An official website of the United States government
Here's how you know

FULL TEXT LINKS

Arch Bone Jt Surg. 2023;11(1):23-28. doi: 10.22038/ABJS.2022.52845.2619.

Accuracy and Reliability of MRI Reports in Diagnosing the Symptomatic Knee in Patients Who Had Bilateral MRI

Tom Joris Crijns [1], Emily Z Boersma [1], Stein Jasper Janssen [2], Melissa D Tonn [3], David Ring [1]

Affiliations

PMID: 36793663 PMCID: PMC9903307 DOI: 10.22038/ABJS.2022.52845.2619

Free PMC article

Abstract

Background: Newly symptomatic chronic musculoskeletal illness is often misinterpreted as new pathology, particularly when symptoms are first noticed after an event. In this study, we were interested in the accuracy and reliability of identifying the symptomatic knee based on bilateral MRI reports.

Methods: We selected a consecutive sample of 30 occupational injury claimants, presenting with unilateral knee symptoms who had bilateral MRI on the same date. A group of blinded musculoskeletal radiologists dictated diagnostic reports, and all members of the Science of Variation Group (SOVG) were asked to indicate the symptomatic side based on the blinded reports. We compared diagnostic accuracy in a multilevel mixed-effects logistic regression model, and calculated interobserver agreement using Fleiss' kappa.

Results: Seventy-six surgeons completed the survey. The sensitivity of diagnosing the symptomatic side was 63%, the specificity was 58%, the positive predictive value was 70%, and the negative predictive value was 51%. There was slight agreement among observers (kappa= 0.17). Case descriptions did not improve diagnostic accuracy (Odds Ratio: 1.04; 95% CI: 0.87 to 1.3; *P=0.65*).

Conclusion: Identifying the more symptomatic knee in adults based on MRI is unreliable and has limited accuracy, with or without information about demographics and mechanism of injury. When there is a dispute concerning the extent of the injury to a knee in a litigious, medico-legal setting such as Workers' Compensation, consideration should be given to obtaining a comparison MRI of the uninjured, asymptomatic extremity.

Keywords: Accuracy; Knee injury; Magnetic resonance imaging; Reliability; Worker's compensation insurance.

LinkOut - more resources

Full Text Sources
Europe PubMed Central
PubMed Central

THE ARCHIVES OF BONE AND JOINT SURGERY ABJS

Arch Bone Jt Surg. 2023; 11(1): 23–28.
doi: 10.22038/ABJS.2022.52845.2619

PMCID: PMC9903307
PMID: 36793663

Accuracy and Reliability of MRI Reports in Diagnosing the Symptomatic Knee in Patients Who Had Bilateral MRI

Tom Joris Crijns, MD,[1] Emily Z. Boersma, MD,[1] Stein Jasper Janssen, MD, PhD,[2] Melissa D. Tonn, MD,[3] and David Ring, MD, PhD[1]

Abstract

Background:

Newly symptomatic chronic musculoskeletal illness is often misinterpreted as new pathology, particularly when symptoms are first noticed after an event. In this study, we were interested in the accuracy and reliability of identifying the symptomatic knee based on bilateral MRI reports.

Methods:

We selected a consecutive sample of 30 occupational injury claimants, presenting with unilateral knee symptoms who had bilateral MRI on the same date. A group of blinded musculoskeletal radiologists dictated diagnostic reports, and all members of the Science of Variation Group (SOVG) were asked to indicate the symptomatic side based on the blinded reports. We compared diagnostic accuracy in a multilevel mixed-effects logistic regression model, and calculated interobserver agreement using Fleiss' kappa.

Results:

Seventy-six surgeons completed the survey. The sensitivity of diagnosing the symptomatic side was 63%, the specificity was 58%, the positive predictive value was 70%, and the negative predictive value was 51%. There was slight agreement among observers (kappa= 0.17). Case descriptions did not improve diagnostic accuracy (Odds Ratio: 1.04; 95% CI: 0.87 to 1.3; *P=0.65*).

Conclusion:

Identifying the more symptomatic knee in adults based on MRI is unreliable and has limited accuracy, with or without information about demographics and mechanism of injury. When there is a dispute concerning the extent of the injury to a knee in a litigious, medico-legal setting such as Workers' Compensation, consideration should be given to obtaining a comparison MRI of the uninjured, asymptomatic extremity.

