Knee osteoarthritis patients can provide useful estimates of passive knee range of motion: development and validation of the Copenhagen Knee ROM Scale

Anne Mørup-Petersen, MD, Pætur Mikal Holm, PT, MSc, Christina Enciso Holm, MD, Tobias Wrenfeldt Klausen, MSc, Statistician, datamanager, Søren Thorgaard Skou, PT, PhD, Michael Rindom Krosgaard, MD, PhD, prof., Mogens Berg Laursen, MD, PhD, Anders Odgaard, MD, D.M.Sc. 1) Corresponding author: anne.moerup-petersen.02@regionh.dk. Copenhagen University Hospital, Gentofte, Kildegaardsvej 28, 2900 Hellerup, DENMARK.

Abstract

Background
Knee arthroplasty surgery does not always require extensive long-term follow-up. If knee range of motion (ROM) could be assessed reliably by patients, some follow-up visits might be replaced by patient-reported outcome measures, and this additional information could be reported directly to registers. We developed and tested the validity and reliability of a simple scale for patients to self-report their passive knee ROM.

Methods
Through an iterative process we created a 2-item scale with 11 illustrations of knee motion in 15° increments. The validity and reliability was tested in knee osteoarthritis and arthroplasty patients at different treatment stages, many with poor ROM. Patient estimates were compared to passive goniometer measurements performed blindly by a physiotherapist and a junior orthopedic surgeon.

Results
The mean difference between 100 patients’ (70.9 y.) estimates and goniometer measurements was -0.7° (SD 12.3°) for flexion and 1.1° (SD 11.6°) for extension, both not significant. Correlation was 0.79 and 0.63, and kappa values at retest were 0.84 and 0.66. For flexion < 110°, sensitivity of patient estimates was 88% and specificity was 88%. For a limit of 100°, values were 95% and 81%. For extension deficits > 10°, sensitivity was 78% and specificity 70%. Values were 100% and 66% for a 15° limit.

Conclusion
The Copenhagen Knee ROM Scale is a patient-friendly and feasible alternative to passive ROM measurement for registers, research and selected clinical use. This scale appears reliable and valid compared to reports of similar tools, and patient estimates are better correlated to goniometer measurements.
Keywords

Range of motion; patient-reported; patient-assessed; validation; knee arthroplasty; knee osteoarthritis.

Manuscript

Introduction

With increasing attention to the advantages of the use of patient-reported outcomes (PROMs) in knee arthroplasty surgery, it has been suggested that PROMs replace some postoperative clinical follow-up visits in uncomplicated cases [1-4]. An important barrier, though, is that information about range of motion (ROM) is not available if patients do not attend a health care clinic in person.

Attempts have been made to have patients self-report ROM and the need for a tool to make this possible has been recognized [1, 2, 4-6]. For surgeons to rely on patient-reported ROM to replace a clinical follow-up visit, the tool must be valid and sensitive. The same applies for use in research and registries. Previous studies have reported promising results [3-6], but we sought to explore whether a new, simple patient-reported ROM tool, in which patient-friendliness was highly prioritized, could provide satisfactory accuracy and sensitivity. The purpose of this study was to develop an illustration-based scale for patients to report passive knee ROM and to test the validity and reliability among knee osteoarthritis and knee arthroplasty patients.

Material and methods

Development process

Our first focus was to design a questionnaire, based on drawings, that patients of any adult age would easily understand and be able to complete unassisted at home. The process was guided by three relevant guidelines: 1) Guidelines for Reporting Reliability and Agreement Studies (GRRAS), 2) The STARD Statement for reporting studies of diagnostic accuracy and 3) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties of health status measurement instruments [7-12].

We met with 18 individual knee osteoarthritis (OA) patients (7 men, 11 women, mean age 69.9 years) who were facing or had just undergone knee arthroplasty surgery. They were asked to show in their own preferable way how much they could bend and straighten their affected knee. We
observed that the majority of patients got up from the chair or bed to show their extension, though some remained seated with their leg stretched out with the heel on the floor, or balancing the leg in the air in front of themselves. To show flexion most patients sat on a chair or remained in bed. Many patients used their hands to pull the ankle backwards.

Through an iterative process of testing and improving our draft illustrations, a total of 34 knee arthroplasty patients (23 pre- and 11 postoperative patients: 13 men, 21 women, mean age 70.4 y.) were shown several drafts of knees in different positions from a lateral view. Patients were asked to describe what they saw in the illustrations and mark the option that fit their knee motion. Some drafts had dotted horizontal and vertical lines to aid estimation of angles. However, with exception of an engineer and a carpenter, most people found the lines more confusing than helpful. Adding a seat and the contralateral leg as navigation points and adding arrows to show the direction of force gradually enhanced the patients’ understanding of the intentions of this tool (fig. 1). Instructions were made short, here regarding flexion: “How much can you bend your knee? Please push your lower leg as far back as possible. You can use your hand to pull your lower leg in the direction of the arrow. Tick the box that fits your situation.”

