Université de Montréal

Validité et fidélité de la combinaison de l'anamnèse et de l'examen physique pour le diagnostic des pathologies communes au genou

par

Simon Décary, pht., M.Sc.

École de réadaptation

Faculté de médecine

Thèse présentée à la Faculté de médecine

en vue de l'obtention du grade de Philosophiae Doctor (Ph. D.)

en Sciences de la réadaptation

31 Août 2017

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Validité et fidélité de la combinaison de l'anamnèse et de l'examen physique pour le diagnostic des pathologies communes au genou

présentée par :

Simon Décary, pht., M.Sc.

A été évaluée par un jury composé des personnes suivantes :

M. Nicolas Dumont, pht., Ph. D.

Président-rapporteur

M. François Desmeules, pht., Ph. D.

Directeur de recherche

Dr. Pascal-André Vendittoli, MD, M.Sc., FRSC

Codirecteur de recherche

Mme. Carole Fortin, pht., Ph. D.

Membre interne du jury

M. Yannick Tousignant-Laflamme, pht., Ph. D.

Examinateur externe

Natacha Trudeau, Ph. D.

Représentante de la doyenne de la FES

Résumé

Les douleurs au genou font partie des motifs de consultation les plus fréquents auprès d'un médecin ou d'un professionnel de la santé. Les pathologies communes au genou incluent celles d'origine traumatique telles les déchirures du ligament croisé antérieur ou les déchirures méniscales et celles d'apparition progressive telles les déchirures méniscales dégénératives, l'ostéoarthose du genou ou le syndrome fémoro-patellaire. Les données probantes démontrent qu'un diagnostic initial valide, basé sur une évaluation musculosquelettique bien accomplie par un intervenant ayant une formation adéquate, permet l'initiation rapide d'un traitement ciblé. Cependant, le manque de connaissances dans l'évaluation musculosquelettique et les erreurs diagnostiques fréquentes dans notre système de santé favorisent l'utilisation inappropriée des tests d'imagerie médicale et augmentent les références non pertinentes en chirurgie orthopédique. Les données probantes sont actuellement limitées et incomplètes concernant la validité et la fidélité de l'évaluation musculosquelettique combinant l'anamnèse et l'examen physique pour orienter le diagnostic différentiel des pathologies au genou. L'amélioration des connaissances en lien avec l'évaluation musculosquelettique des adultes souffrant de douleurs au genou est donc nécessaire afin d'améliorer l'efficience de nos soins de santé.

Cette thèse propose deux objectifs : 1- évaluer l'accord diagnostique entre un physiothérapeute utilisant une évaluation musculosquelettique standardisée sans imagerie et des médecins experts pour les différentes pathologies communes au genou; 2- évaluer la validité de la combinaison de l'anamnèse et de l'examen physique afin de développer une série d'outils valides permettant d'orienter le diagnostic différentiel de quatre pathologies communes au genou.

Nous avons évalué deux cent soixante-dix-neuf participants, présentant 359 diagnostics primaires et secondaires incluant : 43 participants ayant une déchirure du ligament croisé antérieur, 80 participants présentant une déchirure méniscale, 129 participants atteints

d'ostéoarthrose du genou, 75 atteints d'un syndrome fémoro-patellaire et 32 participants présentant une autre pathologie au genou.

Le physiothérapeute qui exécutait une évaluation musculosquelettique standardisée a démontré un excellent accord diagnostique avec des médecins experts dont le diagnostic était basé sur une évaluation musculosquelettique combinée aux résultats des tests d'imagerie (κ=0,89; IC95%: 0,83-0,94). Sur la base de l'évaluation musculosquelettique réalisée indépendamment par le physiothérapeute, nous avons pu développer une série de combinaisons d'éléments de l'anamnèse et de tests de l'examen physique pour orienter le diagnostic différentiel des pathologies au genou. Nous avons démontré qu'un individu qui consultait pour une douleur au genou dont l'origine est un traumatisme du genou en pivot, ayant ressenti un « pop » au moment du traumatisme et chez qui les tests de Lachman ou du *pivot shift* sont positifs, a une probabilité élevée de souffrir d'une déchirure du ligament croisé antérieur avec un rapport de vraisemblance positif de 38,4 (IC95%: 16,0-92,5). Si un individu décrit un traumatisme au genou avec pivot accompagné d'une douleur localisée au côté médial du genou et qu'il démontre une douleur à la palpation de l'interligne articulaire interne, celui-ci a une probabilité élevée de souffrir d'une déchirure méniscale avec un rapport de vraisemblance positif de 8,9 (IC95%: 6,1-13,1).

Un individu présentant une douleur d'apparition progressive au côté médial du genou présente dans les activités qui nécessitent des pivots et qui démontre à l'examen physique un alignement neutre du membre inférieur ou une flexion passive complète du genou, celui-ci a une probabilité modérée d'être atteint d'une déchirure méniscale dégénérative avec un rapport de vraisemblance positif de 6,4 (IC95%: 4,0-10,4). Si toutefois un individu consultant pour une douleur d'apparition progressive est âgé de plus de 50 ans, qu'il a un indice de masse corporel supérieur à 30 et qu'il démontre à l'examen physique la présence d'un alignement au genou en varus ou en valgus, des crépitements à la palpation du genou ou une limitation de l'amplitude articulaire passive en extension, il a une probabilité élevée d'être atteint d'ostéoarthrose du genou avec un rapport de vraisemblance positif de 13,6 (IC95%:

6,5-28,4). Finalement, un individu qui présente une douleur isolée en antérieur du genou accompagnée de difficultés dans les escaliers, d'une douleur à la palpation des facettes rotuliennes et d'une extension passive complète du genou, celui-ci a une probabilité élevée de souffrir d'un syndrome fémoro-patellaire avec un rapport de vraisemblance positif de 8,7 (IC95%: 5,2-14,6).

En résumé, les résultats de cette thèse démontrent qu'il existe un accord diagnostique élevé entre un physiothérapeute qui procède à une évaluation musculosquelettique sans recours à des tests d'imagerie et des médecins experts qui basent leur diagnostic sur l'examen physique et les résultats des tests d'imagerie chez des adultes souffrant de douleurs au genou. Les combinaisons des éléments de l'anamnèse et de l'examen physique développées dans cette thèse permettent d'orienter le diagnostic différentiel de différentes pathologies communes au genou avec une validité considérée modérée à élevée. Ces combinaisons devront être validées dans d'autres contextes cliniques, notamment en première ligne, avant une utilisation clinique répandue.

Mots clés : genou, diagnostic, anamnèse, examen physique, ligament croisé antérieur, ménisque, ostéoarthrose, syndrome fémoro-patellaire.

Abstract

Knee complaints are among the most common reasons for consulting a healthcare practitioner. Common knee disorders include those of traumatic onset such as anterior cruciate ligament or meniscal tears, or those of progressive onset such as degenerative meniscal tears, knee osteoarthritis or patellofemoral pain syndrome. An early diagnosis to guide toward an efficient management is advocated to prevent persistence of pain, functional limitations and loss of quality of life in affected individuals. However, evidence currently shows an overreliance on medical imaging tests and inappropriate orthopaedic surgery referrals, thus delaying initiation of treatment. Evidence is currently limited concerning the validity and reliability of combining history elements and physical examination tests to support the differential diagnosis of common knee disorders.

To answer this evidence gap, this thesis had two objectives: 1- to assess inter-rater diagnostic agreement between a physiotherapist using only a standardized musculoskeletal examination and expert physicians using a musculoskeletal examination in combination with imaging tests for the diagnosis of common knee disorders; 2- to assess the diagnostic validity of clusters combining history elements and physical examination tests and produce a series of tools to support the differential diagnosis of four common knee disorders.

We prospectively recruited two hundred and seventy-nine participants presenting 359 primary and secondary diagnoses including: 43 participants with an anterior cruciate ligament tear, 80 subjects had a meniscal tear, 129 suffered from knee osteoarthritis, 75 were diagnosed with patellofemoral pain syndrome and 32 presented other knee diagnoses. The physiotherapist achieved high inter-rater agreement with the expert physicians for the diagnosis of common knee disorders (κ= 0.89; 95%CI:0.83-0.94).

Multiple clusters combining history elements and physical examination tests were developed to support the differential diagnosis of knee disorders. Our results show that an individual consulting for a knee complaint following trauma during a pivot and describing a “popping

sensation” during the trauma as well as showing a positive Lachman or pivot shift tests has a high probability of having an anterior cruciate ligament tear with a positive likelihood ratio of 38.4 (95%CI: 16.0-92.5). If this individual consulting for a knee complaint following trauma during a pivot experience medially located knee pain, confirmed with medial joint line tenderness at palpation, the subject has a high probability of having a meniscal tear of traumatic origin with a positive likelihood ratio of 8.9 (95%CI: 6.1-13.1).

An individual consulting for a complaint of progressive onset with isolated medially located pain during activities requiring a pivot, who also presents a normal knee alignment or full passive knee flexion, has a moderate probability of having a degenerative meniscal tear with a positive likelihood ratio of 6.4 (95%CI: 4.0-10.4). If this individual complaining of knee pain of progressive onset is older than 50 years old, has a body mass index higher than 30 and at physical examination presents knee crepitus at palpation, a valgus or varus knee misalignment or a restricted passive knee extension, he has a high probability of having symptomatic knee osteoarthritis with a positive likelihood ratio of 13.6 (95%CI: 6.5-28.4). Lastly, an individual consulting for isolated anterior knee pain with difficulty descending stairs, palpable patellar facets tenderness and full passive knee extension has a high probability of having a patellofemoral pain syndrome with a positive likelihood ratio of 8.7 (95%CI: 5.2-14.6).

Overall, the results demonstrated high inter-rater diagnostic concordance between providers and that combining selected history elements and physical examination tests was moderately to highly valid to support the differential diagnosis of common knee disorders. The proposed diagnostic clusters will require external validation in various clinical contexts, including primary care, before widespread implementation in clinical practice.

Key words: knee, diagnosis, history elements, physical examination, anterior cruciate ligament, meniscus, knee osteoarthritis, patellofemoral pain syndrome.

Tables des matières

Liste des tableaux

Article 1

Article 2

Article 3

Article 4

Article 5

Article 6

Article 7

Liste des figures

Liste des sigles et abréviations

Français

IRM : Imagerie par résonnance magnétique
LCA : Ligament croisé antérieur
OA : Ostéoarthrose
RV- : Rapport de vraisemblance négatif
RV+ : Rapport de vraisemblance positif
SFP : Syndrome fémoro-patellaire
Se : Sensibilité
Sp : Spécificité
VPN : Valeur prédictive négative
VPP : Valeur prédictive positive

Anglais

AMSTAR: Assessment of methodological quality of multiple systematic reviews
ACL: Anterior cruciate ligament injuries
DOR: Diagnostic odds ratio
LR: Likelihood ratio
MA: Meta-analysis
NPV: Negative predictive value
OA: Osteoarthritis
SOA: Symptomatic knee osteoarthritis
PCL: Posterior cruciate ligament injuries
PFP: Patellofemoral pain
PPV: Positive predictive value
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Se: Sensitivity
SMT: Symptomatic meniscal tear
Sp: Specificity
SR: Systematic-review

Dédicace

À Annie et Xavier.
Merci de m'offrir la plus belle vie au monde.
À ma famille.
Merci pour votre support inébranlable.

Remerciements

La rédaction d'une thèse doctorale est l'aboutissement d'un long processus au cours duquel le candidat se développe tant sur le plan professionnel que personnel. J'aimerais débuter par remercier ma partenaire de vie, Annie, de me soutenir, de me calmer, mais aussi de m'aider à catalyser mes idées les plus folles. Merci d'être une maman merveilleuse pour notre petit Xavier. Mon fils, merci d'être arrivé dans nos vies, tu nous remplis de bonheur et nous rappelles quotidiennement les valeurs importantes de la vie. J'aimerais aussi remercier mes parents, mon frère et ma sœur pour leur compréhension lorsque je ne suis pas toujours présent. Merci d'avoir respecté mes idées, mes projets et mes folles ambitions pendant toutes ces années.

Mes premiers remerciements professionnels vont à mon directeur de recherche, professeur François Desmeules. Merci d'avoir fait confiance à l'électron libre surexcité que je suis. La liberté que tu m'as offerte m'a permis de développer ma confiance professionnelle pour devenir, je l'espère, un professeur enthousiaste et dévoué pour ses étudiants. Merci pour ton pragmatisme et ton encadrement qui m'ont donné les outils nécessaires pour mener à terme ce projet. J'aimerais ensuite remercier mon codirecteur, Dr Pascal-André Vendittoli de m'avoir fait confiance et de m'avoir ouvert les portes de l'univers des orthopédistes. Merci pour ton regard intéressé sur mes travaux, qui m'oriente et m'amène à me dépasser. Merci à mes mentors, professeure Debbie Feldman et Dre. France Légaré. Merci pour votre temps et votre écoute. Merci d'être des guides pour ma jeune carrière, vous êtes une source d'inspiration et un modèle d'excellence pour la nouvelle génération de chercheurs.

Cette thèse n'aurait aussi tout simplement pas eu lieu sans l'apport inconditionnel des collaborateurs au projet. Merci à Dr Michel Fallaha, Dr Bruno Pelletier, Dr Pierre Frémont et Dr Sylvain Belzile d'avoir si gentiment accepté de partager votre expérience et expertise. Cela a été un honneur d'être votre étudiant et j'espère bientôt pouvoir être votre collègue et poursuivre nos projets et idées pour améliorer la santé des patients. Merci à tous les patients d'avoir accordé leur confiance à un jeune professionnel et d'avoir partagé leurs mille et une

histoires. Un merci spécial à Sylvie pour ton support au quotidien et sans qui l'organisation de ce projet aurait été impossible. Finalement, merci à Véronique, à Philippe et à Roula. Vous êtes les collègues de travail en qui j'ai le plus confiance.

Finalement, j'aimerais remercier tous les organismes qui ont accepté de financer un jeune étudiant chercheur passionné par son travail, mais n'ayant pas encore fait ses preuves. Les Fonds de Recherche en Santé du Québec (FRQS), Les Instituts de Recherche en Santé du Canada (IRSC), la Faculté des Études Supérieures (FESP), l'École de Réadaptation et le Réseau provincial de recherche en adaptation-réadaptation (REPAR). Lorsque vous acceptez de nous financer ou lorsque vous acceptez de donner des prix aux étudiants, vous contribuez à faire grandir notre confiance tant personnelle que professionnelle.

Chapitre 1: Introduction

Près d'un individu sur deux souffrira de douleur au genou au cours de sa vie, ce qui en fait une raison de consultation fréquente en médecine et en réadaptation [1-4]. Les pathologies communes au genou incluent celles d'origine traumatique telles les déchirures du ligament croisé antérieur (LCA) ou les déchirures méniscales [5, 6] ainsi que celles d'apparition progressive telles les déchirures méniscales dégénératives, l'ostéoarthrose du genou (OA) ou le syndrome fémoro-patellaire (SFP) [6]. Un diagnostic initial valide est proposé comme étant une composante essentielle d'une prise en charge efficiente pour ces pathologies [7-23].

Malheureusement, le diagnostic initial est souvent incertain ou erroné chez un nombre important d'individus, ce qui résulte en des trajectoires de soins non optimales [6, 8, 9, 18, 24]. Une première trajectoire non optimale concerne l'utilisation inappropriée des tests en imagerie médicale, principalement l'utilisation de l'imagerie par résonnance magnétique (IRM) [9]. L'utilisation de l'IRM a triplé au cours de la dernière décennie, les genoux représentant 44% de tous les motifs de consultation en IRM pour la catégorie « extrémités » et jusqu'à 50% des examens d'IRM pour les pathologies au genou seraient non justifiés [25]. En plus d'engendrer des coûts importants pour les systèmes de santé, ce test d'imagerie a aussi un fort potentiel de surdiagnostic [9, 25].

Cette incertitude diagnostique mène aussi fréquemment à la référence en spécialité, principalement en chirurgie orthopédique. En effet, on estime qu'entre 55% et 90% des références en chirurgie orthopédique ne seraient pas justifiées puisque ces patients ne nécessiteraient pas de traitements chirurgicaux, mais plutôt une approche conservatrice [26-28]. Dans ce contexte, le chirurgien orthopédique agit principalement à titre de consultant expert pour émettre le diagnostic approprié ce qui amène des délais de prise en charge aux niveaux d'un traitement conservateur médical ou en réadaptation [26, 27, 29]. En effet, en combinant les temps d'attente pour l'obtention d'un examen d'imagerie et subséquemment celui d'un rendez-vous avec un orthopédiste, les délais de prise en charge atteignent souvent

plus de 6 à 12 mois [25]. Cette situation expose les patients à une chronicisation de leur pathologie et une prolongation de leur invalidité.

Ces trajectoires sont souvent causées par une évaluation musculosquelettique initiale incomplète ou erronée en première ligne de soins [25]. Or, il est démontré que dans bien des cas la validité diagnostique de l'évaluation musculosquelettique obtenue par des experts pourrait être équivalente à celle obtenue par les tests d'imagerie pour les pathologies au genou [6]. Afin d'améliorer cet écart dans la capacité diagnostique des cliniciens experts et ceux notamment de première ligne, plusieurs auteurs ciblent le développement d'outils d'aide au diagnostic des pathologies au genou par l'identification de combinaisons valides d'éléments de l'anamnèse et de l'examen physique qui constituent l'évaluation musculosquelettique [6, 9, 25]. Cette thèse explorera donc la validité et la fidélité de l'évaluation musculosquelettique combinant l'anamnèse et l'examen physique afin de développer une série d'outils permettant d'orienter le diagnostic différentiel des pathologies communes au genou.

Rôle du candidat et structure de la thèse

Cette thèse débutera par une revue de la littérature (Chapitre 2) qui inclut deux revues systématiques sur la validité (Article 1) et la fidélité (Article 2) des tests de l'examen physique pour le diagnostic des pathologies au genou. Le Chapitre 3 présentera le protocole utilisé. Le Chapitre 4 présentera les résultats sous la forme de cinq articles scientifiques (Articles 3 à 7). Finalement, une discussion sera présentée au Chapitre 5.

Chapitre 2 : Revue de la littérature

La revue de la littérature dressera d'abord un portrait des données probantes sur la validité et la fidélité de l'évaluation musculosquelettique des pathologies au genou afin d'identifier les limites de la littérature qui mèneront à l'élaboration des objectifs de la thèse.

2.1 Épidémiologie des pathologies communes au genou

La douleur au genou serait présente chez plus de 20% de la population et représente le symptôme le plus commun pour plusieurs pathologies au genou [3, 25, 30]. Ces pathologies sont généralement classifiées entre celles d'origine traumatique ou celles d'apparition progressive [6]. Parmi les pathologies d'origine traumatique, les deux structures du genou les plus couramment lésées sont le ligament croisé antérieur (LCA) et les ménisques [6]. Le LCA est une structure stabilisatrice du genou qui protège contre les forces excessives en translation antérieure et en rotation interne du tibia [19, 31]. La déchirure du LCA correspond à 4% de toutes les atteintes au genou en première ligne [6]. Elle survient, dans environ 70% des cas, lors d'une activité sportive chez des jeunes adultes [19, 32-36]. Une autre structure couramment lésée lors d'un traumatisme au genou est les ménisques. Ceux-ci sont responsables de la distribution des forces de compression et de pivot, améliorant ainsi la stabilité du genou lors des mouvements [37]. Les déchirures méniscales traumatiques correspondent à approximativement 10% de toutes les pathologies au genou en première ligne et surviennent fréquemment dans un contexte de blessures sportives chez des jeunes adultes de façon concomitante à une déchirure du LCA [6, 38].

Parmi les pathologies d'apparition progressive, les ménisques peuvent aussi subir des lésions considérées d'usure, ou dégénératives, typiquement chez une population adulte plus âgée, et correspondent à environ 30% de toutes les pathologies au genou [6, 11, 37, 39-41]. Outre le fait que les déchirures méniscales dégénératives soient associées à une symptomatologie spécifique, certaines données probantes placent cette pathologie dans un continuum concomitant et progressif vers le développement de l'ostéoarthrose du genou (OA) [13, 42]. L'OA du genou correspond à environ 35% de toutes les pathologies et affecte 12,5% des

adultes âgés de plus de 45 ans [43-45]. Cette prévalence est d'ailleurs en augmentation en lien avec le vieillissement de la population et l'augmentation de la prévalence de l'obésité [43-45]. Contrairement aux autres pathologies du genou, l'OA, bien qu'elle puisse affecter préférentiellement les compartiments fémoro-tibiaux ou fémoro-patellaire, est considérée comme une pathologie qui affecte l'articulation dans son ensemble [46]. Une dernière pathologie d'apparition progressive est le SFP. Cette pathologie correspond à environ 20% de toutes les pathologies au genou et se retrouve principalement chez une population de jeunes adultes [47-50]. Les patients atteints rapportent de la douleur au niveau de la rotule lors d'activités qui augmentent les forces de compression à l'articulation fémoro-patellaire, telles la montée ou la descente d'escaliers, la marche ou la course à pied [47-50]. Des données de la littérature démontrent que plus de 50% des individus atteints développeraient des douleurs chroniques [7, 8, 47, 51-55].

Les pathologies au genou ont un impact significatif sur la qualité de vie des individus puisque la douleur engendre fréquemment des limitations fonctionnelles dans les activités de la vie quotidienne, la vie professionnelle ainsi que dans les activités physiques ou sportives [14-21, 56-59]. Dans ce contexte, un diagnostic initial valide basé sur une évaluation musculosquelettique combinant l'anamnèse et l'examen physique est requis afin de prendre en charge les individus atteints de façon efficiente et ciblée.

2.2 Validité et fidélité de l'évaluation musculosquelettique

L'objectif des revues systématiques est de déterminer les propriétés psychométriques des éléments de l'anamnèse et de l'examen physique pour le diagnostic des pathologies communes au genou. Ces propriétés incluent la validité et la fidélité des éléments constituant l'évaluation musculosquelettique. La validité réfère à la précision d'un élément pour identifier les cas et les patients n'ayant pas une pathologie d'intérêt [60]. La fidélité réfère à la reproductibilité des éléments par un ou plusieurs évaluateurs. Puisque l'étude de la validité et de la fidélité nécessite des méthodes distinctes, ces propriétés sont traitées séparément dans deux revues systématiques (Articles 1 et 2).

Article 1: Diagnostic validity of physical examination tests for common knee disorders: an overview of systematic reviews and meta-analysis.

Simon Décary [1, 2]
Philippe Ouellet [1, 2]
Pascal-André Vendittoli [2, 3]
Jean-Sébastien Roy [4, 5]
François Desmeules [1, 2*]

[1] School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada
[2] Orthopaedic Clinical Research Unit, Maisonneuve-Rosemont Hospital Research Center, University of Montreal Affiliated Research Center, Montreal, Quebec, Canada
[3] Department of Surgery, Maisonneuve-Rosemont Hospital, University of Montreal, Montreal, Quebec, Canada
[4] Department of Rehabilitation, Faculty of Medicine, Laval University, Quebec City, Quebec, Canada
[5] Centers for Interdisciplinary Research in Rehabilitation and Social Integration, Quebec City, Canada

Article publié dans la revue *Physical Therapy in Sport*. 2016 Aug 5. Doi: 10.1016/j.ptsp.2016.08.002.

En tant que premier auteur, j'ai réalisé le développement du protocole de cette revue systématique, la collecte des données, l'analyse des résultats ainsi que la rédaction du manuscrit. FD, PAV et JSR ont contribué au développement du protocole, à l'analyse des résultats et à la révision du manuscrit. PO a contribué à la collecte de données, à l'analyse, à la rédaction et à la révision du manuscrit.

ABSTRACT

Rationale:

More evidence on diagnostic validity of physical examination tests for knee disorders is needed to lower frequently used and costly imaging tests.

Objective:

To conduct a systematic review of systematic reviews (SR) and meta-analyses (MA) evaluating the diagnostic validity of physical examination tests for knee disorders.

Methods:

A structured literature search was conducted in five databases until January 2016. Methodological quality was assessed using the AMSTAR.

Results:

Seventeen reviews were included with mean AMSTAR score of 5.5±2.3. Based on six SR, only the Lachman test for ACL injuries is diagnostically valid when individually performed (Likelihood ratio (LR+): 10.2, LR-:0.2). Based on two SR, the *Ottawa Knee Rule* is a valid screening tool for knee fractures (LR-:0.05). Based on one SR, the EULAR criteria had a post-test probability of 99% for the diagnosis of knee osteoarthritis. Based on two SR, a complete physical examination performed by a trained health provider was found to be diagnostically valid for ACL, PCL and meniscal injuries as well as for cartilage lesions.

Conclusions:

When individually performed, common physical tests are rarely able to rule in or rule out a specific knee disorder, except the Lachman for ACL injuries. There is low-quality evidence concerning the validity of combining history elements and physical tests.

Introduction

Knee disorders and injuries are common reasons for consultation in primary care [4]. The lifetime prevalence of knee pain is 45%, and at least 31% of the affected individuals will consult a health care practitioner [1, 2]. Common knee disorders include traumatic injuries such as meniscal injuries [61], anterior cruciate ligament (ACL) injuries [5], fractures [62] or overuse or degenerative disorders like osteoarthritis [59], patellofemoral pain (PFP) [63] and tendinopathies [64]. Knee disorders often result in disabilities as well as in a decrease in health-related quality of life and may lead to workplace absenteeism [58, 65]. Efficient management of patients suffering from knee disorders is often lacking because the initial diagnosis is either erroneous or incomplete. Too often clinicians rely on medical imaging which in turn increase healthcare costs and may incur unnecessary delays in diagnosis and initiation of care [29]. Moreover, evidence suggests that medical imaging may be less valid than a complete physical examination in a large proportion of cases [6].

Clinicians rely on thorough patient history elements and physical examination tests to make a diagnosis where the patient's responses and findings are combined to make a valid diagnosis. This process remains the cornerstone for optimal, fast and efficient management of patients with musculoskeletal disorders [6]. However, evidence indicates that the ability to make a valid diagnosis for common knee disorders based on a complete physical examination in primary care remains a challenge [66]. Therefore, it is important to synthesize the evidence of the diagnostic validity of physical examination to better help clinicians in making a valid diagnosis.

In recent years, systematic reviews of primary diagnostic studies for all diagnoses and tests for hip [67, 68] and shoulder disorders [69] have been published. Also, many new systematic reviews (SR) and meta-analysis (MA) regarding the diagnosis of meniscal injuries [38, 70-72], ACL injuries [34, 73], knee osteoarthritis [46] and patellofemoral pain (PFP) [74] have been published. However, evidence for knee disorders has not been synthesized in one clinically useful format. Therefore, the objective of this study is to provide updated information to

clinicians working with individuals affected by knee disorders by systematically reviewing all the SR and MA reporting the diagnostic validity of physical examination tests performed individually or in combination with patient history elements for diagnosis of common knee disorders.

Methods

1.Study design

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used to guide the design of this systematic review [75, 76]. The tool is a list of 27 items, which assesses the adequate transparent reporting of the results of a systematic review or a meta-analysis.

2. Literature search and study identification

A literature search was performed in five bibliographical databases: Pubmed, Medline, CINAHL, Embase, the Cochrane Database of Systematic reviews and Pedro using relevant and MESH based keywords. The keywords were adapted to the various databases, as demonstrated in Appendix 1. Databases were searched from their date of inception to January 2016. References lists of included studies and important textbooks on musculoskeletal diagnosis were also investigated to verify the completeness of the current search [77, 78].

3. Data extraction and quality assessment

Study selection and data extraction

Each article, title and abstract were screened by two independent reviewers to determine eligibility. To be included, articles needed to 1- be a systematic review or a meta-analysis, 2- report on the diagnostic properties of at least one physical test for at least one knee disorder and 3- be written in English or French.

Data extraction of the selected SR/MA included: the study design (i.e: SR or MA), the physical tests evaluated to diagnose a specific knee disorder, the number of studies included in the SR, the number of studies pooled for a MA, and the total number of participants included in the review. The diagnostic properties of the clinical tests under study were extracted and included (when reported): sensitivity (Se), specificity (Sp), positive and negative predictive values (PPV/NPV), positive and negative likelihood ratios (LR+/-) and diagnostic odds ratio (DOR (LR+/LR-)).

Assessment of methodological quality of the included systematic reviews (AMSTAR)

The methodological quality of the included SR and MA was assessed with the AMSTAR tool [79]. The AMSTAR is a validated and highly reliable tool aimed at assessing the quality of systematic reviews [80, 81]. Each of the 11 methodological items are marked as “yes”, “no”, “cannot answer” (when unclear or insufficient information to answer) and “not applicable” [79]. Two raters independently assessed the methodological quality of each included review, then compared ratings and resolved any differences if present. In order to achieve consensus, a structured process was employed, where rechecking of the facts in the text was initially performed, followed by a discussion of the adherence to standards as well as the use of an independent third rater in case of persistent disagreement.

A total score out of 11 can be calculated adding the number of “yes” answers [80, 81]. To objectively synthesize and formulate recommendations, the methodological quality was used to establish the strength of evidence. A review with an AMSTAR score $\geq$ 8/11 was considered of high-quality, between 5 and 7/11 was considered of moderate-quality and $<$ 5 was considered of low-quality. No systematic reviews or meta-analysis were excluded based on methodological quality.

4. Data analyses

The mean methodological score of each review was calculated. Cohen’s kappa was used to calculate pre-consensus inter-rater agreement on individual methodological items of the

AMSTAR tool. Overall results and related 95% confidence intervals for each diagnostic property statistic extracted from the reviews were directly extracted when pooled results were presented in the original papers. If no pooled results for a given diagnostic property statistic were presented, the range of estimates reported in the SR was extracted. For each relevant diagnostic test, the range of point estimates for Se, Sp, LR+/- and DOR (where applicable) across all included SR/MA was reported for a qualitative assessment of the evidence.

Sensitivity (Se) and specificity (Sp) relate respectively to the proportion of true positive and true negative when a test is performed and do not inform on the probability of a patient having a disorder if a test is positive or negative, while PPV/NPV inform on that probability considering the prevalence of the studied sample [71, 82]. Because of these issues, the positive and negative likelihood ratio (calculated using both Se and Sp and representing the odds for a patient of having or not having a disorder) is advocated to guide clinical decision-making [70, 82]. Although no universal agreement exists, a test was considered valid if it reaches a LR+ of 5 or more and a LR- of 0.2 or less in order to formulate recommendations regarding its validity [82]. These values produce a moderate shift in post-test probability, signifying that a disorder is present if the test is positive and absent if the test is negative, and when a test performs up to these values, it is useful to make a valid diagnosis [74, 82].

Results

1. Overall description of included reviews

As shown in Figure 1, 6750 potential studies were initially identified, 6669 articles were excluded and 17 reviews were ultimately included. Table 1 presents the overall characteristics of the included reviews. Eleven reviews included a MA while the six others were SR without a MA. Eight reviews evaluated meniscal injuries, and the most common tests were the McMurray, the Apley's manoeuver and the joint line tenderness tests (Table 3). Six reviews assessed ACL injuries, and the most common tests evaluated were the Lachman, the anterior

drawer and the pivot shift tests (Table 4). One clinical prediction rule for the screening of knee fractures was evaluated in two reviews (Table 5). One clinical prediction rule and one set of diagnostic criteria to diagnose knee osteoarthritis were the focus of two reviews (Table 6). Combinations of tests for ACL and posterior cruciate ligament (PCL) tears, meniscal injuries and for cartilage defects were appraised in two reviews (Table 7).

2. Methodological quality of included reviews

The AMSTAR scores for the assessment of methodological quality of the SR/MA are presented in Table 2. AMSTAR score ranged from 1 to 8 (out of 11) with a mean of 5.5 ± 2.3 indicating a moderate methodological quality. Seven reviews reached a score of 7 or higher. More than ten out of seventeen reviews performed a comprehensive literature search (item 3), provided a list of studies (item 5), provided characteristics of the included studies (item 6) and assessed the scientific quality of the included studies (item 7). Between seven and ten out of seventeen reviews duplicated study selection and data extraction (item 2), used the status of publication as an inclusion criteria (item 4), used the scientific quality of the included studies in the conclusion (item 8), used appropriate methods to combine findings (item 9) and stated potential conflict of interest (item 11). However, only one review assessed the likelihood of publication bias (item 10) while no reviews provided an “a priori” design (item 1). The average inter-rater reliability for individual item was substantial (kappa=0.69). The majority of items reached substantial agreements (κ>0.6) (items 1, 4, 5, 6, 7, 8, 9, 11). All other items reached a moderate agreement (κ>0.4) (items 2, 3, 10). After discussion between the two raters, consensus was always achieved.

Figure 1: Bibliographic search flowchart

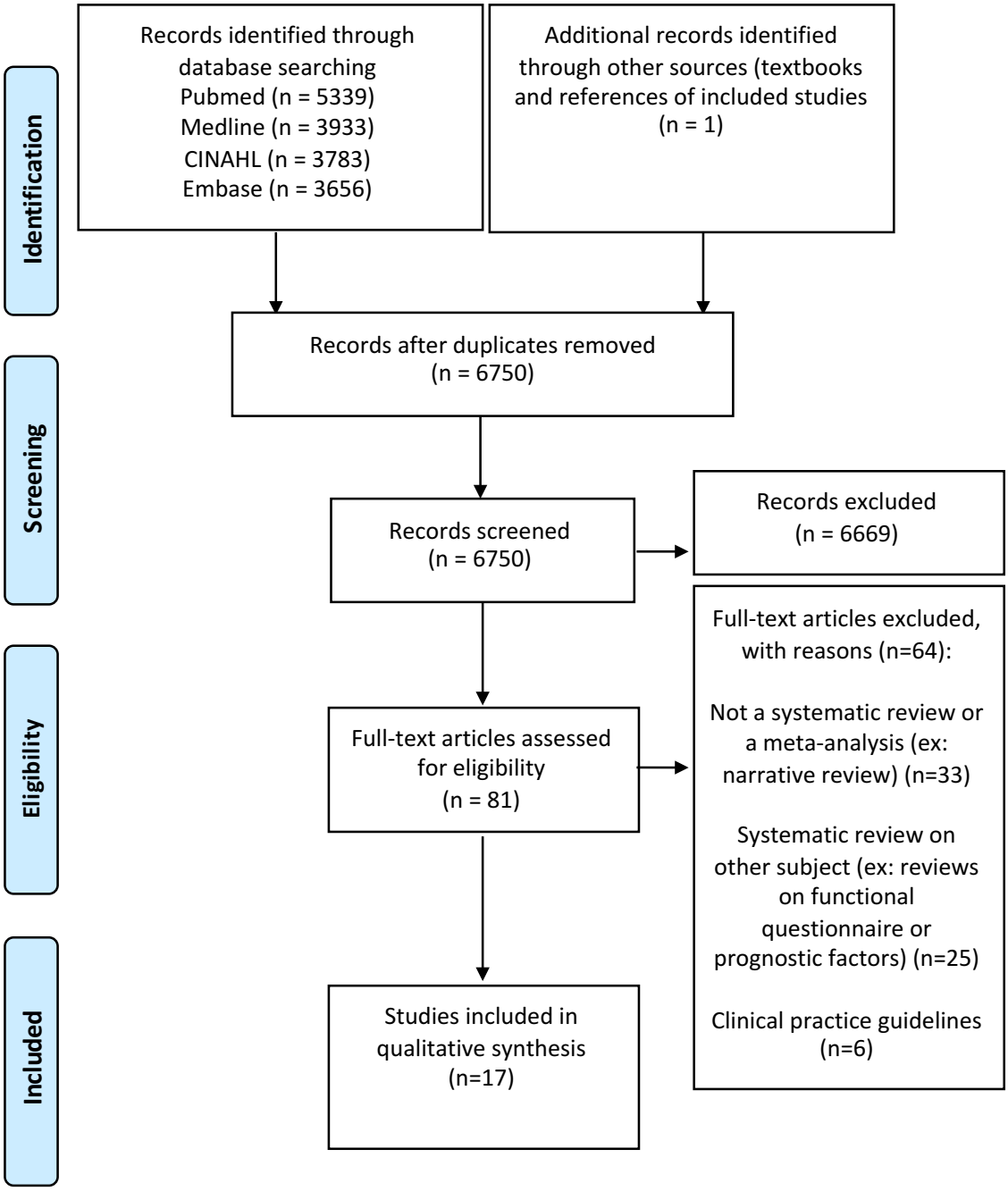


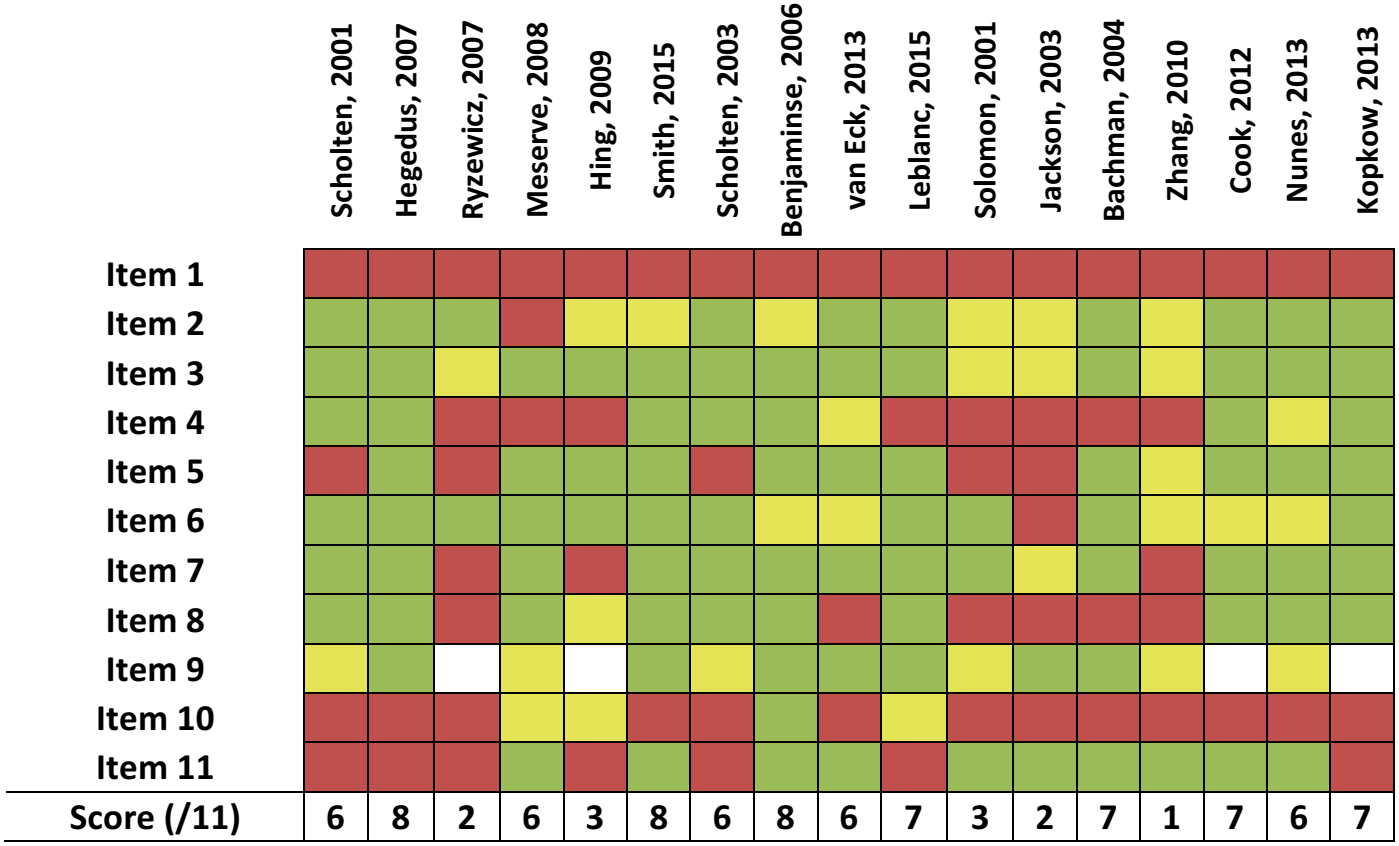
Table 1: Characteristics of the included reviews

Authors	Study Design	Disorders	Gold standard from primary diagnostic studies	Number of studies evaluated	Number of studies selected for pooled analysis	Number of participants included in pooled analyses	AMSTAR score
Scholten, 2001 [83]	Systematic review and Meta-analysis	Meniscal injuries	Arthroscopy or MRI	N=13	McMurray's n = 11 Joint line tenderness n = 10 Apley's Test n = 4	N=2231	6/11
Hegedus, 2007 [70]	Systematic review and meta-analysis	Meniscal injuries	Arthroscopy or MRI	N=18	McMurray's n = 14 Joint line tenderness n=14 Apley's test n = 7	N=2670	8/11
Ryzewicz, 2007 [72]	Systematic Review	Meniscal injuries	Arthroscopy or MRI	N=32	McMurray's n = 5 Apley's n = 2 Joint line tenderness n = 4	-	2/11
Meserve, 2008 [38]	Systematic review and Meta-analysis	Meniscal injuries	Arthroscopy (MRI accepted in combination)	N=11	McMurray's n = 8 Joint line tenderness n = 8 Apley's Test n = 3	McMurray's n = 1232 Joint line tenderness n = 1354	6/11
Hing, 2009 [71]	Systematic Review	Meniscal injuries	Arthroscopy or MRI	N=11	-	-	3/11
Smith, 2015 [84]	Systematic Review and Meta-analysis	Meniscal injuries	Arthroscopy or MRI	N=9	N=9	n=1234	8/11
Scholten, 2003 [85]	Systematic review	Anterior Cruciate Ligament injuries	Arthroscopy or MRI	N=17	Anterior Drawer Test n=6 Lachman Test n=6 Pivot Shift Test n=4	Variable	6/11
Benjaminse, 2006 [34]	Systematic review and meta-analysis	Anterior Cruciate Ligament injuries	Arthroscopy or MRI	N=28	N=28	Variable	8/11
Van Eck, 2013 [73]	Systematic review and meta-analysis	Anterior Cruciate Ligament injuries	Arthroscopy or MRI	N=20	Anterior Drawer Test n=17 Lachman Test n=13 Pivot Shift Test n=12	Anterior Drawer Test n=1579 Lachman Test n=934 Pivot Shift Test n=1192	6/11

Leblanc, 2015 [86]	Systematic review and meta-analysis	Anterior Cruciate Ligament injuries	Arthroscopy or MRI	N=8	Lachman n=5 Pivot Shift n=4	n=1196	7/11
Solomon, 2001 [87]	Systematic Review	Anterior Cruciate Ligament injuries and Meniscal Injuries	Arthroscopy or MRI	N=23	ACL n=15 Meniscal n=9	Variable	3/11
Jackson, 2003 [6]	Systematic review and meta-analysis	Acute Knee Disorders, Anterior Cruciate Ligament injuries and Meniscal Injuries	Arthroscopy/MRI/radiography/ clinical diagnosis	N=35	ACL n=11 Meniscal n=4	Variable	2/11
Bachmann, 2004 [88]	Systematic review and meta-analysis	Knee fracture with the Ottawa Knee Rule	Radiography or follow-up	N=11	N=6	N=4249	7/11
Zhang, 2010 [46]	Systematic Review, Meta-anaysis and Delphi consensus	Knee Osteoarthritis	Clinical features and radiographs	N=313	Variable	Variable	1/11
Cook, 2012 [74]	Systematic review	Patellofemoral pain	Arthroscopy or clinical or imaging (accepted by authors)	N=9	-	-	7/11
Nunes, 2013 [89]	Systematic review and meta-analysis	Patellofemoral pain	Unreported	N=5	N=2	N=145	6/11
Kopkow, 2013 [90]	Systematic review	Posterior Cruciate Ligament injuries	Arthroscopy or MRI	N=11	-	-	7/11

MRI: magnetic resonance imaging

Table 2: Assessment of the methodological quality of the included systematic reviews (AMSTAR)



Item 1: Was an "a priori" design provided?; **Item2:** Was there duplicate study selection and data extraction?; **Item 3:** Was a comprehensive literature search performed?; **Item 4:** Was the status of publication used as an inclusion criterion?; **Item 5:** Was a list of studies provided?; **Item 6:** Were the characteristics of the included studies provided?; **Item 7:** Was the scientific quality of the included studies assessed?; **Item 8:** Was the scientific quality of the included studies used in conclusion?; **Item 9:** Were the methods used to combine the findings appropriate?; **Item 10:** Was the likelihood of publication bias assessed?; **Item 11:** Was the conflict of interest stated? **Green = Yes; Red = No; Yellow = can't say/unclear; Blank = N/A**

3. Summary of findings

Meniscal Injuries

Table 3 presents the diagnostic properties of the physical tests used for the diagnosis of meniscal injuries. Eight SR/MA provided data on the diagnosis of meniscal injuries [6, 38, 70-72, 83, 84, 87] with AMSTAR scores ranging from 2 to 8 out of 11 with a mean of 4.8. Based on the data extracted from the review by *Hegedus et al.* [70], the highest quality review with the most studies included in the meta-analyses (8/11, n=18), the McMurray's test demonstrated the highest Se with a score of 70.5% (95% CI: 67.4-73.4%) [70]. The point estimates for Se varied across all included SR/MA (range: 52.0-70.5%). Also based on the data by *Hegedus et al.*, the joint line tenderness demonstrated the highest Sp with 77.4% (95% CI: 75.6-79.1%) [70]. Again, the point estimates for the Sp of this test varied across all included SR/MA (range: 29.0-83.0%). Based on the data extracted from *Smith et al*. [84], the most recent and highest quality review to provide data for the LR+/- (8/11, n=9), the joint line tenderness also demonstrated both the highest LR+ with 4.0 (95% CI: 2.1-7.5) and the lowest LR- with 0.23 (95% CI: 0.12-0.44). The point estimates of the joint line tenderness LR+/- varied across others included SR/MA for both LR+/- (range LR+: 0.9-4.0; range LR-: 0.23-1.1). The Thessaly's test diagnostic validity was also presented for the first time by the review of *Smith et al*. [84], with a reported LR+ of 5.6 (95% CI: 1.5-21.0) and LR- of 0.28 (95% CI: 0.11-0.71). Overall, authors' recommendations from all included SR/MA were that clinicians should not use these tests individually because of their poor diagnostic validity and advised combining the results of the tests even though no evidence was presented to support this approach [6, 38, 70-72, 83, 84, 87].

Anterior Cruciate Ligament Injuries

Table 4 presents the diagnostic properties of the physical examination tests used for the diagnosis of ACL injuries. Six SR/MA provided data on the diagnosis of ACL injuries [6, 34, 73, 85-87] with AMSTAR scores ranging from 2 to 8 out of 11 with a mean of 5.3. Based on the data extracted from *Benjaminse et al.* [34], the highest quality meta-analysis with the most studies included in analysis (8/11, n=28) to also provide 95% CI for LR+/-, the Lachman test

demonstrated the highest Se with 85.0% (95% CI: 83.0-87.0%). For Se, the point estimates were relatively similar across all included SR/MA (range: 81.0-89.0%). In terms of LR, the Lachman test also reached the highest LR+ and lowest LR- with 10.2 (95% CI: 4.6-22.7) and 0.20 (95% CI: 0.10-0.30) respectively [34]. The point estimates for the LR+ varied across the included SR/MA (range: 4.5-42.0), but was more consistent for the LR- (range: 0.10-0.22). The Pivot Shift test demonstrated the highest Sp with a score of 98.0% (95% CI: 96.0-99.0%) [34], but important variations in the point estimates across all included SR/MA was observed (range: 81.0-98.0%). Overall, authors' recommendations from all included SR/MA were that the Lachman test is of high diagnostic value both to rule in or out an ACL injury while a positive Pivot Shift test can be used to rule in an ACL injury [6, 34, 73, 85, 87]. One review concluded that the anterior drawer test may be used to rule in an ACL injury, but not rule out, and could be used instead of the Lachman in situations where the evaluator has less training performing the test [34].

Knee Fractures

Table 5 presents the diagnostic properties for the *Ottawa Knee Rule*, a clinical prediction rule used to exclude a knee fracture and avoid unnecessary radiograph of the knee at the emergency [88]. Two reviews providing data on this clinical prediction rule with AMSTAR score of 2 and 7 out of 11 [6, 88]. Both reviews reported similar diagnostic properties for the rule. *Bachmann et al.* calculated 98.5% (95% CI: 93.2-100%) for Se, 48.6% (95% CI: 43.4-51.0%) for Sp and 0.05 (95% CI: 0.02-0.23) for LR- [88]. Overall, the authors' recommendations from both included SR/MA are that this clinical prediction rule, with its low LR-, is considered useful to exclude a fracture if all the rule's criteria are negative; if this is not the case, the clinician cannot rule out a fracture and should order a knee radiograph (Table 5).

Knee osteoarthritis

Table 6 presents the diagnostic properties of one set of diagnostic criteria and one clinical prediction rule for the diagnosis of knee OA. Two reviews were found providing data on such tools with AMSTAR scores of 1 and 2 out of 11 [6, 46]. The clinical criteria from the *American*

College of Rheumatology for the diagnosis of knee OA include: age ≥ 50 years, stiffness ≤ 30 minutes, crepitus, bony tenderness, bony enlargement, and no palpable warmth [6]. If at least three criteria are present, the Se is 95.0% and Sp is 69.0%. If a fourth criterion is present, the Se drops to 84.0%, but the Sp increases to 89.0%. The study by *Zhang et al.* (EULAR rule) is presented as both a MA with the results of the validation of the rule in two cross-sectional studies [46]. For the clinical prediction rule, they reported a post-test probability of 99% for the diagnosis of OA (Kellgren Lawrence ≥2) if six criteria were present based on an estimated prevalence of 12.5% in adults aged ≥ 45 years [46]. The criteria include three symptoms: knee pain, limited morning stiffness, functional limitations, and three signs: knee crepitus, restricted knee range of motion and bony enlargements (table 6). In both reviews, the authors concluded that a clinical diagnosis of knee OA can be done with the clinical criteria, but knee radiography remains necessary to assess the radiological grading of OA.

Patellofemoral pain

Two reviews evaluated the diagnostic validity of 25 tests for PFP (AMSTAR score: 7 and 6/11) [74, 89]. Both reviews used the same five studies and one included four more for a total of nine studies [74, 89]. Overall, five tests had a LR+≥5: the active instability test (LR+: 249.0), pain during stair climbing (LR+: 11.6), Clarke's sign (LR+: 7.4), pain during prolonged sitting (LR+: 7.5), and patellar tilt (LR+: 5.4, 95%CI: 1.4-20.8) [74, 89]. They also reported that pain during squatting demonstrated a LR-=0.20 (95%CI: 0.1-0.4) [74, 89]. Combining tests had a mitigated effect on improving the diagnosis of PFP [74, 89]. However, they acknowledged that the primary diagnostic studies included in their SR reported heterogeneous results with an overall high risk of bias [74, 89]. The authors of the included reviews proposed to view PFP as a diagnosis of exclusion [74, 89]. No individual tests can be recommended at this time to diagnose or exclude a PFP.

Posterior cruciate ligament injuries

The SR by *Kopkow et al.* evaluated the diagnostic validity of 11 tests for PCL injuries (AMSTAR score: 7/11) [90]. In their review including eleven studies, they reported that the posterior

drawer test was the most frequently studied test with a Se ranging from 22% to 100% [90]. They reported that the included primary study in their SR with the lowest risk of bias (i.e: moderate) showed a LR+ of 50.1 (95% CI: 7.1-351.7) and a LR- of 0.11 (95% CI: 0.03-0.40) for this test [90]. The authors also reported that the quadriceps active test appeared to be the most specific test with a Sp ranging from 96% to 100% based on two included studies from their SR [90]. *Kopkow et al.* concluded that at this time, evidence was insufficient to recommend any physical tests to diagnose or exclude a PCL injury [90].

