Canadian Orthopaedic Association
Association canadienne d'orthopédie

Physicians' accuracy and interrator reliability for the diagnosis of unstable meniscal tears in patients having osteoarthritis of the knee

Geoffrey F. Dervin, MD;* Ian G. Stiell, MD;† George A. Wells, PhD;‡ Kelly Rody, BScN;§ Jenny Grabowski, MSc§

Objective: To determine clinicians' accuracy and reliability for the clinical diagnosis of unstable meniscus tears in patients with symptomatic osteoarthritis of the knee. **Design:** A prospective cohort study. **Setting:** A single tertiary care centre. **Patients:** One hundred and fifty-two patients with symptomatic osteoarthritis of the knee refractory to conservative medical treatment were selected for prospective evaluation of arthroscopic débridement. **Intervention:** Arthroscopic débridement of the knee, including meniscal tear and chondral flap resection, without abrasion arthroplasty. **Outcome measures:** A standardized assessment protocol was administered to each patient by 2 independent observers. Arthroscopic determination of unstable meniscal tears was recorded by 1 observer who reviewed a video recording and was blinded to preoperative data. Those variables that had the highest interobserver agreement and the strongest association with meniscal tear by univariate methods were entered into logistic regression to model the best prediction of resectable tears. **Results:** There were 92 meniscal tears (77 medial, 15 lateral). Interobserver agreement between clinical fellows and treating surgeons was poor to fair ($\kappa < 0.4$) for all clinical variables except radiographic measures, which were good. Fellows and surgeons predicted unstable meniscal tear preoperatively with equivalent accuracy of 60%. Logistic regression modelling revealed that a history of swelling and a ballottable effusion were negative predictors. A positive McMurray test was the only positive predictor of unstable meniscal tear. "Mechanical" symptoms were not reliable predictors in this prospective study. The model was 69% accurate for all patients and 76% for those with advanced medial compartment osteoarthritis defined by a joint space height of 2 mm or less. **Conclusions:** This study underscored the difficulty in using clinical variables to predict unstable medial meniscal tears in patients with pre-existing osteoarthritis of the knee. The lack of interobserver agreement must be overcome to ensure that the findings can be generalized to other physician observers.

Objectif : Déterminer chez les cliniciens la précision et la fiabilité du diagnostic clinique des déchirements instables du ménisque chez les patients présentant une arthrose symptomatique du genou. **Conception :** Étude prospective de cohorte. **Contexte :** Un seul centre de soins tertiaires. **Patients :** On a sélectionné 152 patients atteints d'une arthrose symptomatique du genou réfractaire au traitement médical usuel pour procéder à une évaluation prospective du débridement arthroscopiquee. **Intervention :** Débridement arthroscopique du genou, y compris résection du déchirement du ménisque et du lambeau chondral sans arthroplastie par abrasion. **Mesures de résultats :** Deux observateurs indépendants ont administré un protocole d'évaluation normalisé à chaque patient. La détermination arthroscopique des déchirements instables du ménisque a été consignée par un observateur qui a étudié l'enregistrement vidéo sans connaître les données préopératoires. Les variables qui présentaient la plus grande uniformité

*From the *Department of Surgery, the †Department of Emergency Medicine and the ‡Department of Epidemiology and Community Medicine, University of Ottawa and the Ottawa Hospital, Ottawa, Ont.*

§Research Coordinator, Ottawa Hospital, General Campus

Presented in part at the annual meeting of the Canadian Orthopaedic Association, Ottawa, Ont., June 22, 1998, and the biennial conference of the International Society of Arthroscopy, Knee Surgery and Orthopaedic Sports Medicine, Washington, June 2, 1999.

Accepted for publication Jan. 12, 2001.

Correspondence to: *Dr. Geoffrey F. Dervin, Ste. 5004, Ottawa Hospital, General Campus, 501 Smyth Rd., Ottawa ON K1H 8L6; fax 613 737-8837*

entre les observateurs et le lien le plus solide avec un déchirement du ménisque établis par la méthode univariée ont été entrées dans une régression logistique visant à modéliser la meilleure prédiction des déchirements résécables. **Résultats :** Il y avait 92 déchirements du ménisque (77 déchirements internes, 15 latéraux). L'uniformité interobservateurs entre les chercheurs cliniciens et les chirurgiens traitants a varié de faible à moyenne ($\kappa < 0,4$) pour toutes les variables cliniques à l'exception des évaluations radiographiques, qui étaient bonnes. Les chercheurs et les chirurgiens ont prédit un déchirement instable du ménisque avant l'intervention avec une précision équivalente de 60 %. La modélisation par régression logistique a révélé que des antécédents d'enflure et d'épanchements ballottables étaient des prédicteurs négatifs. Un résultat positif au test de McMurray a constitué le seul prédicteur positif de déchirements instables du ménisque. Les symptômes «mécaniques» n'ont pas constitué des prédicteurs fiables dans le contexte de cette étude prospective. Le modèle a présenté une précision de 69 % chez tous les patients et de 76 % chez ceux qui présentaient une arthrose avancée du compartiment interne, définie par un espace articulaire de 2 mm ou moins. **Conclusions :** Cette étude souligne la difficulté que présente l'utilisation de variables cliniques pour prédire des déchirements instables de la partie interne du ménisque chez des patients qui ont une arthrose préalable du genou. Il faut surmonter l'absence de convergence entre observateurs pour assurer qu'il est possible de généraliser les constatations à d'autres observateurs médecins.