**Fig. 1:** Examples of illustrations of flexion and extension respectively. Full questionnaires are available in Appendix A.

For flexion, we found six pictures to be appropriate: 60, 75, 90, 105, 120 and 135°. For extension, five illustrations of 45, 30, 15, 0 and -15° were found suitable. We chose 15° increments between the pictures for three reasons: Firstly, only differences above 5-10° represent a true difference in ROM [13, 14]. Secondly, with 10° intervals even the authors of this paper were unable to tell the difference between neighbor illustrations. In the development phase, we noted that many patients
exaggerated their ROM, both in terms of good and bad results. Particularly flexion contractures were overestimated. Collins et al. [5] reported the same tendency when using 5-10° intervals. Therefore we considered more options to be redundant, as patients would use the scale widely no matter the underlying intervals between measures. Thirdly, our goal of making the questionnaire very easy to overview would be compromised with a higher number of illustrations.

We deliberately chose not to write the underlying angle on each picture because we wanted patients to report their unbiased perception of pictures. If angles had been shown there would be a risk of priming patients to a certain answer, in case they had recently been told their exact ROM measure, e.g. by their physiotherapist.

Options were placed in one row with the best score last to show a logical direction of motion. To meet patients with locked bandages or extremely limited motion, we made extra options named (for flexion) “Impossible. I am not able to bend my knee as much as in picture no. 1”.

Though all illustrations show left side knees, no patients were in doubt of which knee to think of. Some patients asked how much it was meant to hurt during testing. However, since pain level varies greatly, this subject could not be fit into instructions in a sensible manner.

The development process ended when there were no longer any new comments to facilitate meaningful changes in layout or wording, and patients understood the task without further explanation. The final version, Copenhagen Knee ROM Scale (CKRS) can be viewed in an English version in Appendix A and is available free of charge (with or without English or Danish text) at www.knee.dk/rom.

**Translation**

Questionnaires were evaluated and tested in the original Danish version. The wording was translated into English for publication, independently by three bilingual persons: One native English layperson and two native Danish doctors (one resident and one orthopedic knee arthroplasty surgeon with four years of experience from English hospitals). The three versions were combined to a final version by the first author. When in doubt, the native English layperson had the final say. The resulting English version of CKRS has not been evaluated among English speaking osteoarthritis patients.
Clinical testing

A patient-reported ROM tool is probably of most clinical value in the follow-up period after knee arthroplasty. However, it may be of great value for patients considering arthroplasty surgery to be informed of what knee motion to expect in the months and years following surgery, compared to the knee motion they have with OA. Therefore, we found the whole spectrum of patients, from the first visit in the arthroplasty clinic to the possible complications several years after arthroplasty surgery, to be of interest when validating this tool. For clinical reliability testing we aimed for a test group of patients reflecting this diversity. Mixing pre- and postoperative patients also gave the opportunity of selecting patients so that all degrees of knee motion were present, and thereby be able to test the validity of CKRS in the entire range of the ROM scale.

Patients were included in both the orthopedic arthroplasty ward and the arthroplasty out-patient clinic over a period of five days. Inclusion criteria were age $\geq$ 40 years and clinical visit due to knee OA or hospital stay due to knee arthroplasty operation of any kind. Patients with bilateral knee osteoarthritis were instructed to answer CKRS for the knee with the most restricted motion. Patients showing signs of dementia or confusion were excluded if they failed a “clock drawing test” [15], as were patients with poor Danish language skills without someone to translate for them. We also excluded hospitalized patients who were unable to get out bed and stand on their own (walking aids were allowed, though).

First, patients filled in the CKRS paper version without the opportunity to ask any questions. Visiting relatives were allowed to stay in order to mimic the situation at home. Completion time was not measured (to avoid stressing the patients), but it was our impression that the far majority of patients completed the form within 1-3 minutes. They were, however, allowed as much time as they needed. Immediately after completion, patients met a junior orthopedic registrar and one of two experienced physiotherapists. Patients were instructed not to reveal their answers, which all obeyed. Sitting on a normal chair, the patient demonstrated his or her maximal flexion once for each examiner, who then filled in the CKRS while the other examiner turned his or her back for blinding purposes. This was repeated for extension with the patient standing up. Patients were told to press on the knee or pull the lower leg with their own hands. Examiners were only permitted to palpate for bony landmarks.

