

Reliability of Radiologic Assessments of Clinically Relevant Growth Remaining in Knee MRI of Children and Adolescents With Patellofemoral Instability

Data From the JUPITER Cohort

Peter D. Fabricant,* MD, MPH, Madison R. Heath, BS, Matthew Veerkamp, BA, Simone Gruber, BA, Daniel W. Green, MD, MS, Sabrina M. Strickland, MD, Eric J. Wall, MD, Douglas N. Mintz, MD, Kathleen H. Emery, MD, JUPITER Study Group, Beth E. Shubin Stein, MD, and Shital N. Parikh, MD

Investigation performed at Hospital for Special Surgery, New York, New York, USA and Cincinnati Children's Hospital, Cincinnati, Ohio, USA

Background: Surgical decision making and preoperative planning for children and adolescents with patellofemoral instability rely heavily on a patient's skeletal maturity. To be clinically useful, radiologic assessments of skeletal maturity must demonstrate acceptable interrater reliability and accuracy.

Purpose: The purpose of this study was to examine the interrater reliability among surgeons of varying experience levels and specialty training backgrounds when evaluating the skeletal maturity of the distal femur and proximal tibia of children and adolescents with patellofemoral instability.

Study Design: Cohort study (diagnosis); Level of evidence, 3.

Methods: Six fellowship-trained orthopaedic surgeons (3 pediatric orthopaedic, 2 sports medicine, and 1 with both) who perform a high volume of patellofemoral instability surgery examined 20 blinded knee radiographs and magnetic resonance images in random order. They assessed these images for clinically relevant growth (open physis) or clinically insignificant growth (closing/closed physis) remaining in the distal femoral and proximal tibial physes. Fleiss' kappa was calculated for each measurement. After initial ratings, raters discussed consensus methods to improve reliability and assessed the images again to determine if training and new criteria improved interrater reliability.

Results: Reliability for initial assessments of distal femoral and proximal tibial physeal patency was poor (kappa range, 0.01-0.58). After consensus building, all assessments demonstrated almost-perfect interrater reliability (kappa, 0.99 for all measurements).

Conclusion: Surgical decision making and preoperative planning for children and adolescents with patellofemoral instability rely heavily on radiologic assessment of skeletal maturity. This study found that initial interrater reliability of physeal patency and clinical decision making was unacceptably low. However, with the addition of new criteria, a consensus-building process, and training, these variables became highly reliable.

Keywords: patella; instability; physis; physes; dislocation; pediatric; skeletal immaturity

Patellofemoral instability is a common orthopaedic condition that is associated with significant functional limitations, pain, arthritis, and diminished quality of life.^{9,20,28} First-time patellar dislocations are more

common in adolescents than any other age group, and these injuries have one of the highest recurrence risks of any injury in orthopaedics.^{9,18,28} The recurrence risk is especially high in younger patients, with an abundance of recent literature indicating that skeletal immaturity increases risk of recurrence after operative and nonoperative treatments for patellar instability.^{2,13-16,18} Management of pediatric and adolescent patients with

The Orthopaedic Journal of Sports Medicine, 9(4), 2325967121991110
DOI: 10.1177/2325967121991110
© The Author(s) 2021

This open-access article is published and distributed under the Creative Commons Attribution - NonCommercial - No Derivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits the noncommercial use, distribution, and reproduction of the article in any medium, provided the original author and source are credited. You may not alter, transform, or build upon this article without the permission of the Author(s). For article reuse guidelines, please visit SAGE's website at <http://www.sagepub.com/journals-permissions>.

patellofemoral instability therefore requires careful consideration of skeletal maturity by the treating orthopaedic surgeon.

Treatment of patellofemoral instability depends on the patient's skeletal maturity, both in first-time dislocators, who might be considered for operative versus nonoperative treatment, and in recurrent dislocators, for whom specific surgical techniques may be modified if the physes are open. Disturbance of the distal femoral physis in skeletally immature patients can lead to growth arrest and deformity.^{5,12,25,26} Therefore, a first step in determining the optimal treatment strategy and planning any potential surgery is to determine if the femoral and tibial physes are open and, if so, whether there is a clinically relevant amount of growth remaining.⁵ An additional goal of surgical intervention for patellofemoral instability is to correct pathologic anatomic abnormalities contributing to the instability, including an increased distance of the tibial tubercle–trochlear groove, severe patella alta, increased valgus alignment, and trochlea dysplasia.^{7,8,27} Procedures that may address the pathologic anatomy, such as tibial tubercle osteotomy and medial patellofemoral ligament reconstruction, involve risk to the tibial tubercle apophysis and distal femoral physis, respectively, which must be avoided in the setting of open physes.^{12,16} Therefore, decision-making algorithms and preoperative planning rely heavily upon assessment of growth remaining in the distal femoral and proximal tibial physes. Specifically, quantifying the amount of remaining physal growth impacts decisions on: (1) tibial tubercle osteotomy versus distal patellar tendon/soft tissue realignment techniques, (2) coronal realignment with osteotomy versus implant-mediated guided growth, and (3) the location and method of femoral fixation of the medial patellofemoral ligament graft.⁵

