (FIP=frequency\*mean importance; the greater the FIP, the more important a symptom or disability is to patients. A high FIP indicates that a symptom or disability is both frequently experienced and most important to patients) |
| Tanner, 2007349 Test: International Knee Documentation Committee Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 4 of the IKDC questions had a mean importance ranking of at least 3 (score on a Likert scale of 0 to 5, with 0 being not experienced and 5 being experienced and very important) |
| Tanner, 2007349 Test: International Knee Documentation Committee Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | At least 51% of the patients with mild to moderate OA endorsed 18 (100%) of the IKDC questions |
| Tanner, 2007349 Test: International Knee Documentation Committee Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 w assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | Two questions had a mean importance ranking of 1 or less |
| Ehrich, 2000437 Test: WOMAC VA 3.1: (pain, stiffness, and physical function scales); VAS (patient walking on a flat surface score); Patient global assessment of disease status; Investigator global assessment of disease status Method:  Anchor | Patient global response to therapy measure | For patient global response to therapy measure: The minimum perceptible clinical improvement (defined as the difference in mean change scores between patients with a "none" response and those with a "poor" response on the patient global response to therapy measure) was 0.43 (on a 0-4 Likert scale) for the investigator global disease status measure, 11.1 (on a 100mm VAS) for the WOMAC pain walking on a flat surface item; 11.7 for the patient global disease status, 9.7 on WOMAC pain, 9.3 on WOMAC physical functioning, and 10.0 on WOMAC stiffness |
| Ehrich, 2000437 Test: WOMAC VA 3.1: (pain, stiffness, and physical function scales); VAS (patient walking on a flat surface score); Patient global assessment of disease status; Investigator global assessment of disease status Method:  Anchor | Investigator global response to therapy measure | For investigator global response to therapy measure: The minimum perceptible clinical improvement (defined as the difference in mean change scores between patients with a "none" response and those with a "poor" response on the investigator global response to therapy measure) was 0.49 (on a 0-4 Likert scale) for the investigator global disease status measure, 12.2 (on a 100mm VAS) for the WOMAC pain walking on a flat surface item; 11.1 for the patient global disease status, 10.8 on WOMAC pain, 7.6 on WOMAC physical functioning, and 10.4 on WOMAC stiffness |
| Ehrich, 2000437 Test: WOMAC: pain walking on a flat surface Method:  Anchor | Patient global response to therapy measure and investigator global response to therapy measure | The minimum perceptible clinical improvement of roughly 10mm is seen as the difference between the median change scores for the "none" and the "poor" groups. |
| Fransen, 2003351 Test: Knee extensor force in Newtons Method:   Anchor | Knee extensor force evaluated on 2 occasions; WOMAC:VAS version | Female (50-59 years) with knee OA would need to show a change of 16.2N in knee extensor force for the clinician to be moderately confident that an actual change had occurred. |
| Fransen, 2003351 Test: Knee extensor force in Newtons Method:  Anchor | Knee extensor force evaluated on 2 occasions; WOMAC:VAS version | Female (60-69 years) with knee OA would need to show a change of 21.6N in knee extensor force for the clinician to be moderately confident that an actual change had occurred. |
| Fransen, 2003351 Test: Knee extensor force in Newtons Method:  Anchor | Knee extensor force evaluated on 2 occasions; WOMAC:VAS version | Female (70-79 years) with knee OA would need to show a change of 18.5N in knee extensor force for the clinician to be moderately confident that an actual change had occurred. |
| Fransen, 2003351 Test: Knee extensor force in Newtons Method:  Anchor | Knee extensor force evaluated on 2 occasions; WOMAC:VAS version | Male (60-69 years) with knee OA would need to show a change of 6N in knee extensor force for the clinician to be moderately confident that an actual change had occurred. |
| Fransen, 2003351 Test: Knee extensor force in Newtons Method:  Anchor | Knee extensor force evaluated on 2 occasions; WOMAC:VAS version | Male (70-79 years) with knee OA would need to show a change of 17N in knee extensor force for the clinician to be moderately confident that an actual change had occurred. |