Subsequently, goniometer ROM measurements were made using the same blinding strategy with the patient lying on an examination table wearing only underwear on the lower body. We used a
long goniometer (30 cm/12 inches, 1° increments) and navigated for bony landmarks; the greater trochanter, the lateral epicondyle of the femur and the lateral malleolus [16]. External hand pressure was applied by the examiner and the patient was told to say stop when it was enough. For extension measurements, we placed a firm cylinder back roll under the Achilles tendon.

Between each examination the knee was left in a relaxed position. The order of surgeon and physiotherapist examination was random. After all measurements were completed by the examiners individually, a consensus measurement was made by both examiners in collaboration.

Reproducibility
Examiners’ CKRS estimates and ROM measures were kept secret to patients. Patients were given a retest questionnaire together with a pre-paid envelope and were instructed to fill in the forms 7-10 days after the first session. Patients who participated during the first days after surgery were asked to perform the retest one or two days later because fast improvement was expected. Patients who were scheduled for surgery between test and retest were omitted from this part of the study.

Before filling in the CKRS again, patients were asked to confirm the affected side (left/right) and answer whether they had experienced any change in knee motion since the first examination. Only patients reporting “no change in knee motion” were included in retest analysis. No goniometer measurement was made on this occasion since the subject of interest was retesting patients’ perception of the scale. Retest questionnaires received later than six weeks after the first testing and retests dated on the day of the first examination were excluded from retest analysis.

Statistical analysis
Based on reports from similar studies, we proposed a sample size of 100 patients [4-6, 17, 18]. Sample size calculation aiming for a power of 0.8 based on an expected Pearson correlation coefficient of 0.7, a null correlation coefficient of 0.5 and two-sided alpha 0.05 suggested a sample size of 80. Since we could not expect a normal distribution of answers more patients were needed so we included 108 patients.

Descriptive statistics was made for all continuous variables including mean differences (mean goniometer measurement minus patient estimate) and 95 % confidence intervals [CI]. Paired t-tests were used for comparisons.

Goniometer measurement was regarded as a gold standard in our calculations [18]. To describe the measurement error of the CKRS tool, we calculated overall 95% limits of agreement (LoA) as mean
difference ± 1.96 x standard deviations (SD). With patients grouped by their CKRS answer, also group mean, SD, range and LoA were calculated (LoA only for groups larger than 15 patients) to ensure clinical applicability, because measurement error was expected to vary with ROM measures.

Sensitivity and specificity for clinically relevant limits were calculated with special consideration to comparability to previously published methods. For the same reason, also Pearson correlation coefficients between methods were calculated. These, however, require equal intervals between answer options. Since we could not guarantee that patients perceived intervals between illustrations to be equal, we also calculated Spearman rank correlation coefficient, which compares only the ranking of subjects.

From the mean goniometer measurement we calculated the “correct” CKRS answer that patients ideally should give. For example, flexion option 5 (120°) should cover the range from 112.5 - 127.5°. Absolute ROM measures (flexion minus extension) were not calculated due to their limited clinical relevance. Test-retest reliability was based on weighted kappa, paired t-tests and percentage agreement between patients’ first estimates and their retest estimates.

P-values below 0.05 were considered significant and were reported when relevant. All P-values were two-sided. Statistical analyses were made in SAS Statistical software (SAS University Edition, version 3.6, Cary, NC, USA). Ethical approval was provided by The National Committee of Health Research Ethics (Jr. no. 16030260) and data management was approved by The Danish Data Protection Agency (Jr. no. 2012-58-0004). Raw data for the primary tests are available in appendix B.

**Results**

A total of 113 patients were asked to participate, but five declined (excused by business or tiredness), so 108 knee OA patients (108 knees) were included (fig. 2). Three were excluded before testing; one had dementia and two were unable to get out of bed on their own. After testing, five more patients were excluded: one because goniometer measurements were performed on the contralateral knee by mistake, one patient failed to answer page two, and three patients had marked more than one option. The final test group ready for data analysis consisted of 100 patients: 59 patients with knee arthroplasty and 41 patients with knee OA. Patient characteristics are presented in table 1.
Fig. 2: Flow diagram of patients in validity and reliability studies.

Table 1: Characteristics of patients (n = 100) in the clinical test group.