Several arthroscopic interventions have been described for the management of symptomatic osteoarthritis of the knee refractory to conservative measures. The prevailing opinion is that resection of unstable meniscal tears has the greatest potential for improvement from arthroscopy,[1–5] although less satisfactory results are expected in knees with more advanced articular damage, severe tibiofemoral malalignment and chronic symptoms.[6] Inasmuch as some patients with unstable meniscal tears improve with arthroscopic resection, it would seem reasonable to identify this subgroup of patients with osteoarthritis of the knee. More accurate selection of these patients could help primary care physicians to refer appropriate patients to be considered for arthroscopy, minimize the need for expensive tests and help stratify the care of these patients to other appropriate procedures.

Several authors have reported on the accuracy of the history and physical examination in the diagnosis of unstable meniscal tears, ranging from 79% to 81%.[7,8] Others have looked at the utility of individual clinical signs such as joint line tenderness and the McMurray test.[9–11] Some excluded patients with severe gonarthrosis, whereas others included a mixed population of patients with or without established osteoarthritis. The clinical diagnosis of unstable meniscal tears may be more difficult in patients with radiographically evident osteoarthritis because of competing sources for the symptoms. This subject was not adequately addressed by our review of the literature. Furthermore, the reliability of the clinical signs and symptoms used in these studies has not been described and raises doubts about the ability to generalize these findings to other clinical settings.[12]

Our study was designed to evaluate prospectively a cohort of patients with symptomatic osteoarthritis of the knee and to develop a standardized clinical assessment protocol that would determine the interrator reliability of commonly used physical signs and symptoms in this condition. We also sought to measure clinicians' accuracy in predicting the presence of unstable tears. Finally, we determined whether diagnostic accuracy could be improved using accepted methodology for developing clinical prediction rules.[13,14]

Methods

Inclusion criteria

All patients aged 40 to 75 years referred to our orthopedic outpatient clinic between March 1995 and November 1997 with primary osteoarthritis of the knee (as defined by the American Rheumatism Association[15]) were considered for the present study. Patients with inflammatory or post-traumatic forms of osteoarthritis were excluded. Patients with prior meniscectomy were also excluded. The age criteria were chosen in accordance with most other published therapeutic studies of osteoarthritis. The clinic is located in a tertiary care teaching institution although the majority of patients referred to the orthopedic outpatient clinic come from general practitioners, and the patient pool is likely similar to that in a community hospital setting. Patients who remained symptomatic despite supervised physical therapy and comprehensive medical management were considered for arthroscopy. The final decision was made between patient and surgeon after a full discussion as to the perceived risk:benefit ratio. The study, which was reviewed and approved by the Research Ethics Board, was explained to all prospective patients, and their informed consent obtained.

Preoperative assessment

A postgraduate orthopedic fellow followed a standardized assessment protocol for each subject in the preadmission clinic of the hospital 7 to 10 days before the surgical procedure. The protocol was developed by consensus of the participating surgeons before the start of the study, to be comprehensive for all possible relevant preoperative predictors, based on a review of the literature. A study manual that described the pertinent physical tests was made available to each participating fellow and surgeon to maximize consistency in examina-

tion. In particular, the original McMurray test was defined: maximal knee flexion and application of external rotation and axial loading and extension of the knee.[16] A positive test was recorded if there was localized joint line pain or a palpable or audible and painful click related to rotation. The test was repeated using internal rotation at full flexion. Valgus or varus loading was not part of this version of the test. When feasible, the operating surgeon repeated the standardized assessment on the day of operation to determine the reliability of the chosen clinical parameters for the study. Radiographs were obtained 1 week preoperatively in both the 1-m standing anteroposterior (AP) and a 45° posteroanterior (PA) flexion weight-bearing projections.[17] The latter has been suggested as a more sensitive technique for joint space narrowing. A foot map was used to normalize rotation for both views and a foam wedge to control flexion for the 45° PA projection. All radiographs were assessed by 2 observers (G.F.D. and K.R.) who were blinded to the arthroscopic and clinical manifestations of the subjects. The features recorded included joint space in the lateral and medial tibiofemoral compartments, anatomical tibiofemoral axis[18] and the presence of tibial or femoral osteophytes. Lateral and tangential patellar views were not used as a means of predicting articular wear. MRI was not used.