Because of the importance of precisely and accurately determining the patency of the distal femoral and proximal tibial physes, it is essential for any physal classification system to demonstrate acceptable interrater reliability among orthopaedic surgeons. The primary purpose of this study was therefore to evaluate interrater reliability among attending surgeons when performing radiologic assessments of physal patency of the distal femur and proximal tibia in pediatric and adolescent patients with patellar instability. Second, this study aimed to standardize physal assessments through peer-to-peer consensus building to achieve higher inter- and intrarater reliability.

METHODS

After institutional review board approval was obtained at each participating institution, a subset of prospectively collected data was selected from a larger cohort of patients with patellofemoral instability in a multicenter study.³ For the current study, complete imaging sets were selected comprising 20 individuals between the ages of 12 and 15 years and equally distributed by sex, to ensure that the participating surgeons assessed an appropriate range of physal patencies. If available, the bone ages were also recorded, as determined from hand and wrist radiographs. For each knee imaging set, the pretreatment anteroposterior and lateral radiographs were reviewed, as were the intermediate-weighted time-to-echo (TE) coronal and sagittal proton density sequences on magnetic resonance imaging (MRI).

Six fellowship-trained orthopaedic surgeons who performed a minimum of 20 patellofemoral instability operations per year examined each imaging set: 3 with pediatric orthopaedic surgery fellowship training (D.W.G., E.J.W.,

*Address correspondence to Peter D. Fabricant, MD, MPH, Pediatric Orthopedic Surgery and Sports Medicine, Hospital for Special Surgery, 535 East 70th St, New York, NY 100219, USA (email: fabricantp@hss.edu).

All authors are listed in the Authors section at the end of this article.

Final revision submitted September 2, 2020; accepted October 9, 2020.

One or more of the authors has declared the following potential conflict of interest or source of funding: P.D.F. has received education payments from Smith & Nephew and hospitality payments from Medical Device Business Services. D.W.G. has received consulting fees from Arthrex; speaking fees from AO Trauma and Arthrex; faculty/speaker fees from Synthes; and royalties from Arthrex, *Current Opinion in Pediatrics*, Pega Medical, and Wolters Kluwer Health. S.M.S. has received research support from JRF and Vericel; consulting fees from DePuy/Medical Device Business Services, Moximed, Pfizer, and Smith & Nephew; speaking fees from Organogenesis and Vericel; royalties from Organogenesis; and hospitality payments from Fidia Pharma and Stryker. E.J.W. has received consulting fees from OrthoPediatrics. D.N.M. has received royalties from Springer. J.M.B. has received education payments and speaking fees from Steelhead Surgical and consulting fees from Smith & Nephew. H.B.E. has received education payments from Pylant Medical; faculty/speaker fees from Smith & Nephew, Pylant Medical, and Synthes; and hospitality payments from Arthrex. J.F. has received research support from Active Implants, Arthrex, Episurf, Fidia, JRF Ortho, Moximed, Novartis, Organogenesis, Samumed, Vericel, and Zimmer Biomet; education payments from Crossroads Orthopedics; consulting fees from Aesculap/B. Braun, Cartiheal, Cook Biotech, DePuy, Exactech, Moximed, Organogenesis, Regentis, RTI Surgical, Samumed, and ZKR Orthopedics; speaking fees from Aastrom Biosciences, Arthrex, Moximed, Organogenesis, and Vericel; royalties from Arthrex, Biopoly, DePuy, Organogenesis, Springer, and Thieme; and hospitality payments from Skeletal Kinetics. J.F. also has stock/stock options in MedShape and Ortho Regenerative Tech. B.E.H. has received education payments from Arthrex and Kairos Surgical, other financial or material support from Allosource and Vericel, and royalties from Springer and has stock/stock options in Imagen Technologies. J.L.K. has received education payments from Medwest and consulting fees from Flexion Therapeutics, has stock/stock options in Acuitive and Marrow Access Technologies, and is an employee of Marrow Access Technologies. D.K. has received education payments from Kairos Surgical and other financial or material support from Arthrex. R.A.M. has received research support from Zimmer, education payments from CDC Medical, and other financial or material support from Arthrex. L.H.R. has received education payments from Arthrex and consulting fees from GLG Consulting and Relief Health and has stock/stock options in Relief Health. S.L.S. has received research support from Arthrex; grant funding from DJO; education payments from Elite Orthopedics; and consulting fees from Arthrex, Ceterix, ConMed Linvatec, Flexion Therapeutics, GLG Consulting, JFR Ortho, Moximed, Olympus, and Vericel. M.T. has received grant support from DJO. P.W. has received research support from AlloSource and Ossur, education payments from Pylant Medical, and royalties from Elsevier. B.E.S.S. has received consulting fees, speaking fees, and royalties from Arthrex. S.N.P. has received education payments from CDC Medical, speaking fees from Synthes, and royalties from Wolters Kluwer Health. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval was obtained from the Hospital for Special Surgery (No. 2015-910).

