

APPENDIX

CONTENT

1. Anchor questions and response categories
2. MIC calculation methods:
 - Predictive modeling approach
 - ROC method
 - Mean change method
3. Sensitivity analysis: Adjusted MIC_{pred} analyses stratified by postoperative follow-up timepoint
 - Table 1
4. ROC analyses
 - Figure 1a-e
 - Table 2
5. Mean change scores for patients having acceptable symptoms, considering the treatment to have failed, or in the undecided intermediate group
 - Tables 3-4
6. Correlation between KOOS change scores and anchor response categories

1. ANCHOR QUESTIONS AND RESPONSE CATEGORIES

Subscale-specific MIC anchor questions:

1. How is your knee **pain** now compared to prior to your operation?
2. How are the other **symptoms** in your knee now compared to prior to your operation (swelling, stiffness, decreased range of motion)?
3. How did the operation affect your ability to perform **daily activities** (sitting, standing, walking, stairs, putting on/taking off socks, household work)?
4. How did the operation affect your ability to perform **sport and recreational activities** (running, jumping, squatting, kneeling, twisting on loaded knee)?
5. How did the operation affect the type of **quality of life** that relates to your knee (trust in knee, life style, how often you think of your knee, etc)?

Response categories:

- Better, an important improvement
- Somewhat better, but enough to be an important improvement
- Very small change, not enough to be an important improvement
- About the same
- Very small change, not enough to be an important worsening
- Somewhat worse, but enough to be an important worsening
- Worse, an important worsening

Acceptable symptoms or treatment failure anchor questions:

Considering your knee function, do you feel that your current state is satisfactory?

With *knee function* you should take into account all activities you have during your daily life, sport and recreational activities, your level of pain and other symptoms, and also your knee-related quality of life.

Yes No

The following question should only be answered if your answer to the former question was NO.

If you answered NO to the former question, would you consider your current state as being so unsatisfactory that you think the treatment has failed?

Yes No

2. MIC calculation methods

Predictive modeling approach

MIC_{pred} values were calculated using logistic regression analyses with the classification of improved or not as the dependent variable and KOOS change score as the independent variable. Patients were classified as being importantly improved when responding “somewhat better” or “better”, and not importantly improved when responding “unchanged”, “very small change”, “somewhat worse” or “worse” to the anchor. MIC_{pred} is the KOOS change score corresponding to a post-test probability of improvement being equal to the pre-test probability of improvement (i.e. the average improvement rate) (Likelihood ratio =1). In other words, when change exceeds the MIC_{pred} value, the patients’ likelihood of being importantly improved is higher than the average improvement rate.³ Moreover, simulation studies have shown that MIC_{pred} equals the mean of the hypothetical latent individual MICs in a group of patients.² However, MIC_{pred} values overestimate the mean individual MIC when the proportion of improved patients is >50%. The MIC_{pred} values were therefore adjusted for this overestimation using the equation $MIC_{adjusted} = MIC_{pred} - (0.090 + 0.103 \times Cor) \times SD_{change} \times \log\text{-odds}(imp)$,² where *Cor* is the point biserial correlation between the change score and the anchor, *SD_{change}* is the change score SD, and *log-odds(imp)* represents the natural logarithm of (proportion improved/[1 - proportion improved]). Bootstrap replications (n=1000) were used to determine 95% confidence intervals (CI) for MIC_{pred}.³

Mean Change Method

MIC_{MeanChange} values were calculated as the KOOS mean change score in the subgroups of patients responding “*Somewhat better*”.¹ The 95% CI for the MIC_{MeanChange} were calculated as $Mean_{change} \pm 1.96 \left(\frac{SD_{change}}{\sqrt{n}} \right)$, where *n* and *SD_{change}* correspond to the subgroup “*somewhat better*”.⁴

ROC method

With the ROC method, the anchors functioned as ‘gold standards’, classifying patients as being importantly improved (“*somewhat better*” or “*better*”) or not (“*unchanged*”, “*very small change*”, “*somewhat worse*” or “*worse*”) and KOOS change as “*test-scores*”. The MIC_{ROC} value was defined by the Youden index, which corresponds to the highest combination of sensitivity and specificity. CI for MIC_{ROC} values were defined using 1000 non-parametric bootstrap replications.^{4,5}

3. SENSITIVITY ANALYSES: ADJUSTED MIC_{PRED} ANALYSES STRATIFIED BY POSTOPERATIVE FOLLOW-UP TIMEPOINT

Methods: To investigate whether adjusted MIC_{pred} values differed with increasing time from surgery, adjusted MIC_{pred} analyses were performed for the 6, 12 and 24 months follow-up timepoint separately.

Results: When stratified by follow-up timepoints, adjusted MIC_{pred} values differed by less than 2.4 points from the adjusted MIC_{pred} values calculated for the total cohort of patients, combining all three follow-up timepoints (Table 1).

Table 1. Sensitivity analyses: Adjusted predicted MIC values stratified by postoperative follow-up timepoints.