Intervention

All patients underwent arthroscopy of the knee with either general or spinal anesthesia. Tourniquet use was according to surgeon preference. A thorough diagnostic arthroscopy was recorded on videotape before and after any intervention. A meniscal tear was considered unstable if it was of full or partial thickness, longitudinal and displaceable, radial or oblique estimated at 3 mm or more, or complex. The surgeon then performed the procedure, which included resection of loose chondral flaps, unstable meniscal tears and synovectomy only when required for visualization using standard, motorized instruments. Abrasion arthroplasty was not performed, although microfracture was performed for full-thickness chondral defects.

Articular cartilage wear was categorized by the method of the French Society for Arthroscopy (SFA),[19] which classifies each tibiofemoral and the patellofemoral compartment by a grade, summarizing the depth of lesion, surface area involvement and exact location. Surface area involvement was estimated as a percentage of compartment involvement and location of the lesion recorded on an articular diagram. The classification has been validated in a population of patients fulfilling the American College of Rheumatology clinical and radiographic criteria for osteoarthritis of the knee,[15,20] similar to the patients enrolled in the present study.

Outcome measures

All physicians were asked to predict to the nearest decile the percent probability of finding an unstable meniscal tear at arthroscopy, based on their clinical evaluation and radiographic review. This was to establish a baseline proficiency measure of accuracy for this diagnosis. The interobserver agreement for all clinical variables and prediction of tears was obtained for 6 clinical fellows versus 7 orthopedic staff, using overall percent agreement and the kappa coefficient (κ).[21] The κ is calculated as the percent agreement expected beyond that of chance and, hence, is a better index of the reliability of the variable being studied.[22] Landis and Koch[21] suggested the following guidelines for establishing the qualitative strength of κ: (0–0.2 = slight, 0.21–0.4 = fair, 0.41–0.6 = moderate, 0.61–0.8 = substantial and 0.81–1.0 = almost perfect). Reliability testing of predictor variables was restricted to interobserver, which is usually smaller than intraobserver reliability but more clinically relevant in the formulation of a prediction rule generalizable to other clinical settings.

The presence of an unstable meniscal tear was selected as the primary surgically relevant outcome in this study since we believed this would be most clinically relevant to surgeons considering arthroscopic débridement. Although much of the available data are retrospective, satisfactory functional outcome after meniscectomy is inversely related to articular cartilage wear.[2,3,23–25] All arthroscopic procedures were recorded on videotape as a permanent record and evaluated by the senior author for the presence of an unstable tear. Ambiguous cases were resolved by a consensus of 3 participants.

Statistical analysis

All analysis was done using SPSS for Windows version 6.1.3 (SPSS Inc., Chicago, 1995). Univariate correlations of clinical signs and symptoms for resectable meniscal tears was carried out using the χ^2 test without continuity correction for nominal data. Continuous and ordinal variables were split into clinically sensible cut points. Interobserver agreement of all clinical variables was measured with a κ.[21] Those factors found to be most reliable and strongly associated with an unstable meniscal tear were entered into a logistic regression to develop the best model for predicting an unstable meniscal tear.

Results

Two hundred and nine patients were referred for admission into the study; 42 did not meet the criteria for established osteoarthritis and were excluded. Of the remaining 167 patients who met eligibility criteria, 15 were ultimately excluded because of unsatisfactory video recording, which could not be reliably interpreted postoperatively. The remaining 152 patients comprised the study cohort. The mean (and

standard deviation) age of the patients was 60.5 (8.5) years and 51% were women. Pain was the most common presenting complaint, particularly with stair climbing and rising from a chair. Ligament stability was largely intact although effusions and tenderness were more prevalent. Fig. 1 indicates the distribution of chondral damage severity, by arthroscopic SFA grading for all cases. The medial compartment had considerably more damage, with 57% showing grade III or IV involvement. In contrast, both the lateral and the patellofemoral compartments were less severely involved, with only 13% and 17%, respectively, showing grade III or IV changes.

Unstable meniscal tears were found in 87 patients. Most were degenerative complex and oblique tears. None was considered to be repairable in the younger patients. Clinical fellows and staff independently showed identical predictive accuracies of only 60% for unstable tears, based on their preoperative clinical and radiographic assessment (Table 1). Accuracies for the staff ranged from 47% to 75%. Experience did not correlate with the predictive accuracy; the 2 surgeons with greatest experience had accuracies of 49% and 71%. Accuracies for the fellows ranged from 40% to 73%. All were of the same clinical experience. Both groups significantly overestimated the frequency of unstable meniscal tears as a reason for persisting knee discomfort, giving rise to low specificities. Although the predictive accuracy rates were identical, the groups demonstrated only fair interobserver agreement with a κ value of 0.24. Otherwise stated, they disagreed on the prediction for 31 of the 115 patients evaluated by both groups of physicians.