| Fransen, 2003351 Test: Knee flexor force in Newtons Method:  Anchor | Knee flexor force evaluated on 2 occasions; WOMAC:VAS version | Female (50-59 years) with knee OA would need to show a change of 14N in knee flexor force for the clinician to be moderately confident that an actual change had occurred. |
| Fransen, 2003351 Test: Knee flexor force in Newtons Method:  Anchor | Knee flexor force evaluated on 2 occasions; WOMAC:VAS version | Female (60-69 years) with knee OA would need to show a change of 7.2N in knee flexor force for the clinician to be moderately confident that an actual change had occurred. |
| Fransen, 2003351 Test: Knee flexor force in Newtons Method:  Anchor | Knee flexor force evaluated on 2 occasions; WOMAC:VAS version | Female (70-79 years) with knee OA would need to show a change of 5.5N in knee flexor force for the clinician to be moderately confident that an actual change had occurred. |
| Fransen, 2003351 Test: Knee flexor force in Newtons Method:  Anchor | Knee flexor force evaluated on 2 occasions; WOMAC:VAS version | Male (60-69 years) with knee OA would need to show a change of15.66N in knee flexor force for the clinician to be moderately confident that an actual change had occurred. |
| Fransen, 2003351 Test: Knee flexor force in Newtons Method:  Anchor | Knee flexor force evaluated on 2 occasions; WOMAC:VAS version | Male (70-79 years) with knee OA would need to show a change of 12.3N in knee flexor force for the clinician to be moderately confident that an actual change had occurred. |
| Tanner, 2007349 Test: Knee, Injury and Osteoarthritis Outcome Score Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 14 of the KOOS questions had a mean importance ranking of at least 3 (score on a Likert scale of 0 to 5, with 0 being not experienced and 5 being experienced and very important) |
| Tanner, 2007349 Test: Knee, Injury and Osteoarthritis Outcome Score Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 9 questions had a top-20 FIP scores (FIP=frequency\*mean importance; the greater the FIP, the more important a symptom or disability is to patients. A high FIP indicates that a symptom or disability is both frequently experienced and most important to patients) |
| Tanner, 2007349 Test: Knee, Injury and Osteoarthritis Outcome Score Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | At least 51% of the patients with mild to moderate OA endorsed 38(90%) of the KOOS questions |
| Tanner, 2007349 Test: Knee, Injury and Osteoarthritis Outcome Score Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | Three questions had a mean importance ranking of 1 or less |
| Bellamy, 2001435 Test: Lequesne Knee Index Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for Lequesne Knee Index =3 |
| Bellamy, 2001435 Test: Likert scale (Physician assessment of physical disability) Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for physicians overall assessment of physical disability on Likert scale =0.68 |
| Bellamy, 2001435 Test: Likert scale (Physician assessment of disease activity) Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for physician's global assessment of disease activity on Likert scale =0.78 |
| Bellamy, 2001435 Test: Likert scale(patient overall assessment of pain) Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for patient overall assessment of pain on the Likert scale =0.78 |
| Bellamy, 2001435 Test: Likert scale (patient overall assessment of morning stiffness) Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for patient overall assessment of morning stiffness on the Likert scale =0.80 |
| Bellamy, 2001435 Test: Likert scale (patient's overall assessment of physical disability) Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for patient’s overall assessment of physical disability on Likert scale =0.8 |
| Bellamy, 2001435 Test: Likert scale (patient's global assessment of disease activity) Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for patient's global assessment of disease activity on the Likert scale =1 |
| Bellamy, 2001435 Test: Likert Scale for pain (physician assessment) Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for physicians overall assessment of pain =0.78 on a Likert scale |
| Bellamy, 2001435 Test: Likert Scale for stiffness (physician assessment) Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for physicians overall assessment of stiffness =0.75 on a Likert scale |