TABLE 1
Surgeon Assessments and Criteria Before and After Consensus Training^a

	Before Consensus Training	After Consensus Training (MRI Only)
Assessment		
Femoral physis decision	Do you think there is significant growth remaining at the femoral physis that you would change your femoral tunnel to avoid the physis?	(Removed)
Femoral physeal patency	Is the femoral physis open or closing/closed?	Is the femoral physis open or closing/closed?
Tibia	Did the patency of the tibial physis alter your surgical decision/plan?	Is the tibial physis open or closing/closed?
Criteria for assessment		
Open physes	(Based on current practice)	If the low signal of the physis on intermediate-weighted TE MRI sequence can be visualized along the entire physis (without a central closure), then physis is classified as open.
Closing/closed physes	(Based on current practice)	If the low signal of the physis on intermediate-weighted TE MRI sequence can NOT be visualized along the entire physis, then physis is classified as closed. Partially closed physes should be included in the closing/closed group. Additionally, the remnant “physeal scar” of the physis, which lacks the characteristic low signal of an active physis, should be classified as closing/closed.

^aAssessments that surgeons were asked to make and the criteria that they were instructed to use when making those assessments before and after consensus training. MRI, magnetic resonance imaging; TE, time-to-echo.

and S.N.P.), 2 with sports medicine fellowship training (B.E.S.S. and S.M.S.), and 1 with both pediatric orthopaedic and sports medicine fellowship training (P.D.F.). Surgeon practice ranged from 3 to 25 years (mean, 16 years). Surgeons were asked to perform 2 rounds of physeal patency assessments based on (1) their current practice without any discussion or consensus training with other surgeons and (2) the strategies discussed and learned during consensus training with other surgeons. Initially, the surgeons were asked to make 3 determinations concerning the distal femoral and proximal tibial physes based on knee radiographs and MRI separately, using the same imaging sets. Before consensus training, radiographs and MRI were used for assessments, but radiographs were removed from assessments after consensus training, owing to indications from this study and previous literature¹⁰ that radiographs were a poor modality for assessing clinically relevant growth remaining in the knee. Table 1 shows the exact wording of the assessments that the surgeons were required to make and the criteria for making those assessments for both rounds of assessments. For the first round of assessments, responses to the first and third questions were dichotomized as *yes* or *no*, and responses to the second question were dichotomized as *open* or *closing/closed*. For the second round of assessments, all responses were dichotomized as *open* or *closing/closed*. Each surgeon was granted access to a deidentified set of images, which were imported into an institutional research picture archiving and communication system database for analysis. Surgeons entered all assessments in an Excel spreadsheet (Microsoft Corp) and sent them to the study coordinator within 1 month.

After the participating surgeons performed the initial set of ratings, with no discrete criteria for physeal patency, the process of physeal assessment was discussed among them. The surgeons subsequently established new physis classifications based on previously validated methodology²² and their consensus discussions. George et al¹⁰ demonstrated that skeletal maturity is significantly overestimated when physes are examined with radiographs rather than MRI; therefore, MRI was chosen exclusively to examine physes in the second rating. Several previous MRI studies of the normal knee described physes using visualization of continuity and thickness of the physeal band.^{6,24} The bone atlas developed by Pennock et al²² provided a thorough review of growth and development of the knee as seen on MRI. Pennock et al described 6 stages for ossification in the distal femur: (1) the presence of the epiphyseal secondary ossification center, (2) complete ossification of the epiphysis, (3) disappearance of the lamellated appearance of the subchondral epiphyseal cartilage (termed the “Oreo” sign), (4) narrowing of the physis, (5) partial closure of the physis, and (6) complete closure of the physis. Additionally, they described 9 stages for ossification in the proximal tibia: (1) the presence of the epiphyseal secondary ossification center, (2) partial ossification of the tibial spine, (3) complete ossification of the tibial spine, (4) tubercle extension of the epiphysis, (5) fusion of the tubercle apophysis ossification center with the epiphysis, (6) partial fusion of the tubercle apophysis, (7) complete ossification of the tibial epiphysis, (8) partial closure of the physis, and (9) complete closure of the physis. These stages had a strong correlation with chronologic age and high interrater reliability, so these images and definitions were used to develop new rating criteria. However, surgical decisions for patellar