History taking and physical examination for the diagnosis of common knee disorders

Table 7 presents the diagnostic properties extracted from reviews on the complete physical examination (including a thorough history and physical tests) for the diagnosis of various knee disorders [6]. We found two reviews providing data on such evaluation for four knee disorders with AMSTAR scores of 2 and 3 out of 11 [6, 87]. For ACL injuries, *Jackson et al*. [6] reported point estimates for LR+ and LR- of 15.0 (95% CI: 5.1-23.0) and 0.27 (95% CI: 0.12-0.42) respectively. For PCL injuries, *Jackson et al*. [6] also reported point estimates for LR+ and LR- of 16.2 (95% CI: 5.2-25.0) and 0.20 (95% CI: 0.13-0.49) respectively. Therefore, a complete physical examination appears valid to diagnose an ACL injury and diagnose or exclude a PCL injury, although the definition of what constitutes a complete physical examination was not provided. Likewise, for the diagnosis of cartilage lesions, the complete physical examination may be considered valid with LR+ of 13.0 (95%CI: 2.7-24.0) and a LR- of 0.51 (95% CI: 0.40-0.62) [6]. For the diagnosis of meniscal injuries, *Solomon et al.* reported in their review a LR+: 2.7 (95% CI: 1.4-5.1) and a LR-: 0.40 (95% CI: 0.20-0.70) [87] and concluded that a complete physical examination is not valid to diagnose or exclude a meniscal injury. However, the review by *Jackson et al*. concluded that a complete physical examination might be valid to identify the presence of lateral meniscal injury or to exclude a medial or lateral meniscus injury [6]. Overall, studies that reviewed the complete physical examination for various knee disorders concluded that a complete physical examination is probably superior to individual tests but it remains unclear how well this approach may perform [6, 87]. Based on this evidence, a complete physical examination may be diagnostically superior to individual tests but further research is needed.

Table 3: Description of diagnostic properties for selected tests for meniscal injuries based on results from included reviews

Tests	Properties	Solomon, 2001	Scholten, 2001	Jackson, 2003	Hegedus, 2007	Ryzewicz, 2007	Meserve, 2008	Hing, 2009	Smith, 2015
McMurray's test	Se	53.0±15.0%[§]	10.0-66.0%[†]	52.0% [95% CI: 35.0-68.0%]	70.5% [95% CI: 67.4-73.4%]	16.0-67.0%[†]	55.0% [95% CI: 50.0-60.0%]	16.0-88.0%[†]	61.0% [95% CI: 45.0-74.0%]
	Sp	59.0 ±36.0%[§]	57.0-98.0%[†]	97.0% [95% CI: 87.0-99.0%]	71.1% [95% CI: 69.3-72.9%]	69.0-98.0%[†]	77.0% [95% CI: 62.0-87.0%]	20.0-98.0%[†]	84.0% [95% CI: 69.0-92.0%]
	LR+	1.3 [95% CI: 0.90-1.7]	1.5-9.5[†]	-	-	-	2.4 [95% CI: 1.3-4.6]	0.82-8.86[†]	3.2 [95% CI: 1.7-5.9]
	LR-	0.80 [95% CI: 0.60-1.1]	0.40-0.90[†]	-	-	-	0.58 [95% CI: 0.46-0.81]	0.24-1.45[†]	0.52 [95% CI: 0.34-0.81]
	DOR	-	-	-	4.5 [95% CI: 3.7-5.4]	-	3.99 [95% CI: 1.04-15.31]	-	-
Apley's Maneuver	Se	-	-	-	60.7% [95% CI: 55.7-65.5%]	16.0-41.0%[†]	22.0% [95% CI: 17.0-28.0%]	-	-
	Sp	-	-	-	70.2% [95% CI: 68-72.4%]	80.0-93.0%[†]	88.0% [95% CI: 72.0-96.0%]	-	-
	DOR	-	-	-	3.4 [95% CI: 2.6-4.4]	-	2.20 [95% CI: 0.27-17.66]	-	-
Joint line tenderness	Se	79.0±4.0%[§]	28.0-95.0%[†]	76.0% [95% CI: 65.0-87.0%]	63.3% [95% CI: 60.9-65.7%]	67.0-97.0%[†]	76.0% [95% CI: 73.0-80.0%]	-	83.0% [95% CI: 73.0-90.0%]
	Sp	15.0 ±22.0%[§]	13.0-95.0%[†]	29.0% [95% CI: 10.0-46.0%]	77.4% [95% CI: 75.6-79.1%]	29.4-87.0%[†]	77.0% [95% CI: 64.0-87.0%]	-	83% [95% CI: 61.0-94.0%]
	LR+	0.90 [95% CI: 0.80-1.0]	0.80-14.9[†]	-	-	-	3.3 [95% CI: 1.6-6.2]	-	4.0 [95% CI: 2.1-7.5]
	LR-	1.1 [95% CI: 1.0-1.3]	0.20-2.1[†]	-	-	-	0.31 [95% CI: 0.23-0.42]	-	0.23 [95% CI: 0.12-0.44]
	DOR	-	-	-	4.5 [95% CI: 3.8-5.4]	-	10.98 [95% CI: 3.02-39.95]	-	-
Number of primary diagnostic studies included in analysis		**N=9**	**N=13**	**N=4**	**N=18**	**N=11**	**N=11**	**N=11**	**N=9**
AMSTAR Score		**3/11**	**6/11**	**2/11**	**8/11**	**2/11**	**6/11**	**3/11**	**8/11**

Se: sensitivity, Sp: specificity, LR: likelihood ratio, DOR: diagnostic odds ratio. § Data presented as mean ± SD (standard deviation), calculated by authors. † Indicates that the authors could not pool the data and did not calculate a mean value with standard deviation. Therefore, we presented the range of values based on the data presented in the article.

Table 4: Description of diagnostic properties for physical examination tests for Anterior Cruciate Ligament (ACL) injuries based on results from included reviews

Tests	Properties	Solomon, 2001	Jackson, 2003	Scholten, 2003	Benjaminse, 2006	van Eck, 2013[¥]	Leblanc, 2015
Lachman Test	Se	84.0±15.0%[§]	87.0% [95% CI: 76.0-98.0%]	86.0% [95% CI: 76.0-92.0%]	85.0% [95% CI: 83.0-87.0%]	81.0%[¥]	89.0% [95% CI: 76.0-98.0%]
	Sp	100%	93.0% [95% CI: 89.0-96.0%]	91.0% [95% CI: 79.0-96.0 %]	94.0% [95% CI: 92.0-95.0 %]	81.0%[¥]	-
	LR+	42.0 [95% CI: 2.7-651.0]	-	2.0-102.1[†]	10.2 [95% CI: 4.6-22.7]	4.5[¥]	-
	LR-	0.10 [95% CI: 0.00-0.40]	-	0.10-0.40[†]	0.20 [95% CI: 0.10-0.30]	0.22[¥]	-
	DOR	-	-	-	70.0 [95% CI: 23.0-206.0]	-	-
Pivot Shift Test	Se	38.0 ±28.0%[§]	61.0% [95% CI: 40.0-82.0%]	18.0-48.0 %[†]	24.0% [95% CI: 21.0-27.0%]	28.0%[¥]	79% [95% CI: 63.0-91.0%]
	Sp	-	97.0% [95% CI: 93.0-99.0%]	97.0-99.0 %[†]	98.0% [95% CI: 96.0-99.0%]	81.0%[¥]	-
	LR+	-	-	8.2-26.9[†]	8.5 [95% CI: 4.7-15.5]	5.35[¥]	-
	LR-	-	-	0.50-0.80[†]	0.90 [95% CI: 0.80-1.0]	0.84[¥]	-
	DOR	-	-	-	12.0 [95% CI: 5.0-31.0]	-	-

Anterior Drawer test	Se	62.0±23.0%[§]	48.0% [95% CI: 38.0-59.0%]	62.0% [95% CI: 42.0-78.0 %]	55.0% [95% CI: 52.0-58.0%]	38.0%[¥]	-
	Sp	67.0 ±42.0%[§]	87.0% [95% CI: 83.0-91.0 %]	88.0% [95% CI: 83.0-92.0%]	92.0% [95% CI: 90.0-94.0 %]	81.0%[¥]	-
	LR+	3.8 [95% CI: 0.70-22.0]	-	1.7-87.9[†]	7.3 [95% CI: 3.5-15.2]	4.52[¥]	-
	LR-	0.30 [95% CI: 0.05-1.50]	-	0.10-0.80[†]	0.50 [95% CI: 0.40-0.60]	0.67[¥]	-
	DOR	-	-	-	21.0 [95% CI: 8.0-53.0]	-	-
Number of primary diagnostic studies included in analysis		**N=15**	**N=11**	**N=17**	**N=28**	**N=20**	**N=8**
AMSTAR Score		**3/11**	**2/11**	**6/11**	**8/11**	**6/11**	**7/11**

Se: sensitivity, Sp: specificity, LR: likelihood ratio, DOR: diagnostic odds ratio. § Data presented as mean ± SD (standard deviation), calculated by authors. † Authors did not pool data. Ranges of values are therefore presented. ¥ 95% CI not presented in review.

Table 5: Description of diagnostic properties for selected tests for knee fractures based on results from included reviews

Test	Properties	Jackson, 2003	Bachmann, 2004
Ottawa Knee Rule†	Se	100.0% [95% CI: 94.0-100.0%]	98.5% [95% CI: 93.2-100%]
	Sp	49.0% [95% CI: 46.0-52.0%]	48.6% [95% CI: 43.4-51.0%]
	LR +	1.96 [95% CI: 1.92-1.99]	-
	LR -	0.11 [95% CI: 0.06-0.18]	0.05 [95% CI: 0.02-0.23]
Number of primary diagnostic studies included in analysis		**N=2**	**N=6**
AMSTAR Score		**2/11**	**7/11**

Se: sensitivity, Sp: specificity, LR: likelihood ratio.

† Ottawa Knee Rule criteria: A knee radiograph examination is required only for patients who have at least ONE of the followings: 1) aged 55 years or older, 2) isolated tenderness of patella (no bone tenderness of knee other than patella), 3) tenderness at the head of fibula, 4) inability to flex knee to 90 degrees, or 5) inability to bear weight both immediately and in the emergency department for 4 steps (unable to transfer weight twice onto each lower limb regardless of limping).

Table 6: Description of diagnostic properties for selected tests for osteoarthritis based on results from included reviews

Test	Properties	Jackson, 2003	Zhang, 2010
American College of Rhumatology **Criteria**[†]	Se (at least 3 criteria)	95.0%[§]	-
	Sp (at least 3 criteria)	69.0%[§]	-
	Se (4 criteria present)	84.0%[§]	-
	Sp (4 criteria present)	89.0%[§]	-
EULAR **Rule**[‡]	Post test probability	-	99.0%[§]
Number of primary diagnostic studies included in analysis		**Unclear**	**Unclear**
AMSTAR score		**2/11**	**1/11**

Se: sensitivity, Sp: specificity. § Data presented directly from authors.

† The clinical criteria from the ACR for the diagnosis of osteoarthritis include: age older than 50 years, stiffness for less than 30 minutes, crepitus, bony tenderness, bony enlargement, and no palpable warmth. Based on a prevalence of 34% in primary care, the presence of at least 3 criteria increases the probability of having osteoarthritis to 62%.

‡ EULAR's clinical prediction rule criteria for the diagnosis of osteoarthritis include: 3 symptoms: knee pain, limited morning stiffness, functional limitations, and 3 signs: crepitus, restricted movement and bony enlargement. Based on a background prevalence of 12.5% in adults age more or equal to 45 years, the post-test probability of having radiographic knee OA (KL≥2) increase with the number of criteria present to 99% when all six criteria are present.

Table 7: Description of diagnostic properties of the history taking and physical examination for selected knee disorders based on results from included reviews

Disorders	Properties	Solomon, 2001	Jackson, 2003
Anterior Cruciate Ligament†	Se	82.0%§	74.0% [95% CI: 60.0-88.0%]
	Sp	94.0%§	95.0% [95% CI: 92.0-98.0 %]
	LR+	25.0 [95% CI: 2.1-306.0]	15.0 [95% CI: 5.1-23.0]
	LR-	0.04 [95% CI: 0.01-0.48]	0.27 [95% CI: 0.12-0.42]
Posterior Cruciate Ligament†	Se	91.0%§	81.0 [95% CI: 63.0-98.0%]
	Sp	98.0%§	95.0 [95% CI: 81.0%-100.0%]
	LR+	21.0 [95% CI: 2.1-205.0]	16.2 [95% CI: 5.2-25.0]
	LR-	0.05 [95% CI: 0.01-0.50]	0.20 [95% CI: 0.13-0.49]
Cartilage lesions†	Se	-	0.51 [95% CI: 0.37-0.65]
	Sp	-	0.96 [95% CI: 0.91-1.0]
	LR+	-	13.0 [95% CI: 2.7-24.0]
	LR-	-	0.51 [95% CI: 0.40-0.62]
Any meniscus†	Se	77.0 ±7.0%§	-
	Sp	91.0 ±3.0%§	-
	LR+	2.7 [95% CI: 1.4-5.1]	-
	LR-	0.40 [95% CI: 0.20-0.70]	-
Medial meniscus†	Se	-	86.0% [95% CI: 79.0-92.0%]
	Sp	-	72.0% [95% CI: 61.0-83.0%]
	LR+	-	3.1 [95% CI: 0.54-5.7]
	LR-	-	0.19 [95% CI: 0.11-0.77]
Lateral meniscus†	Se	-	88.0% [95% CI: 77.0-99.0%]
	Sp	-	92.0% [95% CI: 89.0-95.0%]
	LR+	-	11.0 [95% CI: 1.8-20.2]
	LR-	-	0.13 [95% CI: 0.00-0.25]
Number of primary diagnostic studies included in analysis		**N=3**	**N=35**
AMSTAR Score		**N=3/11**	**2/11**

Se: sensitivity, Sp: specificity, LR: likelihood ratio. § Data presented as mean ± SD (standard deviation), calculated or presented directly from the authors. † The examination tests used in the primary diagnostic studies on the complete physical examination varies greatly and are often partially or not reported [6].

Discussion

The objective of this study was to systematically review all SR and MA evaluating the diagnostic validity of physical examination tests for common knee disorders performed individually or in combination. This review found a mean AMSTAR score of 5.4±2.4 (95% CI: 4.2-6.6). *Gagnier et al*. recently evaluated the methodological quality of seventy-six orthopaedic SR and MA using the AMSTAR tool and the analysis of this review mirrors their findings (AMSTAR score: 5.4 vs 5.9) [91, 92]. Therefore, in accordance with other published literature, the mean AMSTAR score is at best moderate for the included SR and MA. However, given the wide range of methodological scores, key reviews of high methodological quality were found and the estimate from these reviews may be used to guide clinical decision-making for the diagnosis of common knee disorders. Moreover, conclusions across SR and MA were most often consistent, independently of the year of publication and number of included primary diagnostic studies, reinforcing to a certain extent the strength of the evidence.

Eight SR/MA provided data for the diagnosis of meniscal injuries [6, 38, 70-72, 83, 84, 87], two SR provided data for the diagnosis of PFP [74] and one for the diagnosis of PCL injuries [90]. Overall, the authors' conclusions regarding the evidence supporting the use of physical tests performed individually to diagnose one of these common knee disorders is inconclusive. It appears that none of the physical tests studied (McMurray's, Apley's manoeuver, joint line tenderness and Thessaly test) are able to make a valid diagnosis of a meniscal injury when performed individually, even in high-quality reviews. Likewise, based on moderate quality reviews when performed individually, relevant physical tests are unable to diagnose or exclude a PFP or a PCL injury.

Six SR/MA provided data for the diagnosis of ACL injuries [6, 34, 73, 85-87]. The evidence suggests that out of the three tests that were investigated (Lachman, anterior drawer test and the pivot shift), the Lachman appears to be valid when individually performed to diagnose or exclude an ACL injury and the Pivot Shift may be used to diagnose an ACL injury because of its high specificity. Of note, one SR included in the present analysis concluded that the use of the

pivot shift for patients under anaesthesia reached higher specificity than the Lachman [73]. Clinically, the evidence indicates that clinicians can rely on the Lachman test to rule in or out an ACL injury. It should be noted that some authors expressed concerns about the use of the anterior drawer test in clinics because of its lower validity compared to the Lachman test even if this test reaches an adequate validity [6].

Two reviews provided data for the screening of knee fracture using the *Ottawa Knee Rule* [6, 88]. The evidence indicates that the *Ottawa Knee Rule* is valid to exclude knee fractures [93] and has been used extensively in clinical practice since its original development two decades ago. The implementation of this tool in clinical practice, namely in emergency departments, has been shown to decrease radiograph use, to lower direct healthcare costs and to diminish time spent by patients in the emergency department, without compromising patients' safety and quality of care. [94, 95]. Clinicians can therefore rely on the rule to avoid unnecessary radiographic evaluation in this setting.

Two reviews provided data for the diagnosis of OA using diagnostic criteria and a clinical prediction rule [6, 46]. The *American College of Rheumatology* has issued criteria for the diagnosis of knee osteoarthritis without the use of radiographs, but only the sensitivity and specificity are reported, a fact that limits clinical interpretation [6]. Moreover, initial development of the criteria mainly classify patients from having OA or rheumatoid arthritis [96] and there remains conjecture in the literature about the use of the criteria in primary care [97]. The recent EULAR prediction rule provides LR+ and LR- and therefore enables the calculation of post-test probability [46]. The post-test probability of this rule is very high if all six symptoms and signs are present. However, to our knowledge, the diagnostic capabilities of this combination when fewer than six criteria are positive has not been presented, which limits its usefulness for clinicians. Clinically, the evidence indicates that clinicians may use the proposed criteria to diagnose clinical knee OA without the use of radiographs as proposed by both medical associations. However, radiographs remain necessary for radiological grading as

can be used to plan a consultation with an orthopaedic surgeon. More evidence is needed to better define the clinical diagnosis of knee OA in the general population [6, 97].

The authors from all the included SR/MA argued that the majority of primary diagnostic studies were of low to medium methodological quality. Commonly cited poor methodological aspects included a low number of patients and/or of cases, a retrospective study design, the lack of blinding, high risk of spectrum bias and inadequate description of index tests and reference standards [6, 70]. These biases have been shown to inflate the diagnostic validity of the index test under study [98]. Moreover, only two reviews reported the reliability of clinical tests [46, 74]. Reliability is an important and commonly overlooked diagnostic quality that ultimately influences the validity of the test and its usefulness for clinicians [27]. More precise discussion on this subject is beyond the scope of this review, but evaluating the available evidence on the reliability of knee physical tests is warranted.

The evidence presented above demonstrates that the diagnosis of common knee disorders cannot be made by relying exclusively on one physical test, with exception of ACL injuries. The presented evidence can inform clinicians of the validity of physical tests that may be useful to diagnose or exclude a disorder. Interestingly, our review presents evidence, although limited and in low quality reviews, that a complete physical examination may be diagnostically superior to individual tests. However, in the present review, the definition of the history elements and physical tests included in a complete physical examination was notably lacking. Indeed, we do know that clinicians rely on combinations of history elements and physical tests to make a diagnosis when patients present with a knee problem. It has been reported that clinicians with extensive training with musculoskeletal disorders are able to diagnose knee disorders with high confidence using a complete physical examination [24, 27, 99]. Our review also shows that when clinical prediction rules are developed they may become useful, valid tools for the diagnosis of common knee disorders such as OA and fractures. Of interest, several clinical prediction rules have been developed in recent publications not presented in the included reviews [100-104] and demonstrated that the combination of specific history

elements and physical tests have improved the diagnostic validity compared to individual tests for specific knee disorders. We believe that more methodologically sound diagnostic studies are needed and should focus on the evaluation of the clinical prediction rules that incorporate well-defined and specific history elements and physical tests.

Strengths and limitations of the present review

This SR has the advantage to provide an overview of the diagnostic validity for common knee disorders. Every included SR/MA was methodologically assessed using the AMSTAR. A limitation of this SR was the difficulty to combine the point estimates of SR and MA. We chose to present the range of the point estimates across included SR/MA to obtain an overall qualitative appraisal of the evidence. Moreover, the analysis of the range of the point estimates across the included SR/MA provided an opportunity to show the frequent wide range and heterogeneity of the evidence.

Future SRs could be improved by following the established reporting guidelines [75, 76]. In particular, presentation of a priori design, meta-analysis, assessment of publication bias and presentation of conflict of interests were rarely present. Improving the quality of primary diagnostic studies would enable the meta-analysis and more precise estimate of the true diagnostic value of certain tests. Lastly, the aim of this SR was to synthesize all SR/MA and therefore some new and recent primary diagnostic studies evaluating new physical examination tests or clinical prediction rules were not part of our analyses. Nonetheless, we believe that our review provides useful recommendations to clinicians and researchers working with patients suffering from common knee disorders.

Conclusions

Many SR and MA are of low to moderate quality, which warrants caution from clinicians when reading these reviews for clinical guidance. However, a few methodologically sound reviews provide high-quality evidence for ACL and meniscal injuries. The evidence suggests that clinicians may diagnose or exclude an ACL injury with the Lachman test, exclude a knee fracture using the *Ottawa Knee Rule* and make a diagnosis of knee OA based on the results of the *American College of Rhumatology and EULAR* rules. For other knee disorders (meniscal injury, PFP, PCL injury and others), the available evidence does not demonstrate that tests used individually are diagnostically valid. Globally, very few clinical tests, when performed individually, can diagnose or exclude a knee disorder. Based on limited and low quality evidence, the combination of history elements and physical tests may be more diagnostically valid. In the context of increasing healthcare costs, the development of clinical prediction rules comprising history elements and physical examination tests from methodologically sound diagnostic studies are necessary to further advance the diagnosis of knee disorders.

Article 2: Reliability of physical examination tests for the diagnosis of knee disorders: Evidence from a systematic review.

Simon Décary [1, 2*]
Philippe Ouellet [1, 2]
Pascal-André Vendittoli [2, 3]
François Desmeules [1, 2]

[1]School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada
[2]Maisonneuve-Rosemont Hospital Research Center, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada
[3]Department of Surgery, Maisonneuve-Rosemont Hospital, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, University of Montreal Affiliated Hospital, Montreal, Quebec, Canada

Article publié dans la revue *Manual Therapy*. 2016 Sep 23;26:172-182. doi: 10.1016/j.math.2016.09.007.

En tant que premier auteur, j'ai réalisé le développement du protocole de cette revue systématique, la collecte des données, l'analyse des résultats et la rédaction du manuscrit. FD et PAV ont contribué au développement du protocole, à l'analyse des résultats et à la révision du manuscrit. PO a contribué à la collecte de données, à l'analyse, à la rédaction et à la révision du manuscrit.

ABSTRACT

Clinicians often rely on physical examination tests to guide them in the diagnostic process of knee disorders. However, reliability of these tests is often overlooked and may influence the consistency of results and overall diagnostic validity. Therefore, the objective of this study was to systematically review evidence on the reliability of physical examination tests for the diagnosis of knee disorders. A structured literature search was conducted in databases up to January 2016. Included studies needed to report reliability measures of at least one physical test for any knee disorder. Methodological quality was evaluated using the QAREL checklist. A qualitative synthesis of the evidence was performed. Thirty-three studies were included with a mean QAREL score of 5.5±0.5. Based on low to moderate quality evidence, the Thessaly test for meniscal injuries reached moderate inter-rater reliability (κ=0.54). Based on moderate to excellent quality evidence, the Lachman for anterior cruciate ligament injuries reached moderate to excellent inter-rater reliability (κ=0.42 to 0.81). Based on low to moderate quality evidence, the Tibiofemoral Crepitus, Joint Line and Patellofemoral Pain/Tenderness, Bony Enlargement and Joint Pain on Movement tests for knee osteoarthritis reached fair to excellent inter-rater reliability (κ=0.29 to 0.93). Based on low to moderate quality evidence, the Lateral Glide, Lateral Tilt, Lateral Pull and Quality of Movement tests for patellofemoral pain reached moderate to good inter-rater reliability (κ=0.49 to 0.73). Many physical tests appear to reach good inter-rater reliability, but this is based on low-quality and conflicting evidence. High-quality research is required to evaluate the reliability of knee physical examination tests.

Introduction

Knee disorders prevalence is estimated at more than 50% in a lifetime and represent a common reason for medical consultation [1]. Clinicians mostly rely on physical examination tests – such as pain provoking, laxity, range of motion or palpation tests – to diagnose knee disorders. These tests are commonly used alone or in combination and in conjunction with patients' history and symptoms. Ample evidence has been published on the diagnostic validity of physical examination tests for common knee disorders such as meniscal injuries [6, 38, 70], anterior cruciate ligament injuries (ACL) [34, 73], osteoarthritis (OA) [46] or patellofemoral pain (PFP) [74]. However, an important and often overlooked factor that may influence the performance of diagnostic testing is the reliability of these tests [105-107]. Intra and inter-rater reliability needs also to be thoroughly assessed and taken into account when reviewing the performance of a diagnostic test as it informs on the consistency of the test [107-110].

Reliability includes two key concepts: intra-rater reliability or test-retest reliability relates to the agreement of the outcome of a physical examination test when done on the same patient, by the same evaluator in a time span where the patient condition is stable [111]. Inter-rater reliability relates to the agreement of the outcome of a physical examination test when executed by two or more different evaluators, again in a time span where the patient condition is stable [111]. Commonly cited factors that influence reliability include: experience of evaluators, variability in test execution, definition of the response of results' criteria, as well as the spectrum of patients evaluated [105].

Only a few authors specifically reviewed reliability of physical tests for the diagnosis of musculoskeletal disorders. *May et al*. demonstrated that physical examination tests for various shoulder disorders were generally unreliable [108]. For knee disorders, currently reviewed data covers only a fraction of all knee disorders and tests [6]. *Lange et al*. reviewed the reliability of ACL injury tests and found that the Lachman test reached the highest intra and inter-reliability, although included studies presented inconsistent results or were

generally of low methodological quality [105]. *Smith et al.* reviewed reliability of the medio-lateral patellar positioning test commonly used in the assessment of PFP and found good intra-rater and variable inter-rater reliability [112]. Several recent textbooks on the diagnosis of musculoskeletal disorders have also included synthesized results on the reliability of knee tests [77, 78, 113]. However, the process may not be systematic and may not present a complete picture based on all relevant studies to expose variability and impact of methodological bias on reliability.

Given the significance of reliability in the decision-making process to obtain an accurate diagnosis, the objective of this study is to systematically review and update evidence on the reliability of physical examination tests for the diagnosis of common knee disorders.

Methods

Literature search and study identification

The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines were used for the design of this review [75, 114]. A literature search was performed in four bibliographical databases: Pubmed, Medline, CINAHL and Embase, using relevant keywords (Appendix 2). The search strategy was elaborated using the guidance of a health science librarian. Databases were searched from their date of inception to January 2016. References lists of included studies and important textbooks on musculoskeletal diagnosis were also searched for possible additional studies [77, 78, 113].

Inclusion and Exclusion Criteria

Eligible studies needed to: 1- assess intra or inter-rater reliability of any physical examination test for the diagnosis of any knee disorder, 2- Include reliability statistics such as raw agreement, Cohen's kappa or intraclass correlation coefficient (ICC), 3- be written in English or in French. Studies were excluded if: 1- they reported reliability data on a healthy population or used cadavers, 2- reported the reliability of performance measures without any diagnostic value, self-reported questionnaires or instruments.

Study Selection

Two independent reviewers (S.D. and P.O.) selected eligible studies on the basis of titles and abstracts. If the study were deemed eligible, the full texts were then read independently to verify if all inclusion criteria were met. Any disagreements were discussed between the two reviewers, and if necessary, a third reviewer intervened (F.D.).

Methodological Quality of Included Studies

The quality of the included studies was appraised with the Quality Appraisal For Reliability Studies checklist (QAREL) [107]. The QAREL has 11 items covering seven domains: spectrum of examiners, spectrum of subjects, examiners blinding, suitability of the time-interval between repeated measures, the order effects of examination, appropriate test interpretation and application, as well as appropriate statistical analyses [115]. Each of these items may be answered “yes”, “no”, “unclear” or, “not applicable”. Two raters (S.D and P.O) independently evaluated the methodological quality of each included study. Differences were resolved by consensus.

Data Extraction

The raters used an extraction form based on the proposed tool by the authors of the QAREL [107]. Data extraction of the selected studies included: study protocols, design and setting, type of reliability, population characteristics, raters, types of tests and statistics reported.

Data Analysis

Mean methodological score of each article was computed as the number of items present (“yes”) out of eleven possible items, higher score representing a lower risk of bias [107]. Cohen’s kappa was used to calculate pre-consensus inter-rater agreement on individual methodological items of the QAREL tool. From the extracted reliability measures for each relevant diagnostic test, the range of point estimate across all the studies was reported summarizing qualitatively the available evidence. The Landis and Koch scale was used for the interpretation of reliability where ≤ 0 indicates no agreement, 0-0.20 slight, 0.21-0.40 fair,

0.41-0.60 moderate, 0.61-0.80 substantial or good and 0.81-1.0 almost perfect or excellent agreement. [116] A physical test with $\kappa \leq 0.40$ is considered unreliable because the probability agreement due to chance is too high to be clinically useful for diagnostic decision-making [108].

Results

Overall description of included studies

As shown in Figure 2, 6750 potential studies were initially identified, 6629 studies were excluded for various reasons and 33 studies were ultimately included. Table 8 presents the characteristics of the included studies. Four studies present results on the reliability of meniscal injuries tests, nine for ACL injuries tests, five for OA related tests, eight for PFP and seven for other knee disorders such as medial collateral injury, patellar tendinopathy, Baker's cyst and knee effusion.

Methodological Quality of Included Reviews

QAREL scores for the appraisal of methodological quality of the included studies are presented in Table 9. QAREL score ranged from 2 to 10 (out of eleven) with a means of 5.5±0.5. Items most often missing were related to the blinding of the raters to clinical information that may influence their interpretation of physical tests findings (i.e: reference standard), but also to the order of examination between rater or time between evaluation [107]. The mean agreement for individual items between the two evaluators was moderate (κ=0.52) with nine out of eleven items reaching moderate agreement or higher (κ>0.40).

Figure 2: Bibliographic search flowchart

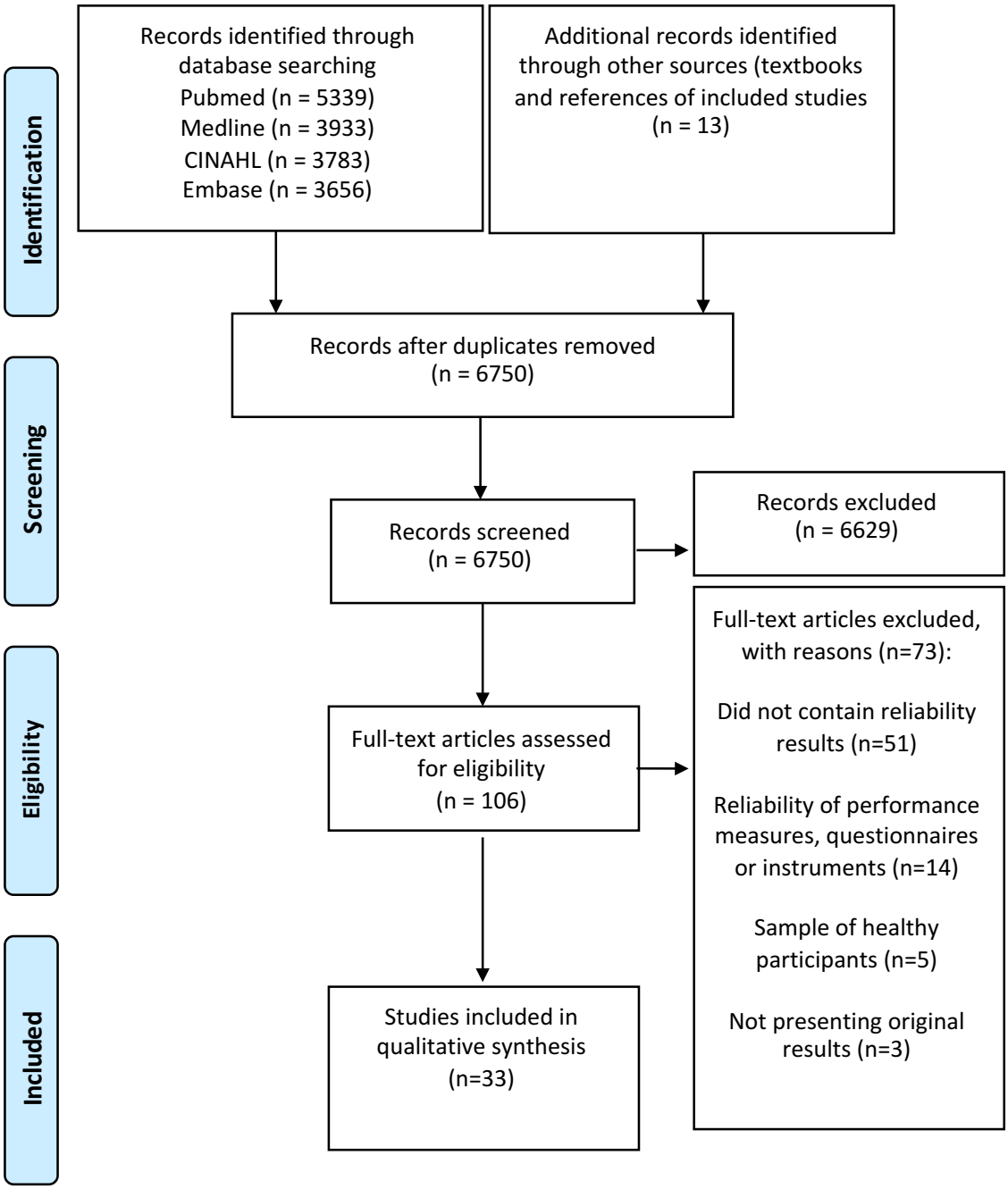


Table 8: Characteristics of included reliability studies according to different knee disorders

Authors	Protocol	Design and Setting	Type of reliability	Population characteristics	Raters	Tests	Agreement statistics
MENISCAL INJURIES							
Evans et al. (1993)	Validity and reliability	Prospective (secondary Care)	Inter-rater	Consecutive patients awaiting arthroscopy, n=104 Confirmed meniscal tear, n=59	Orthopaedic surgeon and medical student	McMurray	Cohen's Kappa
Dervin et al. (2001)	Validity and reliability	Prospective (tertiary care)	Inter-rater	Patients with symptomatic osteoarthritis and suspected unstable meniscus, n=115 Confirmed meniscal tear, n=92 Mean age (y) ± SD: 60.5±8.5	Orthopaedic surgeons and orthopaedic fellows	Joint line tenderness and McMurray	Cohen's Kappa
Galli et al. (2013)	Validity and reliability	Prospective (tertiary care)	Inter-rater	Suspected non-acute meniscal lesion, n=56 Confirmed meniscal tear, n=35 Mean age (y) ± SD: 29.7±11.6 Mean symptom duration (d) ± SD: 52±15	Orthopaedic surgeon, orthopaedic resident and trained medical student	Joint line tenderness and McMurray	Cohen's-Fleiss Kappa
Snoeker et al. (2015)	Validity and reliability	Prospective (primary care)	Inter-rater	Patients consulting for knee symptoms, n=121 Confirmed meniscal tear, n=52 Mean age (y) ± SD: 43.2±12.2 Median symptom duration (w, min-max): 7.0 (0.5-27.0)	Physiotherapists	Deep Squat Test, Joint line tenderness and Thessaly test	Cohen's Kappa
ANTERIOR CRUCIATE LIGAMENT INJURIES							
Cooperman et al. (1990)	Validity and reliability	Prospective (secondary care)	Intra and inter-rater	Patients with unilateral knee problem, n=32 Confirmed ACL injury, n=13 Mean age (y) ± SD: 26.0±9.1	Orthopaedic surgeons and physical therapists	Lachman test	Cohen's and weighted Kappa
Duggan et al. (1991)	Reliability	Retrospective (secondary care)	Inter-rater	Patients with suspected ACL injury, n=16 Confirmed ACL injury, n=14	Physical therapists	Lachman test	ICC and Pearson r
Johnson et al. (2004)	Reliability	Retrospective (secondary care)	Intra and Inter-rater	Patients with confirmed ACL injury, n=102 Mean age (y, range): 33, 19-63	Not reported	Lachman test	Cohen's Kappa

Wiertsema et al. (2008)	Reliability	Prospective (tertiary care)	Intra and inter-rater	Consecutive patients with confirmed complete ACL injury, n=20. Time since injury at least 6 w Mean age (y) ± SD: 30.0±8.1	Physical therapists	Lachman test	ICC
Peeler et al. (2010)	Validity and reliability	Retrospective (secondary care)	Inter-rater	Surgically confirmed ACL injury, n=112	Athletic therapists, family physicians and orthopaedic surgeons	Lachman, anterior drawer and pivot shift	Coefficient of agreement
Labbe et al. (2011)	Reliability	Prospective (secondary care)	Inter-rater	Patients with knee instability following confirmed ACL injury, n=8; asymptomatic patients, n=4 Time since injury at least 3 m Mean age (y, range): 32.9 (18-57)	Orthopaedic surgeons and orthopaedic resident	Pivot shift test	Fleiss-Weighted Kappa and ICC
Mulligan et al. (2011)	Validity and reliability	Prospective (secondary care)	Inter-rater	Consecutive patients with knee pain, n=52 Confirmed ACL injury, n=23 Mean age (y) ± SD: 34.3±12.0 Mean time since injury (d) ± SD: 195±130	Physical therapists	Prone and supine Lachman test	Cohen's Kappa
Yoon et al. (2014)	Validity and reliability	Prospective (secondary care)	Intra- and inter-rater	Patients with confirmed ACL injury, n= 201 Mean age (y, range): 30.9 (16-65) Mean time from trauma to surgery (month): 22.4	Orthopaedic surgeon, fellowship knee surgeon and	Anterior drawer Lachman and pivot shift	Cohen's Kappa
Mulligan et al. (2015)	Validity and reliability	Prospective (secondary care)	Inter-rater	Consecutive patients with knee pain, n= 45 Confirmed ACL injury, n=17 Mean age (y)± SD: 40.7±14.0 Mean time since injury (d) ± SD: 238±157	Physical therapists	Lachman test	Cohen's Kappa
OSTEOARTHRITIS							
Cushnaghan et al. (1990)	Reliability	Retrospective (secondary care)	Intra and inter-rater	Patients with confirmed knee OA, n=8 Mean age (y): 64 Mean symptom duration (y): 12.5	Consultants and trainees (professional qualification N/R)	Complete physical examination for knee OA	Cohen's Kappa

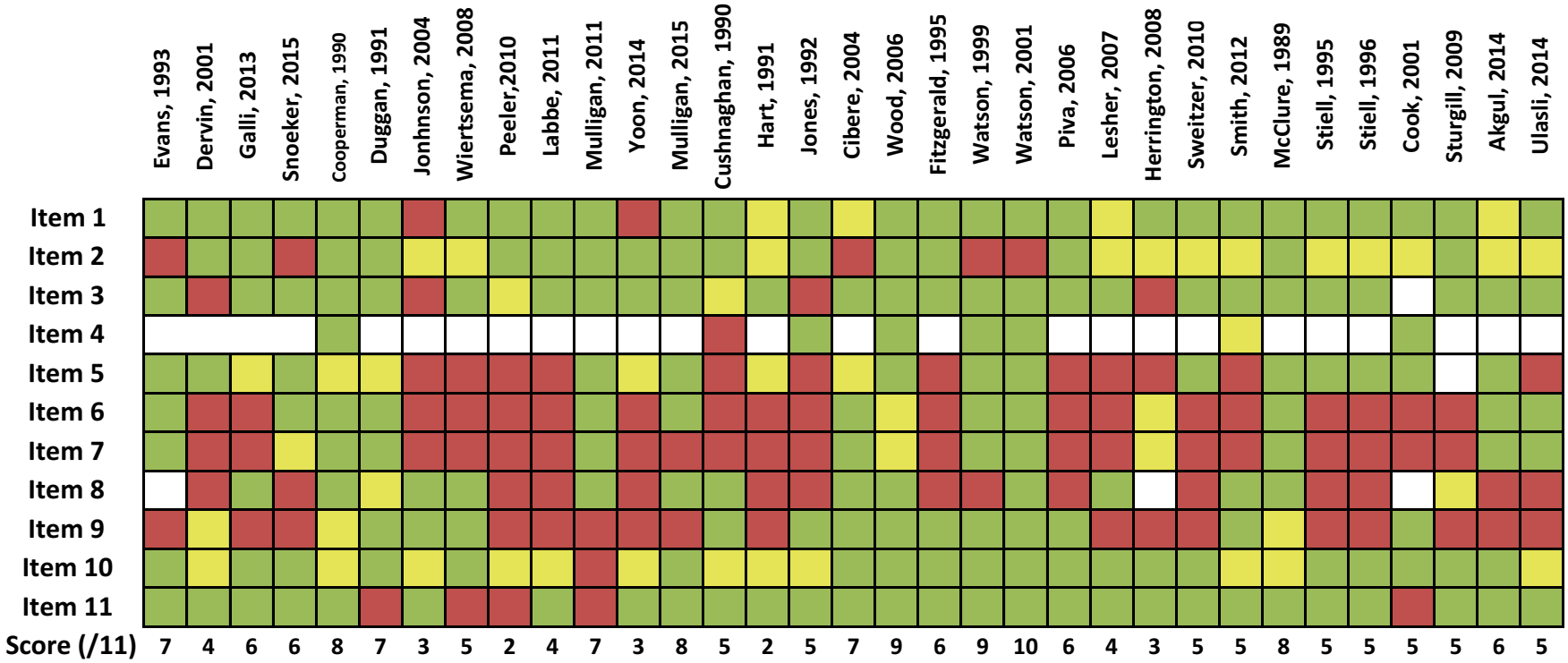
Hart et al. (1991)	Reliability	Prospective (primary care)	Inter-rater	Unselected women from general practice, n=41 Confirmed knee OA N/R. Mean age (y): 53	N/R	Complete physical examination for knee OA	Cohen's Kappa
Jones et al. (1992)	Reliability	Prospective (secondary care)	Intra and inter-rater	Patients consulting a clinic for knee OA, n=49 Confirmed knee OA N/R Mean age (y, range): 79 (50-92)	Rheumatologist, geriatrician, and general practitioner	Complete physical examination for knee OA	Cohen's Kappa
Cibere et al. (2004)	Reliability	Retrospective (secondary care)	Inter-rater	Patients with confirmed knee OA, n=6 Median age (y, range): 62 (44-74) Median symptom duration (y, range): 8 (3-20)	Rheumatologist	Complete physical examination for knee OA	PABAK and reliability coefficient
Wood et al. (2006)	Reliability	Prospective (primary and secondary care	Intra and inter-rater	Patients aged at least 50 y having consulted for a knee complaint in general practices, outpatient rheumatology clinic and waiting list for knee replacement surgery, n=58 Confirmed knee OA N/R Mean age (y, range): 63 (50-86)	Physical therapist	Complete physical examination for knee OA	Quadratic and Weighted Kappa ICC
PATELLOFEMORAL PAIN							
Fitzgerald et al. (1995)	Reliability	Prospective (secondary care)	Inter-rater	Patients with possible PFP, n=66 Confirmed PFP, n=40 Mean age (y)±SD: 9.7±13.1	Physical therapists	Patellofemoral alignment tests	Cohen's Kappa
Watson et al. (1999)	Reliability	Prospective (secondary care)	Intra and inter-rater	Symptomatic patients and healthy participants, n=56 Confirmed PFP, n=17 Mean age (y)±SD: 29±8	Senior physical therapy students	McConnell classification tests	Cohen's Kappa
Watson et al. (2001)	Reliability	Prospective (secondary care)	Intra and inter-rater	Symptomatic patients and healthy participants, n=55 Confirmed PFP, n=26 Mean age (y): 27.8	Senior physical therapy students	Patellar mobility tests	Cohen's Kappa
Piva et al. (2006)	Reliability	Prospective (secondary care)	Inter-rater	Patients with confirmed PFP, n=30 Mean age (y)±SD: 29.1±8.4	Physical therapists	Complete physical examination for PFP	Cohen's Kappa and ICC

Lesher et al. (2007)	Validity and reliability	Prospective (secondary care)	Inter-rater	Patients with confirmed PFP, n=50 Mean age (y)±SD: 22.8±4.2 Mean duration of symptoms (d)±SD: 75.4±123.8	N/R	Complete physical examination for PFP	Cohen's Kappa and ICC
Herrington et al. (2008)	Reliability	Prospective (secondary care)	Intra-rater	Symptomatic and asymptomatic patients, n=24 Confirmed PFP, n=12 Mean age (y)±SD: 21.9±2.6	N/R	Medio-lateral patella position test	ICC
Sweitzer et al. (2010)	Validity and reliability	Prospective (secondary care)	Inter-rater	Consecutive patients with anterior knee pain, n=82 Confirmed PFP, n=59 Mean age (y): 51.2	Orthopaedic surgeons	Patellar mobility tests	Cohen's Kappa
Smith et al. (2012)	Reliability	Prospective (secondary care)	Intra and inter-rater	Patients with confirmed patellar instability, n=5 Mean age (y, range): 26.6 (18-38) Mean duration of symptoms (y, range): 11 (2-26)	Orthopaedic surgeons	Complete physical examination for patellar instability	Weighted and Fleiss' kappa
OTHERS							
McClure et al. (1989)	Reliability	Prospective (secondary care)	Inter-rater	Patients with suspected medial collateral ligament injury, n=50 Confirmed medial collateral ligament injury, n=5 Mean age (y)±SD: 30±11 Mean duration of symptoms (w)±SD: 43.0±59.8	Physical therapists	Medial collateral ligament test	Weighted kappa
Stiell et al. (1995)	Reliability	Prospective (primary care)	Inter-rater	Patients with acute knee injury, n=127 Confirmed fractures in reliability sample N/R Mean age (y)±SD: 36±25 Mean time since injury (hours)±SD: 9±29	Emergency physicians	Complete physical examination for knee fracture	Cohen's Kappa
Stiell et al. (1996)	Reliability	Prospective (primary care)	Inter-rater	Patients with acute knee injury, n=124 Confirmed fractures in reliability sample N/R Mean age (y)±SD: 37±16	Emergency physicians	Complete physical examination for knee fracture	Cohen's Kappa

Cook et al. (2001)	Validity and reliability	Prospective (primary care)	Intra-rater	Symptomatic and asymptomatic junior athletes, n=29 Confirmed patellar tendinopathy in reliability sample N/R Mean age (month)±SD: 197.2 ±12.2	N/R	Palpation for tenderness of the patellar tendon	Pearson's r
Sturgill et al. (2009)	Reliability	Prospective (secondary care)	Inter-rater	Patients with unilateral knee dysfunction, n=75 Confirmed knee effusion N/R Mean age (y)±SD: 31.4±13.4	Physical therapists	Stroke test for effusion	Cohen's Kappa
Akgul et al. (2014)	Validity and reliability	Prospective (tertiary care)	Inter-rater	Patients with confirmed knee OA and non-OA, with asymptomatic popliteal fossae, n=110 Confirmed Baker's cyst, n=17 Age (y, range): 45-66 Mean duration of symptoms (y)±SD: 7.9±7.4	Experienced physicians	Palpation of popliteal fossa for Baker's cyst	Cohen's Kappa
Ulasli et al. (2014)	Validity and reliability	Prospective (tertiary care)	Inter-rater	Patients with confirmed knee osteoarthritis, n=86 Mean age (y) ± SD: 56.2±10.2	Residents (speciality not specified)	Bulge sign, ballottement and patellar tap	Cohen's Kappa

y: years; mo: month; w: week; d: day; ACL: anterior cruciate ligament; OA: osteoarthritis; PFP: patellofemoral pain; SD: standard deviation; N/R: not reported; ICC: intraclass correlation coefficient. PABAK: Prevalence and bias adjusted kappa.

Table 9: Methodological quality of included studies using the QAREL checklist



Green = **Present;** Red = **Absent;** Yellow = **Unclear; Blank = Not Applicable**

Item 1: Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended for the results to be applied?; **Item 2:** Was the test performed by raters who were representative of those to whom the authors intended for the results to be applied?; **Item 3:** Were raters blinded to the findings of other raters during the study?; **Item 4:** Were raters blinded to their own prior findings of the test under evaluation?; **Item 5:** Were raters blinded to the results of the accepted reference standard or disease status for the target disorder (or variable) being evaluated?; **Item 6:** Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design?; **Item 7:** Were raters blinded to additional cues that were not part of the test?; **Item 8:** Was the order of examination varied?; **Item 9:** Was the stability of the variable being measured taken into account when determining the suitability of the time-interval between repeated measures?; **Item 10:** Was the test applied correctly and interpreted appropriately?; **Item 11:** Were appropriate statistical measures of agreement used?

Summary of findings

Meniscal injuries

Four studies evaluated the inter-rater reliability of meniscal injuries tests with mean QAREL score of 5.8±1.3 (Table 10) [117-120]. No study evaluated intra-rater reliability. For the McMurray test, inter-rater reliability kappa values ranged from 0.16 to 0.38. For the Medial and Lateral Joint Line Tenderness, inter-rater kappa values ranged from 0.11 to 0.25. For the Thessaly test, the authors reported an inter-rater kappa of 0.54 (95% CI: 0.37-0.72) [120]. These results suggest, based on low to moderate level quality studies, that the McMurray and Joint Line Tenderness are unreliable ($\kappa \leq 0.40$), but based on one study of moderate quality, the Thessaly test may be considered reliable.

Anterior Cruciate Ligament injuries

Nine studies evaluated the reliability of ACL injury tests with mean QAREL score of 5.2±2.3 (Table 11) [121-129]. The anterior drawer and the pivot shift tests were evaluated in three low-quality studies and demonstrated moderate to excellent inter-rater reliability (anterior drawer test: P_0=0.57, κ=0.96 and pivot shift: P_0=0.53, κ=0.83-0.87) [125, 127, 129]. For the Lachman, intra-rater kappa values ranged from 0.33 to 0.51 in one high-quality study [121] and was of 0.29 in one low-quality study [123]. Inter-rater reliability kappa values ranged from 0.19 to 0.93 in five studies [121, 123, 124, 128, 129] and inter-rater reliability was found to be moderate or excellent in the three highest-quality studies. Although the quality of the evidence is heterogeneous, these results suggest that commonly used tests for ACL injuries may be considered reliable.

Osteoarthritis

Five studies evaluated the reliability of nine knee OA tests with a mean QAREL score of 5.6±2.6 [130-134] (Table 12). Overall, intra-rater reliability kappa values ranged from 0.50 to 0.91 demonstrating moderate to excellent intra-rater reliability for all tests. For inter-rater reliability, results varied from 0.09 to 0.94. Based on two moderate quality studies of very

small sample sizes [130, 131], Tibiofemoral Crepitus was found to have good to excellent reliability (κ = 0.64 to 0.91). For Joint Line or Patellofemoral pain/tenderness, Medial Joint Line Tenderness reached fair to excellent reliability (4 out of 5 studies, κ = 0.35 to 0.95), Lateral Joint Line Tenderness reached fair to excellent reliability (3 out of 4 studies, κ = 0.29 to 0.85) and Patellofemoral Tenderness reached moderate to excellent reliability (2 out of 4 studies, κ = 0.43 to 0.94). Bony Enlargement presented moderate to excellent reliability (2 out of 4 studies, κ = 0.55 to 0.97). The studies were of low to moderate quality and two of these studies were of limited small sample sizes. The Joint Pain on Passive or Active Movement inter-rater kappa values ranged from 0.72 to 0.92, indicating good to excellent agreement in two low to moderate quality studies [132, 133].

Patellofemoral pain

Eight studies evaluated the reliability of PFP tests with a mean QAREL score of 6.0±2.4 (Table 13) [135-142]. Of note, the study of *Smith et al*. relates to patellar instability, considered a subgroup of PFP in the presence of pain [8]. For all tests, intra-rater kappa or ICC [140] values ranged from -0.06 to 0.86. Lateral Tilt reached moderate intra-rater reliability in one high-quality study (κ=0.44-0.50) [138], while Patellar Compression and Crepitus reached moderate to good intra-rater reliability (κ = 0.56-0.66) in one moderate quality study with a small number of patients [136].

Inter-rater reliability kappa values ranged from -0.03 to 0.73. Lateral/Medial Glide reached moderate to good inter-rater reliability (2 out of 5 studies, κ = 0.59-0.73). Lateral Tilt reached moderate to good inter-rater reliability (2 out of 5 studies, κ = 0.49-0.71). Lateral Pull and Quality of Movement reached moderate to good inter-rater reliability each in one study (κ = 0.53-0.67). For all these tests, positive results, concluding on adequate reliability, came from low to moderate quality studies, while tests evaluated in both high-quality studies showed poor inter-rater reliability [138, 139].