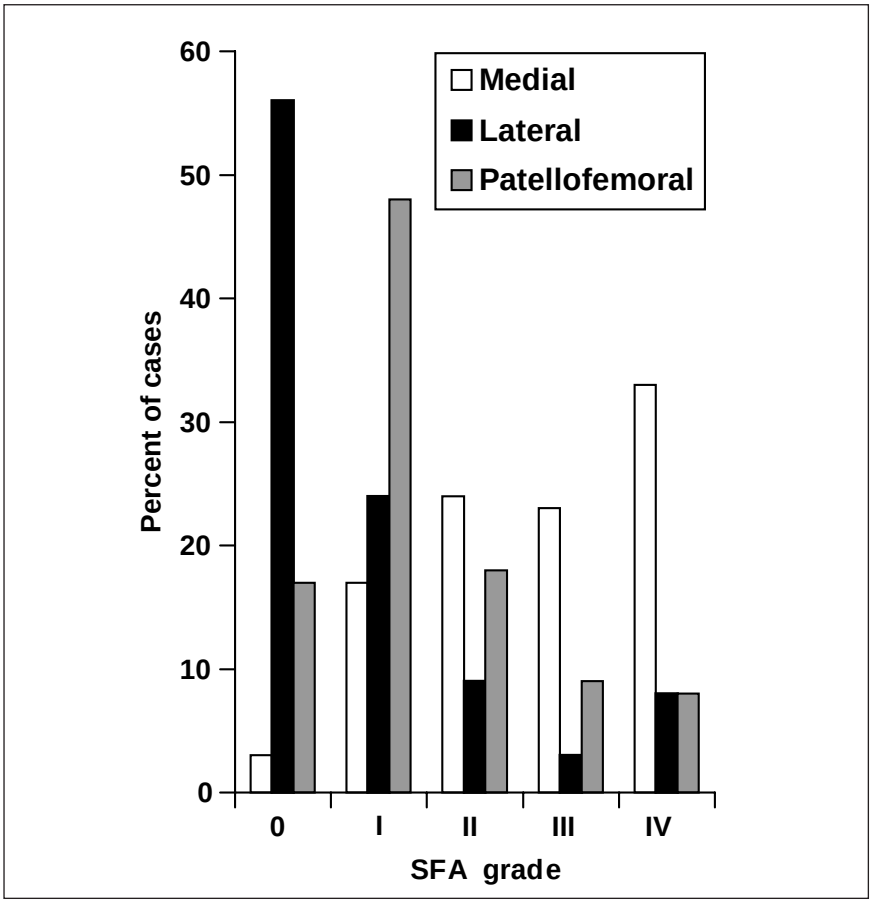

FIG. 1. Articular cartilage wear in the study patients, according to the French Society for Arthroscopy grading system.

The highest level of κ agreement for individual predictor variables by history and physical examination was only 0.44 for a history of locking, despite the production of a study manual that described each clinical variable in detail (Table 2). We found low κ values of interobserver agreement for commonly cited tests such as the McMurray test (0.16), medial joint line tenderness (0.21) and a visible effusion (0.28). The extent of disagreement was disappointing as these variables are among the most often used clinical signs for evaluating internal derangements of the knee. Agreement was much better for radiographic indices, which were split for easier clinical application.

The distribution of the 92 meniscal tears was significantly skewed in favour of medial tears (77 medial v. 15 lateral). Given that many of the predictor variables were localized to one area of the joint (i.e., flexion pain, extension pain, joint line tenderness and other special tests), we surmised that there would be greater clinical relevance and sensibility to rules specific for a medial meniscal tear, which is more consistent with clinical practice, in which a specific diagnosis is sought. The small number of lateral meniscal tears would not allow sufficient statistical power to permit a stable logistic model and therefore no attempt was made to develop a separate decision rule for these tears.[26]

Univariate measures of association for variables predicting unstable medial meniscal tears are listed in Table

Table 1

Percentage Accuracy by Fellows and Staff in Predicting Unstable Meniscal Tear, Using 50% As the Probability Threshold

Prediction	Fellows	Staff
Sensitivity	87	88
Specificity	21	20
Accuracy	60	60
Positive predictive value	61	62
Negative predictive value	54	53

3. Few features on the history appeared to be helpful, and most patients could not report an injury preceding the symptoms. Medial joint line tenderness was found in 90% of all patients and did not discriminate those patients with unstable tears. Posteromedial pain with forced full flexion and extension and a positive McMurray test showed a trend to association with an unstable medial meniscal tear.