| Tanner, 2007349 Test: Modified Lysholm Knee Scoring Scale Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 2 of the Lysholm questions had a mean importance ranking of at least 3 (score on a Likert scale of 0 to 5, with 0 being not experienced and 5 being experienced and very important) |
| Tanner, 2007349 Test: Modified Lysholm Knee Scoring Scale Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | 7 questions had a top-20 FIP scores (FIP=frequency\*mean importance; the greater the FIP, the more important a symptom or disability is to patients. A high FIP indicates that a symptom or disability is both frequently experienced and most important to patients) |
| Tanner, 2007349 Test: Modified Lysholm Knee Scoring Scale Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | At least 51% of the patients with mild to moderate OA endorsed 2 (25%) of the Lysholm questions |
| Tanner, 2007349 Test: Modified Lysholm Knee Scoring Scale Method:  Anchor | A questionnaire of 111 items was developed by combining 222 patient-directed questions from the 11 knee-specific quality-of-life instruments. Patients were asked to rate the importance of the described symptom or disability using a 6-point Likert scale. The Tegner Activity Scale was administered as a separate tool to determine the activity level of the participating patients at the time the study questionnaire was administered. A patient’s activity score from 0 to 10 was assigned, with corresponding definitions ranging from “on sick leave/disability” to “participation in competitive sports such as soccer at a national or international elite level.” | Four questions had a mean importance ranking of 1 or less |
| Salaffi, 2004336 Test: Numerical Rating Scale Method:  Anchor | Patients were instructed to draw a single mark on a horizontally oriented, graduated 10-cm NRS bounded by the descriptors ‘‘no pain’’ at the far left and ‘‘worst possible pain’’ at the far right. At a 3 months follow-up, patients were asked by the same data collector to repeat the measurement on an NRS, without access to any previous NRS ratings. The NRS score at the end of follow-up minus the score at baseline examination prior to the intervention defined the effect measured by NRS. | A raw change of -1.0 (AUC 0.889 +- 0.008) and percent change of -15% (AUC 0.881 +- 0.011) were optimal cut-off point associated with the PGIC (patients global assessment of change) category of "slightly better". When using NRS changes best associated with "much better", a raw change of -2.0 (AUC 0.909 +- 0.010) and percent change of -33% (AUC 0.956 +- 0.008) were shown. |
| Salaffi, 2004336 Test: Numerical Rating Scale Method:  Anchor | Patients were instructed to draw a single mark on a horizontally oriented, graduated 10-cm NRS bounded by the descriptors ‘‘no pain’’ at the far left and ‘‘worst possible pain’’ at the far right. At a 3 months follow-up, patients were asked by the same data collector to repeat the measurement on an NRS, without access to any previous NRS ratings. The NRS score at the end of follow-up minus the score at baseline examination prior to the intervention defined the effect measured by NRS. | For NRS baseline score >7 to 10, a raw change of -2.8 (AUC 0.916+-0.011) and percent change of -40% (AUC 0.904 +- 0.012) were optimal cut-off point associated with the PGIC (patients' global assessment of change) category of "much better". When using NRS changes best associated with "slightly better", if the baseline NRS score is >7 to 10, a raw change of -1.6 (AUC 0.838 +- 0.009) and percent change of -21% (AUC 0.846 +- 0.009) were shown. |
| Salaffi, 2004336 Test: Numerical Rating Scale Method:  Anchor | Patients were instructed to draw a single mark on a horizontally oriented, graduated 10-cm NRS bounded by the descriptors ‘‘no pain’’ at the far left and ‘‘worst possible pain’’ at the far right. At a 3 months follow-up, patients were asked by the same data collector to repeat the measurement on an NRS, without access to any previous NRS ratings. The NRS score at the end of follow-up minus the score at baseline examination prior to the intervention defined the effect measured by NRS. | If the NRS baseline score is <=4, a raw change of -0.7 (AUC 0.968+-0.013) and percent change of -17% (AUC 0.971 +- 0.013) were optimal cut-off point associated with the PGIC (patients' global assessment of change) category of "much better". When using NRS changes best associated with "slightly better", if the baseline NRS score is <=4, a raw change of -0.6 (AUC 0.795 +- 0.008) and percent change of -10.5% (AUC 0.804 +- 0.011) were shown. |