Other disorders or physical tests

Three moderate quality studies [130, 133, 143] and one high-quality study [134] evaluated the inter-rater reliability of palpation for the detection of a Baker's cyst in the popliteal fossa, with only one of the moderate quality study reaching good agreement (PABAK=0.66) [130]. One moderate quality study evaluated the intra-rater reliability of tendon palpation for the diagnosis of patellar tendinopathy with a raw agreement of 85% [144]. One high-quality study reached close to moderate inter-rater reliability for the tibiofemoral joint abduction test (κ=0.40) for medial collateral ligament injury [145]. Two moderate quality studies evaluated knee effusion tests, one finding good inter-rater reliability (κ=0.61; 95% CI: 0.54-0.81) [146] while the other did not (κ=0.25) [147]. Two moderate quality studies found good to excellent inter-rater reliability for the tenderness of head of the fibula (κ=0.64-0.92), tenderness of the patella (κ=0.69-0.76) and inability to bear weight for four steps (κ=0.68-0.75), all included in the *Ottawa Knee Rule* for the exclusion of knee fractures [148, 149].

Table 10: Summary of reliability measures from included studies evaluating meniscal injury tests

Physical tests	Reliability measures	Included studies			
		Evans et al. (1993)	**Dervin et al. (2001)**	**Galli et al. (2013)**	**Snoeker et al. (2015)**
McMurray	Raw Agreement	-	59%	-	-
	Inter-rater kappa	-0.10-0.38	0.16 [-0.01-0.33]	0.21-0.37	-
Medial Joint Line Tenderness	Raw Agreement	-	79%	-	62%
	Inter-rater kappa	-	0.21 [0.01-0.41]	0.11	0.17 [-0.02-0.36]
Lateral Joint Line Tenderness	Raw Agreement	-	70%	-	-
	Inter-rater kappa	-	0.25 [0.07-0.43]	0.11	-
Thessaly	Raw Agreement	-	-	-	77%
	Inter-rater kappa	-	-	-	0.54 [0.37-0.72]
Total number of patients (number of cases)		n=104 (n=59)	n=115 (n=92)	n=56 (n=35)	n=121 (n=52)
QAREL score (/11)		7	4	6	6

Brackets indicate 95% confidence intervals.

Table 11: Summary of reliability measures from included studies evaluating anterior cruciate ligament injury tests

Physical tests	Reliability measures	Included studies								
		Cooperman et al. (1990)	**Duggan et al. (1991)**	**Johnson et al. (2004)**	**Wiertsema et al. (2008)**	**Peeler et al. (2010)**	**Labbe et al. (2011)**	**Mulligan et al. (2011)**	**Yoon et al. (2014)**	**Mulligan et al. (2015)**
Lachman	Raw agreement	71-76%	-	-	-	-	-	79-90%	-	65-91%
	Intra or inter-rater ICC	-	Inter: 0.25-0.34	-	Intra: 1.0 [1.0-1.0] Inter: 0.77 [0.50-0.91]	-	-	-	-	-
	Intra-rater kappa	0.33-0.51	-	0.29	-	-	-	-	-	-
	Inter-rater kappa	0.19-0.42	-	0.23	-	P_0=0.45	-	0.60-0.81	0.93	0.42-0.72
Anterior Drawer	P_0 or inter-rater kappa	-	-	-	-	P_0=0.57	-	-	κ=0.96	-
Pivot Shift	P_0 or inter-rater kappa	-	-	-	-	P_0=0.53	κ=0.83	-	κ=0.87	-
Total number of patients (number of cases)		n=32 (n=13)	n=16 (n=14)	n=102 (n=102)	n=20 (n=20)	n=112 (n=112)	n=12 (n=8)	n=52 (n=23)	n=201 (n=201)	n=45 (n=17)
QAREL score (/11)		8	7	3	5	2	4	7	3	8

P_0=average coefficient of agreement; k=kappa; ICC=intraclass correlation coefficient; Raw agreement is for inter-rater. Brackets indicate 95% confidence intervals.

Table 12: Summary of reliability measures from included studies evaluating osteoarthritis physical tests

Physical tests	Reliability measures	Included studies				
		Cushnaghan et al. (1990)	**Hart et al. (1991)**	**Jones et al. (1992)**	**Cibere et al. (2004)***	**Wood et al. (2006)**
Tibiofemoral crepitus	Raw agreement	-	65.0%	-	-	60.7%
	Intra-rater kappa	0.68 [0.44-0.92]	-	0.78 [0.60-0.96]	-	0.53 [0.24]
	Inter-rater kappa	0.64 [0.48-0.80]	0.14	0.09 [0.00-0.26]	0.75-0.91*	0.22 [-0.08]
Patellofemoral crepitus	Intra-rater kappa	0.50 [0.26-0.74]	-	0.75 [0.59-0.91]	-	-
	Inter-rater kappa	0.24 [0.10-0.38]	-	0.10 [0.00-0.22]	0.73-0.92*	-
Medial joint line pain/tenderness	Raw agreement	-	97.5%	-	-	64.9%
	Intra-rater kappa	0.64 [0.40-0.88]	-	0.60-0.76 [0.47-0.72]	-	0.48 [0.17]
	Inter-rater kappa	0.40 [0.26-0.54]	0.74	0.35-0.48 [0.24-0.45]	0.83-0.94*	0.26 [-0.04]
Patellofemoral pain/tenderness	Raw agreement	-	-	-	-	71.9%
	Intra-rater kappa	0.41 [0.30-0.52]	-	0.61-0.68 [0.43-0.78]	-	0.54 [0.26]
	Inter-rater kappa	0.35 [0.20-0.50]	-	0.27-0.33 [0.05-0.48]	0.93-0.94*	0.43 [0.16]

Lateral joint line pain/tenderness	Raw agreement	-	-	-	-	68.4%
	Intra-rater kappa	0.50 [0.26-0.74]	-	0.60-0.73 [0.44-0.74]	-	0.41 [0.08]
	Inter-rater kappa	0.43 [0.28-0.58]	-	0.29-0.44 [0.14-0.44]	0.85*	0.29 [-0.02]
Non-bony swelling	Intra-rater kappa	0.67 [0.45-0.89]	-	0.54 [0.27-0.81]	-	-
	Inter-rater kappa	0.28 [0.12-0.44]	0.25	0.13 [0.00-0.68]	-	-
Bony enlargement	Raw agreement	-	90.2%	-	-	54.6-72.7%
	Intra-rater kappa	0.74 [0.52-0.56]	-	-	-	0.55-0.61 [0.28-0.33]
	Inter-rater kappa	0.55 [0.40-0.70]	0.10	-	0.91-0.97*	0.08-0.44 [-0.23-0.15]
Warmth	Raw agreement	-	-	-	-	70.6%
	Intra-rater kappa	-	-	0.71 [0.41-1.00]	-	0.91 [0.69]
	Inter-rater kappa	-	-	0.23 [0.00-1.00]	0.14-0.24*	0.18 [-0.25]
Joint pain on passive or active movement	Raw agreement	-	92.7-97.5%	-	-	-
	Intra-rater kappa	-	-	0.84 [0.69-1.0]	-	-
	Inter-rater kappa	-	0.85-0.92	0.72 [0.67-0.76]	-	-
	Total number of patients (number of cases)	n=8 (n=8)	n=41 (n=41)	n=49 (n=N/R)	n=6 (n=6)	n=58 (N/R)
	QAREL score (/11)	5	2	5	7	9

k=kappa; *Prevalence and Bias Adjusted Kappa (PABAK) or reliability coefficient (1-Variance); Raw agreement is for inter-rater. N/R = Not reported; Brackets indicate 95% confidence intervals, for *Wood et al.* brackets are 99% lower bound of confidence intervals.

Table 13: Summary of reliability measures from included studies evaluating common patellofemoral pain tests

Physical tests	Reliability measures	Included studies						
		Fitzgerald et al. (1995)	Watson et al. (1999)	Watson et al. (2001)	Piva et al. (2006)	Lesher et al. (2007)	Sweitzer et al. (2010)	Smith et al. (2012)
Lateral/medial glide	Raw agreement	44%	70%	-	-	-	-	-
	Intra-rater kappa	-	0.11-0.35	-	-	-	-	0.34 [0.16-0.54]
	Inter-rater kappa	0.10	0.02	-	-	0.73	0.59 [0.42-0.72]	0.11 [-0.6-0.27]
Anterior/posterior tilt	Raw agreement	71%	30%	-	-	-	-	-
	Intra-rater kappa	-	0.03-0.23	-	-	-	-	-
	Inter-rater kappa	0.24	0.04	-	-	-	0.48 [-0.28-0.61]	-
Lateral tilt for tightness of lateral retinaculum	Raw agreement	59%	47%	47-62%	93%	-	-	-
	Intra-rater kappa	-	0.28-0.33	0.44-0.50	-	-	-	0.05 [-0.14-0.26]
	Inter-rater kappa	0.21	0.19	0.20-0.35	0.71 [0.57-0.86]	0.49	-	0.08 [-0.12-0.29]
Medial/lateral rotation	Raw agreement	61%	91%	-	-	-	-	-
	Intra-rater kappa	-	-0.06-0.00	-	-	-	-	-
	Inter-rater kappa	0.36	-0.03	-	-	0.29	-	-
Lateral pull for patellar tracking	Raw agreement	-	-	62%	-	-	-	-
	Intra-rater kappa	-	-	0.39-0.47	-	-	-	0.28 [0.06-0.51]
	Inter-rater kappa	-	-	0.31	-	-	-	0.53 [0.28-0.77]
Patellar compression test	Intra-rater kappa	-	-	-	-	-	-	0.66 [0.45-0.84]
	Inter-rater kappa	-	-	-	-	-	-	-0.06 [-0.27-0.14]

Crepitus	Intra-rater kappa	-	-	-	-	-	-	0.56 [0.29-0.83]
	Inter-rater kappa	-	-	-	-	-	-	0.34 [0.06-0.62]
Quality of movement (motor control during step down task)	Raw agreement				80%			
	Inter-rater kappa	-	-	-	0.67 [0.58-0.76]	-	-	-
	Total number of patients (number of cases)	n=66 (n=40)	n=56 (n=17)	n=55 (n=26)	n=30 (n=30)	n=50 (n=50)	n=82 (n=59)	n=5 (n=5)
	QAREL score (/11)	6	9	10	6	4	5	5

k=kappa; ICC: intraclass correlation coefficient; Raw agreement is for inter-rater. Brackets indicate 95% confidence intervals.

Discussion

The current systematic review highlights inconsistent evidence concerning the reliability of physical examination tests for the diagnosis of knee disorders. At the present time, the literature does not demonstrate that common tests for meniscal injuries are reliable, except for the Thessaly test. For ACL injury tests, the Lachman was found reliable, if used with a dichotomous positive/negative outcome. Osteoarthritis tests were reliable when used by the same evaluator, while Tibiofemoral Crepitus, Joint Line and Patellofemoral pain, Bony Enlargement and Pain on Movement may reach moderate to excellent inter-rater reliability – although not consistently across the included studies. For PFP tests, Lateral Tilt, Patellar Compression and Crepitus may reach moderate intra-rater reliability, while Lateral Glide, Lateral Tilt, Lateral Pull and Quality of Movement may reach moderate to good inter-rater reliability – again not consistently across studies.

The results presented here are very heterogeneous and caution is warranted since for specific physical tests, conclusions are only based on a few studies, often with limited sample sizes. A total of thirty-three studies were included with a mean QAREL score of 5.5±0.5, indicating a moderate to high risk of bias. The most common risk of bias were related to the domains of rater blinding to the reference standard or clinical information (i.e: it was unknown if the raters had knowledge or not of the real disorder for which the patient was evaluated), order of examination and time between evaluations by the different raters [107]. These bias may influence the patients' response to the physical tests and the interpretation of these tests by the raters [107]. Of note, only 19 studies out the 33 included were specifically designed to assess the reliability of physical examination tests and this may in part explain the overall low methodological quality related to the reliability part of these studies.

Based on low to moderate quality evidence, the Lachman was found highly reliable especially when used with a dichotomous outcome. *Lange et al*. had previously published a SR on the reliability of physical examination tests for ACL injuries and they concluded that the Lachman in prone position reached the highest interrater reliability, based on only one high-quality

study [105]. Because of heterogeneous methodological quality, the authors could not recommend any other tests [105]. Compared to their results, our review included three other studies, including one recent high-quality study on the reliability of the Lachman test, which may explain the difference in our conclusion [127-129].

Still, the inconsistent results presented in our current review is concerning and two factors are often proposed to explain the varying level of reliability of certain tests: the evaluators' experience and the methods of standardization of the tests. In one study for OA tests, *Cibere et al*. specifically assessed the influence of standardization on reliability and the authors demonstrated improved inter-rater reliability after the standardization process [130]. Their process included training the evaluators to execute the physical tests in a systematic manner while also reaching a consensus between evaluators to minimize variability between their usual practice [130]. For other included studies for OA tests, although their design did not formally evaluated the impact of professional training or standardization, it appears that these factors may also explain higher intra and inter-rater reliability, where experienced rheumatologists were more consistent than trainees or other non-specialist practitioners [130, 131, 133, 134]. For ACL injuries tests, experience of the evaluators appears to be associated with improved reliability [121, 122, 126-128], while high standardization combined with simpler response criteria (i.e: positive/negative instead of grading scale with end-feel) appears associated with moderate to excellent agreement [34, 122, 124, 127-129]. However, for meniscal injuries tests, both experience and standardization did not appear to improve reliability of the associated tests [118, 119]. Also, for PFP, experience of the evaluator did not consistently resulted in improved reliability [135, 140, 142] – even when comparing the inter-rater reliability of five international experts [136].

Other factors may also influence reliability and were often overlooked in the included studies. The time interval between evaluation for inter-rater reliability was commonly unreported. Characteristics of cases and stability of the condition may also impact reliability where acute and severe cases may make pain provocation tests more easily reliable [117]. The nature of

the tests (i.e: pain provocation compared to movement quality or laxity tests) may influence their reliability as some tests' responses are dependent on the skills or the manual pressure applied by the evaluator during the test. These factors may need to be more closely monitored to allow acceptable reliability.

In light of the current results, it is difficult to make specific recommendations concerning the exclusion of some tests based on their inadequate reliability. Many tests showed variable reliability. As previously mentioned, because of the methodological quality of certain studies, the results may be biased and caution is warranted before concluding that certain tests are systematically unreliable. Only the synthesis of consistent replicated results from high-quality studies in various populations will allow definite conclusion on the reliability of physical tests for knee disorders.

Strengths and limitations

We systematically reviewed and updated the evidence on the reliability of physical examination tests for all knee disorders. This review used a validated methodological assessment tool specifically for reliability studies (QAREL) and a wide initial search strategy [111]. An emphasis was made on physical examination tests for the diagnosis of knee disorders and other possibly reliable measures were not included (ex: performance measures, functional questionnaires) [106, 111]. A limitation could be considered for the inter-rater reliability cut-off used ($\kappa \leq 0.40$), as other studies have used higher cut-offs (inter-rater reliability $\kappa=0.85$) [108]). Using a more stringent cut-off value for reliability in our review would not have changed our conclusions. It was decided not to include a meta-analysis because of high heterogeneity of reliability measures, lack of description of physical examination tests, different grading scales, difference in the evaluators' experience and training and low overall methodological quality [105]. Our SR was restricted to English and French language.

Conclusions

The evidence presented in this systematic review raises concerns about the reliability of physical examination tests for knee disorders. Meniscal tests were unreliable except for the Thessaly. The Lachman was reliable in both low and high-quality studies. All OA tests demonstrated moderate intra-rater reliability, but tests that reached moderate inter-rater reliability came from low-quality evidence. Only certain tests for PFP reached moderate intra or inter-rater reliability, again from low-quality evidence. Evaluators' experience and high quality of standardization may improve reliability for ACL injury, OA and PFP tests – although not consistently. At present, clinicians should be aware that relying on specific tests may be unwise for certain knee disorders and that considering the whole clinical picture may be the priority to obtain a reliable knee diagnosis. Further research should continue to evaluate the reliability of physical examination tests using high-quality design and by implementing strategies most likely to improve reliability.

2.3 Validité de la combinaison de l'anamnèse et des tests de l'examen physique.

Les revues systématiques ont démontré que la validité et la fidélité des tests de l'examen physique lorsqu'exécutés individuellement sont souvent insuffisantes pour orienter le diagnostic des pathologies au genou. Les résultats ont démontré qu'une évaluation musculosquelettique complète pourrait être valide, mais les données probantes sont limitées et incomplètes concernant la validité de la combinaison des éléments spécifiques de l'anamnèse et de l'examen physique. Cette section présentera plus en profondeur les données probantes sur ce sujet.

2.3.1 Déchirures partielles ou complètes du LCA

Deux études ont évalué la validité diagnostique de la combinaison spécifique d'éléments de l'anamnèse et de l'examen physique pour le diagnostic des déchirures du LCA [101, 150]. Une étude a évalué un échantillon de 134 patients qui présentaient un traumatisme récent au genou, de ces patients, 28 étaient atteints d'une déchirure partielle ou complète du LCA [101]. Cette étude a démontré que la combinaison d'un épanchement articulaire immédiat, d'une sensation de « pop » au moment du traumatisme, de la présence de symptômes de dérobade et d'un test du tiroir antérieur positif présentait un rapport de vraisemblance (RV) positif de 4,2 (IC95%: 2,4-7,5) et de 7,2 (IC95%: 3,6-14,4) pour diagnostiquer une déchirure partielle ou complète du LCA, tout en classifiant respectivement 63% et 65% des cas [101]. Une autre étude a évalué un échantillon de 60 patients qui avaient subi une arthroscopie diagnostique du genou; 22 patients souffraient d'une déchirure complète du LCA [150]. Les auteurs de cette étude ont conclu que la combinaison d'un épanchement immédiat, d'une sensation de « pop », de symptômes de dérobade ainsi que des réponses positives à trois tests du LCA (Lachman, Pivot shift et tiroir antérieur) présentait un RV positif de 24,0 pour diagnostiquer une déchirure complète du LCA [150].

2.3.2 Déchirures méniscales

Trois études ont évalué la validité diagnostique de la combinaison d'éléments de l'anamnèse et de l'examen physique pour le diagnostic des déchirures méniscales [11, 102, 151]. Sur le

même échantillon préalablement présenté de 134 patients qui présentaient un traumatisme récent au genou, 47 étaient aussi atteints d'une déchirure méniscale d'origine traumatique [102]. Les auteurs ont démontré que la combinaison d'un traumatisme avec mise en charge suivie d'une impossibilité de poursuivre l'activité et de la présence de douleur à la flexion passive complète du genou chez les patients âgés de plus de 40 ans révélait un RV positif de 5,8 (IC95%: 1,3-26,8) pour le diagnostic d'une déchirure méniscale. Seulement 15 % des patients ont été correctement identifiés à l'aide de cette combinaison [102]. Dans une autre étude, les auteurs ont recruté un échantillon de 174 patients âgés de plus de 45 ans qui présentaient une douleur au genou d'apparition progressive dont 64 d'entre eux souffraient d'une déchirure méniscale dégénérative [11]. Les auteurs ont conclu que les éléments suivants étaient utiles pour poser un diagnostic de déchirure méniscale dégénérative : la présence de douleur à la palpation des interlignes articulaires médiale ou latérale, une flexion passive complète du genou, l'apparition de douleur depuis moins d'un an, l'absence de désalignement du genou en varus, l'absence de pieds plats ainsi que l'absence d'amincissement des cartilages fémoro-tibial sur radiographies simples [11]. Différentes combinaisons de ces éléments présentaient une spécificité jusqu'à 98 % pour diagnostiquer une déchirure méniscale dégénérative et classifiaient correctement 28 % des cas [11]. Lors d'une dernière étude, 121 patients ont été recrutés et consultaient en première ligne pour une douleur au genou; 51 participants avaient une déchirure méniscale soit d'origine traumatique ou dégénérative [151]. Les auteurs de cette étude ont conclu que la combinaison d'être âgé de moins de 28 ans, d'être une femme, de ne pas avoir subi un traumatisme en mise en charge, de ne pas présenter une chaleur à la palpation, un épanchement ou un changement coloration de la peau au genou suivant un traumatisme ainsi qu'être en mesure de faire un squat complet sans douleur pouvait permettre l'exclusion d'une déchirure méniscale avec une sensibilité de 86%, tout en identifiant correctement 45% des patients n'ayant pas de déchirure méniscale [151].

2.3.3 Ostéoarthrose du genou

Le consensus d'experts de l'*European League Against Rheumatism (EULAR)* a émis une proposition pour définir les critères diagnostiques de l'OA du genou; ces critères ont été validés à partir de deux cohortes populationnelles (n=745, n=3456) [46]. Ce consensus indique qu'un individu âgé de plus de 45 ans a une probabilité de 99% d'être atteint d'OA radiologique de grade 2 et plus sur l'échelle de Kellgren-Lawrence s'il rapporte une douleur persistante au genou, une raideur matinale de courte durée, des limitations fonctionnelles en lien avec son genou et présente des crépitements et des ostéophytes à la palpation du genou, ainsi que des restrictions d'amplitude articulaire en flexion et/ou en extension [46].

2.3.4 Syndrome fémoro-patellaire

Deux études ont évalué la validité diagnostique de la combinaison d'éléments pour le diagnostic du SFP [100, 137]. Une étude a évalué un échantillon de 76 patients qui consultaient spécifiquement pour une douleur antérieure au genou, parmi ceux-ci, 52 patients présentaient un SFP [100]. Les auteurs rapportent que la combinaison de douleur lors de l'extension résistée isométrique, de douleur lors d'un squat et de douleur à la palpation des facettes rotuliennes présente un RV positif de 4,0 (IC95%: 1,8-10,3) pour le diagnostic de SFP, tout en permettant de classifier correctement 60% des cas [100]. Dans une seconde étude, d'autres auteurs ont démontré que la combinaison d'une série de tests de mobilité patellaire tel que la mobilité médio-latérale, supéro-inférieure, l'inclinaison du pôle inférieur de la rotule et la mobilité du tendon patellaire n'atteignait un RV positif que de 1,9 (IC95%: 0,5-7,7) [137].

2.4 Validité et fidélité inter-évaluateur du diagnostic émis par un physiothérapeute.

Les auteurs d'une revue systématique sur le rôle du physiothérapeute en pratiques avancées rapportent que l'accord inter-évaluateur entre le diagnostic émis par un physiothérapeute et celui émis par un chirurgien orthopédique varie de bon à presque parfait (kappa de Cohen (κ)=0,69 à 1,0) [26]. D'autres auteurs ont évalué l'accord entre un diagnostic basé sur l'évaluation musculosquelettique réalisé par différents professionnels de la santé et celui

proposé par l'IRM. Ils ont démontré que les physiothérapeutes avaient un diagnostic équivalent à celui des chirurgiens orthopédiques et cet accord était supérieur à celui d'autres cliniciens de première ligne non spécialisés en musculosquelettique (accord diagnostique entre l'évaluation musculosquelettique et l'IRM: physiothérapeutes : 74,5 %, orthopédistes : 80,8 %, médecins ou infirmières non spécialisés : 34,5 % [24]). Dans le contexte spécifique du diagnostic des pathologies au genou, une étude québécoise a évalué une cohorte de 109 patients [27]. Les auteurs ont démontré que l'accord diagnostique entre un physiothérapeute et des chirurgiens orthopédiques était presque parfait pour différentes pathologies communes au genou (κ=0.87; IC95% : 0,79-0,94) [27].

Les physiothérapeutes inclus dans les études précédentes possédaient une expérience clinique de plusieurs années et avaient aussi accès aux résultats des tests d'imagerie médicale pour émettre leur diagnostic [26, 27]. Ceci ne représente pas le contexte de pratique actuel de la plupart des physiothérapeutes qui évaluent une pathologie au genou en exécutant uniquement une évaluation musculosquelettique sans imagerie pour émettre une impression diagnostique et orienter le plan de traitement des patients qu'ils prennent en charge [27, 152]. Les recherches futures au sujet de la validité et de la fidélité inter-évaluateur du diagnostic émis par le physiothérapeute devraient se concentrer sur l'évaluation musculosquelettique sans l'utilisation de l'imagerie et dans divers contextes cliniques.

2.5 Conclusions de la revue de la littérature et limites des études dans le domaine.

La revue systématique portant sur la validité des tests de l'examen physique au genou a permis de conclure qu'aucun test de l'examen physique exécuté de façon individuelle, à l'exception du test de Lachman pour les déchirures du LCA, ne semble être suffisamment valide [153]. La revue systématique sur la fidélité de l'examen physique au genou a permis de conclure que l'atteinte d'une fidélité inter-évaluateur modérée pour la plupart des tests de l'examen physique était possible et pouvait dépendre de la standardisation des techniques et de la définition précise de ces tests [154].

Ces deux revues systématiques ont aussi permis de mettre en évidence des biais méthodologiques fréquents touchant les études diagnostiques qui y étaient incluses. Parmi ces biais, on retrouve notamment l'utilisation d'un devis rétrospectif et/ou cas-contrôle, l'absence d'indépendance entre les évaluateurs en lien avec le standard de référence et les tests index ainsi que de l'incorporation des tests index dans le standard de référence (biais d'incorporation) [98]. Ces biais peuvent augmenter les estimations de validité diagnostique, le biais cas-contrôle étant le plus important avec une surestimation pouvant atteindre 50% [98]. Le recrutement d'un échantillon de patients non-représentatif de la population cible (biais de spectre), la description inadéquate du standard de référence et des tests index ainsi que le recours à différents standards de référence (biais de différentiation ou de vérification partielle) font aussi partie des limites méthodologiques communes des études antérieures [98]. La plupart des études semblent aussi souffrir d'une faible taille d'échantillon avec un nombre limité de cas pour les analyses statistiques proposées, ce qui se traduit par une diminution de la précision des estimations de validité diagnostique [60].

Les données probantes sur la combinaison de l'anamnèse et de l'examen physique sont actuellement limitées et incomplètes. Par contre, on observe en clinique que des cliniciens experts ont recours à une combinaison d'éléments de l'anamnèse et l'examen physique pour émettre un diagnostic valide [6]. Ainsi, des études de meilleure qualité touchant l'évaluation musculosquelettique permettraient potentiellement de développer des outils valides pouvant améliorer le diagnostic différentiel des pathologies communes au genou, ce qui pourrait potentiellement mener à une prise en charge plus efficiente de ces patients.

2.6 Objectifs et hypothèses de la thèse de doctorat.

Deux objectifs sont donc proposés à l'intérieur de cette thèse:

1. Évaluer l'accord diagnostique entre un physiothérapeute utilisant une évaluation musculosquelettique standardisée sans imagerie et des médecins experts pour les différentes pathologies communes au genou.
2. Évaluer la validité de la combinaison de l'anamnèse et de l'examen physique afin de développer une série d'outils valides permettant d'orienter le diagnostic différentiel des pathologies communes au genou.

Deux hypothèses seront vérifiées :

1. Basé sur les études antérieures traitant de l'accord diagnostique entre un physiothérapeute en pratiques avancées et des chirurgiens orthopédiques [26, 27], il est attendu que le physiothérapeute utilisant une évaluation musculosquelettique standardisée sans imagerie atteindra un accord diagnostique élevé avec les médecins experts pour les différentes pathologies communes au genou.
2. Basé sur les études combinant les éléments de l'anamnèse et l'examen physique [101, 102, 149], il est attendu qu'il sera possible de développer une série d'outils valides permettant d'orienter le diagnostic différentiel des pathologies communes au genou.

Les objectifs spécifiques découlant de chacun des cinq articles présentés dans la section résultats et répondant aux deux objectifs sont :

Article 3 :

1. Évaluer l'accord diagnostique entre un physiothérapeute utilisant une évaluation musculosquelettique standardisée sans imagerie et des médecins experts pour les différentes pathologies communes au genou.
2. Évaluer l'accord entre un physiothérapeute et des médecins experts concernant le triage chirurgical des patients suite à une première consultation.

Articles 4 à 7 :

3. Évaluer la validité de la combinaison de l'anamnèse et de l'examen physique afin de développer une série d'outils valides permettant d'orienter le diagnostic différentiel de la déchirure partielle et complète du ligament croisé antérieur (Article 4), de la déchirure méniscale d'origine traumatique ou dégénérative (Article 5), de l'ostéoarthrose du genou (Article 6) et du syndrome fémoro-patellaire (Article 7).

Chapitre 3 : Méthodologie

3.1 Devis de l'étude et milieux de recrutement

La méthodologie de l'étude ainsi que la présentation des résultats suivent les normes en vigueur (http://www.equator-network.org) pour un devis d'étude diagnostique prospective tel que proposé par les outils d'évaluation méthodologique *QUADAS* [98, 110, 155, 156] et *Standards for Reporting Diagnostic Accuracy Studies 2015 (STARD)* [157, 158]. Le comité d'éthique de l'Hôpital Maisonneuve-Rosemont (Centre intégré universitaire de santé et de services sociaux (CIUSSS) de l'Est de l'Île de Montréal) ainsi que le comité d'éthique du CIUSSS de la Capitale-Nationale (Québec) ont approuvé cette étude multicentrique (annexe 20). Tous les participants ont signé un formulaire d'information et de consentement (annexe 21).

Tous les nouveaux patients référés à l'un des médecins experts associés au projet pour une douleur ou des symptômes au genou étaient éligibles pour participer à l'étude. Nous avons recruté consécutivement des nouveaux patients de novembre 2014 à août 2016. Ceux-ci consultaient dans deux cliniques externes de chirurgie orthopédique (Hôpital Maisonneuve-Rosemont, Montréal, et Centre Hospitalier de l'Université Laval (CHUL), Québec), ainsi que deux unités de médecine familiale (UMF Maisonneuve-Rosemont, Montréal, et UMF Laurier, Québec). Aussi, en septembre 2015, nous avons envoyé par courriel une invitation aux membres de la communauté universitaire de l'Université Laval, incluant les étudiants et les membres du personnel, afin de recruter des volontaires qui souhaitaient consulter un professionnel de la santé pour une douleur ou des symptômes au genou et pour lesquels ils n'avaient pas encore reçu de diagnostic.

3.2 Sélection des participants

Les critères d'éligibilité étaient les suivants : 1- être âgé de 18 ans et plus ; 2- avoir besoin de consulter pour une douleur ou des symptômes au(x) genou(x) ; 3- comprendre et parler le français ; 4- être résident de la province de Québec et couvert par la Régie de l'Assurance Maladie du Québec (RAMQ) ; 5- être en mesure de consentir légalement à la participation au projet. Les participants étaient exclus si : 1- ils avaient déjà reçu un diagnostic pour leur

douleur ou symptômes au genou par un des médecins experts associés au projet ; 2- s'ils avaient subi une chirurgie au membre inférieur dans les six derniers mois ; 3- s'ils étaient porteurs d'une prothèse au genou, 4- s'ils présentaient plus de deux autres pathologies aux membres inférieurs ; 5- s'ils souffraient d'un trouble inflammatoire systémique diagnostiqué (ex. : polyarthrite rhumatoïde).

3.3 Collecte des données

3.3.1 Portrait clinique des participants et anamnèse

Les caractéristiques sociodémographiques des participants ont été collectées par le physiothérapeute (annexe 3) et incluaient : le sexe, l'âge, le statut d'emploi, les demandes physiques de l'emploi et la présence de comorbidités. Les patients devaient rapporter les éléments de l'anamnèse (annexe 4) suivants : le/les genou(x) douloureux (gauche, droit, bilatéral), l'apparition d'origine traumatique ou d'apparition progressive, la durée de la douleur (en semaine ou en mois), la localisation de la douleur (antérieure, postérieure, médiale, latérale, diffuse), le temps de marche avant l'apparition de la douleur (en minutes) et la sensation d'engourdissement ou de picotement au membre inférieur. Le type de mécanisme traumatique proposé pouvait être une histoire de chute sur le genou en mise en charge, un traumatisme avec force externe au genou, un traumatisme suivant un saut, un traumatisme avec pivot, un traumatisme avec le pied bloqué au sol, une sensation de « pop » au moment du traumatisme, une douleur immédiate ou d'apparition tardive, la possibilité ou non de poursuivre les activités suite au traumatisme et la présence ou l'absence d'un épanchement articulaire immédiat suite au traumatisme [101].

Le physiothérapeute a ensuite administré aux participants le *Knee Injury and Osteoarthritis Outcome Score* (KOOS) (annexe 5). Ce questionnaire comprend 42 questions réparties en cinq domaines qui évaluent la douleur, les symptômes, la fonction dans les activités de la vie quotidienne, la fonction dans les activités physiques et sportives et la qualité de vie [159-162]. Ce questionnaire est validé pour différentes pathologies au genou contrairement à d'autres questionnaires pour des pathologies uniques au genou [161]. Les participants ont aussi

répondu au questionnaire Kessler-6 qui évalue sommairement la présence d'une détresse psychologique sévère (annexe 6) [163]. Les réponses aux questions individuelles des questionnaires ont été analysées afin de vérifier leur validité pour les différents diagnostics. Les réponses aux éléments de l'anamnèse sont détaillées aux annexes correspondantes (annexes 3 à 6). Les données anthropométriques (poids et taille) afin de calculer l'indice de masse corporelle (IMC=poids (kg)/taille (m^2)) ont aussi été recueillies. Les amplitudes articulaires actives et passives en flexion et en extension du genou (en degrés) à l'aide d'un goniomètre ont aussi été mesurées; celles-ci ont été ensuite dichotomisées comme étant limitées ou non en comparaison avec le membre inférieur non affecté [164, 165].

3.3.2 Sélection des tests de l'examen physique

Nous avons développé l'examen physique standardisé sur la base des revues systématiques (Chapitre 2), de livres de référence sur le diagnostic musculosquelettique [77, 78, 113], ainsi que de l'expertise clinique des médecins associés au projet. L'utilisation de diverses sources a permis de considérer plusieurs techniques et définitions pour les éléments de l'anamnèse et de l'examen physique. Précédant la collecte des données, les médecins experts ont rencontré l'équipe de recherche afin de standardiser les techniques d'évaluation des tests de l'examen physique ainsi que la définition des diagnostics d'intérêts. Nous avons produit un guide d'évaluation qui détaille les techniques et les résultats aux différents tests (annexe 7).

Le physiothérapeute a exécuté tous les tests sur tous les patients, indépendamment des résultats à l'anamnèse. Les médecins experts pouvaient sélectionner seulement les tests cliniquement nécessaires pour obtenir un diagnostic, mais ces tests devaient être exécutés selon la méthode standardisée. Les évaluateurs avaient la possibilité d'utiliser d'autres tests non standardisés s'ils n'étaient pas en mesure de poser un diagnostic qu'ils considéraient valide suivant leur évaluation initiale.

L'examen physique standardisé comprenait les tests suivants : un examen palpatoire (reproduction des symptômes douloureux) des structures suivantes : tendon patellaire, apex

patellaire, facettes patellaires médiale et latérale, interlignes articulaires médiales et latérales, tendons et bourse de la patte d'oie, tête du péroné, creux poplité, tendons des ischio-jambiers (semi-membraneux, semi-tendineux et biceps fémoral), la présence de chaleur à la palpation, une évaluation visuelle de l'alignement des membres inférieurs (valgus/varus/recurvatum/normal) et une vérification de l'alignement patellaire (rotation interne/externe, alta, normal). Les évaluateurs ont eu recours à des tests spécifiques pour les différentes pathologies du genou. Pour évaluer un épanchement intra-articulaire, ils ont utilisé le test du flot et le test du glaçon [103]. Pour évaluer une déchirure du LCA, ils ont utilisé le test du tiroir antérieur, le test de Lachman et le test du pivot shift [34] . Pour évaluer le ligament croisé postérieur et les ligaments collatéraux médial et latéral, ils ont eu recours au test du tiroir postérieur [90] et aux tests de stress en varus et en valgus à 0° et 30° de flexion [104]. Pour évaluer une déchirure méniscale, ils ont exécuté le test de McMurray et le test de Thessaly [70]. Pour évaluer l'OA du genou, ils ont évalué la palpation de crépitements dans les compartiments fémoro-patellaire et tibio-fémoral lors de la mobilisation du genou [6, 46]. Pour le SFP, ils ont évalué la présence de douleur à la contraction résistée en extension à 90° et 30° degrés de flexion du genou, la présence de douleur à la compression patellaire, la présence d'hypermobilité patellaire médio-latérale, la présence de crépitements à la mobilisation patellaire, le signe de Clark, le test d'appréhension patellaire et le signe en J à la contraction patellaire en extension [74, 100]. Les évaluateurs ont aussi utilisé le squat et la descente d'une marche pour évaluer la douleur et le contrôle moteur du membre inférieur lors d'une tâche fonctionnelle (aptitude à garder un bon alignement pendant le mouvement). Les tests étaient cotés positifs, négatifs, incertains ou non évalués.

3.3.3 Procédure pour la collecte des données de l'examen physique

Deux évaluateurs ont évalué chaque participant de façon indépendante : un physiothérapeute (SD) et un des médecins experts (MF, SB, PF, BP, PAV). Les deux cliniciens étaient à l'aveugle aux résultats de leurs évaluations respectives. Le physiothérapeute demeurait en tout temps à l'aveugle des résultats aux tests d'imagerie. Ils ont procédé à leurs

évaluations, l'une à la suite de l'autre, pour chacun des participants. Le physiothérapeute a procédé à son évaluation avant celle du médecin expert pour tous les patients. Suite à l'évaluation par le physiothérapeute, le patient devait indiquer si cette évaluation avait augmenté sa douleur (pas du tout ou un peu, modérément, beaucoup plus élevée). Pour assurer des conditions équivalentes entre le physiothérapeute et le médecin, la douleur rapportée devait être « pas du tout » ou « un peu plus élevée » suite à une pause de 15 minutes en comparaison avec la douleur au début de l'évaluation, sinon le patient était alors exclu de l'étude. Aucun des patients évalués n'a été exclu de l'étude. Suite à son évaluation musculosquelettique standardisée, le physiothérapeute devait indiquer son diagnostic primaire et si nécessaire son diagnostic secondaire, ainsi que le type de prise en charge proposée, soit une prise en charge chirurgicale ou une approche conservatrice (triage).

Le physiothérapeute possédait une maitrise professionnelle en physiothérapie et un an d'expérience clinique au moment de débuter la collecte de données. Il a développé le protocole de standardisation de l'évaluation musculosquelettique en collaboration avec les médecins experts. Cinq médecins experts ont participé au développement du protocole et à l'évaluation des patients, dont trois chirurgiens orthopédiques et deux médecins du sport, chacun possédant plus de 20 ans d'expérience.

3.4 Définition du standard de référence

Nous avons comparé les combinaisons d'éléments de l'anamnèse et de l'examen physique recueillis indépendamment par le physiothérapeute au standard de référence du diagnostic final des médecins experts. Les médecins experts ont émis leur diagnostic en trois étapes. D'abord, ceux-ci procédaient de façon indépendante à leur évaluation musculosquelettique, et ce à l'aveugle des résultats du physiothérapeute et des tests d'imagerie médicale. Ensuite, les tests d'imagerie et rapports des radiologistes étaient présentés au médecin expert afin qu'il puisse faire sa propre analyse de ses résultats. Tous les participants devaient avoir une radiographie du genou en mise en charge qui comprenait les vues antéro-postérieures, latérales et tangentielles de la rotule [166]. Un test d'imagerie par résonnance magnétique

(IRM) était prescrit pour confirmer tous les diagnostics de déchirure ligamentaire, de déchirure méniscale ou pour exclure tout autre diagnostic suite à l'évaluation musculosquelettique des médecins [167]. Si des résultats d'IRM (images et rapports) d'au plus six mois et de trois mois pour les radiographies étaient disponibles au dossier au moment de la consultation, les médecins pouvaient utiliser ceux-ci. La qualité de l'image devait aussi être suffisante pour permettre aux médecins de procéder à leur propre analyse, sinon ceux-ci devaient demander un nouveau test d'imagerie.

La dernière étape du diagnostic du médecin expert consistait à combiner l'information obtenue lors de l'anamnèse, de l'examen physique et de l'imagerie médicale pour émettre un diagnostic final primaire et un diagnostic secondaire si pertinent. Pour qu'un diagnostic secondaire soit jugé valide, le médecin devait considérer que ce diagnostic était cliniquement important et modifiait le pronostic ou la prise en charge. Le diagnostic final du médecin expert était considéré comme le standard de référence [168-170].

3.5 Analyses statistiques

Nous avons utilisé des statistiques descriptives, incluant les tests de t de Student et de Chi-carré pour comparer les caractéristiques des participants pour chacun des diagnostics. Nous avons classifié tous les diagnostics émis par le physiothérapeute et les médecins experts selon les pathologies au genou suivantes : 1- déchirure partielle ou complète du LCA, 2- déchirure méniscale d'origine traumatique ou dégénérative, 3- OA du genou, 4- SFP et 5- autres diagnostics.

3.5.1 Évaluation de la fidélité inter-évaluateur

Nous avons évalué la fidélité inter-évaluateur entre le physiothérapeute et les médecins experts pour le diagnostic des différentes pathologies du genou ainsi que pour les résultats aux différents tests de l'examen physique. Nous avons calculé l'accord brut (%), le coefficient kappa de Cohen (κ) et le coefficient kappa ajusté pour la prévalence ou le biais d'évaluateur (*Prevalence and Bias Adjusted Kappas* (PABAK)) ainsi que les intervalles de confiance à 95%

[171]. Le coefficient kappa de Cohen présente l'accord inter-évaluateur en corrigeant pour la possibilité d'obtenir une réponse concordante entre deux évaluateurs simplement par chance [171]. Le coefficient PABAK corrige aussi les situations où les données présentent une proportion élevée de réponses semblables au test à l'étude; la faible variabilité de réponses à un test pouvant amener une distorsion du coefficient de kappa [105, 171]. Cette analyse permet aussi de vérifier la présence d'un biais d'évaluateur, c'est-à-dire si un évaluateur démontre une tendance à répondre systématiquement de façon différente des autres évaluateurs [105, 171]. L'interprétation de la fidélité inter-évaluateur et de ses coefficients est basée sur l'échelle de *Landis et Koch* [116, 172]. Selon cette échelle, un coefficient kappa de 0 indique un accord inexistant, 0-0,2 faible, 0,2-0,4 passable, 0,4-0,6 modéré, 0,6-0,8 substantiel ou bon et >0,8 presque parfait ou excellent [116, 172]. Certains auteurs proposent un seuil de $\kappa \geq 0,6$ pour l'accord sur le diagnostic et de $\kappa \geq 0,4$ pour l'accord à un test de l'examen physique comme représentant un seuil minimal cliniquement acceptable pour la variabilité inter-évaluateurs [116, 172].

3.5.2 Développement des modèles diagnostiques combinant l'anamnèse et l'examen physique

Une particularité du diagnostic des pathologies musculosquelettiques est qu'il requiert la combinaison de plusieurs éléments de l'anamnèse et de l'examen physique pour faire une prédiction valide [82, 173]. Le cadre conceptuel d'analyse statistique proposé est celui du développement des « règles de prédiction clinique » (RPC) visant à développer le meilleur modèle prédictif pour identifier un sous-groupe ayant un diagnostic d'intérêt [60, 174-176]. Le développement d'une RPC suit une séquence débutant par la dérivation initiale et la validation interne d'un modèle combinant plusieurs éléments [60]. Ainsi, nous avons combiné deux analyses statistiques de façon séquentielle pour effectuer la dérivation des modèles diagnostiques combinant l'anamnèse et l'examen physique [177]. La première étape consistait à appliquer un modèle de régression logistique pénalisée (Least Absolute Shrinkage and Selection Operator (LASSO)) sur toutes les variables de l'évaluation musculosquelettique afin d'identifier les variables démontrant la meilleure association avec les diagnostics

d'intérêts [177, 178]. En comparaison avec une régression logistique non pénalisée, l'utilisation du LASSO est proposée pour identifier les meilleurs prédicteurs parmi un large éventail de variables [177]. Dans la présente étude, le physiothérapeute a collecté toutes les questions de l'anamnèse et tous les tests de l'examen physique pour tous les participants [98]. Ce faisant, tous les patients démontraient une caractéristique commune : celle de consulter pour une douleur ou des symptômes au genou. Ceci contraste avec les protocoles dans lesquels des patients sont comparés avec des sujets sains, un biais potentiel pouvant surestimer la validité diagnostique [98]. Nous avons obtenu la sélection finale des variables du LASSO en appliquant un coefficient de pénalité *lambda (*λ) aux coefficients bêta (ß) de la régression logistique. Cette méthode diminue graduellement les coefficients ß vers zéro pour les variables n'étant pas les meilleurs prédicteurs des diagnostics d'intérêts [179]. Nous avons utilisé l'aire sous la courbe maximale (Area Under the Curve, AUC) comme facteur pour identifier le coefficient de pénalité *lambda* approprié lors de la validation croisée (« *cross-validation* ») [179]. Ceci a permis d'obtenir une série de variables qui démontraient la plus forte association avec les diagnostics d'intérêts [179].

La deuxième étape du développement des modèles diagnostiques consistait à trouver la combinaison optimale des variables identifiées précédemment pour inclure ou exclure chacun des diagnostics d'intérêts. Pour ce faire, nous avons utilisé la méthode de partition récursive, aussi connue sous le nom de *Classification and Regression Trees* (CART) [60, 180]. La partition récursive utilise un algorithme automatisé qui identifie la meilleure séquence hiérarchique d'une série de variables pour classifier les cas et les autres pathologies (non-cas) [177, 181]. L'algorithme construit ainsi un « arbre » décisionnel en créant des « branches » qui contiennent les résultats aux différentes variables et classifient un groupe de patients le plus homogène à l'aide de leurs réponses aux variables [177, 180, 182]. Chaque branche correspond donc à une combinaison d'éléments de l'anamnèse et/ou de l'examen physique et permet de classifier les cas et les non-cas. Plusieurs combinaisons peuvent être possibles lorsque plusieurs variables possèdent une valeur hiérarchique semblable dans la partition récursive. Nous avons aussi sélectionné les différentes combinaisons présentées suivant une

réflexion sur leur applicabilité clinique et leur fidélité inter-évaluateur [60, 183]. Chaque article contient une schématisation des différents modèles pour faciliter le diagnostic différentiel.

3.5.3 Évaluation de la validité des modèles diagnostiques

Nous avons mesuré la validité de l'évaluation musculosquelettique complète exécutée par le physiothérapeute en comparant le diagnostic émis sur la base de son évaluation sans imagerie au diagnostic combiné émis par les médecins experts pour chacun des diagnostics d'intérêts. Nous avons mesuré la validité des modèles diagnostiques combinant l'anamnèse et l'examen physique, développés à l'aide des éléments spécifiques extraits de l'évaluation musculosquelettique du physiothérapeute, en comparant la classification des modèles au standard de référence du diagnostic combiné des médecins experts pour chacun des diagnostics d'intérêts. Nous avons calculé les statistiques diagnostiques suivantes : la sensibilité (Se), la spécificité (Sp), les valeurs prédictives positive et négative (VPP/VPN) et les rapports de vraisemblance positifs et négatifs (RV+/-) [60, 184]. La Se et Sp sont mesurées respectivement à l'aide de la proportion entre les vrais positifs et les faux négatifs et entre les vrais négatifs et les faux positifs [71, 82, 185]. Ces deux statistiques correspondent respectivement au pourcentage d'identification des cas et des non-cas d'un test ou d'une combinaison de tests [71, 82, 185]. Les valeurs prédictives positives et négatives (VPP/VPN) représentent plutôt la probabilité d'un patient d'être atteint ou non de la pathologie s'il répond positivement ou négativement à un test ou à une série de tests [70, 82, 184, 185]. Les estimations de ces statistiques sont directement influencées par la prévalence des pathologies, donc la probabilité pré-test d'un échantillon, ce qui peut limiter l'applicabilité dans divers contextes cliniques [70, 82, 184, 185]. Les rapports de vraisemblance (RV) positifs et négatifs représentent aussi la probabilité d'un patient d'être atteint ou non de la pathologie lorsqu'une combinaison d'éléments est positive ou négative. Par contre, ces statistiques sont particulièrement utiles pour l'interprétation clinique des résultats et pour guider la prise de décision [70, 82, 184, 185]. Dans le cadre du présent projet, nous avons considéré qu'une combinaison d'éléments était valide si elle atteignait un RV positif supérieur

à 5 pour inclure une pathologie, ou inférieur à 0,2 pour l'exclure. Ces valeurs produisent un changement modéré de la probabilité post-test d'au moins 30% [70, 82].

En dernière étape, nous avons utilisé la technique « *randomForest »* pour effectuer la validation interne des différentes combinaisons d'éléments [177, 186]. Cette méthode statistique de type « bootstrapping » est spécifiquement conçue pour la partition récursive [177, 186]. Elle utilise un sous-échantillonnage (approximativement 10%) de l'échantillon initial pour produire une série de multiples modèles avec les variables initiales (ici, n=1000). Ceci permet de calculer une estimation globale de la validité interne des différentes combinaisons possibles pour un diagnostic d'intérêt, auquel les estimations des modèles choisis et leurs intervalles de confiance à 95% sont comparés [177, 186-188]. Les analyses statistiques ont été produites à l'aide du logiciel R version 3.3.0 (packages : epiR, rpart, glmnet and randomForest; http://cran.r-project.org/).

3.6 Taille d'échantillon

Aucun standard n'est actuellement disponible concernant la méthode optimale pour le calcul de la taille d'échantillon nécessaire pour les études diagnostiques [189-191]. Les propositions actuelles sont basées sur la précision des estimations de validité diagnostique et leurs intervalles de confiance à 95%. Les auteurs proposent d'obtenir un minimum de 30 à 60 cas d'un diagnostic d'intérêt pour être en mesure d'effectuer une régression logistique non pénalisée [60]. Cette estimation implique un minimum de 10 à 15 cas par variable incluse dans un modèle diagnostique qui combine trois ou quatre variables indépendantes [60]. La méthode LASSO et la partition récursive sont des méthodes statistiques qui permettent la classification des cas et des non-cas sur des échantillons encore plus restreints [177]. La méthode de recrutement utilisée dans ce projet ne procédait pas à une sélection préalable des participants, la taille d'échantillon devenant donc dépendante du diagnostic le moins prévalent soit les déchirures du LCA [27]. Par contre, un patient peut fournir plus d'un diagnostic si le médecin expert détermine la présence d'un diagnostic secondaire cliniquement pertinent, permettant ainsi de diminuer le nombre de patients total à recruter.

Chapitre 4 : Résultats

Les résultats de cette thèse sont présentés sous la forme de cinq articles scientifiques (article 3 à article 7). L'article 3 présente les résultats sur la validité et la fidélité inter-évaluateur du diagnostic émis par le physiothérapeute sur la base d'une évaluation musculosquelettique standardisée sans imagerie comparé au diagnostic émis par des médecins experts comprenant leur interprétation des tests d'imagerie. Les articles 4 à 7, basés sur la même cohorte de participants, présentent les résultats sur la validité de la combinaison des éléments de l'anamnèse et de l'examen physique spécifiquement pour les pathologies du genou déjà décrites. Les articles 4 et 5 traitent des pathologies d'origine traumatique, soit des déchirures partielle ou complète du LCA et des déchirures méniscales traumatiques. La 2^{e} partie de l'article 5 amorce l'exploration des pathologies d'apparition progressive par les déchirures méniscales dégénératives. L'article 6 traite de l'OA du genou. L'article 7 termine la section des résultats et traite du SFP. Les modèles diagnostiques développés ainsi que la fidélité inter-évaluateur des différents tests de l'examen physique inclus dans ces modèles sont présentés pour chacun des articles.

Article 3: Diagnostic validity and triage concordance of a physiotherapist compared to physicians' diagnoses for common knee disorders

Décary S, Fallaha M, Pelletier B, Frémont P, Martel-Pelletier J, Pelletier J-P, Feldman DE, Sylvestre M-P, Vendittoli P-A, Desmeules F.

Simon Décary[1,2]
Michel Fallaha[3]
Pelletier, Bruno[3]
Frémont, Pierre[4]
Martel-Pelletier, Johanne[6]
Pelletier, Jean-Pierre[6]
Feldman, Debbie[1]
Sylvestre, Marie-Pierre[5]
Pascal-André Vendittoli[2,3]
François Desmeules[1,2]

1- School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada.
2- Orthopaedic Clinical Research Unit, Maisonneuve-Rosemont Hospital Research Center, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
3- Department of Surgery, Maisonneuve-Rosemont Hospital, University of Montreal, Montreal, Quebec, Canada, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
4- Department of Rehabilitation, Faculty of Medicine, Laval University, Quebec City, Quebec, Canada.
5- University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada. Department of Social Preventive Medicine, School of Public Health, Université de Montréal
6- Osteoarthritis Research Unit, University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada.