Those variables with best univariate association ($p < 0.2$) and κ agreement (≥0.15) were made available to a forward stepwise logistic regression analysis. The analysis yielded 3 statistically significant variables (Table 4). An odds ratio less than 1 implies a relative decreased probability of association with an unstable tear whereas values greater than 1 signify an increased probability. Thus, a history of swelling and the presence of a ballottable effusion were negative predictors of an unstable medial meniscal tear, whereas a positive McMurray sign was a positive predictor. The odds ratios were even stronger for the subgroup of patients with more severe medial compartment arthrosis defined as medial joint space of 2 mm or less on the 45° PA radiograph. Using this cutoff point, 83% of patients with a medial joint space of 2 mm or less had arthroscopically proven chondropathy of SFA grades III and IV. The accuracy of meniscal tear prediction using these 3 variables was 69% for the entire cohort and 76% for the subgroup, slightly better than physicians' preoperative predictive accuracy of 60% using their clinical judgement based on a review of all the variables.

Discussion

The findings of this study suggest that the clinical diagnosis of unstable meniscal tears in patients with established osteoarthritis is less accurate than has been reported for other groups of patients. To our knowl-

Table 2

Total Interobserver and Kappa (κ) Agreements Between Participating Staff and Fellows for Predictor Variables in 115 Patients Having Symptomatic Osteoarthritis of the Knee

	Agreement			Agreement	
Variable	Overall, %	κ (and 95% confidence interval)	Variable	Overall, %	κ (and 95% confidence interval)
History			Special tests		
Acute injury	61	0.21 (0.03–0.39)	McMurray	59	0.16 (–0.01–0.33)
Swelling	69	0.33 (0.17–0.49)	Steinman	52	0.05 (–0.11–0.21)
Giving way	60	0.12 (–0.04–0.28)	Circumduction	63	0.21 (0.05–0.37)
Locking	80	0.44 (0.26–0.62)	Instability		
			Varus	93	0 (–0.18–0.18)
Pain			Valgus	92	0.05 (–0.13–0.23)
Generalized	60	–0.03 (–0.15–0.21)			
Focal	74	0.11 (–0.08–0.30)	Lachman	96	–0.08 (–0.12–0.04)
At rest	56	0.16 (0.0–0.32)	Abnormal extension	50	0.07 (–0.05–0.19)
Rising from chair	78	0.25 (0.05–0.45)	Effusion		
			Visible	72	0.28 (0.10–0.46)
Climbing stairs	90	0.21 (–0.06–0.48)	Ballotment	62	0.19 (0.02–0.36)
Physical findings			Sweep	58	0.14 (–0.04–0.32)
Tenderness					
Medial joint line	79	0.21 (0.01–0.41)			
Lateral joint line	70	0.25 (0.07–0.43)	**Radiographic signs**		
			1-m standing		
			Tibiofemoral varus (≤4°)	75	0.29 (0.13–0.45)
Pain with passive motion			Medial joint space ≤2 mm	80	0.63 (0.49–0.77)
Full flexion:	67	0.09 (–0.07–0.25)			
Anteromedial	59	0.14 (–0.05–0.33)	Lateral joint space ≤3 mm	86	0.72 (0.49–0.95)
Posteromedial	52	0.10 (–0.09–0.29)	45° Posteroanterior weight bearing		
			Tibiofemoral varus (≤4°)	88	0.63 (0.51–0.75)
Anterolateral	65	0.18 (0.0–0.37)	Medial joint space ≤2 mm	83	0.58 (0.44–0.72)
Posterolateral	72	0.05 (–0.17–0.27)	Lateral joint space ≤3 mm	91	0.80 (0.62–0.98)
Full extension:	57	0.10 (–0.08–0.28)	Clinician prediction*	73	0.24 (0.04–0.44)
Anteromedial	57	0.06 (–0.14–0.26)			
Posteromedial	48	–0.02 (–0.22–0.18)			
Anterolateral	71	0.02 (0.0–0.40)			
Posterolateral	76	0.01 (–0.17–0.19)			

*Clinician prediction of unstable meniscal tear after preoperative evaluation

edge, this is the first attempt to predict meniscal tears in this group of carefully defined patients. Experienced physicians in this study could accurately classify only 60% of patients preoperatively. Daniel and associates[7] showed that experienced clinicians were 79% accurate in diagnosing medial meniscal tears when relying on the clinical examination in patients without radiographically proven arthritis. They did not specify which features of the examination were most important; rather they allowed clinicians to express a probability level of a meniscal problem. Terry and colleagues[8] showed similar accuracy of clinical diagnosis (81%) for all cases of internal derangement, excluding patients with significant ligamentous instability or severe gonarthrosis. Abdon and associates[10] performed a prospective study of patients undergoing arthroscopy for suspected meniscal disorders and used discriminant analysis of a variety of clinical parameters. They found that joint line tenderness and a history of locking were positive predictors, whereas pain at rest, sick leave at the time of surgery and tenderness of the medial patellar facet were negative predictors. Fowler and Lubliner[9] prospectively studied 161 patients with knee pain and a suspected meniscal disorder and found a combi-