| Salaffi, 2004336 Test: Numerical Rating Scale Method:  Anchor | Patients were instructed to draw a single mark on a horizontally oriented, graduated 10-cm NRS bounded by the descriptors ‘‘no pain’’ at the far left and ‘‘worst possible pain’’ at the far right. At a 3 months follow-up, patients were asked by the same data collector to repeat the measurement on an NRS, without access to any previous NRS ratings. The NRS score at the end of follow-up minus the score at baseline examination prior to the intervention defined the effect measured by NRS. | If the NRS baseline score is >4 to <=7, a raw change of -2.1 (AUC 0.941+-0.012) and percent change of -32.7% (AUC 0.883+-0.931) were optimal cut-off point associated with the PGIC (patients' global assessment of change) category of "much better". When using NRS changes best associated with "slightly better", if the baseline NRS score is >4 and <=7, a raw change of -1 (AUC 0.878 +- 0.904) and percent change of -16.9% (AUC 0.844 +- 0.014) were shown. |
| Salaffi, 2004336 Test: Numerical Rating Scale Method:  Anchor | Patients were instructed to draw a single mark on a horizontally oriented, graduated 10-cm NRS bounded by the descriptors ‘‘no pain’’ at the far left and ‘‘worst possible pain’’ at the far right. At a 3 months follow-up, patients were asked by the same data collector to repeat the measurement on an NRS, without access to any previous NRS ratings. The NRS score at the end of follow-up minus the score at baseline examination prior to the intervention defined the effect measured by NRS. | MCID is defined as the difference in mean change from baseline in NRS between patients with no response therapy ("no change", "slightly worse" and "much worse") and patients with next higher level of response (slightly better). |
| Salaffi, 2004336 Test: Patient's Global Assessment of Change Method:  Anchor | The ‘‘transition questionnaire’’ investigated the current pain intensity, related to the rheumatic disease at the 3 month follow-up compared to the pain intensity 3 months earlier (at baseline examination) by the question: Please imagine how you would have described your pain intensity three months ago. How do you feel today as compared to three months earlier as far as your musculoskeletal pain is concerned? The possible replies were ‘‘much better’’, ‘‘slightly better’’, ‘‘no change’’, ‘‘slightly worse’’, or ‘‘much worse’’ (Jaeschke et al., 1989). This five-point categorical questionnaire was proposed to the patient at the 3 months follow-up and assessed the change of pain in written format | A one unit difference at the lowest end of the PGIC ("slightly better") was used to define MCID as it reflects the minimum and lowest degree of improvement that could be detected. |
| Bellamy, 2001435 Test: ROM (Range of Motion) Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for range of movement =15 |
| Kennedy, 2005346 Test: Six Minute Walk Test Distance Method:  Distribution | At 3 points, during the performance of the 6MWT, patients were instructed to cover as much distance as possible during the 6 minute time frame with opportunity to stop and rest if required. | The MDC90 was at 61.34m |
| Mangione, 2010438 Test: Six-Minute Walk Test Method:  Distribution | At 2 times patients were tested: Each participant was instructed to cover as much distance as possible in 6 minutes. | The MDC 90 was 65m |
| Mangione, 2010438 Test: Short Physical Performance Battery Method:  Distribution | At 2 times patients were tested for 3 timed tests: chair rise for 5 repetitions, without the use of arms; standing balance in positions of side-by side stance, semi-tandem stance, and full tandem stance; and walking speed over a 2,44-m (8-ft) course. | The MDC 90 was 2.9 points |
| Kennedy, 2005346 Test: Stair Time Method:  Distribution | At 3 points, patients had to ascend and descend 9 stairs (step height, 20 cm) in their usual manner, and at a safe and comfortable pace. | The MDC90 was at 5.49s |
| Bellamy, 2001435 Test: Time between arising and improvement in stiffness Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for duration of morning stiffness (measured from the time between arising and improvement in stiffness) =0.23 |
| Bellamy, 2001435 Test: Clock time from awaking to when stiffness begins to wear off Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for duration of morning stiffness (measured from the clock time from awaking to when stiffness begins to wear off) =20 |