Article publié dans la revue *BMC Musculoskeletal disorders*. 2017. doi: 10.1186/s12891-017-1799-3.

En tant que premier auteur, j'ai réalisé le développement du protocole de cette étude, la collecte des données, l'analyse des résultats ainsi que la rédaction du manuscrit. FD et PAV ont contribué au développement du protocole, à l'analyse des résultats, à la rédaction et à la

révision du manuscrit. MF, BP et PF ont contribué au développement de l'évaluation musculosquelettique standardisée et ont agi à titre de médecin experts (standard de référence) pour tous les patients. JMP, JPP, DF et MPS ont contribué au développement du protocole, à la vérification des méthodes d'analyse des données et à la présentation des résultats. Tous les auteurs ont contribué à la révision de la version finale du manuscrit.

ABSTRACT

Background: Emergence of more autonomous roles for physiotherapists warrants more evidence regarding their diagnostic capabilities. Therefore, we aimed to evaluate diagnostic and surgical triage concordance between a physiotherapist and expert physicians and to assess the diagnostic validity of the physiotherapist's musculoskeletal examination (ME) without imaging.

Methods: This is a prospective diagnostic study where 179 consecutive participants consulting for any knee complaint were independently diagnosed and triaged by two evaluators: a physiotherapist and one expert physician (orthopaedic surgeons or sport medicine physicians). The physiotherapist completed only a ME, while the physicians also had access to imaging to make their diagnosis. Raw agreement proportions and Cohen's kappa (k) were calculated to assess inter-rater agreement. Sensitivity (Se) and specificity (Sp), as well as positive and negative likelihood ratios (LR+/-) were calculated to assess the validity of the ME compared to the physicians' composite diagnosis.

Results: Primary knee diagnoses included anterior cruciate ligament injury (n=8), meniscal injury (n=36), patellofemoral pain (n=45) and osteoarthritis (n=79). Diagnostic inter-rater agreement between the physiotherapist and physicians was high (k=0.89; 95%CI:0.83-0.94). Inter-rater agreement for triage recommendations of surgical candidates was good (k=0.73; 95%CI:0.60-0.86). Se and Sp of the physiotherapist's ME ranged from 82.0 to 100.0% and 96.0 to 100.0% respectively and LR+/- ranged from 23.2 to 30.5 and from 0.03 to 0.09 respectively.

Conclusions: There was high diagnostic agreement and good triage concordance between the physiotherapist and physicians. The ME without imaging may be sufficient to diagnose or exclude common knee disorders for a large proportion of patients. Replication in a larger study will be required as well as further assessment of innovative multidisciplinary care trajectories to improve care of patients with common musculoskeletal disorders.

INTRODUCTION

Knee disorders are a common reason for seeking diagnosis and management in primary care and can significantly impact quality of life of individuals [1, 6, 192]. However, evidence shows the limited ability of medical providers to perform an appropriate physical examination to make a diagnosis [3, 24]. This has led to an overreliance on imaging or inappropriate referral to specialists to confirm a diagnosis, which incurs increasing health care costs and unnecessary delays to initiate conservative care [3, 9, 24, 27, 193]. Models of care in which physiotherapists act as first contact providers have been proposed [3, 99, 194, 195]. In these models, physiotherapists act as consultants who evaluate the patient, make a diagnosis and offer conservative care or refer to other providers [3, 99, 194, 195].

To adequately take on these autonomous roles, physiotherapists need to be able to provide a valid clinical diagnostic impression and be able to refer accurately patients to other providers or surgical candidates to orthopaedic surgeons; they need do this in manner that is as effective as physicians with expertise in musculoskeletal disorders would do [6]. *Moore et al.* demonstrated the equivalence between physiotherapists and orthopaedic surgeons for the clinical diagnosis of common musculoskeletal disorders [24]. When compared to magnetic resonance imaging (MRI) results, the diagnostic agreement of the physiotherapists was almost as high as the orthopaedic surgeons (raw agreement: 74.5% compared to 80.8%), and it was superior to non-orthopaedic providers such as primary care physicians (35.4%) [24]. A systematic review reported that, based on moderate quality studies, inter-rater agreement kappa values ranged from 0.69 to 1.00 for diagnostic agreement between physiotherapists and orthopaedic surgeons and kappa values ranging from 0.52 to 0.70 for the triage of surgical candidates, indicating moderate to high agreement between providers [26].

However, in many primary care settings, imaging may be difficult to obtain rapidly or physiotherapists may not be allowed to order imaging and must therefore rely exclusively on musculoskeletal examination (ME) when assessing patients. *Jackson et al*. concluded in a meta-analysis including 35 diagnostic studies, not specific to physiotherapists, that a

complete ME demonstrates adequate validity to include or exclude common knee disorders when compared to imaging or arthroscopic findings [6, 196]. However, it is not known whether this also applies to physiotherapists.

Therefore, the objectives of this study were to 1) evaluate agreement on the diagnosis and surgical triage between a physiotherapist using a standardized ME without the use of imaging results and physicians and 2) to assess the validity of the physiotherapist's ME to diagnose common knee disorders.

METHODS

Study design and settings

This study is part of a larger multi-center prospective diagnostic cohort study that aims to identify the optimal combination of elements from the history and physical examination for the diagnosis of common knee disorders. Recruitment took place in an outpatient orthopaedic clinic and a primary care family medicine clinic. All consecutive patients consulting one of the participating physicians for a new knee complaint between November 2014 and January 2016 were recruited. Also, we included participants from a university community (students, teaching staff and other personnel) if they sought a diagnosis and care for a current knee complaint. These participants received an email invitation to participate from September 2015 to January 2016. The present study, its design, methodology and reporting of results is based on the Standards for Reporting Diagnostic Accuracy Studies 2015 (STARD) [157, 158]. The study was approved by the hospital's ethics committee. The study was explained by the physiotherapist to all participants and written informed consent was obtained from them prior to consultation.

Participants

Inclusion criteria were: 18 years of age or older, consulting for a knee complaint for which they sought diagnosis, and being able to understand and speak French. Patients previously diagnosed and treated by one of the participating physicians were excluded to ensure that

the patient did not reveal their previous diagnosis to the physiotherapist. We also excluded patients who had undergone lower limb surgery in the past six months, patients with a knee arthroplasty or who presented with more than two lower limb pathologies in addition to the one for which they were consulting or if they were diagnosed with any systemic inflammatory disorder.

Data collection procedure

Patients' characteristics and history elements

All clinical settings had the same data collection procedure. Upon arrival at the clinic, participants answered a questionnaire which included age, sex and anthropometric data (weight and height) to allow calculation of body mass index (BMI), duration of symptoms, history of the lesion (traumatic or non-traumatic), and presence of bilateral knee pain. Participants also completed the Knee Injury and Osteoarthritis Outcome Score (KOOS), a validated 42-item self-report questionnaire that assesses pain, symptoms, function in daily living, function in sport and recreation and knee-related quality of life [160-162]. Psychological distress was assessed using the Kessler-6 screening scale for serious mental disorders [163, 197].

Physical examination

Each participant was then independently assessed by two evaluators: a physiotherapist and one of the four physicians. The two evaluations were completed on the same day with a fifteen-minute interval between each. The physiotherapist always evaluated the participants prior to the physicians in a separate room. Both the physiotherapist and the physicians were blinded to each other results. Following the physiotherapist's ME, the patients' pain was evaluated using a three-point likert scale (light, moderate, severe) and they were withdrawn from the study if their pain was moderately or severely increased compared to the start of the evaluation. The physician then proceeded to his independent history taking and physical examination.

Diagnosis, reference standard and triage options

After independently seeing the patient, the physiotherapist and the physician each completed a separate form where they indicated their primary and, when applicable secondary diagnosis. The physiotherapist was blinded to imaging results and therefore determined his diagnosis on the sole basis of his ME [155].

As the reference standard, the physicians had access to imaging to establish their diagnosis and they performed their own analysis of the relevant imaging results. All participants were required to have a radiograph of their knee that included the following three views: weight-bearing antero-posterior view with lateral and skyline views [166]. Magnetic resonance imaging (MRI) was required when the physician suspected a ligament injury, a meniscal injury or any other uncertain diagnoses. If participants already had recent radiographs within 3 months of their participation or MRI results within 6 months with suitable views or scans that allow adequate interpretation and grading by the physician, these results were used. If the physician doubted that the imaging did not reflect the current stage of pathology another test was ordered. The physician made a final primary and secondary (if necessary) composite diagnosis based on the patient's history, physical tests and imaging results [198]. This final composite diagnosis was considered the reference standard against which the physiotherapist's diagnoses were compared to all participants. We also compared the physiotherapist's diagnoses to the imaging diagnoses only as a secondary reference standard [50, 168-170].

Lastly, both the physiotherapist and physicians independently selected the triage option - conservative, surgical or undecided - for the patients. For patients seen by the sports medicine physicians in the family medicine unit or from the university community, the patients were considered surgical cases if the physician considered that all other options would not be adequate and that requesting a surgical consultation was the proper conduct.

Standardisation and evaluators' experience

Before the start of the study, physicians met with the research personnel to familiarize with the study protocol and verify if their usual practice differed from the other evaluators to improve concordance. All practitioners participated in the standardization of the techniques, interpretation of the physical tests and definition of the related diagnoses and all agreed to comply with the proposed definitions during their respective evaluation. The physiotherapist had one year of clinical experience. The four participating physicians (two orthopaedic surgeons and two sports medicine physicians [6]) each had more than 20 years of experience in the diagnosis and management of knee disorders.

Sample size

We calculated the sample size to detect an overall inter-rater Kappa value for the overall diagnostic concordance greater than 0.80 assuming a two-tailed null hypothesis for a Kappa equal to 0.4 or less [172, 199, 200]. We estimated the proportions of positives agreement for knee disorders diagnoses and expected Kappa values based on a previous cohort from a similar setting recuited by our team [27]; the required sample size is set at 71 patients considering a 80% power [172, 199, 200].

Statistical Analysis

We used descriptive statistics to present the participants' characteristics. All primary diagnoses were classified using five common categories: 1- ACL injury; 2- meniscal injury; 3- patellofemoral pain; 4- osteoarthritis; 5- others [152, 171]. In the event where the physiotherapist and the physician disagreed on the primary diagnosis, the secondary diagnoses were taken into account to further evaluate diagnostic concordance. To measure the inter-rater agreement for the diagnostic categories and triage recommendations between the physiotherapist and the physicians, proportions of raw agreement and Cohen's Kappas with associated 95% confidence intervals (CI) were calculated. Because Kappa values may become biased due to high or low prevalence of concordant cases compared or non-concordant cases, bias index and prevalence index were calculated where 0 indicates no bias

and 1 a high bias [105, 171]. *Prevalence and Bias Adjusted Kappas* (PABAK) were calculated for each diagnostic category to correct for these potential biases [105, 171]. Interpretation of inter-rater agreement was made according to the Landis and Koch scale in which 0 indicates poor agreement, 0-0.2 slight, 0.2-0.4 fair, 0.4-0.6 moderate, 0.6-0.8 substantial or good and >0.8 almost perfect or high agreement [171, 172]. For the validity of the ME, we compared the final diagnosis proposed by the physiotherapists' ME with the physicians' composite final diagnosis (reference standard) based on the ME, radiographs and MRI results when needed. Sensitivity (Se), specificity (Sp) and likelihood ratios with 95% CIs were calculated [60, 184]. Se and Sp relate respectively to the proportion of true positives and true negatives when a test is performed [71, 82]. Positive and negative likelihood ratios were used to evaluate the diagnostic validity of the physiotherapist's physical examination compared to the physicians' composite diagnosis and the following cut-offs were used: 1-to include a disorder a LR+ ≥5 and 2- to exclude a disorder a LR- ≤0.2 as they are reported to produce at least a moderate shift in post-test probability of having or not a certain disorder [82, 173]. Analysis was performed using SPSS version 21 (SPSS Inc., Chicago) and R version 3.2.3 (packages *epiR, irr* and *psych*, http://cran.r-project.org/).

RESULTS

Table 14 presents the characteristics of participants. Out of 198 eligible patients, five (2.5%) refused to participate, 14 (7.1%) were excluded before consultation and 179 (90.4%) were included in the study (Appendix 8). None were excluded following the physiotherapist's evaluation because of increased pain or for any reasons. Mean age was 49.9±16.1 years old and most participants were female (63.7%) with a mean BMI of 29.1± 6.5 kg/m^2. The majority of participants were recruited from the orthopaedic clinic (79.3%) and consulted for a non-traumatic disorder (73.7%). Most participants had pain for over 3 months at the time of consultation (90.5%). KOOS *Sports* and *Quality of life* domains were most severely affected (31.4±24.8 and 40.9±20.3).

Table 14. Characteristics of participants (n=179)

Characteristics	n (%)	mean (SD)
Age (years)		49.9 (16.1)
Sex		
Female	114 (64)	
Male	65 (36)	
Body Mass Index (Kg\m^2)		29.1 (6.5)
Recruitment site		
Orthopaedic clinic	142 (80)	
Family medicine unit	15 (8)	
University community	22 (12)	
History of trauma	47 (26)	
Bilateral knee pain	39 (22)	
Duration of pain at time of consultation		
<3 months	17 (10)	
3-12 months	45 (25)	
≥ 12 months	117 (65)	
KOOS- Knee Injury and Osteoarthritis Outcome Score (%)		
Pain		58.6 (19.7)
Symptoms		71.0 (19.6)
Activity of Daily Living		66.1 (21.8)
Sports		31.4 (24.8)
Quality of Life		40.9 (20.3)
K6 psychological distress scale (/30)		26.0 (4.5)

SD=standard deviation; KOOS: 0 indicates a severe condition and 100 indicates a normal knee; K6: 6 indicates serious mental illness and 30 indicates no mental illness.

Primary clinical diagnoses made by the participating physicians (using ME and imaging) included: anterior cruciate ligament injury (ACL) (n=8), meniscal injury (n=36), patellofemoral pain (PFP) (n=45), osteoarthritis (OA) (n=79) and other diagnoses (n=11) (Table 15). All participants (n=179) had radiograph results and 70 participants had an MRI scan. Based on imaging results only, diagnoses included: OA (n=96), meniscal tears (n=54), ACL tears (n=16) or others (n=5).

Table 15. Clinical and imaging diagnoses of participants (n=179)

	n (%)
Primary clinical composite diagnoses	
Anterior cruciate ligament injury	8 (5)
Meniscal injury	36 (20)
Patellofemoral Pain Syndrome	45 (25)
Osteoarthritis	79 (44)
Other knee diagnoses	11 (6)
Imaging findings and diagnoses	
Osteoarthritis	96 (56)
K-L Grade 1	14 (15)
K-L Grade 2	36 (38)
K-L Grade 3	19 (20)
K-L Grade 4	13 (14)
Meniscal tears (n)	54 (32)
Medial meniscus	49 (91)
Lateral meniscus	8 (15)
Anterior cruciate ligament tears	16 (9)
Complete	6 (37)
Partial	8 (50)
Unclear	2 (13)
Posterior cruciate ligament tear	1 (1)
Soleus tear	1 (1)
Hamstring tendinopathy	2 (1)
Medial collateral ligament tear	1 (1)

SD=standard deviation; Clinical diagnoses are composite diagnoses made by physicians using both musculoskeletal examination and imaging; Others knee diagnoses included: contusion of the tibial plateau (n=2), PCL tear (n=1), soleus tear (n=1), psychosomatic origin (n=1), muscular spasms linked to multiple sclerosis (n=1), hamstring tendinopathy (n=3), medial collateral ligament injury (n=1), functional instability without meniscal or ACL injury (n=1); Imaging diagnoses are based on imaging studies using radiograph or magnetic resonance imaging; Grades are for Kellgren-Lawrence scale in the most affected compartment; Radiographic OA was defined as K-L$\geq$ 1.

Table 16 presents the concordance between the diagnosis made by the physiotherapist using only the ME and the composite diagnosis made by physicians using both ME and imaging or with the imaging diagnoses only. The overall raw agreement between the physiotherapist and the physicians' diagnosis was 92.2% with an high inter-rater agreement (κ=0.89, 95% CI : 0.83-0.94). Inter-rater agreement for specific knee disorders ranged from κ= 0.88 to 0.94. ACL injury and other diagnoses had fewer cases (8/179 and 11/179) which translated into a high prevalence index (0.91 and 0.89, respectively). However, all PABAK estimates were included in the Cohen's kappa 95% confidence intervals and were therefore not significantly different, which indicates that even when bias were present (i.e: prevalence of ACL injuries), this did not influence the Kappa estimate [171]. When comparing the physiotherapists' diagnosis with imaging only, raw agreement was slightly lower at 84.4% and inter-rater agreement was good (κ= 0.77; 95% CI: 0.68-0.85).

Table 16. Concordance between the physiotherapist and physicians' composite or imaging only diagnoses (n=179)

	Raw agreement	Cohen's kappa	95% CI	Bias Index	Prevalence Index	PABAK
Overall concordance with the physicians' composite diagnoses	92.2% (165/179)	0.89	0.83-0.94	-	-	-
ACL injury	100.0% (8/8)	0.94	0.82-1.00	0.01	0.91	0.99
Meniscal injury	97.2% (35/36)	0.88	0.80-0.93	0.03	0.57	0.92
Patellofemoral pain	91.1% (41/45)	0.88	0.80-0.96	0.05	0.50	0.91
Osteoarthritis	91.1% (72/79)	0.89	0.82-0.95	0.02	0.14	0.89
Other knee disorders	81.8% (9/11)	0.89	0.75-1.00	0.01	0.89	0.98
Overall concordance with imaging only diagnoses	84.4% (151/179)	0.77	0.68-0.85	-	-	-

ACL: anterior cruciate ligament. 95% CI : 95% confidence interval. PABAK is only calculated only for 2x2 tables. Others knee diagnoses included: contusion of the tibial plateau (n=2), PCL tear (n=1), soleus tear (n=1), psychosomatic origin (n=1), muscular spasms linked to multiple sclerosis (n=1), hamstring tendinopathy (n=3), medial collateral ligament injury (n=1), functional instability in the absence of ACL or meniscal injury (n=1).

Table 17 presents the diagnostic validity of the physiotherapist's standardized ME compared to the reference standard (physician's composite diagnosis) to discriminate between each knee disorders. Sensitivity ranged from 82 to 100% and was lowest for *Others* knee disorders. Specificity ranged from 96 to 100%. Positive likelihood ratio ranged from 23.2 to 267.6 and all 95% CI lower bounds were above 10.0. Negative likelihood ratio ranged from 0.00 to 0.18 and all 95% CI upper bounds were below LR-$\leq$ 0.20, except for *PFP* (LR-=0.23) and *Others* (LR-=0.64). This indicates that the standardized ME moderately to highly increases post-test probability to diagnose or exclude common knee disorders.

Table 17. Diagnostic validity of the musculoskeletal examination performed by the physiotherapist compared to the physicians' composite diagnosis

	Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio (95% CI)	Negative Likelihood Ratio (95% CI)
ACL injuries (n=8)	100.0% (52.0-100.0)	99.0% (97.0-100.0)	171.0 (24.2-1207.0)	0.00 (0.00-0.00)
Meniscal injuries (n=36)	97.0% (85.0-100.0)	96.0% (91.0-98.0)	23.2 (10.6-50.8)	0.03 (0.00-0.20)
Patellofemoral pain (n=45)	91.0% (79.0-98.0)	97.0% (93.0-99.0)	30.5 (11.6-80.5)	0.09 (0.04-0.23)
Osteoarthritis (n=79)	91.0% (83.0-96.0)	97.0% (91.0-99.0)	30.4 (10.0-92.8)	0.09 (0.04-0.19)
Others knee disorders (n=11)	82.0% (48.0-98.0)	100.0% (97.0-100.0)	267.6 (16.6-4325.6)	0.18 (0.05-0.64)

ACL: anterior cruciate ligament. 95% CI : 95% confidence interval.

Table 18 presents the concordance between the physiotherapist and the physician for the triage recommendation following consultation. Only 23 participants were considered as surgical candidates and six participants as uncertain by the physicians. Of the 23 patients considered surgical candidates, twenty were evaluated by the orthopaedic surgeons and three by the sports medicine physicians. Among those deemed surgical candidates, five had an ACL tear, seven a meniscal tear and eleven had an OA diagnosis. The overall agreement between the physiotherapist and all physicians was 91.6% with an inter-rater kappa of 0.73 (95% CI : 0.60-0.86). Raw agreement for surgical cases was 91.3% with only two of 23 surgical cases misclassified by the physiotherapist as conservative (Appendix 10). Raw agreement for conservative care was 92.6% with 11 of 150 conservative care cases misclassified by the physiotherapist as surgical cases.

Table 18. Concordance between the physiotherapist and physicians for the triage recommendation following consultation (n=179)

	Raw agreement	Cohen's kappa	95% CI
Overall	91.6% (164/179)	0.73	0.60-0.86
Surgical candidates	91.3% (21/23)		
Conservative care candidates	92.6% (139/150)		
Uncertain	66.7% (4/6)		

95% CI : 95% confidence interval.

DISCUSSION

The objectives of our study were to evaluate the diagnostic and surgical triage agreement between a physiotherapist and physicians to assess the validity of the physiotherapist's musculoskeletal examination (ME) without the use of medical imaging to diagnose common knee disorders. We found high diagnostic agreement and good triage agreement as well as high diagnostic validity for the ME performed by the physiotherapist in patients suffering from common knee disorders and consulting in primary and secondary care settings.

Our results compare well with two previous studies in orthopaedic settings where high inter-rater diagnostic agreement between a physiotherapist and orthopaedic surgeons for the diagnosis of common knee disorders were reported (κ= 0.80 and κ= 0.87 (95% CI: 0.79-0.94)) [27, 152]. Of note, in both these studies, the physiotherapist also had access to imaging results to support his ME which was not the case in our study [27, 152]. Only fourteen patients (n=14) out of 179 (7.8%) were discordant between the physiotherapist using ME compared with the physicians' composite diagnosis. Discordant patients included one meniscus injury, four patellofemoral pain, seven osteoarthritis, one patellar tendinopathy and one functional instability without meniscal or ACL injuries (Appendix 9). A possible cause includes a potentially more complex presentation (history and physical examination). Also, because our composite reference standard is done by only one expert, it is possible that the physiotherapist may not be the source of discordance. When comparing the physiotherapist's diagnosis to imaging diagnoses only, the agreement was somewhat lower supporting the notion that imaging results need to be corroborated with clinical findings from the ME and that these findings may be more important to make a diagnosis [152].

Another objective of our study was to evaluate the diagnostic validity of a ME performed without imaging support. This objective is important in the context where the need to rely on imaging may delay care either by the physiotherapist or a physician or the ordering of a given imaging may be altogether unnecessary to make a diagnosis and initiate appropriate care. Our results show that a ME performed by a physiotherapist without imaging could reach

moderate to high diagnostic validity to diagnose or exclude common knee disorders and these results are comparable to already published evidence. *Jackson et al*. reported in their meta-analysis LR+≥ 10 for lateral meniscus, ACL injuries and cartilage lesions and LR-≤ 0.20 for lateral and medial meniscus and LR- = 0.27 for ACL injury confirming the validity of a complete clinical examination performed by orthopaedic surgeons or sports medicine physicians [6]. What remains to be established is what is the optimal combination of history questions and physical examination tests results in the ME that is helpful to support the differential diagnosis of common knee disorders [100-104]. Nonetheless, the use of imaging may be warranted in more complex cases or to diagnose or exclude uncommon disorders when the expected recovery after initiation of care is not as predicted. In this situation, it is interesting to see that published evidence support physiotherapists to refer patients autonomously and appropriately to imaging [24, 27].

Our findings regarding concordance for triage recommendations of surgical candidates are consistent with three previous studies demonstrating good triage agreement with an orthopaedic surgeon (raw agreement: 87% and 91.8%; κ=0.77; 95%CI: 0.65-0.88, respectively) [27, 152, 201]. As stated above, the physiotherapist had access only to the ME without imaging to make the triage recommendation. Recent evidence proposes that pain, functional limitations and clinical symptoms should be used as surgical eligibility criteria for ACL injuries [14], meniscal injuries [202, 203] and knee OA [204, 205] and not systematically rely on imaging results. In our study, 82% (116/142) of secondary care participants and 92% (34/37) of primary care participants were referred to conservative care after their first consultation. Almost all of these patients (92.6%) would have been appropriately triaged to conservative care directly by the physiotherapist based only on the ME, making their care trajectory likely more efficient. Interestingly, this suggests that our cohort spectrum and representativeness is balanced between primary care and a pure secondary or tertiary surgical setting. Therefore, a well-executed ME may provide appropriate findings to guide patients to the appropriate care and these results suggests the role of physiotherapists as qualified musculoskeletal experts for knee disorders [201, 206].

Strengths and limitations

Our prospective cohort was recruited from three different settings, both in primary and secondary care, allowing for a broad variety of patients with various knee disorders, thus limiting spectrum bias commonly encountered in other diagnostic studies [207]. However, most patients were recruited in orthopedic clinics and this may limit the applicability of the findings to other settings. Evaluators met prior to the initiation of the study to standardize techniques and interpretation of the physical tests and related diagnoses, but no formal evaluation of their skills was undertaken. Also, the physiotherapist always evaluated the patients prior to the physician and this may have increased the diagnostic concordance by sensitizing the patients even though evaluators remained blinded to each others results. Our composite reference standard included both musculoskeletal examination and imaging interpretation by experienced medical experts, which is considered clinically relevant in the study of musculoskeletal disorders [50, 168-170]. However, only one physiotherapist and one of four medical experts evaluated each given participant and this limits the generalizability of our results. It must be noted that the physiotherapist in our study had only one year of clinical experience, which is suggestive of the appropriateness of physiotherapy training programs to adequately train therapists in musculoskeletal examination, but will require confirmation with more physiotherapists of diverse level of experience.

CONCLUSIONS

High diagnostic agreement and good triage of surgical candidates agreement was found between the physiotherapist and experienced physicians for various knee problems. Musculoskeletal examination without imaging performed by trained musculoskeletal providers may yield high diagnostic validity to discriminate between common knee disorders. This suggests the potential role of healthcare professionals such as physiotherapists in the development of multidisciplinary evaluation and triage strategies in the context of innovative and potentially more efficient care trajectories for patients with common musculoskeletal disorders. These results will require confirmation with a larger study, in other primary care settings, with a greater number of physiotherapists and for other musculoskeletal disorders.

Article 4: Diagnostic validity of combining history elements and physical examination tests for partial or complete anterior cruciate ligament tears.

Décary S, Fallaha M, Belzile S, Martel-Pelletier J, Pelletier J-P, Feldman D, Sylvestre M-P, Vendittoli P-A, Desmeules F.

Décary, Simon, PT, PhD candidate[1,2]
Fallaha, Michel, MD[3]
Belzile, Sylvain, MD[4]
Martel-Pelletier, Johanne, PhD[5]
Pelletier, Jean-Pierre, MD, PhD[5]
Feldman, Debbie, PhD[1]
Sylvestre, Marie-Pierre, PhD[6]
Vendittoli, Pascal-André, MD, MSc, FRSC[2,3]
Desmeules, François, PT, PhD[1,2]

1- School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada.
2- Orthopaedic Clinical Research Unit, Maisonneuve-Rosemont Hospital Research Center, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
3- Department of Surgery, Maisonneuve-Rosemont Hospital, University of Montreal, Montreal, Quebec, Canada, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
4- Department of Surgery, Laval University Hospital Center (CHUL), Laval University, Quebec, Quebec, Canada.
5- Osteoarthritis Research Unit, University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada.
6- University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada.
Department of Social Preventive Medicine, School of Public Health, Université de Montréal

Article en processus de révision dans la revue *PLoS ONE*, Mai 2017.

En tant que premier auteur, j'ai réalisé le développement du protocole de cette étude, la collecte des données, l'analyse des résultats ainsi que la rédaction du manuscrit. FD et PAV ont contribué au développement du protocole, à l'analyse des résultats, à la rédaction et à la révision du manuscrit. MF et SB ont contribué au développement de l'évaluation musculosquelettique standardisée et ont agi à titre de médecins experts (standard de

référence) pour tous les patients. JMP, JPP, DF et MPS ont contribué au développement du protocole, à la vérification des méthodes d'analyse des données et à la présentation des résultats. Tous les auteurs ont contribué à la révision de la version finale du manuscrit.

ABSTRACT

Objective: To assess the diagnostic validity of clusters combining history elements and physical examination tests to diagnose partial or complete anterior cruciate ligament (ACL) tears.

Design: Prospective diagnostic study.

Settings: Orthopaedic clinics (n=2), family medicine clinics (n=2) and community-dwelling.

Participants: Consecutive patients with a knee complaint (n=279) and consulting one of the participating orthopaedic surgeons (n=3) or sport medicine physicians (n=2).

Interventions: Not applicable.

Main Outcome Measures: History elements and physical examination tests performed independently were compared to the reference standard: an expert physicians' composite diagnosis including history elements, physical tests and confirmatory magnetic resonance imaging. Penalized logistic regression (LASSO) was used to identify history elements and physical examination tests associated with the diagnosis of ACL tear and recursive partitioning was used to develop diagnostic clusters. Diagnostic accuracy measures including sensitivity (Se), specificity (Sp), predictive values and positive and negative likelihood ratios (LR+/-) with associated 95% confidence intervals (CI) were calculated.

Results: Forty-three individuals received a diagnosis of partial or complete ACL tear (15.4% of total cohort). The Lachman test alone was able to diagnose partial or complete ACL tears (LR+: 38.4; 95%CI: 16.0-92.5). Combining a history of trauma during a pivot with a "popping" sensation also reached a high diagnostic validity for partial or complete tears (LR+: 9.8; 95%CI: 5.6-17.3). Combining a history of trauma during a pivot, immediate effusion after trauma and a positive Lachman test was able to identify individuals with a complete ACL tear (LR+: 17.5; 95%CI: 9.8-31.5). Finally, combining a negative history of pivot or a negative popping sensation during trauma with a negative Lachman or pivot shift test was able to exclude both partial or complete ACL tears (LR-: 0.08 (95%CI: 0.03-0.24)).

Conclusion: Diagnostic clusters combining history elements and physical examination tests can support the differential diagnosis of ACL tears compared to various knee disorders.

INTRODUCTION

The anterior cruciate ligament (ACL) is a major stabilizing structure of the knee against excessive anterior translation and internal rotation of the tibia [19, 31]. An ACL tear, either partial or complete, is a significant knee injury and an early diagnosis is advocated to guide toward accelerated rehabilitation protocols or surgical management to prevent sustained complaints of instability in daily and sporting activities and to potentially reduce subsequent risk of meniscal tear or chondral damage [14-19]. However, less than 15% of patients with an ACL tear are correctly diagnosed upon initial presentation, questioning the validity of a diagnosis based on history and physical examination [18].

Diagnostic validity of physical examination tests for ACL tears has been extensively studied [153]. The three most studied tests are the Lachman, the pivot shift and the anterior drawer test [153]. These three tests reach high specificity and can be used alone to make a valid ACL diagnosis, but only the Lachman reached moderate sensitivity [34, 153]. Combining multiple history elements and physical examination tests is advocated to increase the validity of the diagnosis of ACL tears. History elements such as a pivoting traumatic event, a “popping sensation” or immediate effusion following trauma can be useful to clinicians to make a valid ACL diagnosis, yet limited evidence is available to support this approach [6].

In the hand of an expert evaluator a clinical examination for ACL tears can be equivalent to magnetic resonance imaging (MRI) findings [6]. Specific combinations of history elements and/or physical examination tests have been reported in a few studies, but results are often incompletely presented and difficult to interpret [6, 34, 208-212]. The studies reporting more complete results, useful to clinicians in their assessment of patients with suspected ACL tears [101, 150], have in turn several methodological limitations [98], including the presence of spectrum bias [150], inconsecutive sampling [150], use of MRI only as the reference standard [101] and small sample sizes [101, 150].

Therefore, the aim of this diagnostic study is to assess the diagnostic validity of combining an extensive set of selected history elements and physical examination tests to diagnose or exclude partial or complete ACL tears compared to other common knee disorders based on a composite reference standard using standardized physical assessment and imaging.

METHODS

Study design and settings

This is a prospective multi-center diagnostic study aimed at developing a series of diagnostic clusters for various common knee disorders. The present paper reports on results specific to ACL tears. We recruited consecutive new patients consulting one of the participating physicians for a knee complaint between November 2014 and August 2016 in two outpatient orthopaedic clinics and two primary care family medicine clinics from two urban centers. We also invited via email persons from a university community (academic staff, students or administration personnel) who had a knee complaint and needed to see a health practitioner for their knee problem.

The present study, its design, methodology and reporting of results conform to the Standards for Reporting Diagnostic Accuracy Studies 2015 (STARD) [157, 158]. The study was approved by all participating ethics committees and participants signed an informed consent form.

Participants

Potential participants were initially screened by telephone to assess preliminary eligibility. Inclusion criteria were: 1- 18 years of age or older; 2- needed to consult a physician or referred to one of the participating clinical settings for a current knee problem; 3- able to understand and speak French. Patients previously treated by the participating physicians were excluded, as well as patients who had undergone lower limb surgery in the past six months, patients with a knee arthroplasty or who presented with more than two other lower limb pathologies or if they suffered from any systemic inflammatory disorder related to their knee complaint.

Data collection

Patients' characteristics and history elements

History elements collected included: gender, age, education level, employment status, comorbidities, affected side, duration of knee symptoms, knee pain location (anterior, posterior, medial or lateral or diffuse knee pain), traumatic or progressive onset of symptoms and use of walking aid. Possible traumatic mechanisms included: falling on knee (weight-bearing), direct external blow to the knee, trauma following a jump landing, pivoting trauma, foot/leg blocked during trauma, "popping" sensation during trauma, immediate/delayed pain, ability/inability to pursue activities and immediate/delayed effusion [101]. Participants also completed the Knee Injury and Osteoarthritis Outcome Score (KOOS), a 42-item questionnaire composed of 5 domains: pain, symptoms, function in daily living, function in sport and recreation and knee-related quality of life [159]. Psychological distress was assessed using the K6 screening scale [163].

Physical examination data collection procedure

Before the start of the study, clinicians met with the research personnel to standardize techniques and interpretation of the physical tests as well as the different knee diagnoses definitions. Each participant was independently assessed by two evaluators on the day of their visit: a physiotherapist and one of the participating physicians. The physiotherapist had a professional masters in physiotherapy and had one year of clinical experience. The five participating physicians (three orthopaedic surgeons and two sports medicine physicians) each had more than 20 years of experience. The physiotherapist always completed data collection prior to the physician consultation. Both the physiotherapist and the physicians were blinded to each other's results and any other clinical information from the start of their respective evaluations.

Physical examination tests

A complete standardized physical examination was performed by the two evaluators. Physical examination tests related to ACL tears included: Lachman, pivot shift and anterior drawer

tests [6]. These tests were noted as positive/negative/uncertain/not evaluated [128, 154]. The technique for these tests and the definition of a positive response are described in Table 19.

Table 19: Description of physical examination tests for ACL tear

Tests	Description of technique and positive outcomes.
Lachman [77, 78, 113].	The patient is in supine position. The clinician positions the patient's knee at 15-30° of flexion while holding the tibial plateau with one hand and the proximal aspect of the knee with the other hand. The clinician applies a swift anterior perpendicular force to the tibia. The test is considered positive when the clinician observes an anterior glide of the tibia and/or palpates an absent end-feel significantly different from the unaffected side.
Pivot shift [77, 78, 113].	The patient is in supine position. The clinician fully extends and internally rotates the patient's knee. The clinician's distal hand is placed at the patient's ankle to maintain internal rotation while his other hand palpates the lateral tibial plateau while inducing a slight valgus stress on the knee. The clinician then slowly flexes the knee. The test is considered positive if, during the first 30° of flexion, the clinician observes or palpates a subluxation and/or gliding of the tibial plateau significantly different from the unaffected side.
Anterior drawer test [77, 78, 113].	The patient is in supine position. The clinician positions the patient's knee at 90° of flexion while stabilizing the limb by sitting on the foot of the patient. Using both hands, the clinician holds the tibial plateau and applies a slow anterior perpendicular force to the tibia. The test is considered positive when the clinician observes an anterior glide of the tibia significantly different from the unaffected side.

Reference standard definition

Following his independent collection of history elements and physical examination tests, the expert physician was presented with any imaging results and radiology reports and performed his own analysis of the relevant imaging tests. All participants were required to have a radiograph of their knee. Magnetic resonance imaging (MRI) was also required for all suspected ligament tears, meniscal tears or to exclude any other knee diagnoses [98]. Based

on patient history, physical tests and imaging tests, the physician made a final primary and secondary (if necessary) composite diagnosis. Diagnoses were classified into 1- partial or complete ACL tear or 2- no-ACL tear. The final composite diagnosis was considered the reference standard against which the index tests, independently collected by the physiotherapist, were compared.

Statistical analysis

Descriptive statistics were used to present the participants' characteristics and Student t-tests as well as Chi-square tests were used to compare participants with an ACL tear to those with other knee disorders. Inter-rater reliability between the physiotherapist and the physicians was measured for all physical examination tests and diagnoses using kappa values (κ) with associated 95% CI [171]. A secondary analysis was conducted to identify only complete ACL tear cases compared to all other knee disorders including partial ACL tears, because these individuals may require a timely surgical consultation compared to those with a partial ACL tear [17, 101, 213].

Diagnostic clusters containing multiple history elements and physical examination test findings were developed using a two-step method [177]. First, Least Absolute Shrinkage and Selection Operator (LASSO) penalized logistic regression was used to identify history elements and physical examination tests predictive of the diagnosis of ACL tear either partial or complete [177, 178]. LASSO is used to select variables with a higher predictive ability in the situation where there is a large initial set of variables [177]. Variable selection by LASSO is obtained through shrinking the beta coefficients of unimportant variables to zero [179]. The degree of shrinkage is determined by a penalty parameter, the value of which is identified through cross-validation to select the set of variables that maximize area under the curve (AUC) [179].

Recursive partitioning was performed on the clinical variables selected using the LASSO to form diagnostic clusters to include and exclude a diagnosis of ACL tear, either partial or

complete, and also only for complete tears [180]. Recursive partitioning allows the best sequence of variables to classify ACL tear from non-ACL tear cases [177]. The gini index was used as the splitting criteria [177, 180]. Overall model classification was compared to the reference standard (physicians' composite diagnosis) and, sensitivity (Se), specificity (Sp), positive and negative predictive values (PPV/NPV) and positive and negative likelihood ratios (LR+/-) with 95% CI were calculated [60, 184]. Physical tests with an inter-rater agreement kappa (κ) value $\leq$0.4 were considered unreliable and not used in final clusters development [116]. Final selection of clusters to include and exclude ACL tear were based on overall diagnostic validity and needed to reach a LR+$\geq$10 or a LR-<0.2. These thresholds are recognized to produce a moderate to substantial shift in post-test probability and therefore clinically useful to diagnose or exclude a disorder [70, 82]. Ease of use and clinical applicability was also considered in final selection of clusters [60]. Internal validity was assessed using a validated technique of bootstrapping for recursive partitioning and estimates with 95% CIs were compared to the proposed clusters [177, 186]. Analyses were performed using R version 3.3.0 (packages: rpart, glmnet and randomForest; http://cran.r-project.org/).

RESULTS

Table 20 presents the characteristics of participants. A total of 279 individuals participated (96.2% of those recruited, appendix 11). Participants were diagnosed with 359 primary and secondary diagnoses as follows: knee osteoarthritis (n=129), patellofemoral pain (n=75), symptomatic meniscal tears (n=80) or other knee diagnoses (n=32). Forty-three individuals (15.4%) received a diagnosis of partial (n=21) or complete (n=22) anterior cruciate ligament (ACL) tear. Overall, individuals with an ACL tear were significantly younger (ACL: 38.6±12.9 years old, others: 51.0±15.6 years old, $p<0.05$) and had a lower body mass index (ACL: 26.9±5.8, others: 29.7±6.6, $p<0.05$) compared to those without an ACL tear. Most individuals with an ACL tear were recruited in the orthopaedic clinics (95.3%) and had an history of trauma as the reason for consultation. (79.1%). More individuals with an ACL tear had pain for less than 3 months (ACL: 23.3%, others: 10.2%, $p<0.05$) and more than half were referred to surgery after consultation (ACL: 53.5%, others: 10.6%, $p<0.05$). For complete ACL tears only, 95.4% of individuals had a history of trauma as the reason for consultation and 68.1% were referred to surgery after consultation.

Table 21 describes the clinical diagnoses and imaging findings for partial or complete ACL tears. Overall, ACL tear was the primary diagnosis in 63% of individuals with partial or complete ACL tears and 73% of those with complete ACL tears only. Forty-two percent of individuals with a partial or complete ACL tear and 59% with a complete tear had a combined symptomatic meniscal tear. Based on the composite reference standard diagnosis, 22 participants had a complete ACL tear, including two who had a re-rupture of a previously reconstructed ACL; 21 participants had a partially teared ACL.

Table 20: Characteristics of participants (n=279)

Characteristics	Partial or complete ACL tears (n=43)		Other knee disorders (n=236)	
	n (%)	mean (SD)	n (%)	Mean (SD)
Age		38.6 (12.9)*		51.0 (15.6)
Body Mass Index (Kg/m^2)		26.9 (5.8)*		29.7 (6.6)
Sex				
Female	21 (48.8)		140 (59.3)	
Male	22 (51.2)		96 (40.7)	
Recruitment site				
Orthopaedic clinics	41 (95.3)*		188 (79.7)	
Family medicine unit/university community	2 (4.7)		48 (20.3)	
History of trauma	34 (79.1)*		53 (22.5)	
Duration of pain and symptoms at time of consultation				
<3 months	10 (23.3)*		24 (10.2)	
3-12 months	15 (34.8)*		57 (24.2)	
≥ 12 months	18 (41.9)*		155 (65.6)	
Referred to surgery after consultation	23 (53.5)*		25 (10.6)	
Knee Injury and Osteoarthritis Outcome Score (KOOS)				
Pain		67.6 (16.7)*		56.9 (20.3)
Symptoms		73.8 (14.5)		69.4 (19.8)
Activity of Daily Living		73.7 (18.3)*		64.4 (22.4)
Sports		28.3 (25.3)		28.7 (25.3)
Quality of Life		28.5 (15.6)*		40.4 (20.0)
K6 psychological distress scale (/30)		26.9 (3.5)		26.3 (4.7)

ACL: anterior cruciate ligament; SD: standard deviation; KOOS: a score of 0 indicates a severe condition and 100 indicates a normal knee; K6: a score of 6 indicates serious mental illness and 30 indicates no mental illness; * indicates a significant difference ($p<0.05$) between participants with an anterior cruciate ligament tear and those without.

Table 21: Description of clinical diagnoses and imaging findings for partial or complete ACL tears participants (n=43)

	Partial or complete ACL tears (n=43)	Complete ACL tears only (n=22)
Description of clinical diagnoses	**n (%)**	**n (%)**
ACL tear primary diagnosis	27 (63)	16 (73)
ACL tear alone with no other knee disorder	12 (28)	7 (32)
ACL tear combined with another knee disorder		
Symptomatic meniscal tear	18 (42)	13 (59)
Patellofemoral pain	6 (14)	1 (5)
Osteoarthritis	3 (7)	1 (5)
Other knee disorders	4 (9)	0 (0)
Imaging findings		
Complete ACL tear	22 (51)	22 (100)
Partial ACL tear	21 (49)	0 (0)

ACL: anterior cruciate ligament; Clinical diagnoses are composite diagnoses made by physicians using history elements, physical examination tests and relevant imaging including magnetic resonance imaging (MRI) confirmation for all ACL tear diagnoses; Imaging findings are from MRI radiologists' reports and/or imaging assessment by the physicians; Others knee disorders combined with ACL tears include: contusion of tibial plateau (n=3) and lateral collateral ligament tear (n=1).

Table 22 presents the clinical variables associated with the diagnosis of ACL tear identified through penalized logistic regression. One hundred and thirty-one different clinical variables were entered in the penalized logistic regression. Following cross-validation, six variables were associated with the diagnosis of partial or complete ACL tear and yielded a maximal area under the curve (AUC) of 0.92 (95%CI: 0.86-0.98). Four variables were history elements. A history of pivot during trauma, a popping sensation during trauma, no history of fall on the knee during trauma and no current knee pain or only monthly knee pain frequency as measured with the KOOS questionnaire were associated with a diagnosis of an ACL tear. Two physical examination tests, the Lachman and the pivot shift, were associated with the diagnosis of ACL tear. For complete ACL tears only, seven variables yielded a maximal AUC of 0.94 (95%CI: 0.91-0.97). The history element immediate effusion after trauma was associated with complete ACL tears only, while popping sensation was not. The Lachman and the pivot shift remained associated for complete ACL tears.

Table 22: Clinical variables associated with the diagnosis of partial or complete ACL tears at maximal AUC identified through penalized logistic regression (n=279)

	Variables associated with a partial or complete ACL tear (n=43)	Variables associated with a complete ACL tear only (n=22)
History elements	Absence of fall on knee at initial trauma History of pivoting on knee at initial trauma "Popping sensation" during trauma No or only monthly knee pain*	3-4 weeks pain duration at time of consultation Absence of fall on knee at initial trauma History of pivoting on knee at initial trauma Impossibility to pursue activity after trauma Immediate effusion after trauma
Physical examination tests	Lachman Pivot shift	Lachman Pivot shift

*ACL: anterior cruciate ligament; LASSO: Least Absolute Shrinkage and Selection Operator; Maximal area under the curve (AUC) was used as the criteria for the final penalty parameter to select variables associated with the presence or absence of ACL tear. Complete ACL tears only was compared to all disorders, including partial tears. *KOOS question for pain frequency: How often do you experience knee pain: never, monthly, weekly, daily, always. Timing of effusion after trauma included: no effusion, immediate effusion, effusion appeared more than 2 hours after trauma. Pain duration at time of consultation included: 1-2 weeks, 3-4 weeks, 5-6 weeks, 7-8 weeks, 9-11 weeks, 3-6 months, 6-9 months, 9-12 months, more than 12 months.*

Table 23 presents the diagnostic validity and reliability of history elements and physical examination tests when individually performed. The Lachman reached inter-rater kappa (κ) value of 0.75 (95%CI: 0.63-0.88). The pivot shift reached inter-rater kappa value of 0.84 (95%CI: 0.65-1.00).

A positive "popping" sensation presented a LR+ of 7.6 (95%CI: 4.6-12.6) to diagnose both partial and complete ACL tears. A negative pivoting traumatic mechanism had a LR- of 0.25 (95%CI: 0.14-0.44) to exclude partial or complete ACL tears. The Lachman test reached LR+ of 38.4 (95%CI: 16.0-92.5) and LR- of 0.19 (95%CI: 0.10-0.36) to diagnose and exclude partial or complete ACL tears. The pivot shift reached a LR+ of 37.5 (95%CI: 14.0-100.4) and LR- of 0.24 (95%CI: 0.13-0.42) to diagnose and exclude partial or complete ACL tears.

Immediate effusion after trauma presented a LR+ of 11.0 (6.3-19.1) to diagnose a complete ACL tear and a negative pivoting traumatic mechanism reached a LR- of 0.06 (95%CI: 0.01-0.38) to exclude a complete ACL tear. The Lachman test presented a LR+ of 9.6 (95%CI: 6.1-14.9) and LR- of 0.20 (0.08-0.48) to diagnose and exclude complete ACL tears compared to all knee disorders including partial ACL tears. The pivot shift reached LR+ of 9.5 (95%CI: 5.8-15.6) and LR- of 0.22 (0.09-0.53) to diagnose and exclude complete ACL tears compared to all knee disorders including partial ACL tears.

Table 23: Diagnostic validity and inter-rater reliability of history elements and physical examination tests when individually performed to diagnose or exclude partial or complete ACL tears (n=43) or complete ACL tears only (n=22)

	Se (95% CI)		Sp (95% CI)		LR- (95% CI)		LR+ (95% CI)		Inter-rater reliability (95% CI)
	Partial or complete	Complete only	Partial or complete	Complete only	Partial or complete	Complete only	Partial or complete	Complete only	
History elements									
Pivoting traumatic mechanism	0.79 (0.64-0.90)	0.95 (0.77-1.00)	0.84 (0.79-0.89)	0.81 (0.75-0.85)	0.25 (0.14-0.44)	0.06 (0.01-0.38)	5.0 (3.6-7.0)	4.9 (3.8-6.4)	-
"Popping" sensation	0.58 (0.42-0.73)	-	0.92 (0.88-0.95)	-	0.45 (0.32-0.65)	-	7.6 (4.6-12.7)	-	-
Immediate effusion after trauma	-	0.68 (0.45-0.86)	-	0.94 (0.90-0.96)	-	0.34 (0.18-0.63)	-	11.0 (6.3-19.1)	-
Physical examination tests									
Lachman	0.81 (0.67-0.92)	0.82 (0.60-0.95)	0.98 (0.95-0.99)	0.91 (0.87-0.95)	0.19 (0.10-0.36)	0.20 (0.08-0.48)	38.4 (16.0-92.5)	9.6 (6.1-14.9)	0.75 (0.63-0.88)
Pivot shift	0.77 (0.61-0.89)	0.80 (0.56-0.94)	0.98 (0.95-0.99)	0.92 (0.87-0.95)	0.24 (0.13-0.42)	0.22 (0.09-0.53)	37.5 (14.0-100.4)	9.5 (5.8-15.6)	0.84 (0.65-1.00)

ACL: anterior cruciate ligament; Se: sensitivity; Sp: specificity; LR+: positive likelihood ratio; LR-: negative likelihood ratio. "Popping" sensation was not calculated for complete ACL tears and immediate effusion was not calculated for partial and complete tears because it did not show an association with LASSO results presented in Table 22. Inter-rater reliability (kappa) was measured only for physical examination tests. The pivot shift could not be evaluated in n=45 patients (n=2 partial ACL tears, n=2 complete ACL tears, 41 non-ACL tears). See 2x2 tables in Appendix 12 and 13.

Diagnostic clusters combining history elements and physical examination tests were identified through recursive partitioning using the clinical variables associated with a diagnosis of ACL tears. Table 24 presents two diagnostic clusters to diagnose partial or complete ACL tears or complete ACL tears only compared to other knee disorders. One cluster combined only two history elements: pivoting traumatic mechanism with a "popping" sensation during trauma to diagnose partial or complete ACL tears. The cluster correctly classifies 25/43 individuals with an ACL tear (see 2x2 table in Appendix 14) and has the following diagnostic statistics: Se of 0.58 (95%CI: 0.42-0.73), Sp of 0.94 (95%CI: 0.90-0.97, PPV of 0.64 (95%CI: 0.47-0.79), LR+ of 9.80 (95%CI: 5.55-17.29).

The other cluster (Table 24) combined a pivoting traumatic mechanism with immediate effusion after trauma and a positive Lachman test to diagnose complete ACL tears only compared to other disorders (including partial tears). The cluster correctly identifies 18/22 of cases (see 2x2 table in Appendix 15) and has the following diagnostic statistics: Se of 0.82 (95%CI: 0.60-0.95), Sp of 0.95 (95%CI: 0.92-0.98), PPV of 0.60 (95%CI; 0.41-0.77) and LR+ of 17.5 (95%CI: 9.8-31.5).

Table 25 presents one diagnostic cluster to exclude partial or complete ACL tears compared to other knee disorders. The cluster combined a negative history of pivot or popping sensation during trauma with a negative Lachman or pivot shift test to exclude an ACL tear. The cluster correctly classifies 206/236 individuals without a partial or complete ACL tear (see 2x2 table in Appendix 14). The cluster shows the following diagnostic statistics: Se of 0.93 (95%CI: 0.81-0.99), Sp of 0.87 (95%CI: 0.82-0.91, NPV of 0.99 (95%CI: 0.96-1.00), LR- of 0.08 (95%CI: 0.03-0.24). Tables 24 and 25 also present internal validation estimates and 95% CIs obtained using bootstrapping (n=1000).