Key Words: Accuracy, Knee injury, Magnetic resonance imaging, Reliability, Worker's compensation insurance

Introduction

Slowly progressive musculoskeletal conditions that are newly symptomatic are often misinterpreted as new pathology, especially when symptoms are first noticed after a (perceived) noxious event. [1] Consequent magnetic resonance imaging (MRI) of the knee often reveals signal abnormalities, although changes (meniscal defects in particular) are also common among asymptomatic people with increasing age. [2–5] In fact, a prior study showed that only 43% of patients with new unilateral knee symptoms associated with a specific event at work have worse pathologic findings on the symptomatic side. [6] MR imaging of the symptomatic side alone may contribute to the misperception that age-related joint changes are the result of acute injury, which may affect decision-making and illness behavior. [7, 8]

In this study, we tested the ability of orthopedic surgeons to determine the symptomatic knee among a consecutive sample of occupational injury claimants when viewing radiology reports of MRIs of both knees. We tested the primary null hypothesis that there is no difference in the accuracy of identifying the symptomatic knee based on bilateral MRI reports between observers who receive information about the patient and symptoms and observers who do not. We tested the secondary null hypotheses that 1) there is no difference in the reliability of identifying the symptomatic knee between observers who receive information about the patient and symptoms and observers who do not; there is no difference in 2) accuracy and 3) reliability of diagnosing the symptomatic knee based on patient age category; and 4) surgeons cannot identify the symptomatic knee more often than expected by random chance.

Materials and Methods

Study design and setting

The protocol for this study was approved by our Institutional Review Board (IRB). We selected a consecutive sample of 30 occupational injury claimants (a subset) from a prior study,6 presenting with unilateral knee symptoms who had bilateral MR imaging on the same date. Patients aged 40 years or older were included in this cohort if symptoms were acute in onset and ascribed to a single event at work. All patients who had prior knee surgery or who had radiographic evidence of fracture were excluded. All MR images were multiplanar T1 and T2 weighted sequences without contrast. Diagnostic reports were dictated by a group of expert musculoskeletal radiologists who were blinded to the patients' clinical history and who were unaware which side was symptomatic.

All protected health information (PHI) was removed. All reports, without information on the symptomatic side, were distributed using an online survey design and distribution tool, SurveyMonkey (Palo Alto, CA, USA). Participants were asked to indicate which side they thought to be symptomatic based on the pathologic findings.

Participants

All members of the Science of Variation Group (SOVG) were invited to participate in our online survey. The SOVG is comprised of several hundred orthopedic, plastic, and trauma surgeons, who contribute to studying variation in care by completing monthly questionnaires. Observers were invited through email and had no financial incentive to participate in our study. All participants were randomized (1:1) to evaluate the bilateral MRI report with or without patient demographics (gender, age, BMI) and mechanism of injury.

Seventy-six surgeons completed the survey, of which sixty-nine (91%) men [Table 1]. Thirty-four surgeons (45%) received a case description in addition to the MRI reports. The majority of surgeons (51%) practice in Europe, and 17 (26%) in North America. Most surgeons subspecialize in orthopedic trauma (80%). The groups were similar except for the fact that the proportion of male observers was higher in the group that did not receive a case description ($P = 0.04$).

Table 1

Surgeon characteristics

Surgeon variables	With case description	Without case description	*P value*
N	34 (45%)	42 (55%)	0.42
Male	28 (82%)	41 (98%)	**0.04**
Continent of practice			0.59
United States	8 (24%)	9 (21%)	
Europe	19 (56%)	20 (48%)	
Other	7 (21%)	13 (31%)	
Years in practice			0.40
0-5	6 (19%)	12 (31%)	
6-10	5 (16%)	7 (18%)	
11-20	10 (31%)	13 (33%)	
21-30	11 (34%)	7 (18%)	
Supervising trainees	30 (94%)	35 (90%)	0.68
Subspecialty			0.071
Orthopedic trauma	29 (91%)	28 (72%)	
General orthopedics	3 (9.4%)	11 (28%)	

Variables as number (percentage).