Table 4

Odds Ratios and 95% Confidence Intervals for Predictor Variables Remaining in Logistic Regression Model Predicting an Unstable Medial Meniscal Tear

Variable	All patients	Joint space ≤2 mm
History of swelling	0.42 (0.19–0.96)	0.19 (0.04–0.84)
Ballotable effusion	0.38 (0.16–0.93)	0.12 (0.02–0.60)
Positive McMurray test	2.21 (0.96–5.1)	13.9 (2.3–85.3)

Table 3

Univariate Association of Staff Predictor Variables Versus Unstable Medial Meniscal Tear at Arthroscopy

	Unstable tear, %				Unstable tear, %		
Clinical finding	Yes, *n* = 66	No, *n* = 64	*p* value	Clinical finding	Yes, *n* = 66	No, *n* = 64	*p* value
Demographics				Full extension:	69	66	0.71
Mean (and SD) age, yr	59 (9)	58 (9)	0.51*				
Male	57	43	0.08	Anteromedial	58	68	0.26
Obesity (BMI >27)	54	50	0.73	Posteromedial	43	30	0.18
History				Anterolateral	21	34	0.14
Acute injury	34	38	0.61				
Swelling	48	69	0.01	Posterolateral	4	11	0.18
Giving way	42	42	0.94	Special tests			
				McMurray	66	53	0.18
Locking	22	22	0.96	Steinman	50	43	0.43
Pain				Circumduction	74	61	0.59
Generalized	64	71	0.41				
Focal	81	92	0.06	Instability			
				Varus	17	18	0.85
At rest	39	55	0.07	Valgus	14	21	0.27
Rising from chair	81	89	0.21	Lachman	6	6	1.00
Climbing stairs	89	92	0.54	Abnormal extension	38	54	0.08
Physical findings							
Tenderness				Effusion			
Medial joint line	89	91	0.75	Visible	28	33	0.48
Lateral joint line	24	31	0.41	Ballotment	19	41	0.007
Pain with passive motion				Sweep	37	52	0.08
Full flexion:	89	85	0.46				
Anteromedial	58	68	0.26	**Radiographic signs**			
				1-m standing			
				Tibiofemoral varus (≤4°)	68	51	0.04
Posteromedial	58	40	0.06	Medial joint space ≤2 mm	41	35	0.63
Anterolateral	24	24	1.00	Lateral joint space ≤3 mm	2	9	0.72
Posterolateral	15	11	0.49	45° Posteroanterior weight bearing			
				Tibiofemoral varus (≤4°)	69	52	0.03
				Medial joint space ≤2 mm	47	40	0.41
				Lateral joint space ≤3 mm	4	16	0.01

*Student's *t*-test

nation of signs most valuable. Joint line pain was sensitive (85%) but not specific (29%). Pain on forced flexion had a 50% sensitivity and 68% specificity, and the McMurray test and block to extension had low sensitivity yet high specificity. The variability of pertinent physical signs in these latter 2 studies is likely attributable to the study population heterogeneity, given that the age ranges were 16 to 66 years and 13 to 67 years respectively, with no statement as to the presence of underlying arthritic change. Several traditional key clinical signs were very common in our cohort of patients who failed medical management and hence are not good discriminators. Pain on climbing stairs (91%), arising from a chair (85%), medial joint line tenderness (90%) and pain on forced flexion (87%) are all pertinent examples.

Another challenge in establishing clinical criteria for the diagnosis of meniscal tears is the poor reliability of classic special tests for meniscal injury in this cohort of patients as measured by κ values. The κ is a more valid measure of interrator concordance than the percent agreement because it is not as sensitive to the underlying prevalence of the variable. The latter measure would be deceptively high by chance alone for variables with high prevalence, as seen with medial joint line tenderness and pain with stair climbing, for example. We found low κ values of interobserver agreement for the McMurray test (0.16), circumduction or rotation at full flexion (0.21) and medial joint line tenderness (0.21). Landis and Koch[21] stated those κ values less than 0.4 are evidence of slight to fair interobserver agreement only. We sought to optimize this agreement by providing a study manual with explicit description of all the variables, although a formal training session was not provided, as we had hoped that the findings could be ultimately generalized to all practising orthopedic surgeons. The examinations were spaced 7 to 10 days apart partly for convenience and partly so that the first examination would not influence an immediate subsequent one. For instance, provocative tests for meniscal irritation could have left a persistent discomfort biasing the next examiner. Bias could have existed in the present study design, however, if the baseline condition had changed between both examinations because of increased activity or aggravation of symptoms preoperatively. Although we thought this bias would be impossible to measure, subjects were asked to moderate their activities in the 2 weeks before the procedure to minimize the change in condition.