Table 24: Diagnostic clusters combining history elements and physical examination tests to diagnose partial or complete ACL tears (n=43) or complete ACL tears only (n=22)

Clusters	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	LR+ (95% CI)
Knee complaints are <u>likely due to a partial or complete ACL tear</u> in individuals with: ➢ Pivoting traumatic mechanism *AND* ➢ "Popping" sensation during trauma	0.58 (0.42-0.72)	0.94 (0.90-0.97)	0.64 (0.47-0.79)	9.80 (5.55-17.29)
Internal validation	0.53 (0.38-0.69)	0.92 (0.87-0.95)	0.56 (0.40-0.72)	7.01 (4.15-11.85)
Knee complaints are <u>likely due to a complete ACL tear</u> in individuals with: ➢ Pivoting traumatic mechanism *AND* ➢ Immediate effusion after trauma *AND* ➢ Positive Lachman test	0.82 (0.60-0.95)	0.95 (0.92-0.97)	0.60 (0.41-0.77)	17.5 (9.8-31.5)
Internal validation	0.82 (0.60-0.95)	0.93 (0.90-0.96)	0.51 (0.34-0.69)	12.37 (7.50-20.39)

Se: sensitivity; Sp: specificity; PPV: positive predictive value; LR+: positive likelihood ratio. Clusters are obtained using recursive partitioning with all variables associated with the diagnosis of an ACL tear. For Lachman, cases coded as "Uncertain" were considered "Positive" for a suspicion of ACL tear. Internal validation was assessed by bootstrapping (n=1000).

Table 25: Diagnostic cluster combining history elements and physical examination tests to exclude partial or complete ACL tears (n=43)

Clusters	Se (95% CI)	Sp (95% CI)	NPV (95% CI)	LR- (95% CI)
Knee complaints are <u>unlikely due</u> to a partial or complete ACL tear in individuals with: ➢ Negative history of pivot or "popping" sensation during trauma *AND* ➢ Negative Lachman or pivot shift test	0.93 (0.81-0.99)	0.87 (0.82-0.91)	0.99 (0.96-1.00)	0.08 (0.03-0.24)
Internal validation	0.91 (0.78-0.97)	0.85 (0.80-0.89)	0.98 (0.95-0.99)	0.11 (0.04-0.27)

Se: sensitivity; Sp: specificity; NPV: negative predictive value; LR-: negative likelihood ratio. Clusters are obtained using recursive partitioning with all variables associated with the diagnosis of an ACL tear. For Lachman and pivot shift, cases coded as "Uncertain" were considered "Positive" for a suspicion of ACL tear. Internal validation was assessed by bootstrapping (n=1000).

DISCUSSION

This study assessed the diagnostic validity of clusters combining history elements and physical examination tests to diagnose or exclude partial or complete ACL tears compared to other knee disorders. We first observed that when individually performed, the Lachman and the pivot shift tests reached substantial positive LR (LR+>10) to diagnose and moderate negative LR (LR- <0.2) to exclude partial or complete ACL tears [82]. In the present cohort, this increased positive post-test probability by 72.6% and negative post-test probability by 12.4% and 10.4% for both tests. These results are comparable with data synthesized in a previous meta-analysis reporting substantial positive LR to diagnose a partial or complete ACL tear using these tests (Lachman: LR+=10.2 (95%CI: 4.6-22.7), pivot shift: LR+=12.0 (95%CI: 5.0-31.0)) and moderate negative LR to exclude an ACL tear (Lachman: LR-=0.2 (95%CI: 0.10-0.30)) [34]. Taken alone these tests can be considered clinically useful to diagnose or exclude a partial or complete ACL tear.

We also demonstrated that combining history elements and physical examination tests improved the probability to diagnose or exclude partial or complete ACL tears. The first cluster, combining only history elements, presented near substantial positive LR (LR+=9.80) [82], increasing positive post-test probability by 48.6% for partial or complete ACL tears. A second cluster combining two history elements and adding a positive Lachman test demonstrated a substantial positive LR to diagnose complete ACL tears only, increasing positive post-test probability by 52.1%. The last cluster combined a negative history of pivot or popping sensation during trauma with a negative Lachman or pivot shift tests, and was able to exclude both partial or complete ACL tears (LR-<0.1) [82]. Combining only history elements accurately identified 58% of all ACL tears cases in this cohort, while combining history elements and physical examination tests accurately identified more than 80% of all cases and non-cases.

Our study compares favorably with previous evidence on the combination of selected history elements and physical examination tests. In a sample of 134 acute traumatic patients

including 28 with a partial or complete ACL tear, *Wagemaker et al*. found that combining immediate effusion, popping sensation, giving way symptoms and a positive anterior drawer test yielded a positive LR of 4.2 for partial or complete tears (95%CI: 2.4-7.5) and a positive LR of 7.2 for complete tears only (95%CI: 3.6-14.4) [101]. In another study by *Geraets et al*., with a sample of 60 patients undergoing knee arthroscopy and including 22 complete ACL tears, the authors found that combining immediate effusion, popping sensation and complaints of giving way as well as positive Lachman, pivot shift and anterior drawer tests yielded a positive LR of 24.0 [150]. Although somewhat comparable, the differences observed in the magnitude of the estimates between our study and these two studies may be explained by several factors. The inclusion of more acute patients in the first study may have made the physical examination tests more difficult to perform and interpret because of knee effusion or muscle guarding [101, 213]. In our study, most participants (77%) had their trauma more than 3 months before their evaluation and this may have made it easier to perform the physical tests and interpret these tests, ultimately increasing diagnostic validity estimates. Nevertheless, combining only history elements, which are not affected by an acute presentation, also showed accurate validity in our study to support the diagnosis of partial or complete ACL tears.

Another factor explaining the difference between our study and the literature may be related to the representativeness of included participants. In both studies by *Wagemaker et al*. and *Geraets et al*., the non-ACL tear patients included only participants with a traumatic history which may limit the external validity of their findings [101, 150]. In our study, we compared patients with a partial or complete ACL tear to other patients with various knee disorders, not only suffering from a traumatic onset. This allowed us to demonstrate that the traumatic onset of symptoms was a diagnostically useful history element when discriminating between several knee disorders. The fact that cases and non-cases were more chronic in our study and not all of them had a history of trauma could have inflated the diagnostic capability of our results, yet it may at the same time limit spectrum bias commonly seen in diagnostic studies

in this field [6, 101, 150, 207]. The external validity of our findings will need to be assessed in a validation study with another cohort of participants.

Previously published diagnostic studies reported the inclusion of the anterior drawer test in their diagnostic models [101, 150]. In our cohort, we did not found an association between a positive anterior drawer test and the diagnosis of ACL tear. This may be because our cohort includes patients with both partial or complete ACL tears combined with meniscal tears which have been shown to lower the diagnostic validity of this test [6, 34, 73, 153, 213]. Based on our results, the anterior drawer test produced twice as much false-negative cases as the Lachman test (data not shown). Even if this test may appear easier to learn by untrained clinicians, the false-negative rate precluded us from proposing the anterior drawer test as a valid test at this time [101]. Moreover, even if the Lachman and pivot shift tests are often found to be harder to perform and interpret, we showed higher inter-rater reliability for both tests than in previous reliability results [105, 154]. A likely explanation may be that we used a dichotomous outcome (positive/negative) as originally described in the initial publication of these tests [128]. Also, evaluators assessed both knees and comparison to the unaffected side may have supported higher confidence in a positive or negative response to the tests, hence higher inter-rater reliability.

Strengths and limitations

We recruited a consecutive cohort from different settings allowing for a broad variety of cases and non-cases. However, most individuals with an ACL tear were found in secondary care and the performance of the clusters may be different in primary care patients with an acute presentation. Our data collection procedure ensured blinding of the evaluators between the index tests and the reference standard. Our composite reference standard by expert physicians included history elements, physical examination tests as well as MRI confirmation for all ACL tears. A composite reference standard is advocated because meta-analyses showed that use of MRI only as the reference standard reached only a Se of 86.5% and a Sp of 95.2% [168-170, 214]. Our statistical approach (LASSO and recursive partitioning)

lead to the identification of simple clinically useful clusters based on an extensive set of patient's personal and clinical variables [177].

CONCLUSIONS

We identified diagnostic clusters combining history elements and physical examination tests that can accurately support the differential diagnosis of partial or complete ACL tears compared to other knee disorders. Given adequate external validation, these clusters could be used by clinicians involved in musculoskeletal care in settings with limited access to MRI and initiate an accelerated rehabilitation protocol or propose a specialty referral when surgery is considered.

Article 5: Diagnostic validity of combining history elements and physical examination tests for traumatic and degenerative symptomatic meniscal tears.

Décary S, Fallaha M, Frémont P, Martel-Pelletier J, Pelletier J-P, Feldman DE, Sylvestre M-P, Vendittoli P-A, Desmeules F.

Décary, Simon, PT, PhD candidate[1, 2]
Fallaha, Michel, MD[3]
Frémont, Pierre, MD, PhD[4]
Martel-Pelletier, Johanne, PhD[5]
Pelletier, Jean-Pierre, MD, PhD[5]
Feldman, Debbie, PhD[1]
Sylvestre, Marie-Pierre, PhD[6]
Vendittoli, Pascal-André, MD, MSc, FRSC[2, 3]
Desmeules, François, PT, PhD[1, 2]

1- School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada.
2- Orthopaedic Clinical Research Unit, Maisonneuve-Rosemont Hospital Research Center, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
3- Department of Surgery, Maisonneuve-Rosemont Hospital, University of Montreal, Montreal, Quebec, Canada, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
4- Department of Rehabilitation, Faculty of Medicine, Laval University, Quebec City, Quebec, Canada
5- Osteoarthritis Research Unit, University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada.
6- University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada. Department of Social Preventive Medicine, School of Public Health, Université de Montréal

Article publié dans la revue *PM&R*. 2017. doi: 10.1016/j.pmrj.2017.10.009.

En tant que premier auteur, j'ai réalisé le développement du protocole de cette étude, la collecte des données, l'analyse des résultats ainsi que la rédaction du manuscrit. FD et PAV ont contribué au développement du protocole, à l'analyse des résultats, à la rédaction et à la révision du manuscrit. MF, BP, SB et PF ont contribué au développement de l'évaluation

musculosquelettique standardisée et ont agi à titre de médecins experts (standard de référence) pour tous les patients. JMP, JPP, DF et MPS ont contribué au développement du protocole, à la vérification des méthodes d'analyse des données et à la présentation des résultats. Tous les auteurs ont contribué à la révision de la version finale du manuscrit.

ABSTRACT

Background: Current approach to the clinical diagnosis of traumatic and degenerative symptomatic meniscal tears (SMT) propose combining history elements and physical examination tests without systematic prescription of imaging investigations, yet the evidence to support this diagnostic approach is scarce.

Objective: To assess the validity of diagnostic clusters combining history elements and physical examination tests to diagnose or exclude traumatic and degenerative SMT compared to other knee disorders.

Design: Prospective diagnostic accuracy study.

Settings: Patients were recruited from two orthopaedic clinics, two family medicine clinics and from a university community.

Patients: 279 consecutive patients consulting for a new knee complaint.

Methods: Each patient was independently assessed by two evaluators. History elements and standardized physical examination tests performed by a physiotherapist were compared to the reference standard: an expert physicians' composite diagnosis including a clinical examination and confirmatory magnetic resonance imaging; participating expert physicians were orthopaedic surgeons (n=3) or sport medicine physicians (n=2). Penalized logistic regression (LASSO) was used to identify history elements and physical examination tests associated with the diagnosis of SMT and recursive partitioning was used to develop diagnostic clusters.

Main Outcome Measures: Diagnostic accuracy measures were calculated including sensitivity (Se), specificity (Sp), predictive values (PPV/NPV) and positive and negative likelihood ratios (LR+/-) with associated 95% confidence intervals (CI).

Results: Eighty patients had a diagnosis of SMT (28.7%), including 35 traumatic tears and 45 degenerative tears. Combining a history of trauma during a pivot, medial knee pain location and a positive Medial Joint Line Tenderness test was able to diagnose (LR+=8.9; 95%CI:6.1-13.1) or exclude (LR-=0.10; 95%CI:0.03-0.28) a traumatic SMT. Combining a history of progressive onset of pain, medial knee pain location, pain while pivoting, absence of valgus or varus knee misalignment or full passive knee flexion was able to moderately diagnose (LR+=6.4; 95%CI:4.0-10.4) or exclude (LR-=0.10; 95%CI:0.03-0.31) a degenerative SMT. Internal validation estimates were slightly lower for all clusters but demonstrated positive LR superior to 5 and negative LR inferior to 0.2 indicating moderate shift in post-test probability.

Conclusion: Diagnostic clusters combining history elements and physical examination tests can support the differential diagnosis of SMT. These results represent the initial derivation of the clusters and external validation is mandatory.

Level of evidence: Diagnosis, Level 2b

INTRODUCTION

The menisci are structures of the knee responsible for distributing compression and pivotal forces to improve joint stability in activities and sports [37]. Symptomatic meniscal tears (SMT) are common knee disorders and a leading reason to consult a health care provider [11, 39]. SMT may be classified as either of traumatic or degenerative onset. A traumatic onset is typically seen in younger individuals and can represent 11% of all acute knee disorders, while a degenerative meniscal tear, of progressive onset, typically seen in older individuals may represent up to 31% of all chronic knee disorders while also possibly being in the causal pathway to knee osteoarthritis over the years [6, 11, 13, 37, 39-41]. Recent evidence proposes that early rehabilitation protocols improve patient outcomes for both traumatic and degenerative meniscal tears [40, 41, 215-217]. Therefore, an early diagnosis of SMT based on history findings and physical examination tests is advocated to guide efficient conservative management [9-13].

Current clinical practice to diagnose a meniscal tear is often based on magnetic resonance imaging (MRI) results, a costly diagnostic test likely to cause delays in management and may even result in an inaccurate diagnosis, up to 76% of individuals with a degenerative meniscal tear on MRI may be asymptomatic [9, 218, 219]. A possible explanation for this overreliance on MRI is that the diagnostic validity of physical examination tests presents conflicting results [153]. Based on low-grade evidence, most commonly studied clinical tests (i.e: Joint Line Tenderness test, McMurray's test, Apley's maneuver and Thessaly's test) are not able to accurately diagnose or exclude SMT when used individually [84, 120, 153, 220, 221]. Yet, it is often reported that medical specialists or physiotherapists with extensive training and expertise in musculoskeletal disorders can make an accurate clinical diagnosis of a SMT based on a complete physical examination combined with the patient's history.

Diagnostic reasoning combines multiple patients' history elements with physical examination tests to make a valid diagnosis of SMT. Only a few studies have assessed the validity of combining history elements and physical examination tests [11, 102, 151, 222]. Although

these studies present encouraging results regarding combination of various history elements and tests, there are significant limitations [98] notably the presence of spectrum bias [222], use of MRI only as a reference standard [102, 151], as well as ultimately providing rules which may lack ease-of-use for clinicians [56, 149]. Therefore, to improve current evidence on the clinical diagnosis of SMT, our objective was to assess the validity of diagnostic clusters combining history elements and physical examination tests to diagnose or exclude traumatic or degenerative SMT compared to other common knee disorders.

METHODS

Study design and settings

This is a multi-center diagnostic study aimed at developing a series of diagnostic clusters for various common knee disorders. The present study reports on results specific to SMT. We recruited consecutive new patients consulting one of the participating physicians for a current knee complaint. Recruitment took place in two outpatient orthopaedic clinics and two primary care family medicine clinics between November 2014 and August 2016. University community participants seeking care for a current knee complaint were also invited to participate, via an email sent in September 2015. The present study, its design, methodology and reporting of results conform to the Standards for Reporting Diagnostic Accuracy Studies 2015 (STARD) [157, 158]. The study was approved by the ethics committees of all recruiting institutions and all participants signed an informed consent form.

Participants

Potential participants were initially screened by telephone to assess preliminary eligibility. Inclusion criteria were: 1- 18 years of age or older; 2- needed to consult a physician or referred to one of the participating clinical settings for a knee problem; 3- able to understand and speak French. Patients previously treated by the participating physicians were excluded, as well as patients who had undergone lower limb surgery in the past six months, patients with a knee arthroplasty or who presented with more than two other lower limb pathologies or if they suffered from any systemic inflammatory disorder related to their knee complaint.

Data collection

Patients' characteristics and history elements

Participants answered a standardized questionnaire and information collected included: gender, age, education level, employment status, comorbidities, duration of knee symptoms, affected side, knee pain location, traumatic or progressive onset of symptoms and use of a walking aid. Knee pain location could be categorized as anterior, posterior, medial, lateral or diffuse. Possible traumatic mechanisms included (categories are not mutually exclusive): falling on the knee (weight-bearing), external force to the knee, trauma during a jump landing, pivoting trauma, foot/leg stuck on the ground, "popping" sensation during trauma, immediate or late pain apparition, inability to pursue activities and immediate or late knee effusion [101]. Participants also completed the Knee Injury and Osteoarthritis Outcome Score (KOOS), a 42-item questionnaire composed of 5 domains: pain, symptoms, function in daily living, function in sport and recreation and knee-related quality of life [159]. Psychological distress was assessed using the K6 screening scale [163]. The individual questions of the KOOS and of the K6 were later used as separate potential relevant history elements to include in the development of the different diagnostic clusters.

Physical examination data collection procedure

Before the start of the study, clinicians met with the research personnel to standardize techniques and interpretation of the physical tests as well as the different knee diagnosis definitions. Each participant was independently assessed by two evaluators on the day of their visit: a physiotherapist and one of the participating physicians. The physiotherapist had a professional masters in physiotherapy and one year of clinical experience. The five participating physicians (three orthopaedic surgeons and two sports medicine physicians) each had more than 20 years of clinical experience. The physiotherapist always completed data collection prior to the physician consultation. Both the physiotherapist and the physicians were blinded to each other results and any other clinical information from the start of their respective evaluations.

Physical examination tests

A complete standardized physical examination was performed by the two evaluators. Physical examination tests related to SMT that were systematically performed included: medial and lateral Joint Line Tenderness, McMurray's test and Thessaly's test at 20° of knee flexion [71, 223]. The tests were rated as positive, negative or uncertain. The technique of these tests and the definition of a positive response are described in Table 26. Other knee tests performed included: active and passive knee range of motion in extension and flexion (restricted or not restricted), visual assessment of lower limb alignment (presence of valgus, varus or recurvatum), visual assessment of lower limb motor control (ability to maintain alignment, yes or no) during squatting or descending a step task or presence of pain during these tasks (yes or no). The definitions, standardization and positive responses to these tests are described in Appendix 17.

Reference standard definition

The physician independently collected his standardized history elements and physical examination tests while blinded to the imaging results**.** He was then presented with any imaging results and performed his own analysis of the relevant imaging tests. All participants were required to have a radiograph of their knee. Magnetic resonance imaging (MRI) was also required for all suspected ligament tears, meniscal tears or to exclude any other knee diagnoses. Radiographs were obtained on the day of their appointment except if participants already had recent radiographs within 3 months of their participation. MRI tests done within 6 months were accepted. If a new MRI was deemed needed by the expert physician to make his composite diagnosis, it had to be performed in the month following the consultation. Imaging needed to have suitable views or scans that allowed adequate interpretation and grading by the physician. If the physician doubted that the imaging did not reflect the current stage of pathology another test was ordered.

The physician made a final primary and secondary (if necessary) composite diagnosis based on their clinical judgement of the patient's history, physical tests and imaging tests results. A

secondary diagnosis was defined as a separate diagnosis which could change the prognosis or required a specific management different from the primary diagnosis. A diagnosis of SMT was defined by the expert physicians as the combination of clinically relevant symptoms and signs related to the patient's complaint with a MRI confirmation of the meniscal tear [11]. A SMT was considered of traumatic origin if the patient identified trauma as the onset of their knee complaint, while it was considered degenerative if the patient reported a progressive onset. Meniscal tears identified on MRI which did not correlate with history or clinical symptoms were not considered SMT [11]. Knee diagnoses were therefore classified into 1- traumatic or degenerative SMT or 2- other knee diagnosis. This final composite diagnosis was considered the reference standard against which the index tests, independently collected by the physiotherapist who remained blinded to the imaging tests, were compared for all participants.

Table 26: Description of physical examination tests for meniscal tear

Tests	Description of technique and positive outcomes.
Medial and Lateral Joint Line Tenderness [77, 78, 113]	The patient is in supine position. The clinician positions the patient's knee at 90° of flexion. The clinician palpates the medial and lateral joint lines of the knee. The tests are considered positive when palpation reproduces the patient's symptoms (pain, tenderness or discomfort) compared to the unaffected side.
McMurray's test [77, 78, 113]	The patient is in supine position. The clinician's distal hand grabs the patient's heel and passively flexes the patient's knee while his proximal hand palpates the knee's joint lines. The clinician then internally or externally rotates the tibia and, while keeping the tibial rotation, he fully extends the patient's knee. Internal rotation assesses the lateral meniscus and external rotation assesses the medial meniscus. The test is considered positive: 1- if the clinician hears or palpates a "click" during the manoeuvre and/or 2- the passive tibial rotation movement reproduces the patient's symptoms compared to the unaffected side.

Thessaly's test [77, 78, 113]	The patient is in the standing position. The clinician supports the patient's arms while the patient stands on one leg with the knee slightly bent at 20° of flexion. The clinician then creates a rotation of the patient's trunk to the left and to the right sides; rotations are repeated three times. The test is considered positive when the rotation movement reproduces the patient's symptoms at the medial or lateral joint lines. Anterior pain at or around the patella when the knee is first flexed is not considered a positive finding.

Statistical analysis

Descriptive statistics were used to present the participants' characteristics and Student t-tests as well as Chi-square tests were used to compare SMT participants to those with other knee disorders. Inter-rater agreement between the physiotherapist and the physicians was evaluated for all physical examination tests and diagnoses using *Prevalence and Bias Adjusted Kappas* (PABAK) with associated 95% CI [171].

Diagnostic clusters combining history elements and physical examination tests findings were developed using a two-step method [177]. First, Least Absolute Shrinkage and Selection Operator (LASSO) penalized logistic regression was used to identify history elements and physical examination tests predictive of the diagnosis of SMT [177, 178]. LASSO is used to select variables with a higher predictive ability in the situation where there is a large initial set of variables [177]. The degree of shrinkage is determined by a penalty parameter, the value of which is identified through cross-validation to select the set of variables that maximize area under the curve (AUC) [179]. Recursive partitioning was performed on the clinical variables selected using the LASSO to form diagnostic clusters to include and exclude a diagnosis of SMT [180]. Recursive partitioning allows the best hierarchical sequence of variables to classify SMT from non-SMT individuals [177]. The gini index was used as the splitting criteria [177, 180]. Overall model classification was compared to the reference standard (physicians' composite diagnosis) and sensitivity (Se), specificity (Sp), positive and negative predictive value (PPV/NPV) as well as positive and negative likelihood ratios (LR+/-) with 95% CI were calculated [60, 184]. Final selection of clusters to rule-in or out SMT were based on overall

diagnostic validity and needed to reach a LR+≥5 or a LR-<0.2. These thresholds are recognized to produce a moderate shift in post-test probability and therefore clinically useful to diagnose or exclude a disorder [70, 82]. Ease-of-use and clinical applicability with input from the expert physicians was also used to select final clusters [60]. Internal validity was assessed using a validated technique of bootstrapping for recursive partitioning and estimates with 95% CIs were calculated [177, 186]. Analyses were performed using R version 3.3.0 (packages: rpart, glmnet and randomForest; http://cran.r-project.org/).

RESULTS

Table 27 presents the characteristics of participants. A total of 279 individuals participated (96.2% of those recruited) and 80 individuals received a diagnosis of SMT, 35 of which were traumatic tears and 45 degenerative tears. Other primary and secondary diagnoses (n=359) were made as follows: knee osteoarthritis (n=129), patellofemoral pain (n=75), anterior cruciate ligament tears (n=43) or other knee diagnoses (n=32). Individuals with SMT, either traumatic or degenerative had significantly shorter pain duration at the time of consultation (percentage of patients with pain for less than 12 months: trauma: 68.6% (p=0.001), degenerative: 51.1% (p=0.001), compared to other knee disorders: 29.7%). The proportion of individuals with traumatic SMT referred to surgery after the consultation was significantly higher (trauma: 34.3% (p=0.002), compared to degenerative SMT: 24.4% and other knee disorders: 12.6%) and they also had a lower KOOS quality of life score (trauma: 29.1±17.2 (p=0.001), compared to degenerative SMT: 40.8±20.3 and other knee disorders: 39.7±19.7).

Table 28 describes the SMT clinical diagnoses and imaging findings. Overall, SMT was the primary diagnosis of 60% of individuals with a traumatic tear, and of 73% of those with a degenerative tear. Thirty-seven percent of individuals with a traumatic SMT also had a diagnosis of anterior cruciate ligament tear and 40% of those with a degenerative SMT had also a diagnosis of osteoarthritis. The medial meniscus was affected in 89% of traumatic SMT and 91% of degenerative SMT. Complex tears reported on MRI results were present in 43% of traumatic SMT and 40% of degenerative SMT.

Table 27. Characteristics of participants with a knee complaint (n=279)

Characteristics	Traumatic SMT (n=35)	Degenerative SMT (n=45)	Other diagnoses (n=199)
	n (%) or mean (SD)	n (%) or mean (SD)	n (%) or mean (SD)
Age	45.4 (13.9)	49.1 (11.6)	49.8 (16.9)
Body Mass Index (Kg/m^2)	28.1 (5.7)	28.8 (4.6)	29.7 (6.9)
Sex			
Female	18 (51.4)	22 (48.9)	121 (60.8)
Male	17 (48.6)	23 (51.1)	78 (39.2)
Recruitment site			
Orthopaedic clinics	35 (100.0) *	40 (88.9)	154 (77.4)
Family medicine unit/university community	0 (0.0) *	5 (11.1)	45 (22.6)
History of trauma	35 (100.0) *‡	0 (0.0) §	52 (26.1)
Duration of pain at time of consultation			
<3 months	9 (25.7) *	5 (11.1)	20 (10.1)
3-12 months	15 (42.9) *	18 (40.0) §	39 (19.6)
≥ 12 months	11 (31.4) *	22 (48.9) §	140 (70.3)
Referred to surgery after consultation	12 (34.3) *	11 (24.4)	25 (12.6)
Knee Injury and Osteoarthritis Outcome Score (KOOS)			
Pain	59.3 (21.7)	63.0 (17.7)	57.4 (20.0)
Symptoms	70.2 (19.1)	73.6 (17.1)	69.2 (19.6)
Activity of Daily Living	65.4 (22.3)	73.0 (20.3) §	64.3 (22.2)
Sports	29.9 (26.8)	31.9 (24.3)	27.7 (25.2)
Quality of Life	29.1 (17.2) *‡	40.8 (20.3)	39.7 (19.7)
K6 psychological distress scale (/30)	26.8 (4.6)	27.4 (2.7) §	26.0 (4.9)

SMT: symptomatic meniscal tear; SD: standard deviation; KOOS: a score of 0 indicates a severe condition and 100 indicates a normal knee; K6: a score of 6 indicates serious mental illness and 30 indicates no mental illness; * indicates a significant difference ($p< 0.05$) between participants with a traumatic SMT and other diagnoses; § indicates a significant difference ($p< 0.05$) between participants with a degenerative SMT and other diagnoses; ‡ indicates a significant difference ($p< 0.05$) between participants with a traumatic and a degenerative SMT.

Table 28: Description of clinical diagnoses and imaging findings for SMT participants (n=80)

	Traumatic SMT (n=35)	Degenerative SMT (n=45)
Description of clinical diagnoses	**n (%)**	**n (%)**
SMT primary diagnosis	21 (60)	33 (73)
SMT alone with no other knee disorder	10 (29)	16 (36)
SMT combined with another knee disorder		
Osteoarthritis	3 (9)	18 (40)
Patellofemoral pain	4 (11)	6 (13)
Anterior cruciate ligament tear	13 (37)	5 (11)
Other knee disorders	5 (14)	0 (0)
Imaging findings		
Medial meniscal tear	31 (89)	41 (91)
Lateral meniscal tear	7 (20)	4 (9)

SMT: symptomatic meniscal tear; Clinical diagnoses are composite diagnoses made by physicians using history elements, physical examination tests and relevant imaging including MRI. Imaging findings are from MRI radiologists' reports and/or assessment by the physicians; Others knee disorders include: contusion of the tibial plateau (n=2), medial collateral ligament tear (n=2), posterior cruciate ligament tear (n=1);

Table 29 presents the clinical variables associated with the diagnosis of SMT identified through penalized logistic regression. We entered 131 different clinical variables in the penalized logistic regression. Following cross-validation, 20 variables were associated with the diagnosis of traumatic SMT and yielded a maximal area under the curve (AUC) of 0.90 (95%CI: 0.86-0.94), while 12 variables were associated with degenerative SMT with AUC of 0.84 (95%CI: 0.76-0.92). For traumatic SMT, 17 variables were history elements including 8 questions from the KOOS. Three variables were physical tests: Medial Joint Line Tenderness, popliteal fossae tenderness and knee and lower limb alignment. For degenerative SMT, eight patient history elements were associated with the diagnosis including 5 questions from the KOOS questionnaires. Four variables were physical tests: passive knee flexion range of motion, medial patellar facet tenderness, knee and lower limb alignment and the Thessaly test.

Table 29: Clinical variables associated with the diagnosis of SMT identified through penalized logistic regression in participants with a knee complaint (n=279)

	Variables associated with traumatic SMT (n=35)	Variables associated with degenerative SMT (n=45)
History elements	Light work physical demands	6 to 9 months pain duration at time of consultation
	Medial knee pain location	Progressive onset of symptoms
	History of falling on knee at initial trauma	Medial knee pain location
	History of external force at initial trauma	Mild to severe knee pain while twisting or pivoting
	History of pivoting on knee at initial trauma	Mild to moderate pain going up/down stairs
	Absence of popping sensation at initial trauma	Moderate difficulty sitting
	Delayed pain onset following trauma	Mild difficulty running
	Impossible activity continuation following trauma	Mild difficulty jumping
	Difficulty to fully straighten the knee[¥]	
	Stiffness after sitting, lying or resting[¥]	
	Moderate pain while pivoting on knee during activities[¥]	
	Moderate, severe or extreme pain going up/down stairs[¥]	
	Moderate, severe or extreme pain at night while in bed[¥]	
	Moderate, severe or extreme difficulty descending stairs[¥]	
	Extreme difficulty getting in/out of the car[¥]	
	Extreme lack of confidence in knee[¥]	
	No depressive feelings in the last 30 days	
Physical examination tests	Positive Medial Joint Line Tenderness test	Full passive knee flexion
	Absence of popliteal fossae tenderness	Absence of medial patellar facet tenderness
	Normal knee and lower limb alignment	Normal knee and lower limb alignment
		Positive Thessaly test

SMT: symptomatic meniscal tears; LASSO: Least Absolute Shrinkage and Selection Operator; Maximal area under the curve (AUC) was used as the criteria for the final penalty parameter to select variables associated with the diagnosis of SMT. [¥]KOOS questions are assessed on a five-point Likert scale: none, mild, moderate, severe or extreme.

Diagnostic clusters combining history elements and physical examination tests were identified through recursive partitioning using the clinical variables previously identified. Table 30 presents the diagnostic cluster both to include or exclude traumatic SMT compared to other knee disorders. The cluster correctly classifies 32/35 of all cases and 219/244 of all non-cases (see 2x2 table in Appendix 16). The cluster classifies individuals based on: 1- a history of falling or pivot traumatic mechanism; 2- isolated medial or diffuse knee pain location and, 3- a positive Joint Line Tenderness test. This cluster has the following diagnostic statistics: Se of 0.91 (95%CI: 0.77-0.98), Sp of 0.90 (95%CI: 0.85-0.93), NPV of 0.99 (95%CI: 0.96-1.00), PPV of 0.56 (95%CI: 0.42-0.69), LR- of 0.10 (95%CI: 0.13-0.28) and LR+ of 8.92 (95%CI: 6.07-13.11). Figure 3 presents graphical representation of the diagnostic cluster to rule-in or rule-out a traumatic SMT.

For degenerative SMT, multiple clusters were required to include or exclude a degenerative SMT to reach adequate accuracy. Table 31 presents high specificity (Sp) diagnostic clusters to rule-in a degenerative SMT. The clusters correctly classify 26/45 individuals with a degenerative SMT (see 2x2 table in Appendix 16). The clusters classify individuals based on a history of progressive onset of pain and isolated medial knee pain location combined with pain while pivoting during activities or combined with no varus or valgus knee misalignment or full passive knee flexion. The clusters yielded the following diagnostic statistics: Se of 0.58 (95%CI: 0.42-0.72), Sp of 0.91 (95%CI: 0.87-0.94), PPV of 0.55 (95%CI: 0.40-0.70) and LR+ of 6.44 (95%CI: 3.99-10.39). Figure 4 presents graphical representation of the diagnostic clusters to rule-in degenerative SMT.

Table 32 presents diagnostic clusters with a high sensitivity (Se) to rule-out degenerative SMT. The clusters correctly classify 153/234 individuals without degenerative SMT (see 2x2 table in Appendix 16). One cluster classifies individuals with a progressive pain onset combined with isolated anterior or posterior knee pain, while another cluster classifies individuals based on a progressive onset without isolated medial pain location, but with pain in stairs or restricted passive knee flexion. A remaining cluster includes individuals with a traumatic pain onset. The

clusters yielded the following diagnostic statistics: Se of 0.92 (95%CI: 0.82-0.99), Sp of 0.65 (95%CI: 0.59-0.71), NPV of 0.98 (95%CI: 0.94-1.00) and LR- of 0.10 (95%CI: 0.03-0.31). Figure 5 presents graphical representation of the diagnostic clusters to rule-out degenerative SMT.

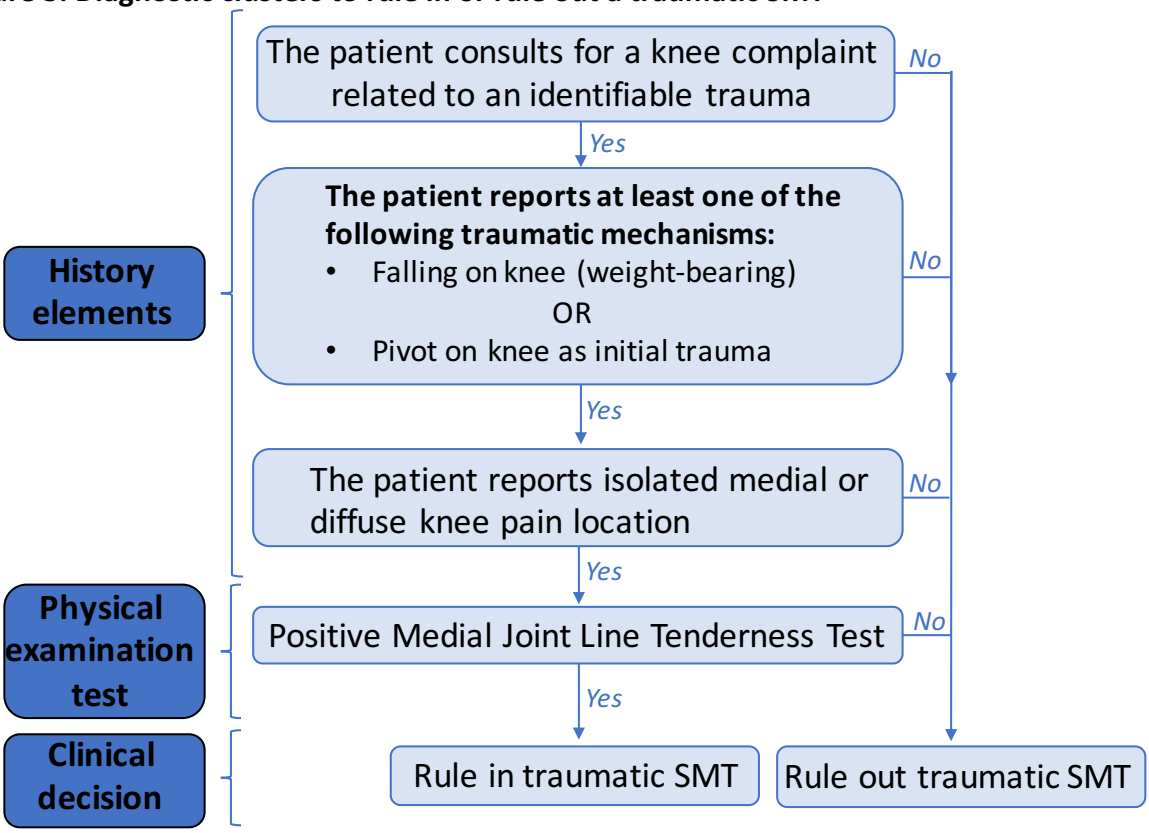
Internal validation estimates and 95% CIs obtained with bootstrapping (n=1000), are presented for all clusters (Tables 30, 31, 32). Internal validation estimates were slightly lower for all clusters but demonstrated positive LR superior to 5 and negative LR inferior to 0.2 indicating moderate shift in post-test probability. We evaluated the inter-rater agreement between the physiotherapist and the expert physicians for the identified physical tests. The following tests with associated inter-rater reliability were included in the various clusters. Knee and lower limb alignment reached a PABAK of 0.73 (95%CI: 0.58-0.84), medial Joint Line Tenderness had a PABAK of 0.77 (95%CI: 0.67-0.84) and passive knee flexion range of motion had a PABAK of 0.86 (95%CI: 0.77-0.92). As for other meniscal tests not included in the clusters, the McMurray's inter-rater reliability was PABAK=0.59 (95%CI: 0.44-0.72) and Thessaly's inter-rater reliability was PABAK = 0.73 (95%CI: 0.42-0.01).

Table 30: Diagnostic clusters combining history elements and physical examination tests to diagnose or exclude traumatic SMT (n=35)

	Se (95% CI)	Sp (95% CI)	NPV (95% CI)	PPV (95% CI)	LR- (95% CI)	LR+ (95% CI)
Knee complaints are <u>likely due</u> to a traumatic SMT in individuals with: • History of fall or pivot on knee at initial trauma *AND* • Isolated medial or diffuse knee pain location *AND* • Positive Medial Joint Line Tenderness test	0.91 (0.77-0.98)	0.90 (0.85-0.93)	0.99 (0.96-1.00)	0.56 (0.42-0.69)	0.10 (0.03-0.28)	8.92 (6.07-13.11)
Internal validation	0.83 (0.66-0.93)	0.88 (0.83-0.92)	0.97 (0.94-0.99)	0.50 (0.37-0.63)	0.19 (0.09-0.40)	6.97 (4.80-10.13)

Se: sensitivity; Sp: specificity; PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR-: negative likelihood ratio. Clusters are obtained using recursive partitioning with all variables associated with the diagnosis of SMT. Possible pain locations included: isolated anterior, posterior, medial, lateral knee pain or diffuse pain. Internal validation was assessed by bootstrapping (n=1000).

Figure 3: Diagnostic clusters to rule in or rule out a traumatic SMT



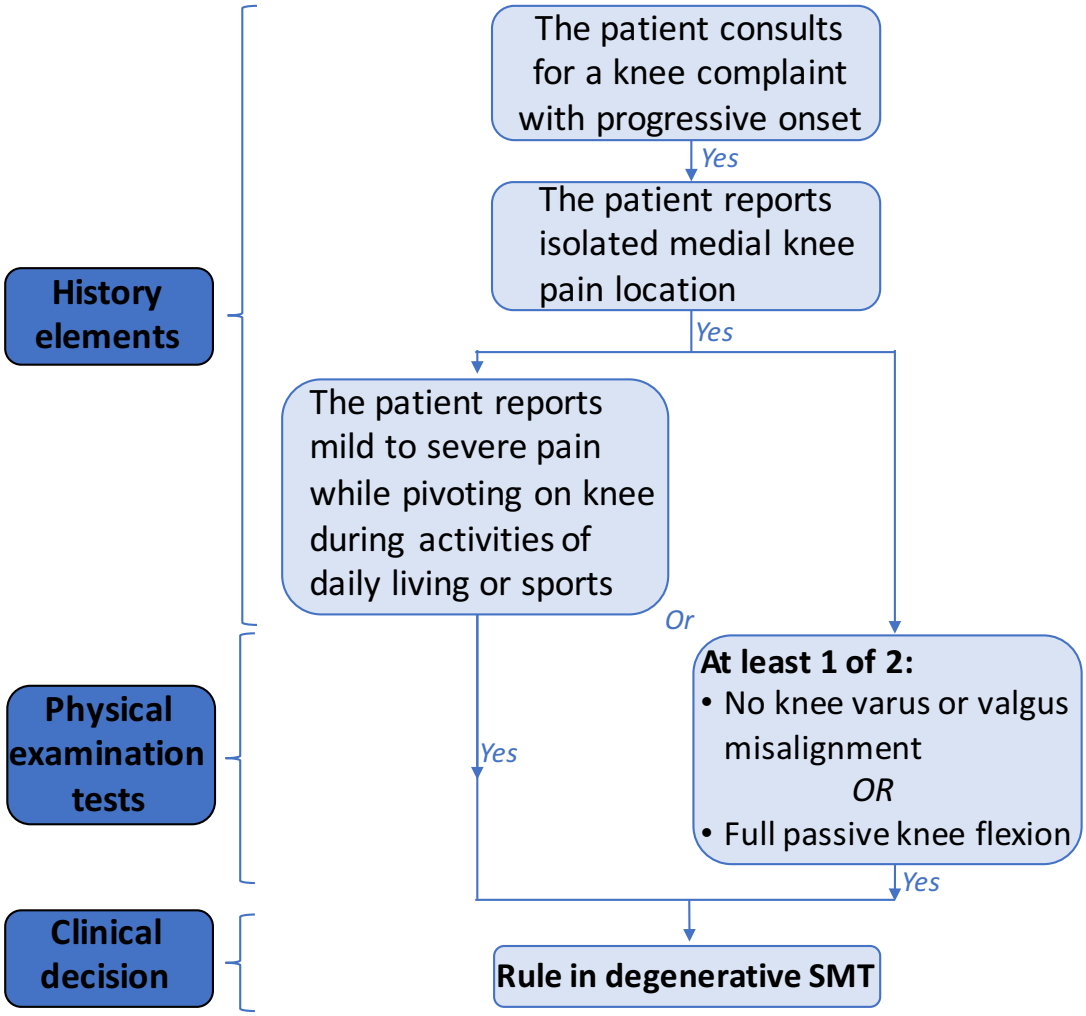
SMT: symptomatic meniscal tear. Clinical decision to rule in or rule out SMT following trauma is based on the clusters of history elements and physical examination test reaching LR+≥5 and LR-≤0.2 indicating at least a moderate change in post-test probability.

Table 31: High specificity diagnostic clusters combining history elements and physical examination tests to diagnose degenerative SMT (n=45)

	Clusters	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	LR+ (95% CI)
Knee complaints are <u>likely due</u> to a degenerative SMT in individuals with:					
Cluster 1	• Progressive onset of pain *AND* • Isolated medial knee pain location *AND* • Mild to severe pain while pivoting on knee during activities or sports	0.58 (0.42-0.72)	0.91 (0.87-0.94)	0.55 (0.40-0.70)	6.44 (3.99-10.39)
Cluster 2	• Progressive onset of pain *AND* • Isolated medial knee pain location *AND* • No knee valgus or varus misalignment *OR* • Full passive knee flexion				
	Internal validation	0.62 (0.47-0.76)	0.89 (0.84-0.93)	0.52 (0.38-0.66)	5.60 (3.65-8.59)

Se: sensitivity Sp: specificity; PPV: positive predictive value; LR+: positive likelihood ratio; Clusters are obtained using recursive partitioning with all variables associated with the diagnosis of degenerative SMT for inclusion with high Sp. Possible pain locations included: isolated anterior, posterior, medial, lateral or diffuse knee pain. Knee limb alignment was assessed visually in standing position with feet close together and included: normal alignment, valgus or varus. Passive knee flexion range of motion was assessed manually in supine position and compared to the healthy side. Knee pain while pivoting during activities or sports was assessed using a five-point Likert scale: none, mild, moderate, severe or extreme. Internal validation was assessed by bootstrapping (n=1000).

Figure 4: Diagnostic clusters to rule in degenerative SMT



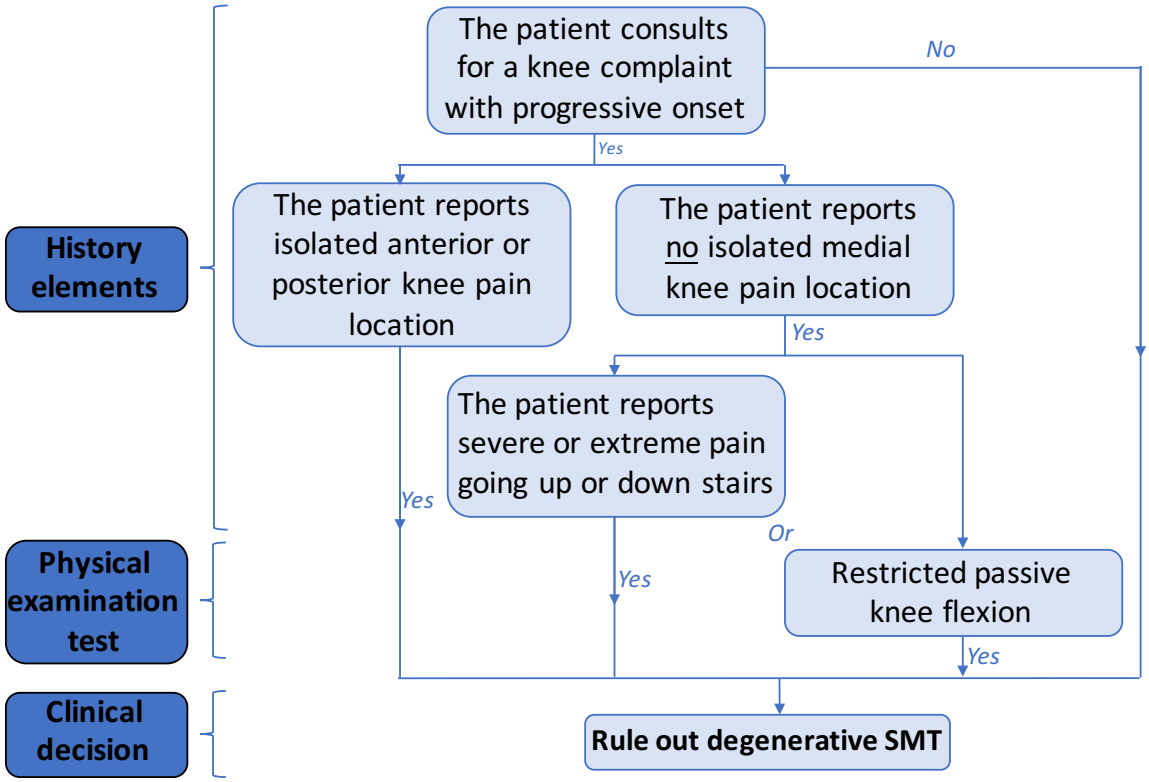
SMT: symptomatic meniscal tear. Clinical decision to rule in degenerative SMT is based on the cluster of history elements and physical examination test reaching LR+≥5 indicating a moderate change in positive post-test probability.

Table 32: High sensitivity diagnostic clusters combining history elements and physical examination tests to exclude degenerative SMT (n=45)

	Clusters	Se (95% CI)	Sp (95% CI)	NPV (95% CI)	LR- (95% CI)
Knee complaints are <u>unlikely due</u> to a degenerative SMT in individuals with:					
Cluster 1	• Progressive onset of pain *AND* • Isolated anterior or posterior knee pain location	0.93 (0.82-0.99)	0.65 (0.59-0.71)	0.98 (0.94-1.00)	0.10 (0.03-0.31)
Cluster 2	• Progressive onset of pain *AND* • No isolated medial pain location *AND* • Severe or extreme pain going up or down stairs *OR* • Restricted passive knee flexion				
Cluster 3	• Traumatic onset of pain				
	Internal validation	0.93 (0.82-0.99)	0.61 (0.55-0.68)	0.98 (0.94-1.00)	0.11 (0.04-0.33)

Se: sensitivity Sp: specificity; NPV: negative predictive value; LR+: negative likelihood ratio; Clusters are obtained using recursive partitioning with all variables associated with the diagnosis of degenerative SMT for exclusion with high Se. Possible pain locations included: isolated anterior, posterior, medial, lateral or diffuse knee pain. Passive knee flexion range of motion was assessed manually in supine position and compared to the healthy side. Pain going up/down stairs was assessed using a five-point Likert scale: none, mild, moderate, severe or extreme. Internal validation was assessed by bootstrapping (n=1000).

Figure 5: Diagnostic clusters to rule out degenerative SMT.



SMT: symptomatic meniscal tear. Clinical decision to rule out degenerative SMT is based on the clusters of history elements and physical examination test reaching LR+≤0.2 indicating a moderate change in negative post-test probability.

DISCUSSION

This study assessed the diagnostic validity of clusters combining history elements and physical examination tests to include or exclude traumatic and degenerative SMT compared to other knee disorders. We developed a diagnostic cluster with the combination of a pivoting or falling traumatic mechanism with isolated medial or diffuse knee pain location and a positive medial Joint Line Tenderness test that was able to diagnose or exclude traumatic SMT. Positive post-test probability increased by 43.5% from a baseline prevalence of 12.5% and negative post-test probability increased by 11.5% from a negative baseline prevalence of 87.5%. The magnitude of the likelihood ratio observed using the developed cluster exceeded cut-off values of LR+=5 and LR-=0.2, indicating a clinically useful moderate change in post-test probability [82]. The cluster correctly classified 91% and 90% of all cases and non-cases. However, there was more false-positive findings than false-negative. Our results present superior accuracy estimates than another study on the diagnostic validity of combining history elements and physical examination tests for traumatic meniscal tears [102]. *Wagermaker et al*. found that combining age over 40 years old, impossible activity continuation, weight bearing during trauma and pain at passive flexion yielded LR+ of 5.8 (95%CI: 1.3-26.8), but correctly classifying only 15% of cases [102]. Compared to our study, they recruited participants with knee pain for less than 5 weeks, while most of our participants were recruited in secondary care and 74.3% had pain for more than 3 months at the time of consultation. An acute presentation of a knee trauma may lower the diagnostic validity of the physical examination tests because of the pain, effusion and muscle guarding, which could explain the difference between our estimates [6]. Another distinction is their use of a MRI-only reference standard in their study, which may have produced more inaccurate findings without an expert clinician's confirmation of SMT [6, 10, 224, 225].

The Medial Joint Line Tenderness test was the only physical test in our diagnostic cluster for traumatic SMT. Previous findings showed that this test was not diagnostically valid when individually performed [119, 153, 220]. Interestingly, the test when combined into a diagnostic cluster with patient's history elements, yields adequate diagnostic accuracy but

only for traumatic SMT. This is perhaps explained because both degenerative SMT and knee osteoarthritis may often exhibit a positive Medial Joint Line Tenderness, making this test less useful to discriminate disorders of progressive origin [226].

The combination of progressive onset of pain and isolated medial pain location with pain while pivoting during activities, absence of varus or valgus misalignment or full passive knee flexion was able to moderately diagnose degenerative SMT, while reporting no isolated medial pain location, pain in stairs or having restricted passive knee flexion excluded degenerative SMT. Positive post-test probability increased by 38.9% from a baseline prevalence of 16.1% and negative post-test probability increase by 14.1% from a negative baseline prevalence of 83.9%. The magnitudes of the likelihood ratios observed using the developed clusters reached cut-off values indicating clinical relevance, but represents only a moderate change in post-test probability [82]. The clusters correctly classified 58% and 65% of all cases and non-cases (see 2x2 tables in Appendix 16). While this can be expected from clustering and combination analyses [11, 151, 222], it may also be explained by the heterogeneous presentation of degenerative SMT which often overlaps with knee osteoarthritis [42], as it was likely the case in our cohort; 40% participants had combined diagnoses of SMT and osteoarthritis [11].