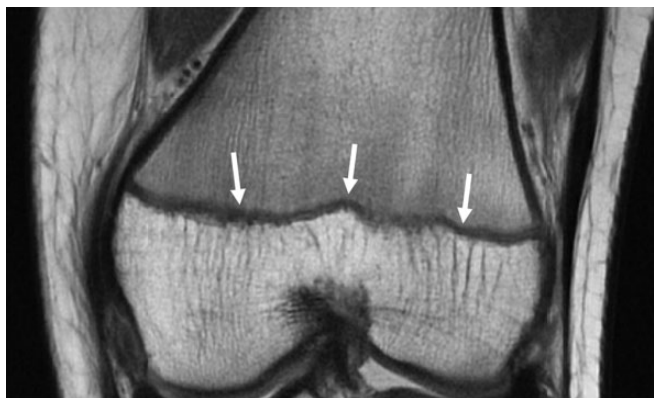


Figure 1. Example of low signal along the entire physis (white arrows) based on new consensus criteria of distal femoral physes that should be classified as “open.”



Figure 2. Example of low signal along the entire distal femoral physis (white arrows) on intermediate-weighted TE coronal MRI sequence, based on the new consensus criteria of an “open” distal femoral physis. The low signal along the proximal tibial physis is interrupted (black arrows), which would be classified as “closing/closed.” MRI, magnetic resonance imaging; TE, time-to-echo.

instability focus on clinically relevant growth remaining in the physis, so the stages were grouped when the new criteria were developed. Stages 1 to 4 for the femur and stages 1 to 7 for the tibia were considered “open,” while the last 2 stages of each were considered “closing/closed.” After this review, criteria were developed for assessing the physis using the intermediate-weighted TE MRI sequence, including sample figures from an atlas of MRI of the knee²² and imaging from other participants in the prospective cohort



Figure 3. (A) The low signal along the proximal tibial physis on intermediate-weighted TE sagittal MRI sequence is interrupted (black arrow), which would be classified as “closing/closed.” (B) The physeal scar along the distal femoral physis without the low signal (black arrows) would be classified as “closing/closed.” MRI, magnetic resonance imaging; TE, time-to-echo.

that was not utilized in this reliability study (Table 1 and Figures 1-3).

Before the second round of ratings using only the MRIs, a training program was implemented using the new criteria and a session of peer-to-peer teaching to ensure that all raters understood the new methodology and criteria for examining physes. Based on the consensus-building process and new criteria, surgeons repeated the physal patency assessments 6 months after the initial assessment (Table 1), and the analyses were rerun for the new responses. Finally, 1 surgeon with fellowship training in pediatric orthopaedic surgery (D.W.G.) and 1 with fellowship training in sports medicine (B.E.S.S.) completed the physal patency assessments a second time 4 months after the previous assessment to establish intrarater reliability for the new assessment methodology.

Data analysis was performed using SPSS Version 22 (IBM). Inter- and intrarater reliability was calculated using Fleiss' kappa. If any of the 6 surgeons deemed that an image was technically inadequate for making a physal patency determination, that image was removed from that variable's analysis. While this decreased the number of included imaging sets for some analyses, it ensured that the results were not confounded by poor or inadequate radiologic images. To determine if additional characteristics of the surgeons influenced their interrater reliability, surgeons were additionally stratified by years of practice and whether they completed a pediatric orthopaedic surgery fellowship. These stratifications were not selected for but were based on characteristics of surgeons already participating in the study. Years of practice for the 6 raters were stratified in thirds: “low” consisted of 2 surgeons with <15 years of practice as attending surgeons; “medium,” 2 surgeons with 15 to 20 years of practice; and “high,” 2 surgeons with 20 to 25 years of practice. Reliability was classified per the work of Landis and Koch¹⁷: 0.0 to 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and 0.81 to 1.0, almost perfect. A kappa score ≤ 0.60 for any given assessment was considered unacceptably low agreement.