| Bellamy, 2001435 Test: Time between arising and when patient is limber Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for duration of morning stiffness (measured from the time between awakening and when patient is limber) =0.3 |
| Bellamy, 2001435 Test: Soft tissue swelling  Method:  Anchor | Delphi exercise to define minimum clinically important differences | The MCID for soft tissue swelling =1.50 |
| Kennedy, 2005346 Test: Timed Up and Go Time (TUG) Method:  Anchor | At 3 points, for the TUG test, patients were required to rise from a standard arm chair, walk at a safe and comfortable pace to a tape mark 3-m away, then return to a sitting position in the chair | The MDC90 was at 2.49s |
| Ren, 1999354 Test: Short form Arthritis Impact Measurement Scale 2 (AIMS2-SF) Method:  Distribution | One generic and one disease -specific health measure. Generic health is measured by a global health question, i.e. patients were asked "In general would you say that your health now is excellent, good, fair, or poor?" Disease-specific health was measured by asking the patient how much of their health problems were due to arthritis, due partly to arthritis, or due to other diseases. | Patients who had mild symptom severity score on AIMS2-SF (score >=68) had a mean core of 8.2 on the upper body limitations domain; 8.2 on the lower body limitations domain; 6.5 on the affect domain; and 6.1 on the social function domain |
| Ren, 1999354 Test: Short form Arthritis Impact Measurement Scale 2 Method:   Distribution | One generic and one disease -specific health measure. Generic health is measured by a global health question, i.e. patients were asked "In general would you say that your health now is excellent, good, fair, or poor?”Disease-specific health was measured by asking the patient how much of their health problems were due to arthritis, due partly to arthritis, or due to other diseases. | Patients who had moderate symptom severity score on AIMS2-SF (score 43-67) had a mean score of 7.7 on the upper body limitations domain; 7.7 on the lower body limitations domain; 5.8 on the affect domain; and 5.5 on the social function domain |
| Ren, 1999354 Test: Short form Arthritis Impact Measurement Scale 2 Method:  Distribution | One generic and one disease -specific health measure. Generic health is measured by a global health question, i.e. patients were asked "In general would you say that your health now is excellent, good, fair, or poor?” Disease-specific health was measured by asking the patient how much of their health problems were due to arthritis, due partly to arthritis, or due to other diseases. | Patients who had severe symptom severity score on AIMS2-SF (score <=42) had a mean score of 6.6 on the upper body limitations domain; 6.0 on the lower body limitations domain; 4.3 on the affect domain; and 4.4 on the social function domain |
| Lequesne, 1997439 Test: Algofunctional index for osteoarthritis Method:  Anchor | Different scores of the index | 1-4 points in the scores for the algofuncitional index corresponds to minor handicap; 5-7 points =moderate handicap; 8-10 points =severe handicap; 11-13 points=very severe handicap and 14 points indicate extremely severe handicap |
| Ornetti, 2011410 Test: Patient NRS: "What is the degree of difficulty you have experienced for the daily activities during the last 48 hours due to your (knee or hip) OA" Method: Anchor Worst to best: 10 to 0 | All patients had to assess their current global state (global PASS) by answering 'Yes' or 'No' in answer to the question 'Taking into account all the activities you have during your daily life, your level of pain, and also your functional impairment, do you consider that your current state is satisfactory?'. | Patients considered their global state as satisfactory if the patient NRS was >3.33 (95% CI: 3.17 to 3.48). Global PASS is defined as the value of measurement beyond which patients consider their global state as satisfactory. |
| Ornetti, 2011410 Test: Physician NRS: Physician's Estimate of functional impairment of each patient Method: Anchor  Worst to best: 10 to 0 | All patients had to assess their current global state (global PASS) by answering 'Yes' or 'No' in answer to the question 'Taking into account all the activities you have during your daily life, your level of pain, and also your functional impairment, do you consider that your current state is satisfactory?'. | Patients considered their global state as satisfactory if the physician NRS was >3.07 (95% CI: 2.94 to 3.21).Global PASS is defined as the value of measurement beyond which patients consider their global state as satisfactory. |