Our diagnostic clusters compare favourably to previous evidence combining history and physical examination for degenerative SMT [11]. *Katz et al*. found that localized pain, ability to fully bend the knee, pain duration of less than a year, no varus alignment, no pes planus, and absence of joint space narrowing on radiographs could be combined to diagnose or exclude meniscal tear [11]. The authors propose an additive index score with cut-off values and Se of up to 98% and Sp of 36% to exclude SMT or Sp of up to 94% with Se of 28% to include SMT. When compared with our study, the cohort was also recruited from secondary care and the authors used an expert diagnosis of degenerative SMT, but not all diagnoses included a confirmatory MRI in addition to a clinical examination [11]. Similar to the results of *Katz et al*., our diagnostic clusters did not include mechanical symptoms or special physical tests that are

traditionally proposed to diagnose internal derangement caused by a meniscal tear (i.e: symptoms of clicking, catching or hanging up or locking, McMurray's test and Thessaly's test). More than 60% of all SMT cases in our cohort reported having no mechanical symptoms as reported on individual questions from the KOOS questionnaire. This result is comparable to other cohorts with meniscal tears [11, 41, 42, 102, 151]. Also, mechanical symptoms, when present, do not arise exclusively from meniscal tears and may be positive in other disorders such as ACL tear or osteoarthritis [6, 11]. This adds to the growing evidence suggesting that these history elements and tests may have limited diagnostic value, as well as a possibly limited use as a surgical criterion [11, 117, 119, 120, 151, 153, 217, 220-222, 227-229].

Strengths and limitations

This study represents the derivation phase of the diagnostic clusters and will require external validation in another cohort before widespread clinical use. We recruited participants in a consecutive cohort of patients from different settings in primary and secondary care, allowing for a broad variety of cases and non-cases. However, most SMT were found in secondary care and thus our results may not be directly generalizable to a primary care population. Nonetheless, only a fraction of cases was severely impaired and required surgical treatment. Our data collection procedure ensured blinding of the evaluators between the index tests and the reference standard. However, since both the physiotherapist and the physicians use the same tests definitions, this could have created an incorporation bias. Our composite reference standard by expert physicians included history elements, physical examination tests as well as MRI confirmation and other imaging tests for all SMT [168-170]. The statistical approach we used was able to identify multiple variables from both history elements and physical examination tests into easy-to-use diagnostic clusters [177]. A limitation is that only one physiotherapist and one expert physician evaluated each participant and this may limit the generalizability of our results. However, we found excellent inter-rater agreement between evaluators for the diagnosis of SMT and for other knee disorders.

CONCLUSIONS

We identified diagnostic clusters combining history elements and physical examination tests that can support the differential diagnosis of SMT compared to other knee disorders. These clusters could be used by clinicians involved in musculoskeletal care to diagnose or exclude SMT in settings with limited access to advanced imaging to initiate early efficient conservative management or to propose an imaging or specialty referral when required. This initial step represents the derivation of the clusters and further research should externally validate these results in other clinical settings before clinical use.

Article 6: Initial derivation of diagnostic clusters combining history elements and physical examination tests for symptomatic knee osteoarthritis.

Décary S, Feldman D, Frémont P, Pelletier J-P, Martel-Pelletier J, Fallaha M, Pelletier B, Belzile S, Sylvestre M-P, Vendittoli P-A, Desmeules F.

Décary, Simon, PT, PhD candidate[1,2]
Feldman, Debbie, PhD[1]
Frémont, Pierre, MD, PhD[3]
Pelletier, Jean-Pierre, MD, PhD[4]
Martel-Pelletier, Johanne, PhD[4]
Fallaha, Michel, MD[5]
Pelletier, Bruno, MD[5]
Belzile, Sylvain, MD[6]
Sylvestre, Marie-Pierre, PhD[7]
Vendittoli, Pascal-André, MD, MSc, FRSC[2,5]
Desmeules, François, PT, PhD[1,2]

1- School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada.
2- Orthopaedic Clinical Research Unit, Maisonneuve-Rosemont Hospital Research Center, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
3- Department of Rehabilitation, Faculty of Medicine, Laval University, Quebec City, Quebec, Canada
4- Osteoarthritis Research Unit, University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada.
5- Department of Surgery, Maisonneuve-Rosemont Hospital, University of Montreal, Montreal, Quebec, Canada, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
6- Department of Surgery, Laval University Hospital Center (CHUL), Laval University, Quebec, Quebec, Canada.
7- University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada. Department of Social Preventive Medicine, School of Public Health, Université de Montréal

Article en processus de révision dans le journal *Physiotherapy*, Novembre 2017.

En tant que premier auteur, j'ai réalisé le développement du protocole de cette étude, la collecte des données, l'analyse des résultats ainsi que la rédaction du manuscrit. FD et PAV

ont contribué au développement du protocole, à l'analyse des résultats, à la rédaction et à la révision du manuscrit. MF, BP, SB et PF ont contribué au développement de l'évaluation musculosquelettique standardisée et ont agi à titre de médecins experts (standard de référence) pour tous les patients. JMP, JPP, DF et MPS ont contribué au développement du protocole, à la vérification des méthodes d'analyse des données et à la présentation des résultats. Tous les auteurs ont contribué à la révision de la version finale du manuscrit.

ABSTRACT

Objective: To assess the validity of clusters combining history elements and physical examination tests to diagnose symptomatic knee osteoarthritis (SOA) compared to other knee disorders.

Design: Prospective diagnostic accuracy study.

Settings: Two orthopaedic clinics, two family medicine clinics and a university community.

Patients: 279 consecutive patients consulting for a knee complaint.

Methods: History elements and standardized physical examination tests were independently obtained by a physiotherapist and compared to an expert physicians' composite diagnosis including clinical examination and imaging. Recursive partitioning was used to develop diagnostic clusters for SOA.

Main Outcome Measures: Diagnostic accuracy measures were calculated including sensitivity(Se), specificity(Sp) and positive and negative likelihood ratios (LR+/-) with associated 95% confidence intervals (CI).

Results: One hundred and twenty-nine patients had a diagnosis of SOA (46.2%). Most cases (76%) had combined tibiofemoral and patellofemoral knee OA and 63% had radiological Kellgren-Lawrence grades of 2 or 3. Different combinations of history elements and physical examination tests were used in clusters to accurately discriminate SOA from other knee disorders. These included age of patients, body mass index, presence of valgus/varus knee misalignment, palpable knee crepitus and limited passive knee extension. Two clusters to rule in SOA reached a LR+ of 13.6 (95%CI:6.5-28.4) and three clusters to rule out SOA reached a LR- of 0.11 (95%CI:0.06-0.20).

Conclusion: Diagnostic clusters combining history elements and physical examination tests could support the differential diagnosis of SOA compared to various knee disorders without systematically relying on imaging. This could support primary care clinicians' role to efficiently manage these patients.

INTRODUCTION

Symptomatic knee osteoarthritis (SOA) is ranked 11th out of 291 conditions as a major cause of disability worldwide and is increasingly prevalent [230]. Current evidence proposes initial conservative management, for which physiotherapists may be uniquely skilled to provide [231]. However, to fully play this role in clinical practice, physiotherapists and other primary care clinicians need to be able to identify SOA patients from a diversity of other knee disorders based on an appropriate musculoskeletal examination and without systematically relying on imaging for which they may not have access or rights to order [46, 232, 233].

The *American College of Rheumatology* (*ACR*) first established the clinical diagnostic criteria of knee OA [96]. These criteria were aimed mainly at distinguishing individuals with OA from individuals with other rheumatoid conditions in rheumatology clinics [96]. This led the criteria to be of limited diagnostic value when applied to a cohort of community dwelling older adults with a sensitivity and specificity of only 41% and 75% [97]. However, other clinical variables could reasonably predict the presence or absence of radiographic OA [234]. In 2010, the *European League Against Rheumatisms (EULAR)* revised the diagnostic criteria for knee OA [46]. It now proposes that the probability of having radiographic knee OA is 99% in individuals aged older than 45 years old when six signs and symptoms are all present: persistent knee pain, limited morning stiffness, reduced function, palpable crepitus, restricted knee movement and knee joint bony enlargement [46].

Current evidence concerning the clinical diagnosis of SOA are thus limited for physiotherapists and primary care clinicians' differential diagnosis needs because they mostly rely on population-based studies to determine a clinical profile associated with radiological OA instead of diagnostic accuracy studies aimed at developing criteria for the differential diagnosis of various common knee disorders. Therefore, the objective of this study is to derivate a clinical tool by assessing the validity of clusters combining history elements and physical examination tests to diagnose SOA compared to other knee disorders.

METHODS

Study design and settings

The present study reports on SOA patients from a multicenter prospective diagnostic cohort aimed at developing a series of diagnostic support tools for common knee disorders. Several diagnostic clusters have been developed for other common knee disorders and have been published [235]. In order to represent a wide spectrum of knee disorders and clinical presentation we recruited prospective patients seeking care for an active knee complaint from two orthopaedic clinics, two primary care family medicine clinics and a university community between November 2014 and August 2016 [235]. The present study, its design, methodology and reporting of results conform to the Standards for Reporting Diagnostic Accuracy Studies 2015 (STARD) [157]. The study was approved by the ethics comity of all participating institutions and all participants signed an informed consent form.

Participants

Inclusion criteria in the prospective diagnostic cohort were: 1- 18 years of age or older; 2- consulting or referred to one of the participating physicians for a knee complaint; 3- able to understand and speak French. Patients previously treated by the participating physicians were excluded, as well as patients who had undergone lower limb surgery in the past six months, patients with a knee arthroplasty or who presented with more than two other lower limb pathologies or if they suffered from any systemic inflammatory disorder related to their knee complaint.

Data collection

Patients' characteristics and history element

Selected history elements were systematically collected including: gender, age, work status, occupation physical demands, comorbidities, affected side, duration of knee symptoms, knee pain location (anterior, posterior, medial, lateral or diffuse knee pain) and traumatic or progressive pain onset. Participants also completed the Knee Injury and Osteoarthritis Outcome Score (KOOS), a 42-item questionnaire that includes 5 domains: pain, symptoms,

function in daily living, function in sport and recreation and knee-related quality of life [159]. Severe psychological distress was screened using the K6 screening scale [163].

Physical examination data collection procedure

Before the start of the study, clinicians met with the research personnel to standardize techniques and interpretation of the physical tests as well as the definition of a SOA diagnosis compared to other knee diagnoses. Each participant was independently assessed by two evaluators on the day of their visit: a physiotherapist and one of the participating physicians. The physiotherapists' training included a clinical master in physiotherapy with one year of clinical experience. The five participating physicians (three orthopaedic surgeons and two sports medicine physicians) each had more than 20 years of experience. The physiotherapist always completed data collection prior to the physicians' consultation. His role was to collect all standardized index tests. Both the physiotherapist and the physicians were blinded to each other's findings.

The physical examination tests used to diagnose SOA or exclude other knee disorders included: medial and lateral joint line tenderness, patellar facets tenderness, popliteal cavity or anserine tenderness, presence of warmth, bulge and ballottement tests for intra-articular effusion, palpation of bony enlargement, palpable crepitus in tibiofemoral or patellofemoral compartments, *McMurray's test*, *Thessaly's test*, patellar tenderness at mobilization or compression, *Clark's sign* and J-sign patellar maltracking during knee extension. These tests were coded as positive/negative/uncertain/not evaluated. We also assessed body mass index (BMI), active and passive range of motion (restricted/not restricted), visual assessment of lower limb alignment (valgus/varus/recurvatum/no misalignment), varus and valgus stress test at 0° and 30° of flexion (normal/painful/medial or lateral collateral ligament pseudo-laxity), visual assessment of motor control and pain during squatting or descending a step (able to maintain alignment during task/painful task/unable to maintain alignment/no pain). All tests could also be rated, when relevant, as uncertain or not evaluated.

Reference standard definition

Following his independent collection of history elements and physical examination tests, the physician was presented with any imaging results and radiologists reports. All participants were required to have a standardized radiograph of their knee that included the following three views: anterior-posterior in weight-bearing, lateral and skyline views [236]. Magnetic resonance imaging (MRI) was obtained when requested by the physician for suspected ligament injury, symptomatic meniscal tear or to include/exclude any other knee diagnoses or pathology. The physician made a final primary and secondary (if necessary) composite diagnosis. This final composite diagnosis, including the patient's history, physical examination and appropriate imaging interpretation, was considered the reference standard against which the index tests, independently collected by the physiotherapist, were compared for all participants. Diagnoses were classified into 1- SOA or 2- other knee disorders based on the physicians' composite diagnosis. Radiographic knee OA (Kellgren-Lawrence grade ≥ 1) without relevant history and symptoms was not considered SOA by the expert physicians.

Statistical analysis

Descriptive statistics were used to present the participants' characteristics and Student t-tests as well as chi-squared tests were used to compare participants with SOA to those with other knee diagnoses. Inter-rater agreement between the physiotherapist and the physician was calculated for physical examination tests using *Prevalence and Bias Adjusted Kappas* (PABAK) with associated 95% CI [172]. Physical tests needed to reach an inter-rater agreement value $\kappa>0.4$ to remain in the final models development [172].

Diagnostic clusters combining history elements and physical examination tests results were developed using a two-step method [177]. First, Least Absolute Shrinkage and Selection Operator (LASSO) penalized logistic regression was used to identify history elements and physical examination tests associated with the diagnosis of SOA [177]. LASSO is used to select variables with a higher predictive ability in the situation where there is a large initial set of variables [177]. Variables selection is obtained through shrinking the beta coefficients of

unimportant variables to zero [177]. The degree of shrinkage is determined by a penalty parameter, the value of which is identified through cross-validation to select the set of variables that maximize area under the curve (AUC) [179]. In the second step, we used recursive partitioning on the clinical variables previously identified to form diagnostic clusters to include or exclude a diagnosis of SOA [177]. Recursive partitioning allows the best hierarchical sequence of variables to classify SOA from non-SOA individuals [177].

Overall model classification was compared to the reference standard (physicians' composite diagnosis) and sensitivity (Se), specificity (Sp), positive and negative predictive values (PPV/NPV) and positive and negative likelihood ratios (LR+/-) with 95% CI were calculated [173]. Final selection of clusters to include and exclude SOA were based on overall diagnostic validity and needed to reach a LR+≥5 or a LR-<0.2. These thresholds are recognized to produce a moderate shift in post-test probability and therefore clinically useful to make or exclude a diagnosis [173]. If more than one cluster yielded such values, ease of use and clinical applicability was also used to select final clusters. Internal validity was assessed using a validated technique of bootstrapping for recursive partitioning and estimates with 95% CIs were compared to the proposed clusters [177]. Analyses were performed using R version 3.3.0 (packages: rpart, glmnet and randomForest; http://cran.r-project.org/).

RESULTS

Table 33 presents the characteristics of participants. A total of 279 individuals participated (Appendix 11) and 359 primary and secondary diagnoses were made as follows: symptomatic meniscal tears (n=80), patellofemoral pain (n=75), anterior cruciate ligament tears (n=43) or other knee diagnoses (n=32). One hundred twenty-nine individuals had a diagnosis of SOA, while the remaining 150 did not. Compared to all individuals with other knee disorders, those with SOA were significantly older (SOA: 59.2±12.2 years, others: 40.5±13.2 years, $p<0.001$) and had a higher BMI (SOA: 31.7± 6.7 kg/m^2, others: 27.4± 5.7 kg/m^2, $p<0.001$). Most individuals with SOA were female (61.2%), were recruited from the orthopaedic clinics (82.2%) and the proportion of individuals sent to surgery at the time of consultation was low (14.7%), but this was not significantly different from other knee disorders (p=0.53). Individuals with SOA had significantly more pain, symptoms, limitation in activity of daily living and sports than their non-SOA counterpart as measured by the KOOS questionnaire ($p<0.05$).

Table 34 describes the clinical diagnoses and imaging findings for SOA. Overall, SOA was the primary diagnosis of 91% of patients and 77% as their sole diagnosis. Seventy-six percent had radiographic OA in two or more compartments with 71% being of Kellgren-Lawrence grade 1 to 3.

Table 33: Characteristics of participants (n=279)

Characteristics	SOA (n=129)	Other knee disorders (n=150)
	n (%) or mean (SD)	n (%) or mean (SD)
Age (years)	59.2 (12.2)*	40.5 (13.2)
Sex (Female)	79 (61.2)	82 (54.7)
Body Mass Index (Kg\m^2)	31.7 (6.7)*	27.4 (5.7)
Recruitment site		
Orthopaedic clinics	106 (82.2)	123 (82.0)
Family medicine unit/university community	23 (17.8)	27 (18.0)
Bilateral knee pain	38 (29.5)*	15 (10.0)
Duration of pain at time of consultation		
<3 months	10 (7.8)*	24 (16.0)
3-12 months	25 (19.4)*	47 (31.3)
≥ 12 months	94 (72.8)*	79 (52.7)
Sent to surgery after consultation	19 (14.7)	29 (19.3)
Knee Injury and Osteoarthritis Outcome Score (KOOS)		
Pain	52.6 (19.5)*	63.7 (18.8)
Symptoms	64.6 (20.0)*	74.8 (17.0)
Activity of Daily Living	59.2 (21.5)*	71.6 (21.0)
Sports	21.4 (22.0)*	34.9 (26.2)
Quality of Life	37.8 (20.8)	39.2 (18.9)
K6 psychological distress scale (/30)	26.3 (4.4)	26.4 (4.8)

SOA: symptomatic knee osteoarthritis; SD=standard deviation; KOOS: 0 indicates a severe condition and 100 indicates a normal knee; K6: 6 indicates serious mental illness and 30 indicates no mental illness. * indicates a significant difference ($p < 0.05$) between participants with SOA and those without.

Table 34: Description of clinical diagnoses and imaging findings for participants with SOA (n=129)

Description of clinical diagnoses	n (%)
SOA as the primary diagnosis	117 (91)
SOA as the sole diagnosis	99 (77)
SOA combined with another knee disorder	
Symptomatic degenerative meniscal tear	18 (14)
Symptomatic traumatic meniscal tear	3 (2)
Patellofemoral pain	4 (3)
Anterior cruciate ligament tear	3 (2)
Other knee disorders	2 (2)
Imaging findings	
Knee compartments with radiographic OA	
2 or more compartments	98 (76)
Isolated medial compartment	21 (16)
Isolated patellofemoral compartment	9 (7)
Isolated lateral compartment	1 (1)
Kellgren-Lawrence grade*	
1	11 (8)
2	42 (33)
3	39 (30)
4	23 (18)

SOA: symptomatic knee osteoarthritis; Clinical diagnoses are composite diagnoses made by physicians using history elements, physical examination tests and relevant imaging; Imaging findings are from radiographs assessment by the physicians; Others knee disorders combined with symptomatic knee OA included: symptomatic medial collateral ligament sprain (n=1) and avascular osteonecrosis of tibial plateau (n=1). Grades are for Kellgren-Lawrence scale in the most affected compartment; *n=14 (11%) are missing data for K-L grades.

Table 35 presents the clinical variables associated with the presence or absence of SOA identified through penalized logistic regression (LASSO). One hundred and thirty-one different clinical variables were entered in the penalized logistic regression. Following cross-validation, 19 variables were predictive of the diagnosis of SOA and yielded a maximal area under the curve (AUC) of 0.92 (95%CI: 0.90-0.94). Of these, 13 were history elements and six physical tests or clinical measures. History elements included: age, work status, occupation physical demands, feeling depressed as measured by the K6 and nine questions were from "pain", "function and daily living", "function, sports and recreational activities" and "quality of life" subscales of the KOOS questionnaire. Physical examination and clinical measures associated variables included: body mass index, palpable crepitus in any compartments, lateral joint line tenderness, lower limb alignment, limited passive knee range of motion in extension and patellar J-sign maltracking during seated knee extension.

Table 35: Clinical variables associated with the diagnosis of SOA identified through penalized logistic regression in participants with a knee complaint (n=279)

	Variables associated with the presence of symptomatic knee OA	**Variables associated with the absence of symptomatic knee OA**
History elements	Age Retired work status Severe pain at night while in bed[¥] Severe pain standing upright[¥] Severe difficulty standing[¥] Mild general difficulty with knee[¥] Feeling depressed in last 30 days	Sedentary occupation Mild pain going up or down stairs[¥] Mild difficulty descending stairs[¥] Moderate difficulty squatting[¥] Mild difficulty pivoting[¥] Moderate difficulty kneeling[¥]
Physical examination and clinical measures	High body mass index Presence of palpable crepitus in any knee compartments Lateral Joint Line Tenderness Limited passive knee range of motion in extension	Normal lower limb alignment Patellar maltracking (J-sign) during knee extension

SOA: symptomatic knee osteoarthritis; Maximal are under the curve (AUC) was used as the criteria for the final penalty parameter to select variables associated with the presence or absence of SOA. [¥]KOOS questions are evaluated using a five-points likert scale: none, mild, moderate, severe or extreme. Passive knee extension range of motion was compared to the unaffected side. Lower limb alignment included the presence of varus, valgus, recurvatum misalignment or normal.

Diagnostic clusters of history elements and physical examination tests were identified through recursive partitioning using the clinical variables associated with the diagnosis of SOA. Table 36 presents two high specificity (Sp) diagnostic clusters to rule in SOA compared to other knee disorders. The clusters accurately classify 82/129 individuals with SOA (see 2x2 table in Appendix 18). One cluster contains individuals aged 50 to 58 years old, considered obese (BMI$\geq$30 kg/m^2) with valgus or varus knee misalignment or with limited passive knee extension. A second cluster contains individuals aged older than 58 years old with palpable crepitus in any compartments. The clusters yielded the following diagnostic statistics: Se of 0.64 (95%CI: 0.55-0.72), Sp of 0.95 (95%CI: 0.91-0.98), PPV of 0.92 (95%CI: 0.84-0.97) and LR+ of 13.62 (95%CI: 6.53-28.41).

Table 37 presents three high sensitivity (Se) diagnostic clusters to rule out SOA compared to other knee disorders. The clusters accurately classify 106/150 individuals without SOA (see 2x2 table in Appendix 18). One cluster contains individuals aged younger than 40 years old. A second cluster contains individuals aged between 40 and 50 years old, without palpable crepitus and with a BMI lower than 35 kg/m^2. The last cluster contains individuals aged between 40 and 58 years old, with palpable crepitus but considered non-overweight (BMI$<$ 26 kg/m^2). The clusters yielded the following diagnostic statistics: Se of 0.92 (95%CI: 0.86-0.96), Sp of 0.71 (95%CI: 0.63-0.78), NPV of 0.91 (95%CI: 0.85-0.96) and LR- of 0.11 (95%CI: 0.06-0.20).

Figures 6 and 7 present the graphical representation of the diagnostic clusters to rule in or rule out SOA. Internal validation estimates and 95% CIs, obtained with bootstrapping (n=1000) are presented for all clusters (Tables 36 and 37). Good inter-rater agreement was found for all included physical examination tests results with PABAK of 0.64 (95%CI: 0.45-0.78, n=94) for palpable crepitus in any compartments, PABAK of 0.73 (95%CI: 0.58-0.84, n=120) for lower limb alignment and PABAK of 0.86 (95%CI: 0.76-0.92, n=184) for passive knee extension range of motion.

Table 36: High specificity diagnostic clusters combining history elements and physical examination tests to rule in SOA

		Se (95% CI)	Sp (95% CI)	PPV (95% CI)	LR+ (95% CI)
Cluster 1	**Knee complaints are <u>likely due</u> to SOA in individuals with:** • Age 50 to 58 years old *AND* • BMI $\geq$ 30 kg/m^2 *AND* • Valgus or varus knee misalignment *OR* • Limited passive knee extension range of motion	0.64 (0.55-0.72)	0.95 (0.91-0.98)	0.92 (0.84-0.97)	13.62 (6.53-28.41)
Cluster 2	• Age > 58 years old *AND* • Palpable crepitus in any compartments				
	Internal validation	0.57 (0.48-0.65)	0.95 (0.90-0.98)	0.90 (0.81-0.96)	10.61 (5.32-21.17)

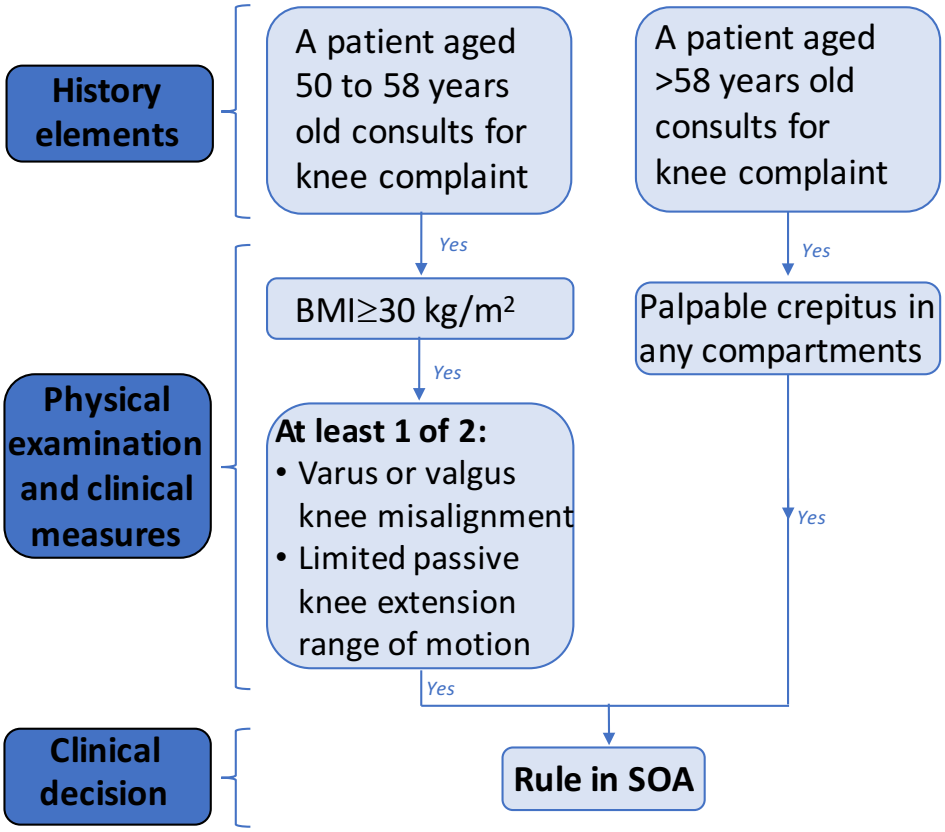
SOA: symptomatic knee OA; BMI: Body mass index; Se: sensitivity; Sp: specificity; PPV: positive predictive value; LR+: positive likelihood ratio; Clusters are obtained using recursive partitioning with all variables associated with the diagnosis of SOA to rule in SOA with high Sp. Lower limb alignment is assessed visually in standing position with feet close together. Passive knee extension range of motion was compared to the unaffected side. Palpable crepitus is assessed manually during active seated knee extension or standing-up with a hand covering tibiofemoral and patellofemoral compartments. Internal validation was assessed by bootstrapping (n=1000).

Table 37: High sensitivity diagnostic clusters combining history elements and physical examination tests to rule out SOA

		Se (95% CI)	Sp (95% CI)	NPV (95% CI)	LR- (95% CI)
Knee complaints are <u>not likely due</u> to SOA in individuals with:					
Cluster 1	• Age < 40 years old	0.92 (0.86-0.96)	0.71 (0.63-0.78)	0.91 (0.85-0.96)	0.11 (0.06-0.20)
Cluster 2	• Age between 40 and 50 years old *AND* • Absence of palpable crepitus in any compartments *AND* • BMI< 35 kg/m^2				
Cluster 3	• Age between 40 and 58 years old *AND* • Palpable crepitus in any compartments *AND* • BMI< 26 kg/m^2				
	Internal validation	0.91 (0.84-0.95)	0.61 (0.53-0.69)	0.88 (0.81-0.94)	0.15 (0.09-0.26)

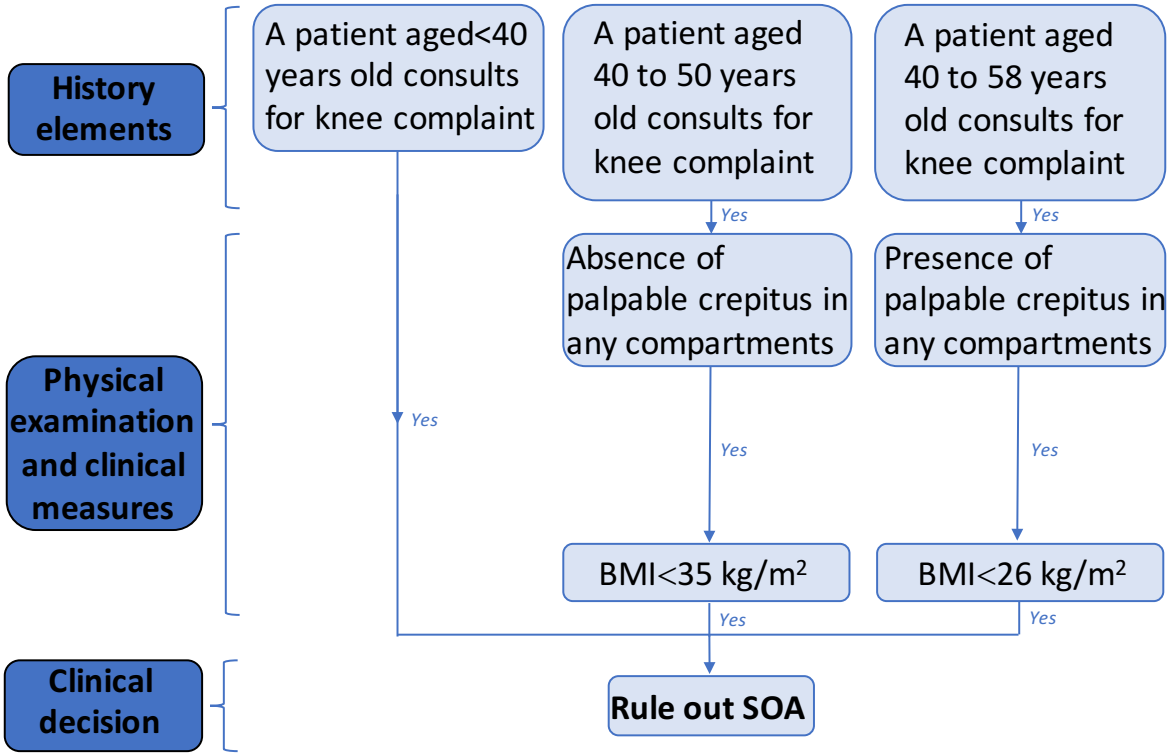
SOA: symptomatic knee OA; BMI: Body mass index; Se: sensitivity; Sp: specificity; NPV: negative predictive value; LR-: negative likelihood ratio. Clusters are obtained using recursive partitioning with all variables associated with the diagnosis of SOA to rule out SOA with high Se. Palpable crepitus is assessed manually during active seated knee extension or standing-up with a hand covering tibiofemoral and patellofemoral compartments. Internal validation was assessed by bootstrapping (n=1000).

Figure 6: Diagnostic clusters to rule in SOA



SOA: symptomatic knee osteoarthritis. Clinical decision to rule in SOA is based on clusters of history elements and physical tests or clinical measures reaching LR+≥5 indicating a moderate change in positive post-test probability.

Figure 7: Diagnostic clusters to rule out SOA.



SOA: symptomatic knee osteoarthritis. Clinical decision to rule out SOA is based on clusters of history elements and physical tests or clinical measures reaching LR+≤0.2 indicating a moderate change in negative post-test probability.

DISCUSSION

The objective of this study was to assess the validity of clusters combining history elements and physical examination tests to diagnose SOA compared to other knee disorders. Our results show that the risk factors for knee OA of older age and higher BMI combined with the presence of valgus or varus knee misalignment, palpable knee crepitus and limited passive knee extension formed accurate diagnostic clusters [237, 238]. Two inclusion clusters reached a substantial positive LR (LR+≥10) to increase post-test probability by 45.8%, to 92.0%, in the present cohort [173]. Three exclusion clusters reached a near substantial negative LR (LR-<0.1) to increase negative post-test probability by 37.2%, to 91.0%. The clusters accurately classified 64% and 71% of all cases and non-cases. This potentially enlightens the heterogeneous clinical presentation of SOA and possible overlap of signs and symptoms with other knee disorders like degenerative meniscal tears [42]. Therefore, clinicians should be aware that misclassification is possible and further assessment of complex cases may be required to guide the decision-making process of imaging and specialty referrals for SOA [218, 239-241]. By example, no high specificity cluster covered the 40 to 50-year-old age group to accurately rule in SOA [10]. This indicates that SOA in this age group may have the most diagnostic uncertainty compared to patients with other knee disorders.

The variables associated with the diagnosis of SOA in this study are generally concordant with the ones identified by other authors in population-based studies [234, 242-248]. These variables included age, BMI, difficulty with stairs, fixed-flexion deformity and knee crepitus [234, 242-248]. These variables yielded an AUC of 0.80 in *Peat et al*. study [234], while we obtained an AUC of 0.92 in our cohort. This may be attributable to differences in the reference standard and characteristics of the population used in the present study. The previous authors' reference standard definition included adults older than 50 years with definite radiographic OA [234]. In the present study, the reference standard was a clinical diagnosis of SOA based on relevant history, physical examination and imaging evaluation by expert physicians. Also, the participants of the present cohort may have pain, symptoms and functional limitations on average higher than community-dwelling older adults from previous

studies [97, 234]. This may be explained by the important proportion of individuals seen upon referral in secondary care who have may have more chronic symptoms [96]. Thus, this may have increased the likelihood of obtaining a positive physical examination in this study [97]. However, these patients may still represent typical patients who seek care for knee complaints as only 15% were surgical candidates.

Other *ACR* and *EULAR* defined diagnostic criteria are missing from the clusters identified in this study [46, 96, 234]. Bony enlargement at palpation was not routinely executed by the expert physicians and thus could not be verified for inter-rater reliability and had to be excluded from our analysis. Other variables not identified were limited morning stiffness, bony tenderness, no palpable warmth and persistent knee pain. A possible explanation for these findings is that our study discriminates SOA from other knee disorders which may also exhibit these signs and symptoms, yielding false positive and false negative findings [98]. Reduced function and quality of life are also a recognized hallmark of SOA in previous diagnostic criteria [46, 249, 250]. In our cohort, nine specific questions from the KOOS were associated with SOA compared to other knee disorders [159]. However, no individual question entered the final diagnostic clusters, including pain in stairs which was previously found to be very sensitive but not specific [245]. This may indicate that individuals with SOA present with various reduced functional profiles. Lastly, recent evidence suggested subjective crepitus as a potential criterion related to SOA, but in this cohort, subjective crepitus yielded both higher false-negative and false-positive rates than palpable crepitus performed by a clinician and suggest that this physical test may have better discriminatory capacity than the subjective criteria. [246, 251].

Strengths and limitations

Our prospective multicenter knee cohort recruited a broad variety of cases and non-cases commonly referred for conservative management [155]. However, the majority of SOA cases were from secondary care and this may limit the generalizability of our findings in primary care when patients exhibit initial symptoms and disability. The data collection procedure

ensured blinding of the evaluators between the index tests and the reference standard. The physiotherapist's main goal was to perform the same standardized physical examination on all participants. Our composite reference standard by the expert physicians included history elements, physical examination tests as well as confirmatory radiographs [168-170]. Even tough high diagnostic concordance was previously found ($\kappa \geq 0.80$) [232], only one physiotherapist and one medical expert evaluated each given participant and this may limit the generalizability of our results.

At this stage of development, the interpretation process may be somewhat difficult as the results present different clusters to rule in or out SOA and some variables include different categories depending on the clusters (i.e.: age and BMI). Refinement of the clinical tool will occur following external validation.

CONCLUSIONS

Age, BMI, valgus or varus knee misalignment, palpable knee crepitus and limited passive knee extension formed accurate clusters to diagnose SOA compared to other knee disorders. Following external validation with another cohort recruited from multiple primary care clinics, these clusters could be used by physiotherapists and primary care clinicians to efficiently identify patients with SOA without systematically relying on imaging.

Article 7: Validity of combining history elements and physical examination tests to diagnose patellofemoral pain.

Décary S, Frémont P, Pelletier B, Fallaha M, Belzile S, Martel-Pelletier J, Pelletier J-P, Feldman D, Sylvestre M-P, Vendittoli P-A, Desmeules F.

Simon Décary, PT, PhD candidate[1,2]
Frémont, Pierre, MD, PhD[3]
Pelletier, Bruno, MD[4]
Michel Fallaha, MD[4]
Belzile, Sylvain, MD[5]
Martel-Pelletier, Johanne, PhD[6]
Pelletier, Jean-Pierre, MD, PhD[6]
Feldman, Debbie, PT, PhD[1]
Sylvestre, Marie-Pierre, PhD[7]
Pascal-André Vendittoli, MD, MSc, FRSC[2,4]
François Desmeules, PT, PhD[1,2]

1- School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada.
2- Orthopaedic Clinical Research Unit, Maisonneuve-Rosemont Hospital Research Center, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
3- Department of Rehabilitation, Faculty of Medicine, Laval University, Quebec City, Quebec, Canada.
4- Department of Surgery, Maisonneuve-Rosemont Hospital, University of Montreal, Montreal, Quebec, Canada, Centre intégré universitaire de santé et de services sociaux de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada.
5- Department of Surgery, Laval University Hospital Center (CHUL), Laval University, Quebec, Quebec, Canada.
6- Osteoarthritis Research Unit, University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada.
7- University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada. Department of Social Preventive Medicine, School of Public Health, Université de Montréal

Article publié dans la revue *Archives of Physical Medicine and Rehabilitation*. 2017. doi: 10.1016/j.apmr.2017.10.014

En tant que premier auteur, j'ai réalisé le développement du protocole de cette étude, la collecte des données, l'analyse des résultats ainsi que la rédaction du manuscrit. FD et PAV ont contribué au développement du protocole, à l'analyse des résultats, à la rédaction et à la révision du manuscrit. MF, BP, SB et PF ont contribué au développement de l'évaluation musculosquelettique standardisée et ont agi à titre de médecins experts (standard de référence) pour tous les patients. JMP, JPP, DF et MPS ont contribué au développement du protocole, à la vérification des méthodes d'analyse des données et à la présentation des résultats. Tous les auteurs ont contribué à la révision de la version finale du manuscrit.

ABSTRACT

Objective: to assess the validity of diagnostic clusters combining history elements and physical examination tests to diagnose or exclude patellofemoral pain (PFP).

Design: prospective diagnostic study.

Settings: orthopaedic outpatients clinics (n=2), family medicine clinics (n=2) and community-dwelling.

Participants: consecutive patients (n=279) consulting one of the participating orthopaedic surgeons (n=3) or sport medicine physicians (n=2) for any knee complaint.

Interventions: not applicable.

Main Outcome Measures: history elements and physical examination tests were obtained by a trained physiotherapist blinded to the reference standard: a composite diagnosis including both physical tests and imaging results interpretation performed by an expert physician. Penalized logistic regression (LASSO) was used to identify history elements and physical examination tests associated with the diagnosis of PFP and recursive partitioning was used to develop diagnostic clusters. Diagnostic accuracy measures including sensitivity (Se), specificity (Sp), predictive values and positive and negative likelihood ratios (LR+/-) with associated 95% confidence intervals (CI) were calculated.

Results: two hundred seventy-nine participants were evaluated and 75 had a diagnosis of PFP (26.9%). Different combinations of history elements and physical examination tests including the age of participants, knee pain location, difficulty descending stairs, patellar facets palpation, and passive knee extension range of motion were associated with a diagnosis of PFP and used in clusters to accurately discriminate between PFP and non-PFP individuals. Two diagnostic clusters to confirm the presence of PFP yielded a LR+ of 8.7 (95% CI: 5.2-14.6) and three clusters to exclude PFP yielded a LR-: of 0.12 (95% CI: 0.06-0.27).

Conclusion: Diagnostic clusters combining common history elements and physical examination tests that can accurately diagnose or exclude PFP compared to various knee disorders were developed. External validation is required before clinical use.

INTRODUCTION

Patellofemoral pain (PFP) is a common condition accounting for 25-40% of all knee disorders [47]. PFP which includes conditions previously referred as chondromalacia, runner's knee or patellofemoral pain syndrome is defined as pain around or behind the patella aggravated by activities that increase loading and compressive forces of the patellofemoral joint such as squatting, ascending and descending stairs, jumping or running [47, 49]. Recent evidence suggests that PFP is not a simple self-limiting condition with more than 50% of individuals developing chronic pain [47, 51-55]. A valid initial diagnosis of PFP is therefore vital for early appropriate management and prevent persistence of symptoms [7, 8].

The diagnosis of PFP is mainly based on patients' history elements and physical examination tests, as there are no specific imaging findings either on radiographs or on magnetic resonance imaging to confirm PFP [252-254]. A clinical examination is considered by experts the cornerstone to make a valid diagnosis, yet the evidence on the diagnostic validity of different physical examination tests for PFP remains limited [153]. Most published diagnostic studies have low to moderate methodological quality [74, 89, 153] and suffer from biases likely resulting in an overestimation of the diagnostic validity of the studied tests [98]. The evidence shows that clinical tests may not be able to accurately diagnose PFP when used individually [74, 89, 153].

Since no single test alone may accurately diagnose PFP, a combination of tests has been proposed, which better reflect the diagnostic process of clinicians. Two diagnostic studies evaluated the combination of selected tests for PFP and presented somewhat greater diagnostic accuracy, but without reaching sufficient post-test probability to diagnose PFP. The study by *Cook et al*. based on a cohort of 76 participants of which 52 were considered to have PFP, concluded that combining a painful patellar facet palpation, pain during squatting or pain during resisted knee extension led to a LR+ of 4.0 (95%CI: 1.8-10.3) [100]. The study of *Sweitzer et al.* based on a cohort of 82 participants of which 59 were considered to have PFP, concluded that combining four patellar mobility tests led to a LR+ of only 1.9 (95%CI: 0.5-7.7)

[137]. To our knowledge, there is no study that specifically combined multiple patients' history elements with physical examination tests to formally evaluate the diagnostic validity of this approach and in effort to better guide clinicians in the differential diagnosis of PFP [153]. Therefore, using predictive clustering statistical methods, our objective was to assess the validity of diagnostic clusters combining history elements and physical examination tests to diagnose or to exclude PFP in a cohort of participants presenting with various knee disorders.

METHODS

Study design and settings

This was a prospective multi-center diagnostic study aimed at developing a series of diagnostic clusters for various common knee disorders. The present paper reports result specific to PFP. We recruited consecutive new patients consulting one of the participating physicians for a current knee complaint. Recruitment took place in two outpatients orthopaedic clinics and two primary care family medicine clinics between November 2014 and August 2016. Also, university community participants were invited to participate, via an email sent in September 2015 if they needed care for a current knee complaint. The present study conforms to the Standards for Reporting Diagnostic Accuracy Studies 2015 (STARD) [157, 158]. The study was approved by the ethics committees of all recruiting institutions and participants signed an informed consent form.

Participants

Potential participants were initially screened by telephone to assess preliminary eligibility. Patients aged 18 years of age or older who were consulting or referred to one of the participating clinical settings for a knee complaint and who were able to understand and speak French were included. Patients previously treated by the participating physicians were excluded, as well as patients who had undergone lower limb surgery in the past six months, patients with a knee arthroplasty or who presented with more than two other lower limb

pathologies or if they suffered from any systemic inflammatory disorder related to their knee complaint.

Data collection

Patients' characteristics and history elements

Selected history elements collected included: gender, age, education level, employment status, comorbidities, affected side, duration of knee symptoms, knee pain location (anterior, posterior, medial, lateral or diffuse knee pain), traumatic or atraumatic onset and use of a walking aid. Patients also completed the Knee Injury and Osteoarthritis Outcome Score (KOOS), a 42-item questionnaire composed of 5 domains: pain, symptoms, function in daily living, function in sport and recreation and knee-related quality of life [159]. Because mental health may influence the response to clinical examination, psychological distress was assessed using the K6 screening scale [163].

Physical examination data collection procedure

Before the start of the study, clinicians met with the research personnel to standardize techniques and interpretation of the physical tests as well as the definition of a PFP diagnosis compared to other knee diagnoses. Each participant was independently assessed by two evaluators on the day of their visit: a physiotherapist and one of the participating physicians. The physiotherapist possessed a master's degree in physiotherapy and had one year of clinical experience. The five participating physicians (three orthopaedic surgeons and two sports medicine physicians) each had more than 20 years of experience. The physiotherapist always completed data collection prior to the physician's examination of the patient. Both the physiotherapist and the physicians were blinded to each other's results and any other clinical information from the start of their respective evaluation.

Physical examination tests

A complete standardized physical examination was independently performed by the two evaluators. We selected tests based on published evidence [74, 89, 153], textbooks [77, 78,

113] and based on the expert physicians' clinical experience. Physical examination tests related to PFP included: tenderness at palpation of the patellar apex or the patellar tendon, the medial or the lateral patellar facets, palpable knee crepitus, resisted painful extension at 30° and 90° in knee flexion, pain at patellar mobilisation or compression, Clark's sign, patellar apprehension test, J-sign patellar maltracking during knee extension. These tests were reported as positive/negative. We also assessed active and passive knee range of motion in extension and flexion (restricted/not restricted), visual assessment of lower limb alignment (normal/abnormal/valgus/varus/recurvatum), patellar position (normal/rotated/tilted), medio-lateral patellar gliding (normal/decreased/increased), visual assessment of motor control and pain during squatting or descending a step (able to maintain alignment during task/painful task/unable to maintain alignment/no pain). All tests could also be coded as uncertain.

Reference standard definition

Following his own collection of history elements and physical examination tests, the physician was presented with reviewed imaging results and radiology reports and performed his own analysis of the relevant imaging diagnoses. All participants were required to have a radiograph of their knee that included the following three views: anterior-posterior with weight-bearing, lateral and skyline views. Magnetic resonance imaging (MRI) was required when the physicians suspected any ligament injury, a meniscal injury or needed to exclude another pathology. The physician made a final primary and secondary (if necessary) composite diagnosis. This final composite diagnosis (history elements, physical examination tests and imaging results), was considered the reference standard against which the index tests (physiotherapists' results) were compared for all participants. Diagnoses were classified into PFP or non-PFP based on the physicians' composite diagnosis. Imaging results were used to exclude other knee disorders including patellofemoral osteoarthritis (PFOA, Kellgren-Lawrence grade ≥ 1) on knee radiographs which was considered a separate diagnosis from simple PFP [7, 100, 252, 255].

Statistical analysis

Descriptive statistics were used to present the participants' characteristics and Student t-tests as well as chi-squared tests were used to compare PFP participants to those with other knee diagnoses. Inter-rater agreement between the physiotherapist and the physicians was calculated for all physical examination tests and diagnoses using *Prevalence and Bias Adjusted Kappas* (PABAK) with associated 95% CI [171].

Diagnostic clusters combining history elements including single questions from the KOOS and physical examination tests results were developed using a two-step method [177]. First, Least Absolute Shrinkage and Selection Operator (LASSO) penalized logistic regression was used to identify history elements and physical examination tests predictive of the diagnosis of PFP [177, 178]. LASSO is used to select variables with a higher predictive ability in the situation where there is a large initial set of variables, including ones of different distributions [177, 179]. The degree of shrinkage was determined by a penalty parameter, the value of which was identified through cross-validation to select the set of variables that maximize area under the curve (AUC) [179]. In the second step, we used recursive partitioning on the clinical variables selected using the LASSO to form diagnostic clusters to include and exclude a diagnosis of PFP [180]. Recursive partitioning allows the best hierarchical sequence of variables to classify PFP from non-PFP individuals [177]. The gini index was used as the splitting criteria [177, 180].

Overall model classification was compared to the reference standard (physicians' composite diagnosis) and sensitivity (Se), specificity (Sp), positive and negative predictive value (PPV/NPV) and positive and negative likelihood ratios (LR+/-) with 95% CI were calculated [60, 184]. Se and Sp express the percentage of true positive cases or true negative non-cases identified by a test or combination [185]. PPV and NPV relate to the probability of having or not the disorder if a test or combination is positive or negative, and is dependent on the initial prevalence of a disorder in a sample [185]. LR+/- represent the odds of having or not a disorder if a test combination is positive or negative and is useful for interpretation across populations with different prevalence [185]. Final selection of clusters to include and exclude

PFP were based on overall diagnostic validity and needed to reach a LR+≥5 or a LR-<0.2. These thresholds are recognized to produce a moderate shift in post-test probability and therefore clinically useful to make or exclude a diagnosis [70, 82]. If more than one cluster yielded such values, ease of use and clinical applicability was also used to select final clusters [60]. Internal validity was assessed using bootstrapping for recursive partitioning and estimates with 95% CIs were compared to the proposed clusters [177, 186]. Analyses were performed using R version 3.3.0 (packages: rpart, glmnet and randomForest; http://cran.r-project.org/).

RESULTS

Table 38 presents the characteristics of participants. A total of 279 individuals participated (96.2% of those approached). They were diagnosed with 359 primary and secondary diagnoses as follows: knee osteoarthritis (n=129), meniscal tears (n=80), anterior cruciate ligament (ACL) tears (n=43) or other knee diagnoses (n=32). Seventy-five individuals received a diagnosis of PFP. Individuals with PFP were significantly younger (PFP: 38.3±13.5 years old, other disorders: 53.1±14.8 years old; $p<0.05$) and had a lower body mass index (PFP: 26.8±5.9 kg/m^2, other disorders: 30.4±6.5 kg/m^2; $p<0.05$). Sixty-seven percent of patients with PFP came from orthopaedic clinics, while 33% came from primary care. The majority of participants with PFP had pain for ≥ 3 months at the time of consultation (92.0%). All KOOS domains indicated that the PFP individuals had significantly less disabilities compared to other knee disorders ($p<0.05$).

Table 39 presents the clinical variables associated with the presence or absence of PFP identified through penalized logistic regression. One hundred and thirty-one different clinical variables were entered in the penalized logistic regression. Following cross-validation, 21 variables were associated with the presence or the absence of PFP and yielded a maximal area under the curve (AUC) of 0.88 (95%CI: 0.86-0.90). Fourteen variables were history elements and seven physical examination tests (Table 39). History elements included: age, anterior knee pain location, feeling depressed or nervous and ten variables were items from

the KOOS questionnaire. Physical examination tests included: patellar apex tenderness, tenderness of medial or lateral patellar facets, patellar J-sign maltracking during seated knee extension, increased medio-lateral patellar glide, medial joint line tenderness and passive knee extension restriction.

Table 38: Characteristics of participants (n=279)

Characteristics	PFP (n=75)		Non-PFP (n=204)	
	n (%)	mean (SD)	n (%)	Mean (SD)
Age		38.3 (13.5)*		53.1 (14.8)
Sex				
Female	46 (61)		115 (56)	
Male	29 (39)		89 (44)	
Body Mass Index (Kg\m^2)		26.8 (5.9)*		30.4 (6.5)
Recruitment site				
Orthopaedic clinics	50 (67)		179 (88)	
Family medicine unit/university community	25 (33)*		25 (12)	
History of trauma	19 (25)		68 (33)	
Bilateral knee pain	17 (23)		36 (18)	
Duration of pain at time of consultation				
<3 months	6 (8)		28 (14)	
3-12 months	15 (20)		57 (28)	
≥ 12 months	54 (72)		119 (58)	
Triaged for surgery at time of consultation	2 (3) *		46 (22)	
Knee Injury and Osteoarthritis Outcome Score (KOOS)				
Pain		64.3 (19.1)*		56.5 (19.8)
Symptoms		76.4 (17.3)*		67.8 (19.3)
Activity of Daily Living		73.2 (21.1)*		63.2 (21.9)
Sports		39.5 (24.7)*		24.6 (24.3)
Quality of Life		43.8 (18.9)*		36.6 (19.8)
K6 psychological distress scale (/30)		25.6 (5.1)		26.6 (4.4)

SD=standard deviation; KOOS: a score of 0 indicates a severe condition and 100 indicates a normal knee; K6: a score of 6 indicates serious mental illness and 30 indicates no mental illness; * indicates a significant difference ($p<0.05$) between PFP participants and non-PFP knee disorders.

Table 39: Clinical variables associated with the presence or absence of PFP at maximal AUC identified through penalized logistic regression (n=279)

	Variables associated with the presence of PFP	Variables associated with the absence of PFP
History elements	Anterior knee pain Knee catching or hanging up when moving Being able to fully bend knee Difficulty squatting when being active Difficulty twisting or pivoting on knee when being active Being troubled with lack of confidence in knee Feeling depressed in last 30 days Feeling nervous in last 30 days	Older age Absence of pain going up or down stairs Pain at night while in bed Extreme difficulty descending stairs Extreme difficulty sitting Extreme difficulty jumping
Physical examination tests	Patellar apex tenderness Medial facet tenderness Any patellar facets (medial or lateral) tenderness Patellar J-sign maltracking during seated knee extension Increased medio-lateral patellar glide	Medial joint line tenderness Restricted passive knee extension

LASSO: Least Absolute Shrinkage and Selection Operator; Maximal area under the curve (AUC) was used as the criteria for the final penalty parameter to select variables associated with the presence or absence of PFP. KOOS questions are assessed on a five point likert scale: none, mild, moderate, severe or extreme.