TABLE 2
 First- and Second-Round Assessments of Distal Femoral and Proximal Tibial Physéal Patency,
 by 6 Fellowship-Trained Orthopaedic Surgeons^a

Variable	First Assessment		Second Assessment	
	No.	Kappa (95% CI)	No.	Kappa (95% CI)
Radiograph				
Q1: Change femoral tunnel	16	0.46 (0.25 to 0.71)		
Q2: Femoral physéal patency	16	0.58 (0.37 to 0.79)		
Q3: Change surgery attributed to tibia	14	0.01 (−0.02 to 0.10)		
MRI				
Q1: Change femoral tunnel	19	0.38 (0.20 to 0.62)		
Q2: Femoral physéal patency	19	0.25 (0.09 to 0.49)	19	0.99 (0.99 to 0.99)
Q3: Change surgery attributed to tibia	19	0.07 (0.00 to 0.21)		
Q4: Tibial physéal patency			19	0.99 (0.99 to 0.99)

^aAll measures demonstrated almost-perfect reliability after a round of consensus training. MRI, magnetic resonance imaging.

TABLE 3
 First Round of Assessments of the Distal Femoral and Proximal Tibial Physéal Patency on Radiographs,
 by the Surgeons' Years of Practice and Fellowship Type^a

Variable	Kappa (95% CI)		
	Q1: Change Femoral Tunnel	Q2: Femoral Physéal Patency	Q3: Change Surgery Attributed to Tibia
Years of practice			
Low	0.59 (0.22 to 0.82)	0.71 (0.40 to 0.88)	0.00 (−0.07 to 0.14)
Medium	0.56 (0.14 to 0.81)	0.56 (0.14 to 0.81)	0.02 (−0.06 to 0.20)
High	0.18 (−0.14 to 0.54)	0.18 (−0.14 to 0.54)	−0.09 (−0.57 to 0.41)
Fellowship			
Pediatrics	0.38 (0.13 to 0.65)	0.49 (0.23 to 0.73)	−0.03 (−0.07 to 0.08)
Sports	0.56 (0.14 to 0.81)	0.56 (0.14 to 0.81)	0.02 (−0.06 to 0.20)

^aAcceptable interrater reliability is in bold.

TABLE 4
 First Round of Assessments of the Distal Femoral and Proximal Tibial Physéal Patency on MRI,
 by the Surgeons' Years of Practice and Fellowship Type^a

Variable	Kappa (95% CI)		
	Q1: Change Femoral Tunnel	Q2: Femoral Physéal Patency	Q3: Change Surgery Attributed to Tibia
Years of practice			
Low	0.00 (−0.43 to 0.43)	0.08 (−0.10 to 0.34)	0.00 (−0.05 to 0.11)
Medium	0.57 (0.17 to 0.81)	0.39 (−0.02 to 0.71)	0.09 (−0.11 to 0.37)
High	0.15 (−0.12 to 0.47)	0.04 (−0.08 to 0.25)	0.00 (−0.05 to 0.11)
Fellowship			
Pediatrics	0.27 (0.08 to 0.53)	0.12 (−0.01 to 0.33)	0.02 (−0.04 to 0.15)
Sports	0.57 (0.17 to 0.81)	0.39 (−0.02 to 0.71)	0.09 (−0.11 to 0.37)

^aThere were no acceptable interrater reliabilities. MRI, magnetic resonance imaging.

RESULTS

Of the individuals in the 20 selected imaging sets, 50% were male, and the age at the time of imaging was 13.9 ± 1.0 years (mean ± SD). For the initial round of assessments, 4 images were removed from the radiographic femoral physéal analysis, 6 from the radiographic tibial physéal analysis, and 1 from MRI analyses for low quality, as determined by at least

1 of the participating surgeon raters. Interrater reliability for all initial physéal assessments was unacceptably low (kappa range = 0.01-0.58) (Table 2). When stratified by years of practice and fellowship type, the surgeons with low years of practice demonstrated acceptable reliability on the femoral physéal assessment on radiographs (kappa = 0.71), while the reliability for all other assessments remained unacceptably low (kappa range = −0.09 to 0.59) (Tables 3 and 4).

TABLE 5
Bone Age Assessed Using the G&P Method
for Each Physeal Patency^a

Sex	Physeal Patency		No.	G&P Bone Age, y	
	Distal Femur	Proximal Tibia		Mean	Range
Female	Open	Open	2	13.5	12-15
	Open	Closing/closed	3	14.8	14.5-15
	Closing/closed	Closing/closed	1	14.5	14.5
Male	Open	Open	8	15	14-17
	Open	Closing/closed	0	NA	NA
	Closing/closed	Closing/closed	1	17	17

^aBone age was not available for 4 of the 19 assessed imaging sets. G&P, Greulich and Pyle; NA, not applicable.