<table>
<thead>
<tr>
<th>(n = 100)</th>
<th>Mean</th>
<th>Mean CI</th>
<th>Median</th>
<th>Range (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y.)</td>
<td>70.9</td>
<td>[69.0; 72.8]</td>
<td>72.6</td>
<td>44.3 – 89.3</td>
</tr>
<tr>
<td>BMI (kg/m²) (n = 99)</td>
<td>28.7</td>
<td>[27.6; 29.9]</td>
<td>27.5</td>
<td>19.6 - 46.7</td>
</tr>
</tbody>
</table>

Sex
- Men: n = 35
- Women: n = 65

Knee arthroplasty
- No: n = 41
- Yes: n = 59 → 45 TKA, 13 mUKA and 1 revision TKA
  Time after surgery: mean 14.8 months
  (median 3 months, range 0 days to 16 years)

(TKA: Total Knee Arthroplasty, mUKA: medial Unicompartmental Knee Arthroplasty)
The goniometer measurements of the surgeon and the physiotherapists were well aligned; mean difference was 0.8° (SD 4.2°, range -13 to 9°, P = 0.06) for flexion and 1.1° (SD 3.0°, range -7 to 7°, P < 0.001) for extension. Therefore, the consensus measurements were deemed unnecessary and left out of analysis as they are not typical of everyday practice. In the following, “goniometer measurement” refers to the mean of the two examiners’ measurements.

**Flexion**

Goniometer measurements of flexion ranged from 62.5 to 150.5° (mean 115.7°, SD 19.6°). In CKRS only one examiner and no patients made use of option 1, and no one marked option 0. 55% of patients had chosen the “correct” picture and 94% were within one adjacent option. No one was further than two options away from the correct answer.

The mean difference between patient estimates in 15° intervals and goniometer measurement was -0.7° [CI: -3.2; 1.7°], P = 0.56. Differences were normally distributed with overall SD 12.3° and total range from -32 to 28°. Hence, overall 95 % limit of agreement (LoA) was 0.7 ± 24.0°.

Patient-reported flexion on CKRS had a strong Pearson correlation of 0.79 [0.70; 0.85] to goniometer measurements. The according Spearman rank correlation was 0.80 [0.71; 0.86]. Fig. 3 shows boxplots of goniometer measurements for patients grouped by their own ROM estimates on CKRS. Measurements, SD and LoA for each group are listed for clinical applicability in table 2.

![Fig. 3: Boxplot of goniometer measurements of flexion grouped by patients’ estimates on Copenhagen Knee ROM Scale.](image-url)
**Table 2**: Goniometer measurements of flexion. Patients are grouped (1-6) by their own estimate of flexion on Copenhagen Knee ROM Scale.

*LoA = Limits of agreement based on the SD for each group (column). LoA is only calculated for groups larger than 15 patients. Standard deviations are in () for the same reason.

<table>
<thead>
<tr>
<th>Patient estimate (picture no.)</th>
<th>Underlying (hidden) measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60°</td>
<td>75°</td>
<td>90°</td>
<td>105°</td>
<td>120°</td>
<td>135°</td>
<td></td>
</tr>
<tr>
<td>Patients (n = 100 total)</td>
<td>0</td>
<td>6</td>
<td>11</td>
<td>19</td>
<td>29</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td><strong>Goniometer measurements (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-</td>
<td>82.5</td>
<td>95.8</td>
<td>100.5</td>
<td>121.4</td>
<td>131.2</td>
<td></td>
</tr>
<tr>
<td>[95% CI of the mean]</td>
<td>[71.7 ; 93.3]</td>
<td>[88.0 ; 103.5]</td>
<td>[93.5 ; 107.5]</td>
<td>[117.2 ; 125.6]</td>
<td>[128.0 – 134.4]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>-</td>
<td>84.5</td>
<td>94.5</td>
<td>101.0</td>
<td>123.0</td>
<td>130.0</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>-</td>
<td>(13.2)</td>
<td>(12.9)</td>
<td>15.2</td>
<td>11.3</td>
<td>9.5</td>
<td></td>
</tr>
<tr>
<td>95% LoA (Mean ± 1.96 SD)*</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>70.7 – 130.3</td>
<td>99.2 – 143.5</td>
<td>112.5 – 149.8</td>
<td></td>
</tr>
<tr>
<td>Total range (min – max)</td>
<td>-</td>
<td>62.5 – 96.5</td>
<td>66.5 – 113.0</td>
<td>73.0 – 125.0</td>
<td>89.0 – 148.0</td>
<td>109.5 – 150.5</td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity and specificity was calculated for clinically relevant values (table 3). For example, if flexion ≥ 100° is considered acceptable, CKRS is able to detect 95% of patients with an unsatisfactory flexion using cut-off between illustration 4 and 5. The specificity in this case is 81%.