KOOS	Timepoint	MIC _{pred}	95% CI low	95% CI high
Pain	Total cohort	2,5	0,4	4,5
	6 months	2,2	-1,0	5,2
	12 months	2,0	-1,8	5,5
	24 months	3,2	-0,7	7,2
Symptoms	Total cohort	-1,2	-3,2	0,8
	6 months	-0,1	-4,2	3,9
	12 months	-2,7	-6,6	0,7
	24 months	-0,6	-4,7	3,2
ADL	Total cohort	2,4	0,7	4,1
	6 months	2,7	0,7	4,8
	12 months	2,8	-0,3	6,0
	24 months	1,3	-2,0	5,0
Sport/Rec	Total cohort	12,1	9,3	14,8
	6 months	10,7	6,1	15,1
	12 months	11,5	6,5	16,3
	24 months	14,4	8,8	20,0
QOL	Total cohort	18,3	16,0	20,6
	6 months	17,5	13,5	21,5
	12 months	16,9	12,8	20,6
	24 months	20,7	17,0	24,0

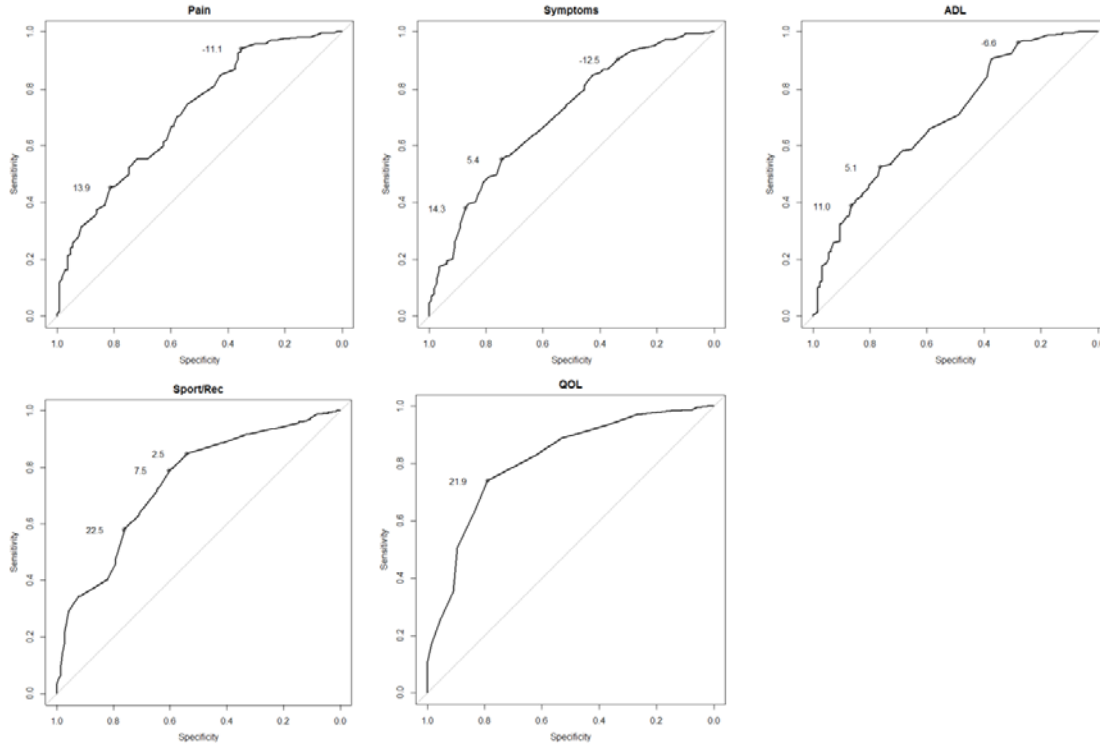
Conclusion: Adjusted MIC_{pred} values calculated using the total patient sample (all three follow-up timepoints combined) are considered applicable for use 6-24 months postoperatively.

4. ROC ANALYSES

Two MIC_{ROC} improvement analyses were performed for each KOOS subscale. The primary analyses were performed with the total sample of patients: with 'unchanged' and 'deteriorated' patients combined in the 'not importantly improved' groups. Sensitivity analyses were performed by excluding 'deteriorated' patients from the analyses.

With the total sample of patients, ROC curves were generally flattened (Figures 1a-e), the overall misclassification rates (the sum of 1-sensitivity + 1-specificity) were generally high for all subscales (lower for the QOL subscale) and AUCs were generally low for Pain, Symptoms, ADL and Sport/Rec (range 0.70-0.74), and higher for QOL (0.81) (Table 2). For the QOL subscale, the results were more reliable. In this

subscale, the ROC curve was more curved, the AUC was higher and the misclassification rate lower. The more reliable results experienced with the QOL subscale were probably due to a higher correlation between the anchor question and KOOS change scores.



Figures 1a-e. ROC curves for each KOOS subscale. The points presented on each curves correspond to the “best” ROC cut-off point and the 95% CI.

Table 2. ROC MIC results for whole sample of patients, with deteriorated patients excluded and when excluding patients with preoperative scores ≥ 85 .