For the second assessment, interrater reliability for all physeal assessments demonstrated almost perfect agreement (Table 2). Regarding intrarater reliability, the 2 participating surgeons demonstrated almost-perfect agreement (intraclass correlation coefficient, 0.99; 95% CI, 0.99-0.99). Based on the final assessment, the number of images from individuals with open distal femoral physes was 17 of 19 (89%), and the number with open proximal tibial physes was 13 of 19 (68%). See Table 5 for bone ages within each physeal patency.

DISCUSSION

The primary purpose of this study was to determine the interrater reliability of radiologic assessments of physeal patency by attending orthopaedic surgeons in pediatric and adolescent patients with patellofemoral instability. We found that before the establishment of any discrete criteria for physeal assessment, interrater reliability for physeal assessment was unacceptably low, even when accounting for years of practice and formal pediatric fellowship training. This poor initial reliability is noteworthy, given the importance of skeletal maturity for treatment decision making and preoperative planning for patients with patellofemoral instability.^{16,19} Because of the importance of skeletal maturity assessments, surgeons must use reliable methods of physeal evaluation to appropriately risk-stratify patients and minimize risk of iatrogenic complications.²⁵

Although an assessment may display good interrater reliability, this does not necessarily indicate that the assessment is accurate. In this study, we attempted to determine how accurate the final physeal patency assessments were by comparing them with the bone ages of the included individuals. We found that while the adolescents with open physes had lower bone ages on average than those with closed physes, there was a wide range of bone ages among participants with the same knee physeal patency. Skeletal age in children and adolescents is typically determined using a left hand/wrist radiograph and the Greulich and Pyle atlas.¹¹ However, for knee pathology and surgery, the skeletal maturation of the knee and the

estimation of remaining growth around the knee are more important than an overall assessment of skeletal age. In support of this study's findings, previous studies have demonstrated significant intraindividual variability between the skeletal maturity of the hand and the knee, with differences ranging from 1.45 to 2.99 years.^{1,29} Thus, for children and adolescents with patellar instability, assessment of physes around the knee is more important than skeletal age estimation, and skeletal age may not be an accurate way to determine clinically relevant growth in these physes. Future studies should confirm the accuracy of these growth plate assessments using follow-up leg-length imaging to determine how much growth occurred after these assessments until the physes completely closed.

The results of the initial physeal evaluation were similar to those in previous orthopaedic studies that demonstrated low interrater reliability of radiologic assessments. A study on osteochondritis dissecans of the knee showed poor interrater reliability on the healing status of the osteochondritis dissecans lesions.²¹ In a trauma-related setting, Butcher et al⁴ reported low interrater reliability between and within institutions on the classification of "polytrauma" in patients in the intensive care unit, an important indicator for severity of injury and urgency of treatment. Similar to the current study however, Riddle et al²³ noted that after minimal training there was significantly improved interrater reliability in surgeons assessing the osteoarthritis status of the knee. These studies show that common classifications important to medical decision-making algorithms may not be reliable among practicing surgeons. Lack of assessment reliability can jeopardize patient care, so it is essential for practicing surgeons to reach a reliable consensus on best practice with their peers. However, these studies also demonstrate that improving reliability can be achieved quickly without large investments of time or resources. Therefore, consensus discussions should be highly considered by all surgeons to improve patient care.

This study has several limitations. Although intermediate-weighted TE coronal and sagittal MRI sequences were specified, no particular slice was specified, and interrater reliability of the slice chosen was not assessed, which could have reduced the interrater reliability of ratings. However, these conditions resembled actual clinical practice and provided a more realistic setting for image evaluation. Exclusion of technically inadequate images was done to ensure that the results were not altered by poor imaging technique. This may also have introduced some nondifferential bias toward overestimation of reliability and decreased generalizability, as real-world imaging is not always technically adequate. However, we felt that it was appropriate, as it is always clinically feasible to repeat imaging that is not technically adequate to improve diagnostic accuracy. Additionally, for the second rating, questions about the influence of physeal patency on clinical decision making were eliminated. This was done to simplify the decision-making process: Open physes of the femur and tibia would imply skeletal immaturity, which would alter or modify medical decisions, whereas closing/closed physes of either the femur or the tibia would imply skeletal maturity and favor adult-like treatment. Furthermore, because the