A few authors have reported on the reliability of physical examination of the knee with variable results. Evans and associates[11] evaluated the κ agreement for the McMurray test, differentiating between a medial thud ($\kappa = 0.35$), sensation of reproducing symptoms ($\kappa = -0.10$) and pain ($\kappa = 0.30$). The patient profile was not thoroughly described in their study but presumably included many patients who did not have established osteoarthritis of the knee. The authors concluded that the medial "thud" was more valid for the less experienced examiner whereas "sensation" and "pain" were more prognostic for the more experienced examiner. Stiell and colleagues[27] found moderate interobserver agreement for medial and lateral joint line tenderness ($\kappa = 0.5$ and 0.45 respectively) and visible effusion ($\kappa = 0.59$) in a cohort of acutely injured patients (mean age 36 years) assessed in an emergency department, although the underlying prevalence of osteoarthritis was unknown. To our knowledge, the only previously documented study of clinical sign reliability in osteoarthritis of the knee is that of Cushnaghan and associates.[28] They studied physical signs in 8 patients with osteoarthritis and found substantial intraobserver agreement for several clinical signs and lower levels of interobserver agreement for joint tenderness ($\kappa = 0.4$) and effusion ($\kappa = 0.28$). Our findings are more consistent with those of Cushnaghan and associates, implying some influence of the study population.

The clinical significance of accurate diagnosis of unstable meniscal tears in osteoarthritic knees merits further scrutiny. Many have described the results of partial meniscectomy in older patients and have differentiated outcome based on the presence of degenerative compartmental changes. Arthroscopic partial meniscectomy is a well-tolerated, effective procedure in patients aged 40 years or older without significant degenerative change.[1–5] The role for resection of degenerative tears with coexisting articular wear is more contentious. Jones and colleagues[25] reviewed partial meniscectomies in patients over 40 years of age and found considerably worse outcome in patients with degenerative tears (absence of trauma and fissured, horizontal cleavage tears) than those with traumatic tears (history of trauma, bucket handle or parrot-beak tears). Although the study was small and retrospective, the authors recommended resection of only traumatic tears causing mechanical symptoms. Lotke and associates[24] reviewed their long-term results (mean 10.8 yr) of open medial meniscectomy grouped according to preoperative radiographs. The outcome was satisfactory in 90% of patients with normal radiographs but in only 21% for those with moderate or marked degenerative change. Jackson and Rouse[23] reported satisfactory short-term results at a mean 2.5 years in 80% of patients who underwent arthroscopic partial meniscectomy in the presence of degenerative chondral change versus 95% in those without degenerative change. These studies suggest that there are patients with established osteoarthritis of knee and unstable meniscal tears who derive benefit from arthroscopic resection. Precise, early identification of these patients would at least allow for a more informed treatment plan. Patients with less se-

vere forms of osteoarthritis may be the best candidates and would merit early consideration for arthroscopy if conservative measures have failed. The participating surgeons in the present study agreed with resection of all unstable meniscal tears in this cohort (which covers the spectrum of severity), although we await the clinical consequences of this when we review our functional outcome results at minimum 2-year follow-up, which will be the subject of a subsequent report.

Conclusions

The methodologic criteria for derivation of a prediction rule for unstable medial meniscal tears in osteoarthritic knees were defined and adhered to in the present study. Unfortunately, a standardized clinical assessment of patients with osteoarthritis of the knee did not yield sufficiently reliable or discriminating variables to ensure a reproducible prediction rule for the clinical diagnosis of unstable meniscal tears. The study did highlight those clinical variables that appear to be most pertinent for this group of patients. Mechanical symptoms, such as locking or giving way, and joint line tenderness did not discriminate for the presence of tears as has been traditionally espoused. Swelling and effusion, in particular, should be noted as negative predictors for unstable tears in this population with articular wear. Caution should be exercised not to apply this rule to nonarthritic patients where swelling and effusion may indeed positively predict meniscal tears. The McMurray test was the most useful positive test for detecting unstable meniscal tears but suffers from only fair interrator reliability. Clinicians should accept these results as a challenge to further refine their diagnostic skills. Noninvasive MRI or ultrasonography may be particularly suited for use in this subgroup of patients in contrast to cases of isolated meniscal derangement in nonarthritic knees.