Statistical analysis

To identify factors associated with diagnostic accuracy of the injured extremity, we constructed a multilevel mixed-effects logistic regression model with a random intercept. Since observers were randomly allocated into two groups, we did not account for surgeon characteristics. Additionally, interobserver agreement was calculated with Fleiss' kappa, using bootstrapping (resamples = 1000) to calculate the standard error and confidence intervals. A kappa value of zero equates to the degree of agreement expected from random chance, while a kappa value of 1.00 represents perfect agreement. We used the Landis and Koch 9 classification system to interpret kappa values: a value of 0.01 to 0.20 indicates slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 0.99, near-perfect agreement. Patients were categorized into two age groups using median split (age = 53), to create an equal number of patients in each group. We compared kappa values with a two-sample z-test between 1) observers who received a case description and those who did not, and 2) between patients in

the older and younger age group. All two-tailed P values below 0.05 were considered statistically significant. We used a binomial test to address whether surgeons more often indicate the correct side than would be expected by random chance.

Results

On average, the surgeons indicated the correct side in 61% of cases; the correct percentage ranged from 12 to 92 percent by case [Table 2]. In a binomial test, surgeons indicated the injured knee slightly more frequently than expected by random chance ($P<0.001$).

Table 2

Patient characteristics

Patient	Age	Sex	Height (Inches)	BMI (kg/m^2)	Symptomatic side	Percentage correct (%)
1	58	Female	66	30.7	Right	21
2	65	Male	67	28.2	Left	92
3	50	Female	63	33.7	Left	62
4	42	Male	75	48.1	Right	90
5	54	Female	70	32.3	Right	88
6	43	Female	66	46.6	Right	44
7	48	Male	63	23.9	Right	71
8	63	Female	62	29.3	Left	49
9	77	Female	59	25.4	Right	71
10	64	Male	77	26.1	Left	57
11	53	Male	69	32.5	Left	63
12	52	Male	72	28.7	Right	56
13	58	Male	67	54.8	Right	52
14	55	Female	62	33.3	Right	57
15	52	Female	62	26.3	Right	58
16	57	Male	72	25.1	Left	66
17	54	Male	62	24.5	Right	92
18	56	Male	66	32.3	Right	76
19	53	Male	74	37.2	Right	52
20	49	Male	67	35.2	Right	62
21	51	Female	66	35.5	Right	52
22	52	Female	65	43.3	Right	43
23	65	Female	57	40.2	Left	66
24	57	Female	66	28.2	Right	81
25	60	Male	70	33.6	Left	36
26	52	Male	67	32.1	Left	81
27	51	Female	64	29.2	Left	72
28	54	Female	62	22.5	Left	12

In multilevel mixed-effects logistic regression, there is no difference in diagnostic accuracy between observers who received information about patient gender, age, and mechanism of injury and those that did not receive this information (Odds Ratio: 1.04; 95% CI: 0.87 to 1.3; *P =0.65*). Among all surgeons, the kappa value was 0.17, which is considered slight agreement, and there was no difference in interobserver agreement between surgeons who received case descriptions and those who did not [Table 3]. The sensitivity of diagnosing the symptomatic side was 63% (CI: 60% to 66%), the specificity was 58% (CI: 54% to 62%), the positive predictive value was 70% (CI: 67% to 72%), and the negative predictive value was 51% (CI: 47% to 54%). Patient age did not affect diagnostic accuracy in multilevel logistic regression; and it did not affect interobserver agreement [Table 3].

Table 3

Interobserver agreement of the symptomatic extremity

	Observed Agreement	Kappa (95% Confidence interval)	*P value*
All	0.59	0.17 (0.098 to 0.24)	
Case description			0.88
Yes	0.60	0.19 (0.11 to 0.27)	
No	0.59	0.18 (0.096 to 0.26)	
Patient age			0.17
53 or younger	0.55	0.10 (0.021 to 0.19)	
54 or older	0.62	0.21 (0.085 to 0.34)	

Bold indicates statistical significance, *P < 0.05*.