second set of questions asked of the surgeons was different from the first set, it is unclear how much the improvement in reliability was from the new questions asked versus the new criteria. Another limitation was that the second set of ratings was performed using the original MRI scans rather than a new set of images, which was done to ensure technical adequacy of the imaging set. This may have introduced bias into the second set of assessments, since the surgeons had already seen them. However, surgeons were not told which images they disagreed on during their consensus discussions, and several months passed between ratings, therefore minimizing any potential bias in this instance. Additionally, only 2 of the 19 distal femoral physes were closing/closed, so raters had few opportunities to assess closing/closed femoral physes. However, this is a common scenario when evaluating adolescent knees, as tibial physes typically close before femoral physes, and the raters were able to assess several closing/closed tibial physes. Also, because of inadequate follow-up imaging, this study was not able to evaluate the accuracy of the physal assessments for clinically relevant growth remaining. Finally, all surgeons were fellowship-trained, high-volume patellofemoral instability surgeons. Because no general orthopaedic surgeons performed the ratings, these findings may not be generalizable to all surgeons. Although using only patellofemoral surgeons may have overestimated the reliability assessment, it is probable that baseline reliability of skeletal maturity is, if anything, even worse in general practice than in the first round of assessments in the current study. This underscores the importance of consensus building and training to maximize interrater reliability of skeletal maturity assessment by all surgeons.

In conclusion, physal assessment without consensus training was poor in this study. With consensus building and training, these assessments demonstrated almost-perfect reliability. While this study focused on patients with patellofemoral instability, these results can be generalized to pediatric patients with other orthopaedic lower extremity injuries (eg, anterior cruciate ligament tears). Lack of measurement reliability and accuracy for knee physal patency can jeopardize pediatric patient care when surgical indications are determined and an appropriate procedure is chosen on the basis of skeletal maturity. Surgeons should focus on using reliable imaging metrics in children and adolescents with patellofemoral instability, and measurements that remain unreliable after consensus building and training should be removed from clinical decision-making algorithms.

AUTHORS

Peter D. Fabricant, MD, MPH (Hospital for Special Surgery, New York, New York, USA); Madison R. Heath, BS (Hospital for Special Surgery, New York, New York, USA); Matthew Veerkamp, BA (Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA); Simone Gruber, BA (Hospital for Special Surgery, New York, New York, USA); Daniel W. Green, MD, MS (Hospital for Special

Surgery, New York, New York, USA); Sabrina M. Strickland, MD (Hospital for Special Surgery, New York, New York, USA); Eric J. Wall, MD (Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA); Douglas N. Mintz, MD (Hospital for Special Surgery, New York, New York, USA); Kathleen H. Emery, MD (Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA); JUPITER Study Group; Jacqueline M. Brady, MD (Oregon Health and Science University Hospital, Portland, Oregon, USA); Henry B. Ellis, MD (Texas Scottish Rite Hospital for Children, Dallas, Texas, USA); Jack Farr, MD (OrthoIndy Hospital, Indianapolis, Indiana, USA); Benton E. Heyworth, MD (Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA); Jason L. Koh, MD, MBA (Orthopedic & Spine Institute, NorthShore University Health System, Evanston, Illinois, USA); Dennis Kramer, MD (Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA); Robert A. Magnusen, MD (The Ohio State University Wexner Medical Center, Columbus, Ohio, USA); Lauren H. Redler, MD (Columbia University Irving Medical Center, New York, New York, USA); Seth L. Sherman, MD (Stanford Health Care, Redwood City, California, USA); Marc Tompkins, MD (TRIA Orthopedic Center, Bloomington, Minnesota, USA); Philip L. Wilson, MD (Texas Scottish Rite Hospital for Children, Dallas, Texas, USA); Beth E. Shubin Stein, MD (Hospital for Special Surgery, New York, New York, USA); and Shital N. Parikh, MD (Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA).

ACKNOWLEDGMENT

The authors acknowledge Drs Matthew Milewski, Yi-Meng Yen, and Adam Yanke for their contributions.

REFERENCES

1. Aicardi G, Vignolo M, Milani S, Naselli A, Magliano P, Garzia P. Assessment of skeletal maturity of the hand-wrist and knee: a comparison among methods. *Am J Hum Biol.* 2000;12(5): 610-615.
2. Arendt EA, Askenberger M, Agel J, Tompkins MA. Risk of redislocation after primary patellar dislocation: a clinical prediction model based on magnetic resonance imaging variables. *Am J Sports Med.* 2018;46(14):3385-3390.
3. Bishop ME, Brady JM, Ling D, Parikh S, Shubin Stein BE. Descriptive epidemiology study of the Justifying Patellar Instability Treatment by Early Results (JUPITE) cohort. *Orthop J Sports Med.* 2019;7(3_suppl): 2325967119S00046.
4. Butcher NE, Enninghorst N, Sisak K, Balogh ZJ. The definition of polytrauma: variable interrater versus intrarater agreement—a prospective international study among trauma surgeons. *J Trauma Acute Care Surg.* 2013;74(3):884-889.
5. Clark D, Metcalfe A, Wogan C, Mandalia V, Eldridge J. Adolescent patellar instability current concepts review. *Bone Joint J.* 2017;99(2): 159-170.
6. Dedouit F, Auriol J, Rousseau H, Rougé D, Crubézy E, Telmon N. Age assessment by magnetic resonance imaging of the knee: a preliminary study. *Forensic Sci Int.* 2012;217(1-3):232.
7. Dejour D, Le Coultre B. Osteotomies in patello-femoral instabilities. *Sports Med Arthrosc.* 2007;15(1):39-46.