In this population, where many patients had poor knee motion, 64% of patients marked ≥ option 5. The negative predictive value, i.e. the probability that a patient who marked ≥ option 5 did in fact have flexion ≥ 100° was 98%. Correspondingly, the positive predictive value of having flexion < 100° was 100% for patients marking ≤ option 2. Similar calculations for other relevant values are listed in table 3.
# Limits of Knee Motion

<table>
<thead>
<tr>
<th>Limit of knee motion</th>
<th>CKRS threshold (between pictures no.)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Positive predictive value (%)</th>
<th>Negative predictive value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexion &lt; 90°</td>
<td>2 - 3</td>
<td>31</td>
<td>98</td>
<td>67</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>3 - 4</td>
<td>46</td>
<td>87</td>
<td>35</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>4 - 5</td>
<td>92</td>
<td>72</td>
<td>33</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>5 - 6</td>
<td>92</td>
<td>72</td>
<td>33</td>
<td>98</td>
</tr>
<tr>
<td>Flexion &lt; 100°</td>
<td>2 - 3</td>
<td>27</td>
<td>100</td>
<td>100</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>3 - 4</td>
<td>55</td>
<td>94</td>
<td>71</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>4 - 5</td>
<td>95</td>
<td>81</td>
<td>58</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>5 - 6</td>
<td>100</td>
<td>45</td>
<td>34</td>
<td>100</td>
</tr>
<tr>
<td>Flexion &lt; 110°</td>
<td>2 - 3</td>
<td>19</td>
<td>100</td>
<td>100</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>3 - 4</td>
<td>47</td>
<td>97</td>
<td>88</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>4 - 5</td>
<td>88</td>
<td>88</td>
<td>78</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>5 - 6</td>
<td>97</td>
<td>50</td>
<td>48</td>
<td>97</td>
</tr>
<tr>
<td>Flexion &lt; 120°</td>
<td>4 - 5</td>
<td>65</td>
<td>96</td>
<td>94</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>5 - 6</td>
<td>90</td>
<td>63</td>
<td>72</td>
<td>86</td>
</tr>
<tr>
<td>Extension ≥ 15°</td>
<td>2 - 3</td>
<td>57</td>
<td>92</td>
<td>36</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>3 - 4</td>
<td>100</td>
<td>66</td>
<td>18</td>
<td>100</td>
</tr>
<tr>
<td>Extension ≥ 10°</td>
<td>2 - 3</td>
<td>50</td>
<td>98</td>
<td>82</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>3 - 4</td>
<td>78</td>
<td>70</td>
<td>36</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>4 - 5</td>
<td>100</td>
<td>26</td>
<td>23</td>
<td>100</td>
</tr>
<tr>
<td>Extension ≥ 5°</td>
<td>2 - 3</td>
<td>22</td>
<td>100</td>
<td>100</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>3 - 4</td>
<td>57</td>
<td>78</td>
<td>72</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>4 - 5</td>
<td>90</td>
<td>31</td>
<td>55</td>
<td>76</td>
</tr>
</tbody>
</table>

*Guide to interpretation of values:*

**Sensitivity:** The chance that a patient with a knee motion worse than the specified limit is identified as having an abnormal knee motion using this Copenhagen Knee ROM Scale (CKRS) threshold.

**Specificity:** The chance that a patient with better knee motion than the specified limit is identified as having acceptable knee motion using this CKRS threshold.
**Positive predictive value:** The chance that a patient reporting knee motion worse than the specified CKRS threshold does have a knee motion worse than the specified limit. (Note that this value is population specific).

**Negative predictive value:** The chance that a patient reporting better knee motion than the specified CKRS threshold does have a knee motion better than the specified limit. (Note that this value is population specific).

**Table 3:** Limits of goniometer measurements of knee motion and the according sensitivities and specificities for detecting problematic knee motion using different thresholds (cut-off values) in Copenhagen Knee ROM Scale. Negative and positive predictive values in the test population are listed. Authors find the numbers in **bold** to be of largest clinical relevance.

### Extension

Passive extension measurements ranged from -8.5 to 35° (mean 5.8°, SD 6.5°). All CKRS illustrations except option 0 (> 45°) were used by the patients. The correct illustration was chosen by 45% of patients, 99% were within one option from the correct and one patient was two from the correct answer.

The mean difference between patient estimate and goniometer measurement was 1.1° [-1.2; 3.4], P = 0.35. Overall differences were normally distributed with an SD of 11.6° and LoA 1.1 ± 22.8° (total range -34 to 22.5°). However, goniometer measurements for each CKRS group reveal how patients perceive only 2-9° intervals between pictures instead of the actual 15° intervals that drawings are measured by (table 4). For example, the mean goniometer measurement for patients marking picture 5 (-15°) was 0.7° and for patients marking picture 4 (0°) it was 4.9°. Boxplots of patient estimates against their goniometer measurements illustrate the same phenomenon: the slope is not as steep as if there was perfect agreement (fig. 4).
### Measurements of extension