	MIC _{ROC}	95% CI ¹	Sensitivity	Specificity	AUC ²
Pain					
Total sample n=490	-11.1	-11.1;13.9	0.94	0.36	0.71
Deteriorated excluded n=461	13.9	-11.1;18.1	0.45	0.82	0.69
Symptoms					
Total sample n=517	5.4	-12.5;14.3	0.55	0.75	0.71
Deteriorated excluded n=485	5.4	-7.1;14.3	0.55	0.74	0.70
ADL					
Total sample n=488	5.1	-6.6;11.0	0.53	0.76	0.70
Deteriorated excluded n=469	5.1	-3.7;11.0	0.53	0.77	0.69
Sport/Rec					
Total sample n=501	7.5	2.5;22.5	0.79	0.60	0.74
Deteriorated excluded n=449	2.5	2.5;42.5	0.84	0.46	0.71
QOL					
Total sample n=519	21.9	21.9;21.9	0.74	0.84	0.81
Deteriorated excluded n=484	21.9	21.9;28.1	0.74	0.78	0.80

1: 95% CI performed by 1000 nonparametric bootstrapped replications and reported as 0.025-0.975 quantiles

2: AUC: Area Under the Curve

Investigating the ROC cut-off points further, we found that the degrees of misclassification were more or less the same for the range of cut-off points within the large confidence intervals (Figures 1a-e). This finding indicates that the discriminative power is more or less the same across a large range of scores.

Furthermore, confidence intervals for the MIC_{ROC} values are much wider than confidence intervals for the MIC_{pred} values, indicating that predictive modeling results in a much more precise estimate of the MIC than the ROC method. This finding is concordant with Terluin et al.'s previous simulations of the predictive modeling approach.^{2,3}

Excluding the deteriorated patients from the sample is sometimes preferred for ROC MIC analyses⁵. In this study, when excluding the deteriorated patients, we found that MIC values increased markedly by 25 points to a value of 13.9 for Pain, and for Sport/Rec the MIC decreased by 5 points to a value of 2.5. For the other subscales, the MIC cut-off points were identical, but confidence intervals were wider for all subscales when the deteriorated patients were excluded from the analyses (Table 2).

Conclusion: Due to the generally flattened ROC curves, the selection of the “best” cut-off point according to the Youden principle became highly vulnerable to minor chance fluctuations in the sample. The MIC_{ROC} values were therefore considered unreliable.

5. MEAN CHANGE SCORES FOR PATIENTS HAVING ACCEPTABLE SYMPTOMS, CONSIDERING THE TREATMENT TO HAVE FAILED, OR BEING UNDECIDED ABOUT TREATMENT OUTCOME

Table 3. Mean change scores for patients achieving acceptable symptoms, considering the treatment to have failed, or in the undecided intermediate group.

KOOS	Acceptable symptoms	Undecided intermediate	Treatment Failure
	Mean ± SD	Mean ± SD	Mean ± SD
Pain	13.91±17.68	6.30±16.55	-6.15±18.26
Symptoms	9.79±17.85	3.57±18.33	-10.71±21.22
ADL	10.88±15.92	5.74±15.61	-3.86±17.74
Sport/Rec	29.07±26.42	11.79±23.52	-11.25±27.67
QOL	36.87±21.40	19.22±18.03	-0.26±17.35

Table 4. Mean change scores for patients experiencing important improvement, no important change and important deterioration.

KOOS	Importantly improved	Not importantly changed	Importantly deteriorated
	Mean ± SD	Mean ± SD	Mean ± SD
Pain	12.81±17.53	0.00±17.11	-5.65±16.62
Symptoms	9.15±17.95	-3.30±17.86	-10.38±20.23
ADL	10.98±15.62	0.22±14.76	-8.36±18.80
Sport/Rec	27.63±26.90	7.17±25.01	-3.94±29.33
QOL	34.69±21.13	11.3±18.91	4.11±20.16

6. CORRELATION BETWEEN KOOS CHANGE SCORES AND ANCHOR RESPONSE CATEGORIES

Table 5. Spearman's correlations between anchor response categories and KOOS change scores and postoperative scores, respectively.

KOOS	Correlation with change score	Correlation with postoperative score
Pain	0.32	0.40
Symptoms	0.39	0.44
ADL	0.32	0.32
Sport/Rec	0.41	0.55
QOL	0.53	0.56

The anchor responses correlated more with the postoperative KOOS scores, than the KOOS change scores for all subscales but the ADL, where the correlation was the same (Table 5).

REFERENCES

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415. doi:10.1016/0197-2456(89)90005-6.
2. Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. *J Clin Epidemiol*. 2017. doi:10.1016/j.jclinepi.2016.12.015.
3. Terluin B, Eekhout I, Terwee CB, de Vet HCW. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. *J Clin Epidemiol*. March 2015. doi:10.1016/j.jclinepi.2015.03.015.
4. Terwee CBB, Roorda LDD, Dekker J, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol*. 2010;63(5):524-534. doi:10.1016/j.jclinepi.2009.08.010.
5. De Vet HCW, Ostelo RWJG, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res*. 2007;16(1):131-142. doi:10.1007/s11136-006-9109-9.