Acknowledgements: We thank the following orthopedic surgeons for their participation in the study and the contribution of their patients: Drs. J. Brunet, J. Bouchard, J.P. Desjardins, R. Feibel, A. Giachino, J. McAuley and G. Moreau. Thanks are also extended to the postgraduate orthopedic fellows who contributed by way of clinical assessment of all subjects: Drs. B. Blastorah, Y. Boudreau, S. Daczak, M. Fiala, R. Greene, J. Ostrowski and S. Tubin. Kim Morin ably helped with preparation of the manuscript.

The Physicians Services Institute (grant #95-33) and the Canadian Orthopaedic Foundation (Hip Hip Hooray) provided grant support for this project.

References

1. Covall DJ, Wasilewski SA. Roentgenographic changes after arthroscopic meniscectomy; five-year follow-up in patients more than 45 years old. *Arthroscopy* 1992; 8:242-6.
2. Matsusue Y, Thomson NL. Arthroscopic partial medial meniscectomy in patients over 40 years old: a 5- to 11-year follow-up study. *Arthroscopy* 1996;12:39-44.
3. McBride GG, Constine RM, Hofmann AA, Carson RW. Arthroscopic partial medial meniscectomy in the older patient. *J Bone Joint Surg [Am]* 1984;66:547-51.
4. Wouters E, Bassett FH 3rd, Hardaker WT Jr, Garrett WE Jr. An algorithm for arthroscopy in the over-50 age group. *Am J Sports Med* 1992;20:141-5.
5. Yang SS, Nisonson B. Arthroscopic surgery of the knee in the geriatric patient. *Clin Orthop* 1995;316:50-8.
6. Novak PJ, Bach BR Jr. Selection criteria for knee arthroscopy in the osteoarthritic patient [review]. *Orthop Rev* 1993;22: 798-804.
7. Daniel D, Daniels E, Aronson D. The diagnosis of meniscus pathology. *Clin Orthop* 1982;163:218-24.
8. Terry GC, Tagert BE, Young MJ. Reliability of the clinical assessment in predicting the cause of internal derangements of the knee. *Arthroscopy* 1995;11:568-76.
9. Fowler PJ, Lubliner JA. The predictive value of five clinical signs in the evaluation of meniscal pathology. *Arthroscopy* 1989; 5:184-6.
10. Abdon P, Lindstrand A, Thorngren K-G. Statistical evaluation of the diagnostic criteria for meniscal tears. *Int Orthop* 1990; 14:341-5.
11. Evans PJ, Bell GD, Frank C. Prospective evaluation of the McMurray test. *Am J Sports Med* 1993;21(4):604-8.
12. Wright JG, Feinstein AR. Improving the reliability of orthopaedic measurements. *J Bone Joint Surg [Br]* 1992;74:287-91.
13. Wasson JH, Sox HC Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985;313:793-9.
14. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277:488-94.
15. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* 1986;29:1039-49.
16. McMurray TP. The semilunar cartilages. *Br J Surg* 1942;29:407-14.
17. Rosenberg TD, Paulos LE, Parker RD, Coward DB, Scott SM. The forty-five-degree posteroanterior flexion weight-bearing radiograph of the knee. *J Bone Joint Surg [Am]* 1988;70:1479-83.
18. Moreland JR, Bassett LW Hanker GJ. Radiographic analysis of the axial alignment of the lower extremity. *J Bone Joint Surg [Am]* 1987;69:745-9.
19. Dougados M, Ayral X, Listrat V, Gueguen A, Bahuaud J, Beaufils P, et al. The SFA system for assessing articular cartilage lesions at arthroscopy of the knee. *Arthroscopy* 1994;10:69-77.
20. Ayral X, Dougados M, Listrat TV, Bonvarlet JP, Simonnet J, Amor B. Arthroscopic evaluation of chondropathy in osteoarthritis of the knee. *J Rheumatol* 1996; 23:698-706.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
22. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology. A basic science for clinical medicine*. 2nd ed. Boston: Little, Brown; 1991. p. 25-31.
23. Jackson RW, Rouse DW. The results of partial arthroscopic meniscectomy in patients over 40 years of age. *J Bone Joint Surg [Br]* 1982;64:481-5.
24. Lotke PA, Lefkoe RT, Ecker ML. Late results following medial meniscectomy in an older population. *J Bone Joint Surg [Am]* 1981;63:115-9.
25. Jones RE, Smith EC, Reisch JS. Effects of medial meniscectomy in patients older than forty years. *J Bone Joint Surg [Am]* 1978;60:783-6.
26. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993; 18:201-10.
27. Stiell IG, Greenberg GH, Wells GA, McKnight RD, Cwinn AA, Cacciotti T, et al. Derivation of a decision rule for the use of radiography in acute knee injuries. *Ann Emerg Med* 1995;26:405-13.
28. Cushnaghan J, Cooper C, Dieppe P, Kirwan J, McAlindon T, McCrae F. Clinical assessment of osteoarthritis of the knee. *Ann Rheum Dis* 1990;49:768-70.