<table>
<thead>
<tr>
<th>Patient estimate (picture no.)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underlying (hidden) measure</td>
<td>45°</td>
<td>30°</td>
<td>15°</td>
<td>0°</td>
<td>-15°</td>
</tr>
<tr>
<td>Patients (n = 100 total)</td>
<td>2</td>
<td>9</td>
<td>28</td>
<td>40</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Goniometer measurements (°)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>23.0</td>
<td>14.4</td>
<td>6.8</td>
<td>4.9</td>
</tr>
<tr>
<td>[95% CI of the mean]</td>
<td>[-1.0; 47.0]</td>
<td>[10.0; 18.9]</td>
<td>[4.8; 8.7]</td>
<td>[3.7; 6.1]</td>
</tr>
<tr>
<td>Median</td>
<td>23.0</td>
<td>14.5</td>
<td>6.5</td>
<td>4.5</td>
</tr>
<tr>
<td>SD</td>
<td>(17.0)</td>
<td>(6.7)</td>
<td>5.1</td>
<td>3.9</td>
</tr>
<tr>
<td>95% LoA (Mean ± 1.96 SD)*</td>
<td>*</td>
<td>*</td>
<td>-3.3 – 16.8</td>
<td>-2.7 – 12.5</td>
</tr>
<tr>
<td>Total range (min – max)</td>
<td>11.0 – 35.0</td>
<td>8.0 – 30.0</td>
<td>-6.0 – 17.5</td>
<td>-5.5 – 13.5</td>
</tr>
</tbody>
</table>

**Table 4:** Goniometer measurements of extension. Patients are grouped (1-5) by their own estimate of extension on Copenhagen Knee ROM Scale.

*95% LoA = Limits of agreement based on the SD for each group (column). LoA is only calculated for groups larger than 15 patients. Standard deviations are in () for the same reason.

![Boxplot of goniometer measurements of extension grouped by patients’ estimates on Copenhagen Knee ROM Scale.](image)

**Fig. 4:** Boxplot of goniometer measurements of extension grouped by patients’ estimates on Copenhagen Knee ROM Scale.

On group level, variation was far lower than the overall variation; SD was 5.1, 3.9 and 4.6°, respectively, for the three pictures covering 89% of patients tested.
Pearson’s correlation coefficient was 0.63 [0.49; 0.73] (moderate) between patient estimates and goniometer measurements and 0.57 [0.42; 0.69] using Spearman’s rank correlation coefficient.

Sensitivity and specificity values are listed in table 3. If 15° is considered acceptable passive extension, a cut-off between option 3 and 4 offers a sensitivity of 100% at the cost of a specificity of 66%. If extension limit is lowered to 10°, the according values are 78 and 70% respectively. In this population, the negative predictive value, i.e. the chance of not having an extension deficit > 10° when answering ≥ option 4 was 93%. By contrast, the positive predictive value of having an extension deficit > 10° was 82% for patients marking ≤ option 2.

Reliability
Retest questionnaires were handed out to the 93 patients that were not awaiting surgery within a week (45 of patients already had a knee arthroplasty). We received 71 answers dated mean 8.7 days later (median 8 days, range 1-24 days). 54 patients had replied “no change in knee motion” and so were eligible for retest analysis (fig. 2). Of these, 32 were arthroplasty patients and 22 were OA patients.

In both flexion and extension there were no overall differences between values in test and retest (P = 0.25 and 0.35 respectively). 45 patients (83%) gave the exact same answer regarding flexion as in their first test. For extension, the number was 36 (68%) and perfect agreement on both parameters was reached in 32 (59%) of the cases. Weighted kappa value for flexion was 0.84 [0.74; 0.94] which represents “almost perfect” test-retest reliability [19]. For extension, weighted kappa reached 0.66 [0.52; 0.80] representing “substantial” reliability.

The role of age, BMI and arthroplasty status
We found no correlation between BMI and measurement error (= absolute difference between patient estimate and goniometer measurement). Nor was there any correlation between increasing age and measurement error. On the contrary, the only significant outcome was for flexion, where a weak Pearson correlation of -0.20 [-0.38; -0.01] (P = 0.04) indicated that older patients made more accurate estimates than younger patients. A comparison of measurement error between the 41 OA patients and the 59 arthroplasty patients revealed no difference in their ability to estimate ROM using this scale (P = 0.75 for flexion and 0.68 for extension using unpaired t-test of unequal variances).
Examiner’s estimates of ROM

Both examiners were aware of the underlying angles behind CKRS illustrations. Their estimates of CKRS prior to measuring agreed well with passive goniometer measurements: the mean difference was 1.6° for flexion (SD 6.7°, Pearson’s r 0.94) and 1.1° for extension (SD 4.6°, r 0.84). Examiners appointed the correct flexion option in 67% of cases, the adjacent option in 32.5% and were two apart in 0.5% of cases. For extension, 77% of estimates were correct and 23% were one away.