Discussion

Knee pathology accumulates with age, yet pathologic findings on MRI do not correspond well with symptoms and limitations and are not indicative of acute injury. [10–13] Prior work has shown that the majority of patients with unilateral knee symptoms after a single event at work do not have worse pathology on the symptomatic side. [6] This study tested whether a large group of surgeons could identify the symptomatic knee based on bilateral MR imaging reports, and whether information about patient age, gender, and mechanism of injury increases agreement on the symptomatic side. We found that surgeons were slightly better at indicating the symptomatic side than random chance. There was no difference in accuracy and reliability between observers who had information about the patients.

The reader should keep the following limitations in mind when considering our work. First, since observers were randomized into two groups to complete the survey either with or without additional case information, it was technically not feasible to randomize the case sequence in conjunction. This may have caused questionnaire fatigue, although the questionnaire was relatively short. It may be more likely that observers quit the survey prematurely, since sixteen observers quit the survey after ten questions (21%), and 25% of initial participants did not complete the last question. Second, there was a greater proportion of male observers in the group that did not receive a case description. To mitigate this difference, we accounted for sex in the multilevel model, which yielded similar results. Third, information about the physical exam might have aided diagnosis. The scope of our current study was to link MRI findings to symptomatology, but findings in the physical exam are usually taken into consideration in a clinical setting. Fourth, observers did not have access to the MRI images, which might have affected the accuracy. Nevertheless, all MRI reports were dictated by experienced musculoskeletal radiologists whose reports typically guide diagnosis and treatment. Finally, surgeons who are participants of the Science of Variation Group may be more academically inclined than the average surgeon, decreasing generalizability.

We found limited diagnostic accuracy and very low reliability of identifying the symptomatic knee, regardless of whether observers received information about the patient and mechanism of injury or not. This is consistent with the larger prior study on occupational injury claimants that found similar pathologic changes on the asymptomatic side. [6] Unilateral MR imaging of the symptomatic extremity may reinforce the misconception of injury among patients with meniscal changes due to age. Use of the word "tear" to describe the pathology reinforces this misconception. [14] Operative treatment of age-related meniscal pathology is no better than sham operative treatment or non-operative treatment. [15–18] This set of circumstances risks misdiagnosis and overtreatment of expected changes in the human knee with age. [19]

Although observers were statistically more likely to indicate the correct side (61%) than expected by random chance (50%), diagnostic accuracy may be considered low. Diagnostic accuracy varies substantially by case and was as low as 12%. This goes to show that structural changes to the knee are similar in the symptomatic and asymptomatic side. As a matter of fact, one of the survey participants contacted the authors indicating that several cases were too similar to choose which side is symptomatic. Our findings are consistent with prior studies that found a substantial proportion of meniscal changes among asymptomatic people. [10–13] An acute incident or event at work, such as twisting the knee or falling, may make people more aware of the structural joint changes and attenuation that accumulate over time.

In this study, we invited a large number of orthopedic surgeons to identify the symptomatic knee based on blinded MRI reports and found that both diagnostic accuracy and reliability were low, independent of patient and surgeon characteristics. This adds to the growing body of evidence that indicates that pathologic changes on radiographic imaging do not correlate well with symptomatology. Although surgeons are slightly better than random chance, for some cases pathologic changes were substantially worse on the asymptomatic side. The results of this study may support rethinking the role that MRI currently has in the diagnosis of knee pain. Future work may help identify patient populations for which MR imaging is contributory towards diagnosis and treatment. The most impactful effect of such findings is to make a more accurate determination of the

extent of an alleged injury in a compensation setting. With a dispute concerning the extent of the injury to a knee in a litigious, medico-legal setting such as Workers' Compensation, consideration should be given to obtaining a comparison MRI of the uninjured, asymptomatic extremity.