| Ornetti, 2011410 Test: Patient NRS: "What is the degree of difficulty you have experienced for the daily activities during the last 48 hours due to your (knee or hip) OA" Method: Anchor  Worst to best: 10 to 0 | PASS for functional state :The PASS of each function scale was defined as the 75th centile of the absolute score among patients who considered their final state as satisfactory | Patients considered their functional state as satisfactory if the patient NRS was >3.3 (95% CI: 3.16 to 3.45).Function PASS is defined as the value of measurement beyond which patients consider their functional state as satisfactory. |
| Ornetti, 2011410 Test: Physician NRS: Physician's Estimate of functional impairment of each patient Method: Anchor  Worst to best: 10 to 0 | PASS for functional state :The PASS of each function scale was defined as the 75th centile of the absolute score among patients who considered their final state as satisfactory | Patients considered their functional state as satisfactory if the physician NRS was >3.03 (95% CI: 2.90 to 3.16).Function PASS is defined as the value of measurement beyond which patients consider their functional state as satisfactory. |
| Ornetti, 2011410 Test: Patient NRS: "What is the degree of difficulty you have experienced for the daily activities during the last 48 hours due to your (knee or hip) OA" Method: Anchor  Worst to best: 10 to 0 | All patients had to assess their degree of improvement of global state (global MCII), on a three-point Likert scale (worsened function, no change, improved function). Among patients who improved, the degree of improvement was scored on a four-point Likert scale (poor, fair, good, excellent) | Patients considered their global state as improved for a change of patient NRS >-2.72 (95% CI: -2.92 to -2.51). Global MCII is defined as the smallest change in global state that signifies an important improvement in a patient's symptoms. |
| Ornetti, 2011410 Test: Physician NRS: Physician's Estimate of functional impairment of each patient Method: Anchor  Worst to best: 10 to 0 | All patients had to assess their degree of improvement of global state (global MCII), on a three-point Likert scale (worsened function, no change, improved function). Among patients who improved, the degree of improvement was scored on a four-point Likert scale (poor, fair, good, excellent) | Patients considered their global state as improved for a change of physician NRS >=2.50 (95% CI: -2.68 to -2.32).Global MCII is defined as the smallest change in global state that signifies an important improvement in a patient's symptoms. |
| Ornetti, 2011410 Test: Patient NRS: "What is the degree of difficulty you have experienced for the daily activities during the last 48 hours due to your (knee or hip) OA" Method: Anchor  Worst to best: 10 to 0 | MCII for functional state: The MCII of each function scale was defined as the 75th centile of the absolute change in score among patients whose final evaluation of response to NSAID was improved (improvement good or excellent). | Patients considered their functional state as improved for a change of patient NRS >=2.79 (95% CI: -3.01 to -2.57).Functional MCII is defined as the smallest change in functional state that signifies an important improvement in a patient's symptoms. |
| Ornetti, 2011410 Test: Physician NRS: Physician's Estimate of functional impairment of each patient Method: Anchor  Worst to best: 10 to 0 | MCII for functional state: The MCII of each function scale was defined as the 75th centile of the absolute change in score among patients whose final evaluation of response to NSAID was improved (improvement good or excellent). | Patients considered their functional state as improved for a change of physician NRS >=2.55 (95% CI: -2.73 to -2.38).Functional MCII is defined as the smallest change in functional state that signifies an important improvement in a patient's symptoms. |
| Mangione, 2010438 Test: Timed "Up & Go" test Method:  Distribution | At 2 times patients were tested: Each participant was asked to walk at his or her "normal" speed across the mat for 2 trials and then as "fast as possible" for 2 trials. The 2 trials of fast speed were averaged each session. | The MDC 90 was 4.0s |