Discussion

Validity

Our aim to develop an easily understandable questionnaire was obtained. The scale measures the intended items and through the whole range of motion there were no severe outliers. Furthermore, measurement error was unaffected by BMI and age.

We consider it a strength that CKRS was tested in a diverse group of knee OA and arthroplasty patients. Our inclusion of many patients with poor ROM has furthermore confirmed the validity of the scale in the whole range of both flexion and extension.

It is an advantage that the whole leg and the contralateral leg are both visible. Arrows clearly indicate which motion is requested and instructions are condensed to a minimum. To enhance patient-relevance we have let patient positioning be directed by patient preferences. Drawings are simplistic and, as opposed to photographs, they are neutral in terms of race, sex and age.

Even though option 0 was not used at all and flexion option 1 was only used once in the test setting, we have kept both options in CKRS as they were suggested by patients in the development phase and therefore relevant. In a prospective study of 1600 knee arthroplasty patients, where CKRS was applied five times during the first year after surgery (unreported, personal information), patients marked option 0 in 0.3% and option 1 in 2.5% of cases.

It can be argued that the 15° increments lower the precision of this tool. But unlike other published methods, that use 5-10° intervals [3, 4, 6], we found no overall systematic difference between patient estimates and goniometer measurements for either flexion or extension. This supports our hypothesis, that simplicity should be prioritized over small increments between options. In our study, Pearson’s correlation coefficient was 0.79 for flexion, which is remarkably higher than the 0.44 reported by Gioe et al. and the 0.35 reported by Khanna et al. [3, 4]. For extension, Gioe found a correlation of 0.31 between patient estimate and measurement; this was 0.13 in Khanna’s study.
and 0.63 in the present study. These differences suggest that CKRS has an advantage over other tools in measuring the item of interest; even though we examined passive ROM, which is normally considered to require help from an examiner, we did not find any overall difference between patient estimates and measurements.

**Measurement error**

In flexion, obviously the variance was larger for patient estimates versus goniometer measurements (SD 12.3°) than between goniometer measurements (SD 4.2°) and also larger than the variance of examiners’ CKRS estimates vs. goniometer measurements (SD 6.7°). It was, however, lower than the 20.6° reported by Borgbjerg et al. [6] and comparable to the 12.4° and 12.8 ° reported by Gioe and Khanna respectively [3, 4].

In extension, though we found no systematic overall mean difference between methods, patients did exaggerate both extension deficits and hyperextension; they used the scale more widely than actual measurements justified. This was expected from our experiences in the development phase and from reports from similar studies [5, 6, 17]. A patient bothered by an extension deficit of 10° may well feel like option no. 2 (30°) illustrates. So, when extension measurements were grouped by CKRS answers, there were systematic differences, but this can be compensated for by using the grouped values (table 4) in the interpretation of clinical results. Here, we have taken the consequence and let patients be the judge of what they see in ROM illustrations. For example, 95 % of patients marking picture 3 (15°) can be expected to have passive extension between -8.4 and 9.8°. Collins et al. reported a quite similar distribution of extension estimates using 5-10° increments (Gioe’s method), regarding active ROM, though [5]. Except for Borgbjerg, who presented Bland-Altman plots, none of the other previously mentioned publications provided information about the distribution of measurement errors in relation to absolute ROM measures [3-6].

The finding of considerable overlap between groups of answers lowers the precision of the patient estimates of extension. The overall SD for patient estimates in CKRS vs. goniometer measurements was 11.6°, whereas SD for examiners’ CKRS estimates vs. goniometer measurement was 4.6° and between goniometer measurements only 3.0°. On group levels, however, SD’s for patient-reported extension on CKRS were only 5.1, 3.9 and 4.6° for the three most used options (table 4). This is lower than Borgbjerg’s SD of 10.6° and quite similar to the 5.0, 4.4 and 4.2-6.7° reported by Gioe,
Khanna and Collins respectively [3-5]. However, patient groups are difficult to compare, since Gioe and Khanna included > 1 year postoperative patients only, with absolute extension measures ranging 1.4 ± SD 4.3° and 0.5 ± 2.5° respectively. Our mix of patients had a wider extension range of 5.8 ± SD 6.5°, which was comparable to Collins’ measures of active extension. To summarize, these findings together with the correlation coefficients lead us to argue, that passive extension is most precisely measured with CKRS. Also, we find no reason to believe that accuracy would increase with smaller increments between illustrations.