References

1. van Hoorn BT, Wilkens SC, Ring D. Gradual Onset Diseases: Misperception of Disease Onset. *J Hand Surg Am.* 2017;42(12):971–977. [PubMed] [Google Scholar]

2. Culvenor AG, Oiestad BE, Hart HF, Stefanik JJ, Guermazi A, Crossley KM. Prevalence of knee osteoarthritis features on magnetic resonance imaging in asymptomatic uninjured adults: a systematic review and meta-analysis. *Br J Sports Med.* 2018;53(20):1268–1278. [PMC free article] [PubMed] [Google Scholar]

3. Beals CT, Magnussen RA, Graham WC, Flanigan DC. The Prevalence of Meniscal Pathology in Asymptomatic Athletes. *Sports Med.* 2016;46(10):1517–24. [PubMed] [Google Scholar]

4. Keng A, Sayre EC, Guermazi A, et al. Association of body mass index with knee cartilage damage in an asymptomatic population-based study. *BMC Musculoskelet Disord.* 2017;18(1):517. [PMC free article] [PubMed] [Google Scholar]

5. Kornick J, Trefelner E, McCarthy S, Lange R, Lynch K, Jokl P. Meniscal abnormalities in the asymptomatic population at MR imaging. *Radiology.* 1990;177(2):463–465. [PubMed] [Google Scholar]

6. Liu TC, Leung N, Edwards L, Ring D, Bernacki E, Tonn MD. Patients Older Than 40 Years With Unilateral Occupational Claims for New Shoulder and Knee Symptoms Have Bilateral MRI Changes. *Clin Orthop Relat Res.* 2017;475(10):2360–2365. [PMC free article] [PubMed] [Google Scholar]

7. Bachoura AM, Vranceanu AP, Ring D. Illness constructs in musculoskeletal medicine. *Orthop J Harvard Med Sch.* 2009;11:115–124. [Google Scholar]

8. Dersh J, Polatin PB, Leeman G, Gatchel RJ. The management of secondary gain and loss in medicolegal settings: strengths and weaknesses. *J Occup Rehabil.* 2004;14(4):267–279. [PubMed] [Google Scholar]

9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174. [PubMed] [Google Scholar]

10. Ring DC, Dobbs MB, Gioe TJ, Manner PA, Leopold SS. Editorial: How the Words We Use Affect the Care We Deliver. *Clin Orthop Relat Res.* 2016;474(10):2079–2080. [PMC free article] [PubMed] [Google Scholar]

11. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med.* 2013;369(26):2515–2524. [PubMed] [Google Scholar]

12. Khan M, Evaniew N, Bedi A, Ayeni OR, Bhandari M. Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis. *CMAJ.* 2014;186(14):1057–1064. [PMC free article] [PubMed] [Google Scholar]

13. Sihvonen R, Englund M, Turkiewicz A, Jarvinen TLN. Mechanical Symptoms and Arthroscopic Partial Meniscectomy in Patients with Degenerative Meniscus Tear: A Secondary Analysis of a Randomized Trial. *Ann Intern Med.* 2016;164(7):449–455. [PubMed] [Google Scholar]

14. Brignardello-Petersen R, Guyatt GH, Buchbinder R, et al. Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. *BMJ Open.* 2017;7(5):e016114. [PMC free article] [PubMed] [Google Scholar]

15. Zanetti M, Pfirrmann CWA, Schmid MR, Romero J, Seifert B, Hodler J. Patients with suspected meniscal tears: prevalence of abnormalities seen on MRI of 100 symptomatic and 100 contralateral asymptomatic knees. *AJR Am J Roentgenol.* 2003;181(3):635–641. [PubMed] [Google Scholar]

16. Bhattacharyya T, Gale D, Dewire P, et al. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. *J Bone Joint Surg Am.* 2003;85(1):4–9. [PubMed] [Google Scholar]

17. Briet JP, Houwert RM, Hageman MGJS, Hietbrink F, Ring DC, Verleisdonk EJJM. Factors associated with pain intensity and physical limitations after lateral ankle sprains. *Injury.* 2016;47(11):2565–2569. [PubMed] [Google Scholar]

18. Becker SJE, Briet JP, Hageman MGJS, Ring D. Death, taxes, and trapeziometacarpal arthrosis. *Clin Orthop Relat Res.* 2013;471(12):3738–44. [PMC free article] [PubMed] [Google Scholar]

19. Schwartzberg R, Reuss BL, Burkhart BG, Butterfield M, Wu JY, McLean KW. High Prevalence of Superior Labral Tears Diagnosed by MRI in Middle-Aged Patients with Asymptomatic Shoulders. *Orthop J Sport Med.* 2016;4(1):2325967115623212. [PMC free article] [PubMed] [Google Scholar]