Diagnostic clusters combining history elements and physical examination tests were identified through recursive partitioning using the 21 clinical variables previously identified. Table 40 presents two high specificity (Sp) diagnostic clusters to include PFP compared to other knee disorders. The clusters correctly classify 48/75 individuals with PFP (see 2x2 table in Appendix 19). One cluster contains individuals under 40 years old who report isolated anterior knee pain or have tenderness of the medial patellar facet. A second cluster contains individuals aged between 40 and 58 years old presenting with all the following: anterior or diffuse knee pain, mild to moderate difficulty descending stairs, medial patellar facet tenderness and full passive knee extension. The clusters present the following diagnostic statistics: Se of 0.64 (95%CI: 0.52-0.75), Sp of 0.93 (95%CI: 0.88-0.96), PPV of 0.76 (95%CI: 0.64-0.86) and LR+ of 8.70 (95%CI: 5.20-14.58). Figure 8 presents graphical representation of the diagnostic clusters to rule-in PPF.

Table 41 presents three high sensitivity (Se) diagnostic clusters to exclude PFP. The clusters correctly classify 132/204 individuals without PFP (see 2x2 table in Appendix 19). One cluster includes individuals aged under 58 years old with isolated medial, lateral or posterior knee pain and without medial or lateral patellar facets tenderness. Another cluster includes individuals aged under 58 presenting with diffuse or lateral knee pain knee pain, medial or lateral patellar facets tenderness and restricted passive knee extension. The last cluster excludes PFP if individuals are aged $\geq$ 58 years old. The clusters yielded the following diagnostic statistics: Se of 0.92 (95%CI: 0.83-0.97), Sp of 0.65 (95%CI: 0.58-0.71), NPV of 0.96 (95%CI: 0.91-0.98) and LR- of 0.12 (95%CI: 0.06-0.27). Figure 9 presents graphical representation of the diagnostic clusters to rule-out PFP.

Internal validation showed comparable estimates and 95% CIs between the initial clusters estimate and the average estimate obtained with bootstrapping (n=1000). Inter-rater agreement between the physiotherapist and physicians for the included tests was good and their reliability would be adequate for clinical use. Medial and lateral patellar facets tenderness reached a PABAK of 0.63 (95%CI: 0.51-0.72, n=224) and 0.62 (95%CI: 0.50-0.72,

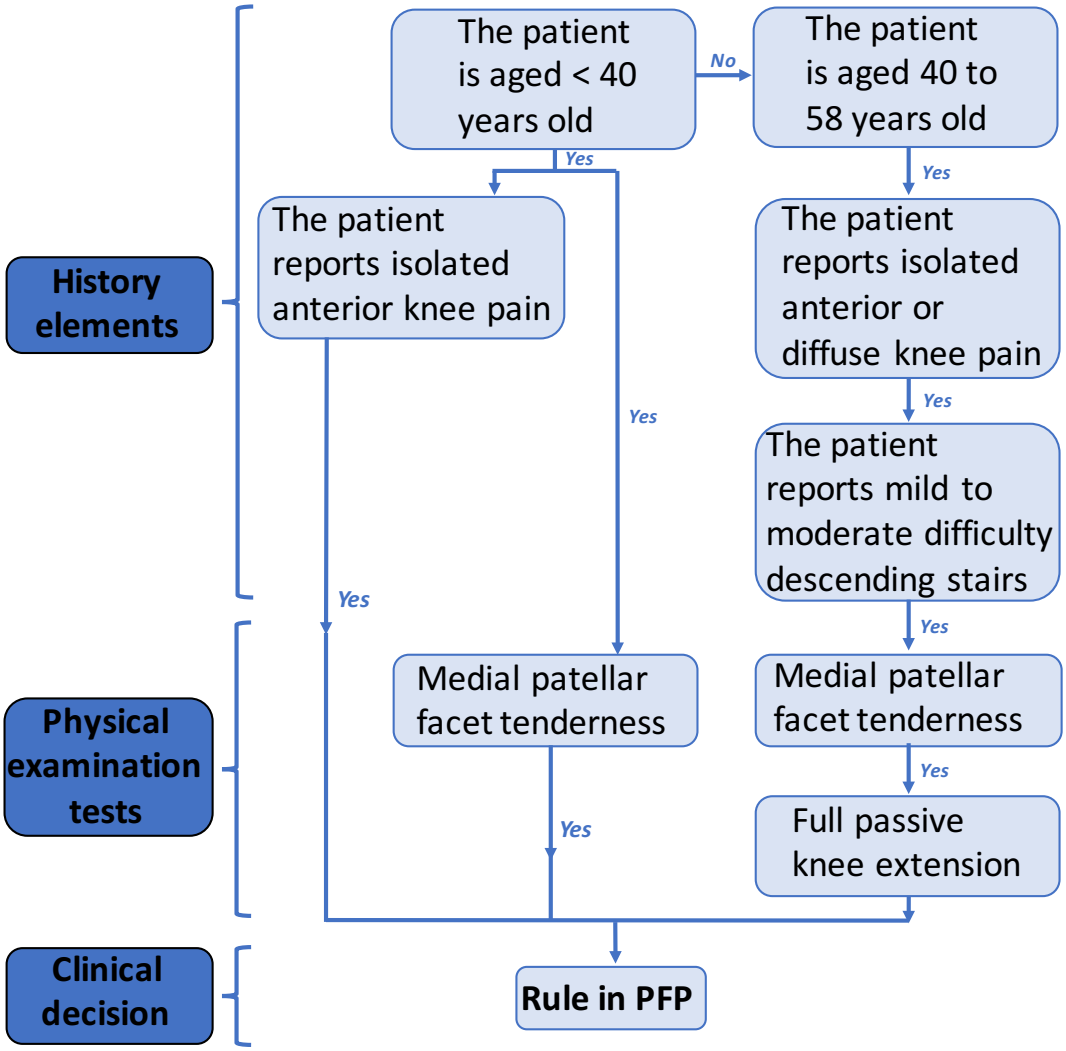
n=220) and passive knee extension range of motion reached PABAK of 0.86 (95%CI: 0.76-0.92, n=184)).

Table 40: High specificity diagnostic clusters combining history elements and physical examination tests to diagnose PFP

		Se (95% CI)	Sp (95% CI)	PPV (95% CI)	LR+ (95% CI)
Knee complaints are <u>likely due</u> to PFP in individuals with:					
Cluster 1	• Age <40 years old *AND* • Isolated anterior knee pain *OR* • Medial patellar facet tenderness	0.64 (0.52-0.75)	0.93 (0.88-0.96)	0.76 (0.64-0.86)	8.70 (5.20-14.58)
Cluster 2	• Age 40 to 58 years old *AND* • Isolated anterior or diffuse knee pain *AND* • Mild to moderate difficulty descending stairs *AND* • Medial patellar facet tenderness *AND* • Full passive knee extension				
Internal validation		0.56 (0.44-0.67)	0.96 (0.92-0.98)	0.84 (0.71-0.93)	14.28 (7.03-28.99)

Se: sensitivity; Sp: specificity; PPV: positive predictive value; LR+: positive likelihood ratio; Clusters are obtained using recursive partitioning with all variables associated with the diagnosis of PFP to include PFP with high Sp. Possible pain locations included: anterior, posterior, medial, lateral or diffuse knee pain. Difficulty descending stairs was evaluated in the KOOS questionnaire using a five-points likert scale: none, mild, moderate, severe or extreme. Passive knee extension range of motion is assessed by visually comparing the affected side to the healthy side. Internal validation was assessed by bootstrapping (n=1000).

Figure 8: Diagnostic clusters to rule in patellofemoral pain.



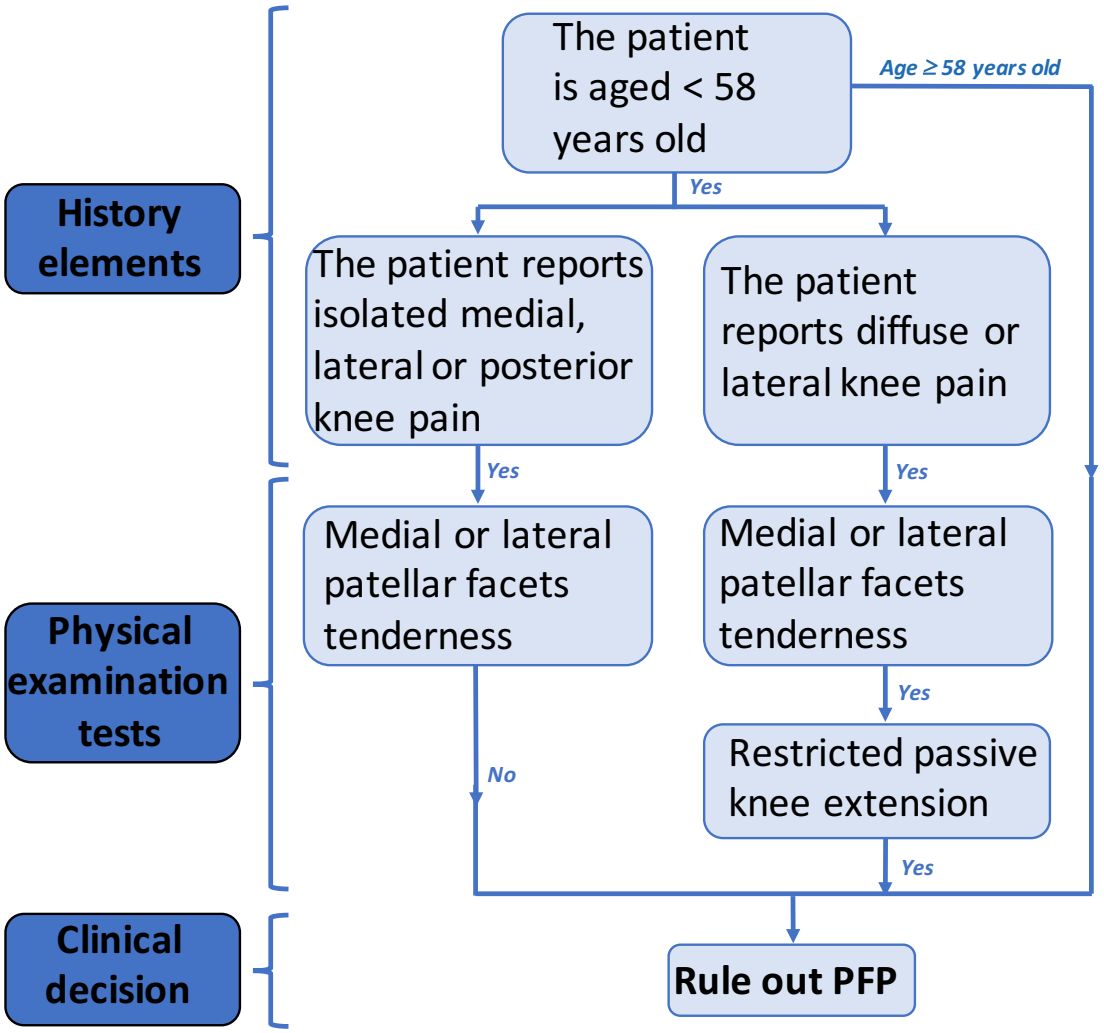
PFP: patellofemoral pain. Clinical decision to rule in PFP is based on clusters of history elements and physical examination tests reaching LR+≥5 indicating a moderate change in positive post-test probability.

Table 41: High sensitivity diagnostic clusters combining history elements and physical examination tests to exclude PFP

		Se (95% CI)	Sp (95% CI)	NPV (95% CI)	LR- (95% CI)
Knee complaints are <u>not likely due</u> to PFP in individuals with:					
Cluster 1	• Age <58 years old *AND* • Medial, lateral or posterior knee pain *AND* • No medial or lateral patellar facets tenderness	0.92 (0.83-0.97)	0.65 (0.58-0.71)	0.96 (0.91-0.98)	0.12 (0.06-0.27)
Cluster 2	• Age <58 years old *AND* • Diffuse or lateral knee pain *AND* • Medial or lateral patellar facets tenderness *AND* • Restricted passive knee extension				
Cluster 3	• Age ≥ 58 years old				
Internal validation		0.97 (0.91-1.00)	0.56 (0.49-0.63)	0.98 (0.94-1.00)	0.05 (0.01-0.18)

Se: sensitivity Sp: specificity; NPV: negative predictive value; LR-: negative likelihood ratio; Clusters are obtained using recursive partitioning with all variables associated with the diagnosis of PFP to exclude PFP with high Se. Possible pain locations included: anterior, posterior, medial, lateral or diffuse knee pain. Passive knee extension range of motion is assessed by visually comparing the affected side to the healthy side. Internal validation was assessed by bootstrapping (n=1000).

Figure 9: Diagnostic clusters to rule out patellofemoral pain.



PFP: patellofemoral pain. Clinical decision to rule out PFP is based on clusters of history elements and physical examination tests reaching LR+≤0.2 indicating a moderate change in negative post-test probability.

DISCUSSION

The objective of this study was to assess the validity of diagnostic clusters combining history elements and physical examination tests to diagnose or exclude PFP in a cohort of 279 primary and secondary care participants with various knee disorders including 75 participants with a diagnosis of PFP. From this cohort of participants, we have developed simple and valid clusters to help diagnose or exclude the presence of PFP.

History elements and physical examination tests associated with the diagnosis of PFP

The set of variables identified using penalized logistic regression is similar to the proposed core criterion for PFP in the *2016 International Patellofemoral Pain Consensus Statement* including: anterior knee pain around or behind the patella, difficulty in knee-flexed activities loading the patellofemoral compartment and tenderness at palpation of the patella [8, 256]. However, our analyses did not find an association with several others reported variables associated with the diagnosis of PFP such as: difficulty with running, pain at mobilisation or compression of the patella, pain during resisted knee extension, patellar tilting, Clark's sign, patellar crepitus, visual assessment of knee control or pain during squatting or step descending [8, 74, 89, 137, 153, 246]. A possible explanation is that our results are derived from a cohort of participants with other knee disorders instead of being based on a cohort of patients with PFP compared to healthy individuals [98]. In this context, some variables may be strongly associated with PFP but also to other knee disorders, limiting their discriminative capacities. Another possible explanation may be that we recruited a sample with a broader age range than previous PFP diagnostic study cohorts [51, 100, 153, 256] that also presented more chronic symptoms (72% had pain for 12 months or more) [51, 153, 256].

Diagnostic clusters combining history elements and physical examination tests for inclusion or exclusion of PFP

Our diagnostic clusters provide new evidence on the validity of combining history elements and physical examination tests to diagnose PFP considering selected age groups, knee pain location and patellar facets tenderness [8]. Interestingly, we also found that knee range of

motion may help guide a PFP diagnosis as part of a diagnostic cluster in older individuals in the present study [8]. The inclusion clusters increased positive post-test probability by 49.1% from a baseline prevalence of 26.9% in this cohort. The exclusion clusters increased negative post-test probability by 22.9% from a negative baseline prevalence of 73.1%. The magnitude of the likelihood ratio observed using the developed clusters surpassed our pre-determined cut-off value of LR+=5 and LR-=0.2, indicating a clinically useful moderate change in post-test probability [82]. These clusters correctly classified 64% and 65% of all cases and non-cases. Therefore, misclassification remains with the developed clusters. While this can be expected from these types of analyses [96, 100, 149], it also potentially enlightens the heterogeneous presentation of PFP. The fact that various etiologic factors, not only related to the knee joint contribute to PFP also supports that notion [257].

Our results are coherent with the other diagnostic studies that combined physical examination tests for PFP [100, 137]. In a sample of 76 patients with complaints of anterior knee pain, including 52 with a confirmed diagnosis of PFP, *Cook et al*. showed that the combination of pain during resisted knee extension and/or pain during squatting and/or painful facet palpation of the patella were indicative of PFP [100]. Compared to their studies, our cohort included various knee disorders, which might explain the differences observed in the included tests and in the magnitude of the positive LR [100]. This also allowed us to demonstrate that complaint of anterior knee pain is an important history element to discriminate PFP from other knee disorders. The study of *Sweitzer et al.* found that combining multiple patellar mobility tests was not clinically useful to diagnose PFP [137]. In our study, we evaluated only medio-lateral gliding, which was associated with the diagnosis of PFP, but was not included in any of our diagnostic cluster, confirming its limited use even in cluster of tests.

Authors have proposed that the diagnosis of PFP is predominantly an exclusion diagnosis made by ruling out other disorders when imaging results are negative [74, 100], yet our study identified clinical factors from history and physical examination that matched experts' clinical

diagnosis of PFP. Although imaging will probably still be warranted in selected patients to rule out other disorders, the proposed clusters, upon further validation, could be useful clinical diagnostic tools to aid clinicians in their diagnostic process while reducing use of imaging.

Strengths and limitations

We recruited from different settings in primary and secondary care, allowing for a broad variety of cases and non-cases [74, 155, 258]. Our data collection procedure ensured blinding of the evaluators between the index tests and the reference standard. The composite reference standard by the expert physicians included history elements, physical examination tests as well as use of imaging [74, 168-170]. Our stepwise statistical approach was able to identify multiple variables from both history elements and physical examination tests that led to the development of simple diagnostic clusters [177]. A limitation of our study is that only one physiotherapist and one medical expert evaluated each given participant. We also included the most common physical tests used by clinicians to evaluate the knee joint but not all tests related to PFP were evaluated in our study and more complex tests or functional lower limb assessments were not included.

CONCLUSIONS

We identified diagnostic clusters combining common history elements and physical examination tests that can accurately diagnose or exclude PFP based on a clinical diagnosis. Refining diagnostic criteria for PFP is important to support decision-making concerning appropriate treatment options, use of imaging or referral to a specialist. External validation is required before widespread clinical use.

Chapitre 5: Discussion générale

Cette thèse a exploré la thématique du diagnostic clinique des pathologies communes au genou [1-4]. Les données probantes démontrent que le diagnostic initial des pathologies au genou est souvent erroné ou incomplet, ce qui conduit les patients sur des trajectoires de soins non optimales [6, 8, 9, 18, 24-26, 216]. Les objectifs de cette thèse étaient de : 1- évaluer l'accord diagnostique entre un physiothérapeute utilisant une évaluation musculosquelettique standardisée sans imagerie et des médecins experts pour les différentes pathologies communes au genou; 2- évaluer la validité de la combinaison de l'anamnèse et de l'examen physique afin de développer une série d'outils valides permettant d'orienter le diagnostic différentiel des pathologies communes au genou.

5.1 Caractéristiques de la cohorte et représentativité

Jusqu'à 45% des consultations en première ligne au Canada seraient reliées aux pathologies musculosquelettiques, incluant les pathologies au genou, ce qui engendre une pression sur le système de santé [3]. Dans la présente étude, nous avons recruté deux cent soixante-dix-neuf patients qui consultaient pour une douleur ou des symptômes au genou. La prévalence des diagnostics dans l'échantillon était de 15% pour les déchirures du LCA, de 29% pour les déchirures méniscales, de 46% pour l'OA du genou, de 27% pour le SFP et de 11% pour les autres pathologies. Ces prévalences semblent différentes d'une clientèle de première ligne [6], mais généralement plus représentatives d'une population qui consulte en clinique externe de chirurgie orthopédique [27].

Outre la comparaison des prévalences, la présente cohorte démontre aussi des distinctions quant au spectre des présentations cliniques des différentes pathologies et des caractéristiques des patients. Certains patients présentaient des déchirures du LCA partielles ou complètes, aiguës ou chroniques et souvent combinées avec d'autres diagnostics. Ceci représente un spectre plus hétérogène de présentations cliniques pour les déchirures du LCA que d'autres cohortes présentées dans d'autres études diagnostiques sur le sujet [34, 101, 150, 213]. Cette situation est aussi présente pour les individus atteints d'OA du genou qui

présentaient une variété de grades radiologiques d'OA selon l'échelle Kellgren-lawrence [96, 97, 234]. Les déchirures méniscales quant à elles touchaient principalement le ménisque interne, ce qui est usuel et comparable aux autres cohortes diagnostiques dans le domaine [6, 10, 11, 102, 151, 224, 225].

Par rapport aux caractéristiques des patients, nous avons observé une grande diversité quant à l'âge, aux symptômes et aux limitations fonctionnelles pour chacune des pathologies au genou. Sur ces aspects, la présente cohorte serait représentative d'un plus large spectre de patients que d'autres cohortes qui incluaient des échantillons plus spécifiques, ce qui augmente potentiellement la validité externe de notre étude. Par contre, les patients atteints d'OA du genou dans la présente cohorte présentaient de la douleur et des atteintes fonctionnelles légèrement plus importantes que les patients de cohortes populationnelles de première ligne atteints d'OA radiologique [96, 97, 234]. Aussi, les patients atteints d'une déchirure du LCA ou d'un SFP avaient des symptômes plus chroniques en comparaison avec les cohortes d'autres études [51, 100, 101, 153, 256]. Ceci peut être expliqué par la sur-représentativité des patients recrutés en clinique orthopédique (82%) qui démontrent des symptômes persistants depuis une plus longue durée qu'une population évaluée en première ligne lors de l'apparition initiale de symptômes. L'impact des caractéristiques des participants de notre étude sur les résultats de cette thèse sera discuté ultérieurement.

5.2 Synthèse des résultats et comparaison avec les données probantes de la littérature

Pour l'objectif 1, le physiothérapeute utilisant une évaluation musculosquelettique standardisée a démontré un excellent accord diagnostique avec les médecins experts pour les différentes pathologies au genou (Chapitre 4, Article 3). Ces résultats sont concordants avec les études antérieures et permettent aussi de suggérer l'utilisation d'une évaluation musculosquelettique pour orienter le diagnostic différentiel des pathologies au genou sans le recours systématique à l'imagerie, et ce, pour la plupart des patients [27, 152].

Pour l'objectif 2, les deux premiers diagnostics d'intérêt pour le développement des combinaisons incluaient les pathologies d'origine traumatique, soit les déchirures partielles ou complètes du LCA (Chapitre 4, Article 4) et les déchirures méniscales d'origine traumatique (Chapitre 4, Article 5). Les variables qui constituaient les différentes combinaisons d'éléments de l'anamnèse et de l'examen physique concordent, en grande partie, avec les études antérieures pour ces deux diagnostics et les modèles proposés ont atteint les seuils fixés pour être considérés modérément à hautement valides [101, 102, 150].

Les estimations de validité diagnostique obtenues pour ces deux pathologies sont plus élevées que les études précédentes publiées dans la littérature. Plusieurs facteurs peuvent potentiellement expliquer cette différence. D'abord, les études antérieures qui portaient sur les déchirures du LCA et méniscales ont étudié la validité diagnostique des combinaisons de tests uniquement sur des échantillons de patients qui présentaient une condition aigüe (moins de cinq semaines post-trauma) [101, 102, 150]. Ceci peut avoir rendu l'exécution et l'interprétation de certains tests de l'examen physique plus difficiles en raison de la présence d'épanchement articulaire ou de co-contractions musculaires de protection reliées à la douleur potentiellement plus importante [101, 213]. La plupart des participants de la présente cohorte avaient subi un traumatisme plus de trois mois avant l'évaluation, ce qui a donc pu faciliter l'exécution et l'interprétation adéquate des tests de l'examen physique et donc réduire les erreurs de classification [6, 10, 224, 225].

Un autre facteur qui a pu affecter les estimations obtenues pour les différentes combinaisons d'éléments de l'histoire et de l'examen physique est la représentativité des participants des études publiées. Les auteurs des études antérieures sur le LCA et les atteintes méniscales avaient aussi recruté des échantillons de patients plus restreints qui incluaient seulement des patients qui souffraient d'une atteinte au genou d'origine traumatique. Ces cohortes comparaient donc principalement des déchirures du LCA, des déchirures méniscales ainsi que des contusions ou des entorses sans déchirure [6, 101, 102]. Cette situation a pu diminuer les estimations dans ces études, car la plus grande similitude des présentations entre ces

diagnostics a pu amener un taux supérieur de faux positifs et de faux négatifs lors des analyses [6, 101, 102]. La présente étude a comparé des patients atteints d'une déchirure du LCA ou d'une déchirure méniscale traumatique à diverses pathologies au genou et incluait aussi des participants qui souffraient d'une atteinte d'apparition progressive. Ceci a permis de démontrer que l'origine traumatique d'une douleur, combinée à d'autres critères, est en soi un élément de l'anamnèse valide pour discriminer entre différentes pathologies du genou. Le fait que notre cohorte était généralement de condition plus chronique pour ces deux types d'atteintes et que tous n'avaient pas une histoire traumatique récente a pu augmenter les estimations diagnostiques en diminuant le taux de faux positifs et de faux négatifs. Par contre, la composition de notre cohorte a aussi pu limiter le biais de spectre souvent présent dans les études diagnostiques musculosquelettiques [6, 101, 102, 150, 207]. La cohorte recrutée se veut donc potentiellement plus représentative de la population cible dont les raisons de consultation incluent autant des symptômes aigus ou chroniques ainsi que d'origine traumatique ou d'apparition progressive.

Les trois autres diagnostics d'intérêt incluaient les pathologies d'apparition progressive soit les déchirures méniscales dégénératives (Chapitre 4, Article 5), l'OA du genou (Chapitre 4, Article 6) et le SFP (Chapitre 4, Article 7). Les variables constituant les différentes combinaisons d'éléments de l'anamnèse et de l'examen physique concordent aussi, en grande partie, avec les études antérieures pour ces trois diagnostics [11, 22, 100, 234]. De plus, les modèles proposés dans cette thèse ont aussi atteint les seuils fixés pour être considérés modérément à hautement valides [11, 100, 234].

Les combinaisons d'éléments de l'anamnèse et de l'examen physique pour les pathologies d'apparition progressive ont permis la classification de 58% à 71% de tous les cas et les non-cas. Ces résultats sont similaires ou supérieurs aux classifications obtenues dans les études précédentes [11, 96, 100, 149, 151, 222]. L'utilisation de la partition récursive peut avoir contribué à classifier plus précisément un nombre important de cas et de non-cas puisque cet algorithme statistique a été développé pour catégoriser des individus [96, 149]. Par contre,

nos résultats signifient aussi que l'obtention d'une seule combinaison pour identifier tous les cas et non-cas pour les pathologies d'apparition progressive est difficile [96, 257]. Ces pathologies étant donc plus à susceptible à des erreurs de classification, cela souligne aussi que l'utilisation du jugement clinique dans la détermination du diagnostic demeure nécessaire lorsque la présentation clinique semble atypique, et ce malgré l'accès à des outils d'aide à la décision diagnostique.

L'autre distinction observée dans nos résultats pour les pathologies d'apparition progressive, principalement pour l'OA du genou et pour le SFP, est que certaines variables proposées dans la littérature comme étant associées avec ces diagnostics n'ont pas été incluses dans nos combinaisons d'éléments [8, 46, 49, 96, 234]. Puisque tous les patients ont reçu une évaluation musculosquelettique standardisée, certaines variables de l'anamnèse et de l'examen physique pouvaient démontrer des réponses positives dans plusieurs diagnostics [98]. Il est donc important de bien distinguer la différence entre une variable clinique associée à un diagnostic, soit faisant partie du profil clinique attendu et la capacité discriminative d'une variable qui vise plutôt à classifier précisément un cas d'un non-cas chez une population d'adulte souffrant d'un problème au genou.

5.3 Forces de l'étude

Le cadre méthodologique utilisé pour construire le devis de l'étude est celui de QUADAS. Cet outil comporte 14 items qui évaluent les aspects de la qualité méthodologique d'une étude diagnostique afin de limiter les biais pouvant affecter les estimations de validité [98, 110, 155, 259, 260]. Nous traiterons les forces de l'étude selon quatre thèmes : le mode de recrutement, le standard de référence, les tests index et la méthode statistique.

5.3.1 Forces liées au mode de recrutement

Une force de cette étude réside dans le recrutement prospectif et, pour la majorité des participants, de façon consécutive dans des milieux cliniques de première et de deuxième ligne. Cette procédure a permis de maximiser les probabilités d'obtenir un échantillon varié et

représentatif de la population cible. Certains de ces aspects ont été discutés préalablement en début de discussion.

Certains auteurs considèrent le mode de recrutement comme le biais pouvant affecter le plus la validité diagnostique [98, 261]. Le devis de type cas-contrôle, lors duquel des patients avec une pathologie sont comparés à des patients sains, pourrait surestimer la validité diagnostique de plus de 50% puisque les sujets sains ont une faible probabilité de répondre positivement aux éléments de l'évaluation [98, 223, 261]. Un autre biais lié au mode de recrutement est la présélection d'un échantillon de population, ce qui augmente la prévalence d'un diagnostic d'intérêt et donc la probabilité d'identifier plus facilement ce diagnostic [262]. Le biais de spectre lié à la présélection d'un échantillon de population peut aussi affecter la représentativité des patients recrutés. Notamment, le spectre de sévérité des pathologies peut augmenter ou diminuer la validité diagnostique puisque par exemple un cas sévère pourrait répondre plus facilement à une manœuvre de l'évaluation musculosquelettique comparativement à un cas présentant des symptômes moindres [262].

Nous avons recruté des participants dans deux cliniques orthopédiques, deux unités de médecine familiale et parmi une communauté universitaire afin d'obtenir l'échantillon de population le plus hétérogène et obtenir un large spectre de présentations cliniques de patients souffrant d'une atteinte du genou et nécessitant des soins. Nous avons recruté les participants de façon prospective et consécutive pour 87% de l'échantillon. Les 13% de patients recrutés dans la communauté universitaire l'ont été sur la base du volontariat. Dans les deux cas, nous n'avons pas effectué une présélection d'un groupe distinct de patients sur la base des diagnostics ou des caractéristiques des patients [98, 261]. Nous avons recruté seulement des patients qui consultaient pour une douleur ou des symptômes au genou et les avons comparés entre eux sur la base de la même évaluation musculosquelettique lors des analyses statistiques. Ceci signifie que les cas ont aussi agi à titre de participants contrôles lors de la comparaison aux autres diagnostics. Ainsi, tous les participants avaient la caractéristique commune de consulter pour des symptômes au genou [98, 261].

5.3.2 Forces liées au standard de référence

La deuxième force de cette étude est le recours à un standard de référence du diagnostic émis par un des cinq médecins experts combinant leur évaluation musculosquelettique indépendante et leur interprétation des tests d'imagerie médicale appropriés.

Le standard de référence est un autre aspect important de la qualité méthodologique des études diagnostiques [98, 261]. Les études antérieures sur le diagnostic des pathologies musculosquelettiques ont utilisé trois types de standards de référence : l'évaluation musculosquelettique par un expert, les résultats de l'imagerie médicale ou encore lors de la chirurgie [169]. Chacun de ces standards de référence pris de façon individuelle est imparfait [169]. L'évaluation par l'expert peut être biaisée par le spectre de sa pratique clinique. L'imagerie médicale, notamment l'IRM, peut amener des faux positifs ou des faux négatifs en plus de proposer le diagnostic d'une pathologie à partir de lésions structurelles possiblement asymptomatiques. Le standard de référence utilisant la chirurgie inclura principalement des cas qui présentent une condition clinique sévère [6, 14, 169, 217, 231, 263].

Dans la présente étude, nous avons utilisé un standard de référence qui combinait l'évaluation musculosquelettique recueillie par le médecin expert ainsi que son interprétation du diagnostic radiologique approprié [168, 170, 198]. Cette procédure réplique la méthode clinique usuelle pour émettre un diagnostic musculosquelettique en combinant l'évaluation musculosquelettique et l'imagerie [168, 170, 198].

Certains biais spécifiques peuvent diminuer la validité d'un standard de référence. Le biais d'incorporation est présent lorsque les tests index font partie de la définition du standard de référence; le biais de différentiation est lui présent si un standard de référence différent est donné en fonction de la réponse aux tests index. Finalement, le biais de vérification partielle est possible si certains patients ne sont pas tous évalués par le même standard de référence [98, 155, 260, 261]. Dans la présente étude, tous les patients ont eu le même standard de référence, soit le diagnostic combiné par l'un des cinq médecins experts. Le standard a été

administré indépendamment et à l'aveugle de la réponse aux tests index exécutés par le physiothérapeute qui avait le rôle principal de collecter ces tests de façon standardisée.

5.3.3 Forces liées à la collecte des tests index

La troisième force de cette étude est que le physiothérapeute a exécuté systématiquement tous les tests index à l'aveugle du standard de référence, le diagnostic émis par le médecin expert incluant les résultats aux tests d'imagerie. Aussi, nous avons standardisé les réponses aux questions de l'anamnèse et aux tests de l'examen physique à l'aide de diverses sources incluant les données probantes des revues systématiques (Chapitre 2), des livres de référence sur le diagnostic musculosquelettique, et ce en accord avec les préférences des médecins experts associés au projet [77, 78, 113, 153].

5.3.4 Forces liées aux méthodes statistiques

La quatrième force de cette étude est l'utilisation séquentielle de deux méthodes statistiques prédictives. Cette procédure a permis l'identification de plusieurs combinaisons valides d'éléments de l'anamnèse et de l'examen physique pour classifier les cas des non-cas des différentes pathologies au genou.

Nous avons combiné deux méthodes statistiques afin d'adresser les problématiques inhérentes au développement des modèles prédictifs qui combinent un nombre important de variables indépendantes [60, 177]. L'utilisation de la régression logistique pénalisée de type LASSO a remplacé la régression logistique univariée et multivariée non pénalisée classiquement utilisée dans la littérature [60]. Le LASSO a permis d'identifier les meilleurs prédicteurs parmi un large éventail de variables indépendantes [177, 178]. La sélection automatisée des variables a aussi permis de traiter la problématique de la multi-colinéarité et de l'interaction en sélectionnant automatiquement la plus forte de deux variables colinéaires tout en limitant les pertes d'information pour les variables à plusieurs niveaux [177-179]. Le mode de sélection des variables lors du LASSO a aussi permis d'atténuer la problématique d'*over-fitting* par rapport à la régression logistique non pénalisée [179, 264]. L'utilisation de

cette méthode a donc constitué une approche rigoureuse de sélection des variables associées avec les diagnostics d'intérêts.

Les résultats obtenus suite au LASSO ne présentent pas la capacité prédictive des variables ni un format directement utilisable en clinique [174]. Pour ce faire, l'utilisation de la partition récursive a permis d'obtenir une série de combinaisons qui permettent la classification des cas et des non-cas pour les différents diagnostics d'intérêts [177, 180-182]. L'aspect séquentiel de la partition récursive réplique le processus de diagnostic différentiel et de prise de décision utilisé en clinique [175, 183]. Cette méthode statistique n'a été que rarement utilisée dans le champ de recherche du diagnostic musculosquelettique contrairement à d'autres champs de la médecine [177]. Or, plusieurs modèles produits à partir de ce type d'analyse statistique ont par la suite été validés et implantés en clinique [60, 93, 95, 96, 149, 174, 177, 265].

Finalement, la validation interne des modèles a permis de vérifier la présence d'*overfitting* afin d'atténuer le « phénomène d'optimisme » qui signifie que la performance d'un modèle est souvent meilleure lors de la dérivation initiale [60, 186, 266]. Dans la présente étude, nous avons utilisé une méthode de « boostrapping » validée spécifiquement pour la partition récursive [177, 267, 268].

5.4 Limites de l'étude

La présente étude comporte des limites imposées principalement pour des raisons de faisabilité. Ces limites n'invalident pas les résultats de l'étude, mais présentent des pistes de réflexion concernant la population étudiée et la généralisation des résultats, l'exactitude du standard de référence et des tests index ainsi que pour la méthode statistique des études diagnostiques.

5.4.1 Biais potentiels et limites liés à la population à l'étude et à la généralisation des résultats

Le spectre de la population à l'étude est un item méthodologique important lié à la validité des études diagnostiques [98, 155, 207]. Nous avons recruté l'échantillon de population de la présente étude dans divers milieux cliniques afin d'obtenir un large spectre de présentations cliniques des différentes pathologies. Malgré cette stratégie de recrutement, nous avons obtenu environ 80% de l'échantillon par les cliniques externes de chirurgie orthopédique. Ceci a pu avoir un impact sur la représentativité des résultats, tel que déjà discuté, mais aussi sur les estimations de validité.

Pour les pathologies du genou d'origine traumatique, la plus grande chronicité des cas peut avoir rendu l'interprétation des tests de l'examen physique moins contraignante à cause d'une diminution de l'épanchement articulaire du genou liée au traumatisme plus ou moins lointain, diminuant ainsi la probabilité de faux positifs ou de faux négatifs [101, 213]. Pour les participants qui présentaient des atteintes d'apparition progressive, la plus grande chronicité ainsi que la sévérité peuvent avoir facilité l'identification des vrais positifs [11, 234]. Ces phénomènes ont pu potentiellement biaiser les estimations de sensibilité et de spécificité de nos modèles [262]. Ainsi, les estimations de validité des différents modèles pourraient être moins élevées lorsqu'appliquées sur un échantillon de première ligne, lorsque les patients consultent pour la première fois pour leurs symptômes aigus ou traumatiques, lorsqu'ils consultent pour une pathologie d'apparition progressive démontrant des symptômes transitoires en début de progression ou lorsqu'ils se présentent avec une autre pathologie présentant une réponse semblable à l'examen clinique que les pathologies communes [6, 101, 102].

Un autre aspect en lien avec la stratégie de recrutement dans différents milieux cliniques concerne la prévalence des pathologies de l'échantillon [262]. La concentration de certains diagnostics dans les cliniques orthopédiques fait en sorte que la prévalence est supérieure dans le présent échantillon en comparaison aux prévalences proposées pour une population qui consulte en première ligne [6, 27, 262]. Un effet de l'augmentation de la prévalence est

que les estimations pour les valeurs prédictives positives et négatives, qui sont basées sur la probabilité pré-test d'un échantillon, ne sont pas directement applicables dans tous les contextes cliniques [173, 185]. Le calcul des rapports de vraisemblance a permis de contourner cette problématique d'interprétation [173, 185]. Un autre impact de la prévalence concerne la composition de l'échantillon de population. Ainsi, notre cohorte ne représente pas le spectre complet de toutes pathologies possibles au genou, qui inclut par exemple certaines tendinopathies [100], certains syndromes comme celui de la bandelette ilio-tibiale, certains types de contusions, ou des entorses simples sans déchirures [101, 103]. L'ajout de toutes ces pathologies dans une étude diagnostique en première ligne pourrait diminuer les estimations de validité diagnostique de l'évaluation musculosquelettique puisque ces pathologies peuvent démontrer des présentations cliniques similaires aux pathologies communes. Aussi, aucun cas de pathologies graves (par exemple des cas de cancer ou d'infection) n'a été diagnostiqué dans cette cohorte et nous avons exclu les patients atteints d'une atteinte rhumatologique connue.

5.4.2 Biais potentiels et limites liés au standard de référence

Le standard de référence utilisé était le diagnostic émis par un médecin expert qui combinait son évaluation musculosquelettique et son interprétation des résultats des tests d'imagerie médicale pour former un diagnostic final [155, 168, 170, 198]. Malgré qu'il ait été administré de façon à minimiser les biais, ce standard de référence contient trois limites potentielles.

La première limite concerne la partie de l'évaluation musculosquelettique du standard de référence. Les médecins experts devaient exécuter une évaluation musculosquelettique indépendante pour poser une hypothèse initiale sans les résultats de l'imagerie. Contrairement au physiothérapeute, ils avaient la possibilité de n'exécuter que les tests requis pour vérifier leur hypothèse suite à leur anamnèse aussi faite de façon indépendante. Ils n'ont donc pas exécuté systématiquement tous les mêmes tests que le physiothérapeute pour chaque patient. Chez certains patients, certains tests supplémentaires auraient pu leur permettre d'obtenir de l'information clinique pour former une nouvelle hypothèse

diagnostique basée sur leur anamnèse et potentiellement diminuer le nombre de patients pour lesquels le diagnostic entre le physiothérapeute et le médecin était non-concordant.

Un autre aspect limitatif relié à l'évaluation musculosquelettique par le médecin expert est la standardisation des tests avec ceux du physiothérapeute. Ceci pourrait représenter une forme de biais d'incorporation puisque bien que les tests index aient été exécutés indépendamment du standard de référence pour le développement des modèles, les tests exécutés par le médecin expert répondaient à la même définition que ceux du physiothérapeute [98].

La deuxième limite concerne l'évaluation de l'imagerie médicale par le médecin expert. Tous les participants devaient avoir reçu une radiographie du genou, et tous les patients pour lesquels le médecin expert suspectait une atteinte ligamentaire, méniscale ou pour exclure tout autre diagnostic suite à l'évaluation musculosquelettique devaient obtenir un test d'imagerie par résonnance magnétique. Ainsi, pour des raisons de faisabilité, les médecins experts ont demandé une IRM seulement pour les cas jugés pertinents sur la base de leur évaluation musculosquelettique [214]. Ceci peut représenter un biais de différentiation ou de vérification partielle puisque les tests d'imagerie appropriés sont déterminés en fonction de l'évaluation musculosquelettique de l'expert (mais non de la réponse aux tests index du physiothérapeute) et parce que tous les patients n'ont pas reçu un test d'IRM [98, 155, 260, 261]. Le fait que les pathologies au genou n'ont pas toutes les mêmes standards quant à l'imagerie médicale appropriée pour émettre le diagnostic atténue cette limite [11, 214, 233].

Un autre aspect limitatif lié à l'obtention des imageries est que nous n'avons pas été en mesure d'obtenir tous les tests du même centre ou de standardiser précisément les protocoles d'imagerie. Lorsqu'un patient se présentait à l'étude avec des tests d'imagerie préalablement obtenus, les IRM de moins de six mois et les radiographies de moins de trois mois étaient acceptées pour des raisons de faisabilité. Un nouveau test d'imagerie était demandé si le médecin expert doutait de la validité d'une imagerie, de la qualité de l'image

obtenue ou s'il considérait que la condition du genou du patient avait pu changer ou que sa condition nécessitait une vérification supplémentaire.

La troisième limite du standard de référence concerne l'émission du diagnostic final par le médecin expert. Celui-ci émettait son diagnostic final sur la base de son évaluation musculosquelettique et de son interprétation des tests d'imagerie. Comme chacune des étapes de la séquence du diagnostic présente un risque d'erreur, le diagnostic final émis a aussi pu être erroné. Ainsi, l'oubli d'un test significatif de l'examen physique ou une erreur d'interprétation pourrait avoir mal orienté l'obtention du test d'imagerie approprié pour un diagnostic d'intérêt. Globalement, les limites du standard de référence pourraient avoir empêché l'identification de certains cas. Pour résoudre cette problématique, deux médecins experts auraient pu évaluer chacun des patients et arriver à un diagnostic conjoint. La faisabilité de cette approche devra être explorée pour améliorer la définition du standard de référence dans des études futures [96]. Cependant, dans la présente étude, l'accord diagnostique entre le physiothérapeute et les médecins experts était tout de même excellent pour tous les diagnostics d'intérêts (κ=0,89; IC95%: 0,83-0,94) et le taux de désaccord brut à posteriori était faible (n=14/179, Chapitre 4 article 3). Ceci suggère que la très grande majorité des participants ont été diagnostiqués correctement.

5.4.3 Biais potentiels et limites liés aux tests index

Le physiothérapeute a exécuté les tests index à l'aveugle du standard de référence. Les techniques d'évaluation des tests index ainsi que la définition des réponses aux différents tests ont été préalablement standardisés selon l'usage reconnu en clinique et dans la littérature. Par contre, certaines limitations pourraient diminuer l'applicabilité clinique des tests index.

La première limite liée à l'exécution des tests index est que ceux-ci ont été exécutés par un seul évaluateur. À cette étape préliminaire du développement des modèles diagnostiques, ceci a assuré un contrôle sur l'exécution systématique des tests et le respect de la

standardisation. Par contre, l'expertise de cet évaluateur n'est pas nécessairement représentative de tous les cliniciens. De ce fait, l'interprétation de la validité et de la fidélité des tests de l'examen physique doit être faite avec prudence. Cependant, chacun des tests intégrés dans les modèles a démontré une fidélité inter-évaluateur au moins modérée ($\kappa \geq 0.4$) entre le physiothérapeute et les médecins experts.

L'exécution systématique des tests index avant l'évaluation par le médecin expert pourrait aussi avoir contribué à augmenter la fidélité inter-évaluateur [107]. Bien qu'aucun patient n'ait été exclu pour une augmentation de la douleur entre les évaluations, le physiothérapeute a pu sensibiliser les patients par son évaluation ce qui a pu rendre plus facile ou plus difficile selon le cas l'obtention d'une réponse concordante par le médecin expert qui le suivait. Ce choix méthodologique assurait que les résultats aux tests index étaient recueillis dans des conditions d'évaluation similaires. Aussi, ceci assurait l'aveuglement des tests index au standard de référence et améliorait la faisabilité de l'étude. Il était demandé aux patients de ne pas divulguer d'information sur leur réponse aux tests entre les évaluateurs.

Un autre aspect qui a pu influencer la validité et la fidélité inter-évaluateur des tests de l'examen physique est que le physiothérapeute a exécuté ceux-ci après la collecte de l'anamnèse. Bien que le physiothérapeute fût à l'aveugle du standard de référence, il débutait son évaluation par l'administration standardisée des questionnaires reliés à l'anamnèse du patient avant d'exécuter son examen physique. Bien que ceci réplique le processus clinique de collecte d'information, la connaissance de l'anamnèse peut influencer l'interprétation des réponses à l'examen physique, augmentant potentiellement les estimations de validité diagnostique ainsi que la fidélité inter-évaluateur par la diminution possible des erreurs d'interprétation [98]. Pour remédier à cette limite, les recherches futures pourraient exécuter de façon indépendante les phases de validité et de fidélité et collecter les éléments de l'anamnèse de façon indépendante à l'examen physique [107].

5.4.4 Biais potentiels et limites liés aux méthodes statistiques

Une limite du modèle LASSO concerne la sélection automatisée des variables [177-179]. Le processus de pénalisation des coefficients ß demeure un processus géré par l'algorithme qui prend lui-même la décision de rejeter certaines variables qui démontrent une association comparable ou une colinéarité qu'il considère moins prédictive sur une base mathématique [177-179]. De ce fait, certaines variables d'intérêt clinique ont pu être éliminées par le LASSO. Afin d'atténuer cette limite, nous avons maintenu la possibilité d'inclure manuellement des variables d'intérêt sur la base du jugement clinique et scientifique [60]. Aussi, nous avons analysé les courbes de validation croisée permettant d'observer le retrait progressif des variables du LASSO afin de vérifier si une variable d'intérêt démontrait tout de même une association. Nous avons aussi déterminé un marqueur commun à tous les modèles pour la sélection du coefficient lambda approprié, soit l'optimisation de l'aire sous la courbe. Pour tous les modèles, l'utilisation de ce marqueur a permis d'obtenir les meilleurs prédicteurs en prévision de l'étape subséquente de combinaison des variables.

Nous avons ensuite utilisé la partition récursive pour développer des combinaisons d'éléments de l'anamnèse et de l'examen physique [174, 177, 180, 268, 269]. À l'instar de la régression LASSO, une limite de la partition récursive est que cette analyse présente un algorithme automatisé qui permet de trouver les meilleures séquences d'éléments pour classifier les cas des non-cas. Ceci rend difficile l'imposition de variables précises dans la séquence hiérarchique proposée par la partition récursive. Par exemple, pour certaines combinaisons, un test de l'examen physique pouvait démontrer une meilleure capacité de classification et donc était proposé en première position dans la séquence, avant les éléments de l'anamnèse. Nous avons présenté les combinaisons dans un ordre logique pour l'utilisation clinique, mais de futures recherches pourraient explorer plus en détail la différence de valeur discriminante des éléments qui constituent une combinaison [151].

Une dernière limite de la partition récursive concerne le nombre de différentes possibilités de combinaisons, malgré une présélection des variables à l'aide du LASSO. Une analyse

approfondie a permis de constater que dans la situation où plusieurs variables étaient considérées comme démontrant une « importance » équivalente dans la séquence hiérarchique de la partition récursive, l'algorithme sélectionnait une seule des variables sur une base mathématique. Ainsi, nous avons pu créer plusieurs combinaisons qui démontraient des statistiques diagnostiques similaires avec différentes variables. Pour cette raison, nous avons décidé de présenter différentes combinaisons orientées vers une haute spécificité ou une haute sensibilité. Aussi, l'équipe a analysé la facilité d'utilisation en clinique et la fidélité inter-évaluateur des tests pour sélectionner les modèles les plus pertinents cliniquement. L'utilisation de l'algorithme *randomForest* pour la validation interne permettra aussi d'explorer la création de multiples modèles, ce qui pourrait permettre d'extraire plusieurs combinaisons permettant de classifier précisément plusieurs catégories de patients avec différents niveaux de validité diagnostique [177, 268]. Ceci pourrait améliorer la validité externe de ces modèles diagnostiques et mieux représenter la réalité clinique dans laquelle plusieurs présentations peuvent mener à un diagnostic avec divers degrés de certitude, surtout lors de la présence d'une forte hétérogénéité des présentations cliniques. La partition récursive est donc un algorithme de classification automatisé qui a permis d'extraire des combinaisons parmi un large éventail de possibilités, certaines qui n'auraient pu être identifiées de façon manuelle [174, 177, 180, 268, 269]. Malgré les limites, ce modèle statistique a permis d'améliorer notre compréhension du diagnostic musculosquelettique des pathologies au genou [96, 149, 265].

5.5 Retombées cliniques

Les objectifs de cette thèse étaient d'évaluer l'accord diagnostique entre un physiothérapeute et des médecins experts pour certaines pathologies communes au genou et d'évaluer la validité de la combinaison des éléments de l'anamnèse et de l'examen physique afin d'orienter le diagnostic différentiel des pathologies communes au genou.

Les résultats démontrent que le physiothérapeute participant à cette étude possède l'expertise pour émettre un diagnostic concordant avec celui de médecins experts pour

certaines pathologies au genou étudiées, et ce basé uniquement sur l'évaluation musculosquelettique. Ces résultats pourraient avoir comme impact de réaffirmer le physiothérapeute comme un acteur pertinent dans un rôle de première ligne pour le diagnostic et la prise en charge initiale des problèmes musculosquelettiques, dont les atteintes au genou.

Les résultats présentent aussi le développement initial d'outils valides pour orienter le diagnostic différentiel des pathologies au genou. Rappelons que ces outils permettent d'orienter le processus diagnostique. Ils ne devraient pas être utilisés pour remplacer une évaluation musculosquelettique complète à ce stade préliminaire de développement ni remplacer l'utilisation appropriée des tests d'imagerie médicale et des références en spécialités en situation d'incertitude diagnostique lorsque la présentation clinique du patient est atypique. Notamment, ces outils permettent de classifier efficacement environ 2/3 des patients avec différentes pathologies d'apparition progressive. Finalement, un diagnostic différentiel complet incluant d'autres pathologies rares et potentiellement graves devrait toujours être considéré lors de l'évaluation musculosquelettique. Plusieurs étapes de recherche seront nécessaires avant la mise en place à l'échelle clinique de ces outils [95].

5.6 Avenues futures de recherche

Cette thèse a présenté la dérivation initiale d'une série de modèles qui orientent le diagnostic différentiel des pathologies communes au genou. Un avantage de ces modèles est la relative facilité d'utilisation en clinique nécessitant un minimum d'investissement en temps pour la maitrise des éléments qui les constituent. Ils pourront potentiellement améliorer la capacité diagnostique des cliniciens qui désirent être impliqués dans le diagnostic et la prise en charge des pathologies au genou, tout en diminuant l'utilisation inappropriée des tests d'imagerie et des références en spécialité [28, 270].