The lack of precision in extension is caused by several factors: Even for healthcare professionals extension measurements are more difficult than flexion measurements [20]. Estimates are affected by the extension ability of the contralateral leg. Compared to measurements of flexion, the relative difference between passive and active extension is larger given the small absolute numbers of degrees in extension. Also, instead of answering the question of passive motion, patients may be answering whether or not extension poses an actual problem for them in daily life. For example, one would expect patients answering CKRS option 4 or 5 (0 and -15°) to be satisfied with their extension. To evaluate this hypothesis would require a new study asking the additional question:”Is the extent to which you can straighten the knee a problem for you?” This might demonstrate whether CKRS offers a better identification of patients who feel a need for e.g. additional physiotherapy than passive goniometer measurements or a simple “yes/no” question would.

**Clinical use**

Copenhagen Knee ROM Scale may be included as extra information in registries and surveys, and it may be a feasible replacement for professional goniometer measurement in some clinical settings. Whether patient-reported ROM using CKRS offers the necessary level of accuracy varies between different settings and must be evaluated for each situation [21-23].

The basis for application of CKRS as a screening tool is that the positive and negative predictive values are acceptable. A limitation to this study is that the predictive values given are not applicable to all populations of knee OA patients, as values change with the distribution of ROM measures. Values for sensitivity and specificities, however, are directly applicable to other knee OA patients. Which thresholds to use when applying CKRS to a patient group solely depends on the purpose of testing (screening, monitoring, surveillance etc.). For example, if a clinic aims to identify postoperative patients in need of intensive physiotherapy or manipulation under anesthesia, e.g.
within three months after surgery, a flexion limit of 95-110° and a CKRS threshold between option 4 and 5, or even 5 and 6 may be appropriate to ensure a high sensitivity [1, 2, 5, 22-25].

With the accelerating advancements in everyday technology, intelligent knee braces and various smartphone goniometer apps are being developed and are likely to become a natural part of future knee patients’ rehabilitation programs [25-29]. One example is the “DrGoniometer” app which has recently been proven valid for knee ROM measurement [30]. A selected group of arthroplasty patients had an accompanying relative or friend take photographs of the knee in a simulated home setting. Photographs were sent to the staff who made the ROM measurements using the app. Though precision was higher than what we have found in the present study, we find that a tool such as CKRS still has some advantages over technical solutions: CKRS requires no professional intervention, it is self-explaining and quickly completed without removal of clothes, and it involves no expensive or technical equipment; factors which may be of particular importance when applied to large groups of elderly patients.

Written instructions in CKRS are brief. However, we recommend careful translation and cultural validation of text and illustrations before use in other languages. Though we have suggested an English translation, ideally, future translations should be based on the original Danish version until the English version is validated in an English speaking patient population (both versions are available at www.knee.dk/rom). Further validation is especially important if the tool is used for other types of knee patients, since younger or more active patients may differ in their perception of both ROM and CKRS illustrations.

In the present study we focused on validity testing against a gold standard (goniometer measurement) in a relevant patient population. Responsiveness testing was not conducted. We do of course welcome further studies on change in CKRS estimates and ROM before and after knee arthroplasty. A mapping of patients’ baseline values and postoperative results after different types of arthroplasty operations could benefit the preoperative expectation alignment, and using this instrument would provide a unique opportunity to compare ROM-values across arthroplasty types, centers and patient populations without the bias of having surgeons or physiotherapists perform measurements on their own patients.

**Conclusion**

Copenhagen Knee ROM Scale is a patient-friendly and feasible tool for knee OA and arthroplasty patients to self-report their passive knee ROM for use in long-term follow-up as well as knee
registries and research when professional goniometer measurement is not a feasible option, or when a virtually unbiased ROM estimate is desirable. We recommend further studies to prove responsiveness to change e.g. after knee arthroplasty operation. With 15° increments between answer options we have reached better correlation with goniometer measurement than what was reported with similar tools using 5-10° increments. Furthermore, we have reached at least the same level of accuracy and strong retest reliability, particularly regarding flexion. We believe this tool meets the appropriate level of ambition in the field of patient-reported passive knee ROM.

Acknowledgements

Authors would like to send special thanks to physiotherapist Hanne Hornshøj. Authors thank the following persons for their contributions to this work: Henrik Morville Schröder, Carsten Bogh Juhl, Jonathan David Comins, Anna Pors Nielsen, Michelle Möger Andersen, Thomas Linding Jakobsen, Camilla Ryge, Karen Dyreborg, and patients and staff at the Departments of Orthopedic Surgery at Gentofte and Naestved Hospitals.

Funding

This work was indirectly supported by the Health Research Fund of the Capital Region of Denmark, grant 2015, part one. The funding source had no influence on the content of the study.

Appendices

Appendix A: Copenhagen Knee ROM Scale (English) (PDF).

Appendix B: Raw data set (Excel).

References