8. Dejour H, Walch G, Nove-Josserand L, Guier C. Factors of patellar instability: an anatomic radiographic study. *Knee Surg Sports Traumatol Arthrosc.* 1994;2(1):19-26.
9. Fithian DC, Paxton EW, Stone ML, et al. Epidemiology and natural history of acute patellar dislocation. *Am J Sports Med.* 2004;32(5):1114-1121.
10. George J, Nagendran J, Azmi K. Comparison study of growth plate fusion using MRI versus plain radiographs as used in age determination for exclusion of overaged football players. *Br J Sports Med.* 2012;46(4):273-278.
11. Greulich W, Pyle S. *Radiographic Atlas of Skeletal Development of the Hand and Wrist.* Stanford University Press; 1959.
12. Harrison MHM. The results of a realignment operation for recurrent dislocation of the patella. *J Bone Joint Surg Br.* 1955;37(4):559-567.
13. Hevesi M, Heidenreich MJ, Camp CL, et al. The Recurrent Instability of the Patella score: a statistically based model for prediction of long-term recurrence risk after first-time dislocation. *Arthroscopy.* 2019;35(2):537-543.
14. Huntington LS, Webster KE, Devitt BM, Scanlon JP, Feller JA. Factors associated with an increased risk of recurrence after a first-time patellar dislocation: a systematic review and meta-analysis. *Am J Sports Med.* 2020;48(10):2552-2562.
15. Jaquith BP, Parikh SN. Predictors of recurrent patellar instability in children and adolescents after first-time dislocation. *J Pediatr Orthop.* 2017;37(7):484-490.
16. Khormae S, Kramer DE, Yen YM, Heyworth BE. Evaluation and management of patellar instability in pediatric and adolescent athletes. *Sports Health.* 2015;7(2):115-123.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159.
18. Lewallen L, McIntosh A, Dahm D. First-time patellofemoral dislocation: risk factors for recurrent instability. *J Knee Surg.* 2015;28(4):303-309.
19. Lewallen LW, McIntosh AL, Dahm DL. Predictors of recurrent instability after acute patellofemoral dislocation in pediatric and adolescent patients. *Am J Sports Med.* 2013;41(3):575-581.
20. Magnussen RA, Verlage M, Stock E, et al. Primary patellar dislocations without surgical stabilization or recurrence: how well are these patients really doing? *Knee Surg Sports Traumatol Arthrosc.* 2017;25(8):2352-2356.
21. Parikh SN, Allen M, Wall EJ, et al. The reliability to determine "healing" in osteochondritis dissecans from radiographic assessment. *J Pediatr Orthop.* 2012;32(6):35-39.
22. Pennock AT, Bomar JD, Manning JD. The creation and validation of a knee bone age atlas utilizing MRI. *J Bone Joint Surg Am.* 2018;100(4):e20.
23. Riddle DL, Jiranek WA, Hull JR. Validity and reliability of radiographic knee osteoarthritis measures by arthroplasty surgeons. *Orthopedics.* 2013;36(1):e25-e32.
24. Sasaki T, Ishibashi Y, Okamura Y, Toh S, Sasaki T. MRI evaluation of growth plate closure rate and pattern in the normal knee joint. *J Knee Surg.* 2002;15(2):72-76.
25. Seitlinger G, Moroder P, Fink C, Wierer G. Acquired femoral flexion deformity due to physeal injury during medial patellofemoral ligament reconstruction. *Knee.* 2017;24(3):680-685.
26. Trionfo A, Shah A, Misaghi A, Arkader A, Angeles L. Acquired distal femoral deformity after MPFL reconstruction. *Journal of the Pediatric Orthopaedic Society of North America.* 2020;2(2):1-4.
27. Vavken P, Wimmer MD, Camathias C, Quidde J, Valderrabano V, Pagenstert G. Treating patella instability in skeletally immature patients. *Arthroscopy.* 2013;29(8):1410-1422.
28. Waterman BR, Belmont PJ, Owens BD. Patellar dislocation in the United States: role of sex, age, race, and athletic participation. *J Knee Surg.* 2011;25(1):51-58.
29. Xi H, Roche AF. Differences between the hand-wrist and the knee in assigned skeletal ages. *Am J Phys Anthropol.* 1990;83(1):95-102.