Par contre, un outil d'aide au diagnostic doit réussir plusieurs étapes de développement avant l'implantation en clinique [60]. L'étape subséquente à la dérivation sera la validation externe

des outils dans d'autres contextes cliniques d'intérêts [271-276]. Tel que discuté dans les limites de la présente étude, la validation externe devra inclure principalement des milieux de première ligne afin d'évaluer les modèles sur une population présentant des symptômes au moment de leur consultation initiale, notamment dans des unités de médecine familiale, des urgences et des cliniques de physiothérapie. Ces contextes de recrutement permettront d'obtenir plus de cas aigus, ainsi qu'une plus grande variété de diagnostics en première ligne. Par contre, le recrutement prospectif en première ligne nécessitera une plus grande taille d'échantillon que la présente étude [95, 155, 187, 273]. Cette nouvelle cohorte bonifiée représentera une opportunité de procéder à certaines analyses secondaires tel que la mesure de l'impact des lésions combinés ou de l'état aigu sur la validité. Aussi, l'utilisation de mesures cliniques supplémentaires (ex : force, mesure de performance, aspects cognitivo-affectifs) pourrait permettre la détermination de sous-catégories diagnostiques [277]. L'étape de validation externe devra aussi inclure plusieurs évaluateurs avec divers niveaux d'expérience, ce qui permettra d'évaluer certains mécanismes liés à une amélioration de la fidélité inter-évaluateur, une limite de la présente étude.

L'étape subséquente à la validation externe est l'évaluation de l'impact de l'implantation des outils d'aide au diagnostic dans divers milieux cliniques [60, 95, 273]. Les études d'implantation nécessitent un devis expérimental différent des études de validation, l'étude clinique randomisée multicentrique représentant la méthode optimale [273]. Suivant ce devis, différentes mesures de résultats pourront être évaluées. Par exemple, des mesures de résultats pertinentes pourraient inclure la diminution du temps d'attente entre le diagnostic initial et une prise en charge appropriée ou une diminution des références en imagerie médicale ou en spécialité [95]. L'analyse coût-efficacité du changement de trajectoires de soins reliées à une amélioration du diagnostic initial représente le point culminant de l'évaluation du processus d'implantation [28, 95].

Finalement, le processus d'implantation devrait aussi suivre un cadre de transfert des connaissances afin d'assurer la participation des cliniciens et pour supporter un changement

de pratique durable [95, 273, 278-281]. De plus, l'amélioration du diagnostic et de la prise en charge des pathologies du genou en première ligne devrait mettre l'emphase sur l'approche centrée sur le patient [282, 283]. Les données probantes sont particulièrement incomplètes et limitées concernant les sources de conflits décisionnels reliés au diagnostic et à la prise en charge des pathologies musculosquelettiques en première ligne [284-287]. L'intégration de la décision partagée permettra de respecter les préférences des patients et favoriser leur engagement dans le processus décisionnel et ainsi favoriser leur adhésion aux recommandations concernant l'utilisation appropriée des tests d'imagerie médicale, des références en spécialité ainsi qu'à la prise en charge [288-290].

5.7 Conclusion générale

Cette thèse démontre qu'un physiothérapeute est en mesure d'émettre un diagnostic concordant avec celui de médecins experts sur la base seule d'une évaluation musculosquelettique standardisée. Ceci a permis par la suite de cibler les éléments clés de l'anamnèse et de l'examen physique qui lorsque combinés permettent de classifier certaines pathologies communes au genou d'origine traumatique ou d'apparition progressive. Au total, les différentes combinaisons requièrent vingt-cinq éléments cliniques, quinze éléments de l'anamnèse et dix de l'examen physique pour classifier les individus. Les combinaisons ont atteint des rapports de vraisemblance positifs et négatifs les rendant modérément à hautement valides pour orienter le diagnostic différentiel des pathologies d'intérêts. Ces outils ne remplacent pas une évaluation musculosquelettique complète ni l'utilisation appropriée des tests d'imagerie ou des références en spécialité et ne permettent pas d'identifier tous les patients. Un diagnostic différentiel complet incluant des pathologies rares et pouvant représenter un risque pour la santé devrait toujours être considéré.

Les recherches futures devront combler les limites de la présente étude. Suivant une validation externe adéquate, ces outils d'aide au diagnostic pourront permettre aux cliniciens d'améliorer la précision du diagnostic initial et ainsi améliorer les trajectoires de prise en charge de patients souffrant de problèmes au genou. Cette méthode diagnostique a le

potentiel de diminuer le risque de chronicisation des symptômes et des limitations fonctionnelles chez les individus atteints d'une pathologie au genou, d'améliorer leur qualité de vie et ultimement d'améliorer l'efficience des soins chez ces populations.

Bibliographie

1. Baker, P., et al., *Knee disorders in the general population and their relation to occupation.* Occup Environ Med, 2003. 60(10): p. 794-7.
2. Jette, A.M. and A. Delitto, *Physical therapy treatment choices for musculoskeletal impairments.* Phys Ther, 1997. 77(2): p. 145-54.
3. Lau, B.H., et al., *Utilization and cost of a new model of care for managing acute knee injuries: the Calgary Acute Knee Injury Clinic.* BMC Health Serv Res, 2012. 12: p. 445.
4. Mody, G.M. and P.M. Brooks, *Improving musculoskeletal health: global issues.* Best Pract Res Clin Rheumatol, 2012. 26(2): p. 237-49.
5. Granan, L.P., et al., *Sport-Specific Injury Pattern Recorded During Anterior Cruciate Ligament Reconstruction.* Am J Sports Med, 2013.
6. Jackson, J.L., P.G. O'Malley, and K. Kroenke, *Evaluation of acute knee pain in primary care.* Ann Intern Med, 2003. 139(7): p. 575-88.
7. Crossley, K.M., M.J. Callaghan, and R. van Linschoten, *Patellofemoral pain.* BMJ, 2015. 351: p. h3939.
8. Crossley, K.M., et al., *2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures.* Br J Sports Med, 2016. 50(14): p. 839-43.
9. Mather, R.C., 3rd, et al., *Cost-effectiveness Analysis of the Diagnosis of Meniscus Tears.* Am J Sports Med, 2015. 43(1): p. 128-37.
10. Deshpande, B.R., et al., *Association of MRI findings and expert diagnosis of symptomatic meniscal tear among middle-aged and older adults with knee pain.* BMC Musculoskelet Disord, 2016. 17: p. 154.
11. Katz, J.N., et al., *The Value of History, Physical Examination, and Radiographic Findings in the Diagnosis of Symptomatic Meniscal Tear among Middle-Age Subjects with Knee Pain.* Arthritis Care Res (Hoboken), 2016.
12. Tuakli-Wosornu, Y.A., et al., *Predictors of Exercise Adherence in Patients With Meniscal Tear and Osteoarthritis.* Arch Phys Med Rehabil, 2016. 97(11): p. 1945-1952.
13. Khan, H.I., et al., *Natural history and clinical significance of meniscal tears over 8 years in a midlife cohort.* BMC Musculoskelet Disord, 2016. 17: p. 4.
14. Anderson, M.J., et al., *A Systematic Summary of Systematic Reviews on the Topic of the Anterior Cruciate Ligament.* Orthop J Sports Med, 2016. 4(3): p. 2325967116634074.
15. Lohmander, L.S., et al., *The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis.* Am J Sports Med, 2007. 35(10): p. 1756-69.
16. Suter, L.G., et al., *Projecting Lifetime Risk of Symptomatic Knee Osteoarthritis and Total Knee Replacement in Individuals Sustaining a Complete Anterior Cruciate Ligament Tear in Early Adulthood.* Arthritis Care Res (Hoboken), 2017. 69(2): p. 201-208.

17. Sanders, T.L., et al., *Is Anterior Cruciate Ligament Reconstruction Effective in Preventing Secondary Meniscal Tears and Osteoarthritis?* Am J Sports Med, 2016. 44(7): p. 1699-707.
18. Parwaiz, H., A.Q. Teo, and C. Servant, *Anterior cruciate ligament injury: A persistently difficult diagnosis.* Knee, 2016. 23(1): p. 116-20.
19. Micheo, W., L. Hernandez, and C. Seda, *Evaluation, management, rehabilitation, and prevention of anterior cruciate ligament injury: current concepts.* PM R, 2010. 2(10): p. 935-44.
20. Zhang, W., et al., *OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence.* Osteoarthritis Cartilage, 2007. 15(9): p. 981-1000.
21. Zhang, W., et al., *OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines.* Osteoarthritis Cartilage, 2008. 16(2): p. 137-62.
22. Zhang, W., et al., *OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009.* Osteoarthritis Cartilage, 2010. 18(4): p. 476-99.
23. Meneses, S.R., et al., *Clinical algorithms to aid osteoarthritis guideline dissemination.* Osteoarthritis Cartilage, 2016.
24. Moore, J.H., et al., *Clinical diagnostic accuracy and magnetic resonance imaging of patients referred by physical therapists, orthopaedic surgeons, and nonorthopaedic providers.* J Orthop Sports Phys Ther, 2005. 35(2): p. 67-71.
25. Institut National d'Excellence en Santé et Services Sociaux, *Utilisation de l'imagerie par résonance magnétique en présence de douleurs musculosquelettiques chez les personnes adultes du Québec.* 2017.
26. Desmeules, F., et al., *Advanced practice physiotherapy in patients with musculoskeletal disorders: a systematic review.* BMC Musculoskelet Disord, 2012. 13(1): p. 107.
27. Desmeules, F., et al., *Validation of an advanced practice physiotherapy model of care in an orthopaedic outpatient clinic.* BMC Musculoskelet Disord, 2013. 14: p. 162.
28. Churchill, L., et al., *The development and validation of a multivariable model to predict whether patients referred for total knee replacement are suitable surgical candidates at the time of initial consultation.* Can J Surg, 2016. 59(6): p. 407-414.
29. Stein, L.A., *Making the best use of radiological resources in Canada.* Healthc Pap, 2005. 6(1): p. 18-23.
30. Jackson, J., et al., *Quantitative Blood Loss Measurement: A Superior Technique for Trainees and Seasoned Clinicians [22K].* Obstetrics & Gynecology, 2016. 127: p. 94S-95S.
31. Logerstedt, D.S., et al., *Knee stability and movement coordination impairments: knee ligament sprain.* J Orthop Sports Phys Ther, 2010. 40(4): p. A1-A37.

32. Wasserstein, D., et al., *Risk factors for recurrent anterior cruciate ligament reconstruction: a population study in Ontario, Canada, with 5-year follow-up.* Am J Sports Med, 2013. 41(9): p. 2099-107.
33. Prodromos, C.C., et al., *A meta-analysis of the incidence of anterior cruciate ligament tears as a function of gender, sport, and a knee injury-reduction regimen.* Arthroscopy, 2007. 23(12): p. 1320-1325 e6.
34. Benjaminse, A., A. Gokeler, and C.P. van der Schans, *Clinical diagnosis of an anterior cruciate ligament rupture: a meta-analysis.* J Orthop Sports Phys Ther, 2006. 36(5): p. 267-88.
35. Dejour, D., et al., *The diagnostic value of clinical tests, magnetic resonance imaging, and instrumented laxity in the differentiation of complete versus partial anterior cruciate ligament tears.* Arthroscopy, 2013. 29(3): p. 491-9.
36. Delince, P. and D. Ghafil, *Anterior cruciate ligament tears: conservative or surgical treatment? A critical review of the literature.* Knee Surg Sports Traumatol Arthrosc, 2012. 20(1): p. 48-61.
37. Jarraya, M., et al., *Meniscus morphology: Does tear type matter? A narrative review with focus on relevance for osteoarthritis research.* Semin Arthritis Rheum, 2016.
38. Meserve, B.B., J.A. Cleland, and T.R. Boucher, *A meta-analysis examining clinical test utilities for assessing meniscal injury.* Clin Rehabil, 2008. 22(2): p. 143-61.
39. Englund, M., et al., *Incidental meniscal findings on knee MRI in middle-aged and elderly persons.* N Engl J Med, 2008. 359(11): p. 1108-15.
40. Ghislain, N.A., J.N. Wei, and Y.G. Li, *Study of the Clinical Outcome between Traumatic and Degenerative (non-traumatic) Meniscal Tears after Arthroscopic Surgery: A 4-Years Follow-up Study.* J Clin Diagn Res, 2016. 10(4): p. RC01-4.
41. Thorlund, J.B., et al., *Patient reported outcomes in patients undergoing arthroscopic partial meniscectomy for traumatic or degenerative meniscal tears: comparative prospective cohort study.* BMJ, 2017. 356: p. j356.
42. Hare, K.B., et al., *Middle-aged patients with an MRI-verified medial meniscal tear report symptoms commonly associated with knee osteoarthritis.* Acta Orthop, 2017: p. 1-6.
43. Chapple, C.M., et al., *Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies.* Arthritis Care Res (Hoboken), 2011. 63(8): p. 1115-25.
44. Cheung, P.P., L. Gossec, and M. Dougados, *What are the best markers for disease progression in osteoarthritis (OA)?* Best Pract Res Clin Rheumatol, 2010. 24(1): p. 81-92.
45. Cross, M., et al., *The global burden of rheumatoid arthritis: estimates from the Global Burden of Disease 2010 study.* Ann Rheum Dis, 2014.
46. Zhang, W., et al., *EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis.* Ann Rheum Dis, 2010. 69(3): p. 483-9.
47. Crossley, K.M., et al., *2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 2: recommended*

physical interventions (exercise, taping, bracing, foot orthoses and combined interventions). Br J Sports Med, 2016. 50(14): p. 844-52.
48. Fredericson, M. and K. Yoon, *Physical examination and patellofemoral pain syndrome.* Am J Phys Med Rehabil, 2006. 85(3): p. 234-43.
49. Witvrouw, E., et al., *The 3rd International Patellofemoral Research Retreat: an international expert consensus meeting to improve the scientific understanding and clinical management of patellofemoral pain.* Br J Sports Med, 2014. 48(6): p. 408.
50. Barton, C.J., et al., *Kinematic gait characteristics associated with patellofemoral pain syndrome: a systematic review.* Gait Posture, 2009. 30(4): p. 405-16.
51. Kastelein, M., et al., *The 6-year trajectory of non-traumatic knee symptoms (including patellofemoral pain) in adolescents and young adults in general practice: a study of clinical predictors.* Br J Sports Med, 2015. 49(6): p. 400-5.
52. Lankhorst, N.E., et al., *Factors that predict a poor outcome 5-8 years after the diagnosis of patellofemoral pain: a multicentre observational analysis.* Br J Sports Med, 2016. 50(14): p. 881-6.
53. Rathleff, M.S., et al., *Is Knee Pain During Adolescence a Self-limiting Condition? Prognosis of Patellofemoral Pain and Other Types of Knee Pain.* Am J Sports Med, 2016. 44(5): p. 1165-71.
54. Wyndow, N., et al., *Is There a Biomechanical Link Between Patellofemoral Pain and Osteoarthritis? A Narrative Review.* Sports Med, 2016. 46(12): p. 1797-1808.
55. Conchie, H., et al., *Adolescent knee pain and patellar dislocations are associated with patellofemoral osteoarthritis in adulthood: A case control study.* Knee, 2016. 23(4): p. 708-11.
56. Shrier, I., M. Boudier-Reveret, and K. Fahmy, *Understanding the different physical examination tests for suspected meniscal tears.* Curr Sports Med Rep, 2010. 9(5): p. 284-9.
57. Neogi, T., *The epidemiology and impact of pain in osteoarthritis.* Osteoarthritis Cartilage, 2013. 21(9): p. 1145-53.
58. Laslett, L.L., et al., *A prospective study of the impact of musculoskeletal pain and radiographic osteoarthritis on health related quality of life in community dwelling older people.* BMC Musculoskelet Disord, 2012. 13(1): p. 168.
59. Tveit, M., et al., *Former male elite athletes have a higher prevalence of osteoarthritis and arthroplasty in the hip and knee than expected.* Am J Sports Med, 2012. 40(3): p. 527-33.
60. Lubetzky-Vilnai, A., M. Ciol, and S.W. McCoy, *Statistical analysis of clinical prediction rules for rehabilitation interventions: current state of the literature.* Arch Phys Med Rehabil, 2014. 95(1): p. 188-96.
61. Snoeker, B.A., et al., *Risk factors for meniscal tears: a systematic review including meta-analysis.* J Orthop Sports Phys Ther, 2013. 43(6): p. 352-67.
62. Bharam, S., M.S. Vrahas, and F.H. Fu, *Knee fractures in the athlete.* Orthop Clin North Am, 2002. 33(3): p. 565-74.

63. Barber Foss, K.D., et al., *Expected prevalence from the differential diagnosis of anterior knee pain in adolescent female athletes during preparticipation screening.* J Athl Train, 2012. 47(5): p. 519-24.
64. Andersen, C.A., et al., *High prevalence of overuse injury among iron-distance triathletes.* Br J Sports Med, 2013. 47(13): p. 857-61.
65. Filbay, S.R., et al., *Quality of life in anterior cruciate ligament-deficient individuals: a systematic review and meta-analysis.* Br J Sports Med, 2015. 49(16): p. 1033-41.
66. Skou, S.T., H. Thomsen, and O.H. Simonsen, *The value of routine radiography in patients with knee osteoarthritis consulting primary health care: A study of agreement.* Eur J Gen Pract, 2013.
67. Reiman, M.P., et al., *Diagnostic accuracy of clinical tests of the hip: a systematic review with meta-analysis.* Br J Sports Med, 2013. 47(14): p. 893-902.
68. Rahman, L.A., et al., *A systematic review of the diagnostic performance of orthopedic physical examination tests of the hip.* BMC Musculoskelet Disord, 2013. 14(1): p. 257.
69. Hegedus, E.J., et al., *Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests.* Br J Sports Med, 2012. 46(14): p. 964-78.
70. Hegedus, E.J., et al., *Physical examination tests for assessing a torn meniscus in the knee: a systematic review with meta-analysis.* J Orthop Sports Phys Ther, 2007. 37(9): p. 541-50.
71. Hing, W., et al., *Validity of the McMurray's Test and Modified Versions of the Test: A Systematic Literature Review.* J Man Manip Ther, 2009. 17(1): p. 22-35.
72. Ryzewicz, M., et al., *The diagnosis of meniscus tears: the role of MRI and clinical examination.* Clin Orthop Relat Res, 2007. 455: p. 123-33.
73. van Eck, C.F., et al., *Methods to diagnose acute anterior cruciate ligament rupture: a meta-analysis of physical examinations with and without anaesthesia.* Knee Surg Sports Traumatol Arthrosc, 2013. 21(8): p. 1895-903.
74. Cook, C., et al., *Best tests/clinical findings for screening and diagnosis of patellofemoral pain syndrome: a systematic review.* Physiotherapy, 2012. 98(2): p. 93-100.
75. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* PLoS Med, 2009. 6(7): p. e1000097.
76. Liberati, A., et al., *The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration.* PLoS Med, 2009. 6(7): p. e1000100.
77. Magee, D.J., *Orthopedic Physical Assessment, 6th Edition*. 2014: Saunders. 1184.
78. Cook, C. and E. Hegedus, *Orthopedic Physical Examination Tests: An Evidence-Based Approach (2nd Edition)*. 2011: Prentice Hall.
79. Shea, B.J., et al., *Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews.* BMC Med Res Methodol, 2007. 7: p. 10.
80. Shea, B.J., et al., *External validation of a measurement tool to assess systematic reviews (AMSTAR).* PLoS One, 2007. 2(12): p. e1350.

81. Shea, B.J., et al., *AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews.* J Clin Epidemiol, 2009. 62(10): p. 1013-20.
82. Jaeschke, R., G.H. Guyatt, and D.L. Sackett, *Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group.* JAMA, 1994. 271(9): p. 703-7.
83. Scholten, R.J., et al., *The accuracy of physical diagnostic tests for assessing meniscal lesions of the knee: a meta-analysis.* J Fam Pract, 2001. 50(11): p. 938-44.
84. Smith, B.E., et al., *Special tests for assessing meniscal tears within the knee: a systematic review and meta-analysis.* Evid Based Med, 2015. 20(3): p. 88-97.
85. Scholten, R.J., et al., *Accuracy of physical diagnostic tests for assessing ruptures of the anterior cruciate ligament: a meta-analysis.* J Fam Pract, 2003. 52(9): p. 689-94.
86. Leblanc, M.C., et al., *Diagnostic accuracy of physical examination for anterior knee instability: a systematic review.* Knee Surg Sports Traumatol Arthrosc, 2015.
87. Solomon, D.H., et al., *The rational clinical examination. Does this patient have a torn meniscus or ligament of the knee? Value of the physical examination.* JAMA, 2001. 286(13): p. 1610-20.
88. Bachmann, L.M., et al., *The accuracy of the Ottawa knee rule to rule out knee fractures: a systematic review.* Ann Intern Med, 2004. 140(2): p. 121-4.
89. Nunes, G.S., et al., *Clinical test for diagnosis of patellofemoral pain syndrome: Systematic review with meta-analysis.* Phys Ther Sport, 2013. 14(1): p. 54-9.
90. Kopkow, C., et al., *Physical examination tests for the diagnosis of posterior cruciate ligament rupture: a systematic review.* J Orthop Sports Phys Ther, 2013. 43(11): p. 804-13.
91. Faggion, C.M., Jr., *Critical appraisal of AMSTAR: challenges, limitations, and potential solutions from the perspective of an assessor.* BMC Med Res Methodol, 2015. 15: p. 63.
92. Gagnier, J.J. and P.J. Kellam, *Reporting and methodological quality of systematic reviews in the orthopaedic literature.* J Bone Joint Surg Am, 2013. 95(11): p. e771-7.
93. Stiell, I.G., et al., *Implementation of the Ottawa Knee Rule for the use of radiography in acute knee injuries.* JAMA, 1997. 278(23): p. 2075-9.
94. Perry, J.J. and I.G. Stiell, *Impact of clinical decision rules on clinical care of traumatic injuries to the foot and ankle, knee, cervical spine, and head.* Injury, 2006. 37(12): p. 1157-65.
95. Stiell, I.G. and C. Bennett, *Implementation of clinical decision rules in the emergency department.* Acad Emerg Med, 2007. 14(11): p. 955-9.
96. Altman, R., 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(8): p. 1039-49.
97. Peat, G., et al., *Clinical classification criteria for knee osteoarthritis: performance in the general population and primary care.* Ann Rheum Dis, 2006. 65(10): p. 1363-7.

98. Whiting, P.F., et al., *A systematic review classifies sources of bias and variation in diagnostic test accuracy studies.* J Clin Epidemiol, 2013. 66(10): p. 1093-104.
99. Daker-White, G., et al., *A randomised controlled trial. Shifting boundaries of doctors and physiotherapists in orthopaedic outpatient departments.* J Epidemiol Community Health, 1999. 53(10): p. 643-50.
100. Cook, C., et al., *Diagnostic accuracy and association to disability of clinical test findings associated with patellofemoral pain syndrome.* Physiother Can, 2010. 62(1): p. 17-24.
101. Wagemakers, H.P., et al., *Diagnostic accuracy of history taking and physical examination for assessing anterior cruciate ligament lesions of the knee in primary care.* Arch Phys Med Rehabil, 2010. 91(9): p. 1452-9.
102. Wagemakers, H.P., et al., *Diagnostic value of history-taking and physical examination for assessing meniscal tears of the knee in general practice.* Clin J Sport Med, 2008. 18(1): p. 24-30.
103. Kastelein, M., et al., *Diagnostic value of history taking and physical examination to assess effusion of the knee in traumatic knee patients in general practice.* Arch Phys Med Rehabil, 2009. 90(1): p. 82-6.
104. Kastelein, M., et al., *Assessing medial collateral ligament knee lesions in general practice.* Am J Med, 2008. 121(11): p. 982-988 e2.
105. Lange, T., et al., *The reliability of physical examination tests for the diagnosis of anterior cruciate ligament rupture--A systematic review.* Man Ther, 2015. 20(3): p. 402-11.
106. Tooth, L.R. and K.J. Ottenbacher, *The kappa statistic in rehabilitation research: an examination.* Arch Phys Med Rehabil, 2004. 85(8): p. 1371-6.
107. Lucas, N.P., et al., *The development of a quality appraisal tool for studies of diagnostic reliability (QAREL).* J Clin Epidemiol, 2010. 63(8): p. 854-61.
108. May, S., et al., *Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review.* Physiotherapy, 2010. 96(3): p. 179-90.
109. Kundel, H.L. and M. Polansky, *Measurement of observer agreement.* Radiology, 2003. 228(2): p. 303-8.
110. Whiting, P., et al., *Sources of variation and bias in studies of diagnostic accuracy: a systematic review.* Ann Intern Med, 2004. 140(3): p. 189-202.
111. Kottner, J., et al., *Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed.* J Clin Epidemiol, 2011. 64(1): p. 96-106.
112. Smith, T.O., L. Davies, and S.T. Donell, *The reliability and validity of assessing medio-lateral patellar position: a systematic review.* Man Ther, 2009. 14(4): p. 355-62.
113. Cleland, K., Su, *Netter's Orthopaedic Clinical Examination 3rd edition*. 2015: Elsevier. 664.
114. Liberati, A., et al., *The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration.* BMJ, 2009. 339: p. b2700.
115. Lucas, N., et al., *The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL).* BMC Med Res Methodol, 2013. 13: p. 111.

116. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. 33(1): p. 159-74.
117. Dervin, G.F., et al., *Physicians' accuracy and interrator reliability for the diagnosis of unstable meniscal tears in patients having osteoarthritis of the knee.* Can J Surg, 2001. 44(4): p. 267-74.
118. Evans, P.J., G.D. Bell, and C. Frank, *Prospective evaluation of the McMurray test.* Am J Sports Med, 1993. 21(4): p. 604-8.
119. Galli, M., et al., *Joint line tenderness and McMurray tests for the detection of meniscal lesions: what is their real diagnostic value?* Arch Phys Med Rehabil, 2013. 94(6): p. 1126-31.
120. Snoeker, B.A., et al., *Detecting Meniscal Tears in Primary Care: Reproducibility and Accuracy of 2 Weight-Bearing Tests and 1 Non-Weight-Bearing Test.* J Orthop Sports Phys Ther, 2015. 45(9): p. 693-702.
121. Cooperman, J.M., D.L. Riddle, and J.M. Rothstein, *Reliability and validity of judgments of the integrity of the anterior cruciate ligament of the knee using the Lachman's test.* Phys Ther, 1990. 70(4): p. 225-33.
122. Duggan, N.P. and J.C. Ross, *Intertester reliability of the Modified Anterior Drawer (Lachman) Test for anterior cruciate ligament laxity.* Aust J Physiother, 1991. 37(3): p. 163-8.
123. Johnson, D.S., W.G. Ryan, and R.B. Smith, *Does the Lachman testing method affect the reliability of the International Knee Documentation Committee (IKDC) Form?* Knee Surg Sports Traumatol Arthrosc, 2004. 12(3): p. 225-8.
124. Mulligan, E.P., J.L. Harwell, and W.J. Robertson, *Reliability and diagnostic accuracy of the Lachman test performed in a prone position.* J Orthop Sports Phys Ther, 2011. 41(10): p. 749-57.
125. Peeler, J., J. Leiter, and P. MacDonald, *Accuracy and reliability of anterior cruciate ligament clinical examination in a multidisciplinary sports medicine setting.* Clin J Sport Med, 2010. 20(2): p. 80-5.
126. Wiertsema, S.H., et al., *Reliability of the KT1000 arthrometer and the Lachman test in patients with an ACL rupture.* Knee, 2008. 15(2): p. 107-10.
127. Labbe, D.R., et al., *Accounting for velocity of the pivot shift test manoeuvre decreases kinematic variability.* Knee, 2011. 18(2): p. 88-93.
128. Mulligan, E.P., et al., *The reliability and diagnostic accuracy of assessing the translation endpoint during the lachman test.* Int J Sports Phys Ther, 2015. 10(1): p. 52-61.
129. Yoon, K.H., et al., *Can physical examination predict the intraarticular tear pattern of the anterior cruciate ligament?* Arch Orthop Trauma Surg, 2014. 134(10): p. 1451-7.
130. Cibere, J., et al., *Reliability of the knee examination in osteoarthritis: effect of standardization.* Arthritis Rheum, 2004. 50(2): p. 458-68.
131. Cushnaghan, J., et al., *Clinical assessment of osteoarthritis of the knee.* Ann Rheum Dis, 1990. 49(10): p. 768-70.
132. Hart, D.J., et al., *Clinical signs of early osteoarthritis: reproducibility and relation to x ray changes in 541 women in the general population.* Ann Rheum Dis, 1991. 50(7): p. 467-70.

133. Jones, A., et al., *Evaluation of a method for clinically assessing osteoarthritis of the knee.* Ann Rheum Dis, 1992. 51(2): p. 243-5.
134. Wood, L., et al., *A study of the noninstrumented physical examination of the knee found high observer variability.* J Clin Epidemiol, 2006. 59(5): p. 512-20.
135. Fitzgerald, G.K. and P.W. McClure, *Reliability of measurements obtained with four tests for patellofemoral alignment.* Phys Ther, 1995. 75(2): p. 84-90; discussion 90-2.
136. Smith, T.O., et al., *The intra- and inter-observer reliability of the physical examination methods used to assess patients with patellofemoral joint instability.* Knee, 2012. 19(4): p. 404-10.
137. Sweitzer, B.A., et al., *The inter-rater reliability and diagnostic accuracy of patellar mobility tests in patients with anterior knee pain.* Phys Sportsmed, 2010. 38(3): p. 90-6.
138. Watson, C.J., et al., *Reliability of the lateral pull test and tilt test to assess patellar alignment in subjects with symptomatic knees: student raters.* J Orthop Sports Phys Ther, 2001. 31(7): p. 368-74.
139. Watson, C.J., et al., *Reliability of McConnell's classification of patellar orientation in symptomatic and asymptomatic subjects.* J Orthop Sports Phys Ther, 1999. 29(7): p. 378-85; discussion 386-93.
140. Herrington, L., *The difference in a clinical measure of patella lateral position between individuals with patellofemoral pain and matched controls.* J Orthop Sports Phys Ther, 2008. 38(2): p. 59-62.
141. Lesher, J.D., et al., *Development of a clinical prediction rule for classifying patients with patellofemoral pain syndrome who respond to patellar taping.* J Orthop Sports Phys Ther, 2006. 36(11): p. 854-66.
142. Piva, S.R., et al., *Reliability of measures of impairments associated with patellofemoral pain syndrome.* BMC Musculoskelet Disord, 2006. 7: p. 33.
143. Akgul, O., Z. Guldeste, and S. Ozgocmen, *The reliability of the clinical examination for detecting Baker's cyst in asymptomatic fossa.* Int J Rheum Dis, 2014. 17(2): p. 204-9.
144. Cook, J.L., et al., *Reproducibility and clinical utility of tendon palpation to detect patellar tendinopathy in young basketball players. Victorian Institute of Sport tendon study group.* Br J Sports Med, 2001. 35(1): p. 65-9.
145. McClure, P.W., J.M. Rothstein, and D.L. Riddle, *Intertester reliability of clinical judgments of medial knee ligament integrity.* Phys Ther, 1989. 69(4): p. 268-75.
146. Sturgill, L.P., et al., *Interrater reliability of a clinical scale to assess knee joint effusion.* J Orthop Sports Phys Ther, 2009. 39(12): p. 845-9.
147. Ulasli, A.M., et al., *Accuracy in detecting knee effusion with clinical examination and the effect of effusion, the patient's body mass index, and the clinician's experience.* Clin Rheumatol, 2014. 33(8): p. 1139-43.
148. Stiell, I.G., et al., *Prospective validation of a decision rule for the use of radiography in acute knee injuries.* JAMA, 1996. 275(8): p. 611-5.
149. Stiell, I.G., et al., *Derivation of a decision rule for the use of radiography in acute knee injuries.* Ann Emerg Med, 1995. 26(4): p. 405-13.

150. Geraets, S.E., et al., *Diagnostic value of medical history and physical examination of anterior cruciate ligament injury: comparison between primary care physician and orthopaedic surgeon.* Knee Surg Sports Traumatol Arthrosc, 2015. 23(4): p. 968-74.
151. Snoeker, B.A., et al., *A clinical prediction rule for meniscal tears in primary care: development and internal validation using a multicentre study.* Br J Gen Pract, 2015. 65(637): p. e523-9.
152. Aiken, A.B. and M.A. McColl, *Diagnostic and treatment concordance between a physiotherapist and an orthopedic surgeon--a pilot study.* J Interprof Care, 2008. 22(3): p. 253-61.
153. Decary, S., et al., *Diagnostic validity of physical examination tests for common knee disorders: An overview of systematic reviews and meta-analysis.* Phys Ther Sport, 2016.
154. Decary, S., et al., *Reliability of physical examination tests for the diagnosis of knee disorders: Evidence from a systematic review.* Man Ther, 2016. 26: p. 172-182.
155. Whiting, P., et al., *The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews.* BMC Med Res Methodol, 2003. 3: p. 25.
156. Whiting, P.F., et al., *QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.* Ann Intern Med, 2011. 155(8): p. 529-36.
157. Bossuyt, P.M., et al., *STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies.* BMJ, 2015. 351: p. h5527.
158. Simel, D.L., D. Rennie, and P.M. Bossuyt, *The STARD statement for reporting diagnostic accuracy studies: application to the history and physical examination.* J Gen Intern Med, 2008. 23(6): p. 768-74.
159. Collins, N.J., et al., *Knee Injury and Osteoarthritis Outcome Score (KOOS): systematic review and meta-analysis of measurement properties.* Osteoarthritis Cartilage, 2016. 24(8): p. 1317-29.
160. Peer, M.A. and J. Lane, *The Knee Injury and Osteoarthritis Outcome Score (KOOS): a review of its psychometric properties in people undergoing total knee arthroplasty.* J Orthop Sports Phys Ther, 2013. 43(1): p. 20-8.
161. Collins, N.J., et al., *Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS).* Arthritis Care Res (Hoboken), 2011. 63 Suppl 11: p. S208-28.
162. Esculier, J.F., J.S. Roy, and L.J. Bouyer, *Psychometric evidence of self-reported questionnaires for patellofemoral pain syndrome: a systematic review.* Disabil Rehabil, 2013. 35(26): p. 2181-90.

163. Kessler, R.C., et al., *Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative.* Int J Methods Psychiatr Res, 2010. 19 Suppl 1: p. 4-22.
164. Piriyaprasarth, P. and M.E. Morris, *Psychometric properties of measurement tools for quantifying knee joint position and movement: a systematic review.* Knee, 2007. 14(1): p. 2-8.
165. Borgbjerg, J., F. Madsen, and A. Odgaard, *Patient Self-Assessed Passive Range of Motion of the Knee Cannot Replace Health Professional Measurements.* J Knee Surg, 2017.
166. Culvenor, A.G., et al., *Patellofemoral osteoarthritis is prevalent and associated with worse symptoms and function after hamstring tendon autograft ACL reconstruction.* Br J Sports Med, 2014. 48(6): p. 435-9.
167. Boks, S.S., et al., *Magnetic resonance imaging abnormalities in symptomatic and contralateral knees: prevalence and associations with traumatic history in general practice.* Am J Sports Med, 2006. 34(12): p. 1984-91.
168. Reitsma, J.B., et al., *A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard.* J Clin Epidemiol, 2009. 62(8): p. 797-806.
169. Henschke, N., et al., *The methodological quality of diagnostic test accuracy studies for musculoskeletal conditions can be improved.* J Clin Epidemiol, 2014. 67(4): p. 416-24.
170. Rutjes, A.W., et al., *Evaluation of diagnostic tests when there is no gold standard. A review of methods.* Health Technol Assess, 2007. 11(50): p. iii, ix-51.
171. Razmjou, H., et al., *Evaluation of an advanced-practice physical therapist in a specialty shoulder clinic: diagnostic agreement and effect on wait times.* Physiother Can, 2013. 65(1): p. 46-55.
172. Sim, J. and C.C. Wright, *The kappa statistic in reliability studies: use, interpretation, and sample size requirements.* Phys Ther, 2005. 85(3): p. 257-68.
173. Hegedus, E.J. and B. Stern, *Beyond SpPIN and SnNOUT: Considerations with Dichotomous Tests During Assessment of Diagnostic Accuracy.* J Man Manip Ther, 2009. 17(1): p. E1-5.
174. Adams, S.T. and S.H. Leveson, *Clinical prediction rules.* BMJ, 2012. 344: p. d8312.
175. Reilly, B.M. and A.T. Evans, *Translating clinical research into clinical practice: impact of using prediction rules to make decisions.* Ann Intern Med, 2006. 144(3): p. 201-9.
176. Laupacis, A., N. Sekar, and I.G. Stiell, *Clinical prediction rules. A review and suggested modifications of methodological standards.* JAMA, 1997. 277(6): p. 488-94.
177. Lu, F. and E. Petkova, *A comparative study of variable selection methods in the context of developing psychiatric screening instruments.* Stat Med, 2014. 33(3): p. 401-21.
178. Tibshirani, R., *Regression shrinkage and selection via the LASSO.* Journal of the Royal Statistical Society, Series B., 1996. 58:267–288.
179. Copas, J., *Regression, Prediction and Shrinkage.* Journal of the Royal Statistical Society, 1983. Series B (Methodological)(Vol 45. No 3. (1983)): p. pp. 311-354.
180. Breiman, L.F., JH.; Alshen, RA.; Stone, CJ, *CART: Classification and Regression Trees.* 1984: Belmont, CA: Wadsworth; .

181. Muller, R. and M. Mockel, *Logistic regression and CART in the analysis of multimarker studies.* Clin Chim Acta, 2008. 394(1-2): p. 1-6.
182. Strobl C, B.A., Augustin T, *Unbiased split selection for classification trees based on the Gini index.* Comput Stat Data Anal, 2007. 52:483.
183. Aguiar, F.S., et al., *Classification and regression tree (CART) model to predict pulmonary tuberculosis in hospitalized patients.* BMC Pulm Med, 2012. 12: p. 40.
184. Simel, D.L., G.P. Samsa, and D.B. Matchar, *Likelihood ratios with confidence: sample size estimation for diagnostic test studies.* J Clin Epidemiol, 1991. 44(8): p. 763-70.
185. Eusebi, P., *Diagnostic accuracy measures.* Cerebrovasc Dis, 2013. 36(4): p. 267-72.
186. Steyerberg, E.W., et al., *Internal validation of predictive models: efficiency of some procedures for logistic regression analysis.* J Clin Epidemiol, 2001. 54(8): p. 774-81.
187. Steyerberg, E.W., et al., *Internal and external validation of predictive models: a simulation study of bias and precision in small samples.* J Clin Epidemiol, 2003. 56(5): p. 441-7.
188. Efron, B., *An introduction to the bootstrap.* New York, NY: Chapman & Hall, 1993.
189. Bachmann, L.M., et al., *Sample sizes of studies on diagnostic accuracy: literature survey.* BMJ, 2006. 332(7550): p. 1127-9.
190. Austin, P.C. and E.W. Steyerberg, *The number of subjects per variable required in linear regression analyses.* J Clin Epidemiol, 2015. 68(6): p. 627-36.
191. Kent, P., et al., *Four hundred or more participants needed for stable contingency table estimates of clinical prediction rule performance.* J Clin Epidemiol, 2017. 82: p. 137-148.
192. van Middelkoop, M., et al., *Knee complaints seen in general practice: active sport participants versus non-sport participants.* BMC Musculoskelet Disord, 2008. 9: p. 36.
193. Swart, N.M., et al., *Efficacy of MRI in primary care for patients with knee complaints due to trauma: protocol of a randomised controlled non-inferiority trial (TACKLE trial).* BMC Musculoskelet Disord, 2014. 15: p. 63.
194. de Gruchy, A., C. Granger, and A. Gorelik, *Physical Therapists as Primary Practitioners in the Emergency Department: Six-Month Prospective Practice Analysis.* Phys Ther, 2015. 95(9): p. 1207-16.
195. Bornhoft, L., M.E. Larsson, and J. Thorn, *Physiotherapy in Primary Care Triage - the effects on utilization of medical services at primary health care clinics by patients and sub-groups of patients with musculoskeletal disorders: a case-control study.* Physiother Theory Pract, 2015. 31(1): p. 45-52.
196. Décary, S., et al., *Diagnostic validity of physical examination tests for common knee disorders: An overview of systematic reviews and meta-analysis.* Physical Therapy in Sport.
197. Kessler, R.C., et al., *Screening for serious mental illness in the general population.* Arch Gen Psychiatry, 2003. 60(2): p. 184-9.
198. Naaktgeboren, C.A., et al., *Value of composite reference standards in diagnostic research.* BMJ, 2013. 347: p. f5605.
199. Flack, V.F., Afifi, A.A., Lachenbruch, P.A., & Schouten, H.J.A., *Sample size determinations for the two rater kappa statistic.* Psychometrika, 1988. 53: p. 321-325.

200. Cantor, A.B., *Sample-size calculation for Cohen's kappa.* Psychological Methods, 1996. 1: p. 150-153.
201. MacKay, C., et al., *Expanding roles in orthopaedic care: a comparison of physiotherapist and orthopaedic surgeon recommendations for triage.* J Eval Clin Pract, 2009. 15(1): p. 178-83.
202. Parker, B.R., et al., *Surgical Trends in the Treatment of Meniscal Tears: Analysis of Data From the American Board of Orthopaedic Surgery Certification Examination Database.* Am J Sports Med, 2016. 44(7): p. 1717-23.
203. Stensrud, S., M.A. Risberg, and E.M. Roos, *Knee function and knee muscle strength in middle-aged patients with degenerative meniscal tears eligible for arthroscopic partial meniscectomy.* Br J Sports Med, 2014. 48(9): p. 784-8.
204. Hawker, G., et al., *Perspectives of Canadian Stakeholders on Criteria for Appropriateness for Total Joint Arthroplasty in Patients With Hip and Knee Osteoarthritis.* Arthritis Rheumatol, 2015. 67(7): p. 1806-15.
205. Skou, S.T., et al., *Criteria used when deciding on eligibility for total knee arthroplasty - Between thinking and doing.* Knee, 2016. 23(2): p. 300-5.
206. Deyle, G.D., *Direct access physical therapy and diagnostic responsibility: the risk-to-benefit ratio.* J Orthop Sports Phys Ther, 2006. 36(9): p. 632-4.
207. Hegedus, E.J. and J. Moody, *Clinimetrics corner: the many faces of selection bias.* J Man Manip Ther, 2010. 18(2): p. 69-73.
208. Oberlander, M.A., R.M. Shalvoy, and J.C. Hughston, *The accuracy of the clinical knee examination documented by arthroscopy. A prospective study.* Am J Sports Med, 1993. 21(6): p. 773-8.
209. O'Shea, K.J., et al., *The diagnostic accuracy of history, physical examination, and radiographs in the evaluation of traumatic knee disorders.* Am J Sports Med, 1996. 24(2): p. 164-7.
210. Rose, N.E. and S.M. Gold, *A comparison of accuracy between clinical examination and magnetic resonance imaging in the diagnosis of meniscal and anterior cruciate ligament tears.* Arthroscopy, 1996. 12(4): p. 398-405.
211. Kocabey, Y., et al., *The value of clinical examination versus magnetic resonance imaging in the diagnosis of meniscal tears and anterior cruciate ligament rupture.* Arthroscopy, 2004. 20(7): p. 696-700.
212. Felli, L., et al., *Comparison of clinical, MRI and arthroscopic assessments of chronic ACL injuries, meniscal tears and cartilage defects.* Musculoskelet Surg, 2016. 100(3): p. 231-238.
213. Leblanc, M.C., et al., *Diagnostic accuracy of physical examination for anterior knee instability: a systematic review.* Knee Surg Sports Traumatol Arthrosc, 2015. 23(10): p. 2805-13.
214. Crawford, R., et al., *Magnetic resonance imaging versus arthroscopy in the diagnosis of knee pathology, concentrating on meniscal lesions and ACL tears: a systematic review.* Br Med Bull, 2007. 84: p. 5-23.
215. Thorlund, J.B., et al., *Arthroscopic surgery for degenerative knee: systematic review and meta-analysis of benefits and harms.* Br J Sports Med, 2015. 49(19): p. 1229-35.

216. Katz, J.N., et al., *Surgery versus physical therapy for a meniscal tear and osteoarthritis.* N Engl J Med, 2013. 368(18): p. 1675-84.
217. Sihvonen, R., et al., *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): p. 449-55.
218. Adelani, M.A., et al., *The Use of MRI in Evaluating Knee Pain in Patients Aged 40 Years and Older.* J Am Acad Orthop Surg, 2016. 24(9): p. 653-9.
219. Bhattacharyya, T., 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-A(1): p. 4-9.
220. Blyth, M., et al., *Diagnostic accuracy of the Thessaly test, standardised clinical history and other clinical examination tests (Apley's, McMurray's and joint line tenderness) for meniscal tears in comparison with magnetic resonance imaging diagnosis.* Health Technol Assess, 2015. 19(62): p. 1-62.
221. Goossens, P., et al., *Validity of the thessaly test in evaluating meniscal tears compared with arthroscopy: a diagnostic accuracy study.* J Orthop Sports Phys Ther, 2015. 45(1): p. 18-24.
222. Lowery, D.J., et al., *A clinical composite score accurately detects meniscal pathology.* Arthroscopy, 2006. 22(11): p. 1174-9.
223. Karachalios, T., et al., *Diagnostic accuracy of a new clinical test (the Thessaly test) for early detection of meniscal tears.* J Bone Joint Surg Am, 2005. 87(5): p. 955-62.
224. Ahn, J.H., S.H. Jeong, and H.W. Kang, *Risk Factors of False-Negative Magnetic Resonance Imaging Diagnosis for Meniscal Tear Associated With Anterior Cruciate Ligament Tear.* Arthroscopy, 2016. 32(6): p. 1147-54.
225. Dufka, F.L., et al., *Accuracy of MRI evaluation of meniscus tears in the setting of ACL injuries.* Knee, 2016. 23(3): p. 460-4.
226. Maricar, N., et al., *Interobserver and Intraobserver Reliability of Clinical Assessments in Knee Osteoarthritis.* J Rheumatol, 2016. 43(12): p. 2171-2178.
227. Niu, N.N., et al., *Development and preliminary validation of a meniscal symptom index.* Arthritis Care Res (Hoboken), 2011. 63(2): p. 208-15.
228. MacFarlane, L.A., et al., *Associations among meniscal damage, meniscal symptoms and knee pain severity.* Osteoarthritis Cartilage, 2016.
229. Galli, M. and E. Marzetti, *Accuracy of McMurray and Joint Line Tenderness Tests in the Diagnosis of Chronic Meniscal Tears: An Ad Hoc Receiver Operator Characteristic Analysis Approach.* Arch Phys Med Rehabil, 2016.
230. Cross, M., et al., *The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study.* Ann Rheum Dis, 2014. 73(7): p. 1323-30.
231. Skou, S.T. and E.M. Roos, *Good Life with osteoArthritis in Denmark (GLA:D): evidence-based education and supervised neuromuscular exercise delivered by certified physiotherapists nationwide.* BMC Musculoskelet Disord, 2017. 18(1): p. 72.
232. Decary, S., et al., *Diagnostic validity and triage concordance of a physiotherapist compared to physicians' diagnoses for common knee disorders.* BMC Musculoskelet Disord, 2017. 18(1): p. 445.

233. Sakellariou, G., et al., *EULAR recommendations for the use of imaging in the clinical management of peripheral joint osteoarthritis.* Ann Rheum Dis, 2017.
234. Peat, G., et al., *Estimating the probability of radiographic osteoarthritis in the older patient with knee pain.* Arthritis Rheum, 2007. 57(5): p. 794-802.
235. Decary, S., et al., *Diagnostic validity of combining history elements and physical examination tests for traumatic and degenerative symptomatic meniscal tears.* PM R, 2017.
236. Damen, J., et al., *The additional value of different radiographic views on identification of early radiographic hip and knee osteoarthritis and its progression: the CHECK (cohort hip and cohort knee) study.* Arthritis Care Res (Hoboken), 2017.
237. Silverwood, V., et al., *Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis.* Osteoarthritis Cartilage, 2015. 23(4): p. 507-15.
238. Tanamas, S., et al., *Does knee malalignment increase the risk of development and progression of knee osteoarthritis? A systematic review.* Arthritis Rheum, 2009. 61(4): p. 459-67.
239. Aggarwal, R., et al., *Distinctions between diagnostic and classification criteria?* Arthritis Care Res (Hoboken), 2015. 67(7): p. 891-7.
240. Karel, Y.H., et al., *Effect of routine diagnostic imaging for patients with musculoskeletal disorders: A meta-analysis.* Eur J Intern Med, 2015. 26(8): p. 585-95.
241. Skou, S.T., H. Thomsen, and O.H. Simonsen, *The value of routine radiography in patients with knee osteoarthritis consulting primary health care: a study of agreement.* Eur J Gen Pract, 2014. 20(1): p. 10-6.
242. Peat, G., et al., *Clinical features of symptomatic patellofemoral joint osteoarthritis.* Arthritis Res Ther, 2012. 14(2): p. R63.
243. Hart, H.F., et al., *Is body mass index associated with patellofemoral pain and patellofemoral osteoarthritis? A systematic review and meta-regression and analysis.* Br J Sports Med, 2016.
244. Lankhorst, N.E., et al., *Incidence, prevalence, natural course and prognosis of patellofemoral osteoarthritis: the Cohort Hip and Cohort Knee study.* Osteoarthritis Cartilage, 2016.
245. Stefanik, J.J., et al., *Diagnostic Performance of Clinical Examination Measures and Pain Presentation to Identify Patellofemoral Joint Osteoarthritis.* Arthritis Care Res (Hoboken), 2017.
246. Schiphof, D., et al., *Crepitus is a first indication of patellofemoral osteoarthritis (and not of tibiofemoral osteoarthritis).* Osteoarthritis Cartilage, 2014. 22(5): p. 631-8.
247. Stefanik, J.J., et al., *The diagnostic performance of anterior knee pain and activity-related pain in identifying knees with structural damage in the patellofemoral joint: the Multicenter Osteoarthritis Study.* J Rheumatol, 2014. 41(8): p. 1695-702.
248. Van Ginckel, A., et al., *Location of knee pain in medial knee osteoarthritis: patterns and associations with self-reported clinical symptoms.* Osteoarthritis Cartilage, 2016. 24(7): p. 1135-42.

249. Hensor, E.M., et al., *Toward a clinical definition of early osteoarthritis: onset of patient-reported knee pain begins on stairs. Data from the osteoarthritis initiative.* Arthritis Care Res (Hoboken), 2015. 67(1): p. 40-7.
250. Case, R., et al., *Prodromal symptoms in knee osteoarthritis: a nested case-control study using data from the Osteoarthritis Initiative.* Osteoarthritis Cartilage, 2015. 23(7): p. 1083-9.
251. Lo, G.H., et al., *Subjective Crepitus as a Risk Factor for Incident Symptomatic Knee Osteoarthritis: Data from the Osteoarthritis Initiative.* Arthritis Care Res (Hoboken), 2017.
252. van der Heijden, R.A., et al., *No Difference on Quantitative Magnetic Resonance Imaging in Patellofemoral Cartilage Composition Between Patients With Patellofemoral Pain and Healthy Controls.* Am J Sports Med, 2016. 44(5): p. 1172-8.
253. Harris, J.D., et al., *Sensitivity of magnetic resonance imaging for detection of patellofemoral articular cartilage defects.* Arthroscopy, 2012. 28(11): p. 1728-37.
254. Pihlajamaki, H.K., et al., *Reliability of clinical findings and magnetic resonance imaging for the diagnosis of chondromalacia patellae.* J Bone Joint Surg Am, 2010. 92(4): p. 927-34.
255. Hinman, R.S., et al., *Is patellofemoral osteoarthritis common in middle-aged people with chronic patellofemoral pain?* Arthritis Care Res (Hoboken), 2014. 66(8): p. 1252-7.
256. Collins, N.J., et al., *Pain During Prolonged Sitting Is a Common Problem in Persons With Patellofemoral Pain.* J Orthop Sports Phys Ther, 2016. 46(8): p. 658-63.
257. Holmes, S.W., Jr. and W.G. Clancy, Jr., *Clinical classification of patellofemoral pain and dysfunction.* J Orthop Sports Phys Ther, 1998. 28(5): p. 299-306.
258. Witvrouw, E., et al., *Patellofemoral pain: consensus statement from the 3rd International Patellofemoral Pain Research Retreat held in Vancouver, September 2013.* Br J Sports Med, 2014. 48(6): p. 411-4.
259. Whiting, P.F., et al., *Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies.* BMC Med Res Methodol, 2006. 6: p. 9.
260. Reitsma JB, R.A., Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ,. *Chapter 9: Assessing methodological quality.* In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration. Available from: http://srdta.cochrane.org/. 2009.
261. Rutjes, A.W., et al., *Evidence of bias and variation in diagnostic accuracy studies.* CMAJ, 2006. 174(4): p. 469-76.
262. Leeflang, M.M., et al., *Variation of a test's sensitivity and specificity with disease prevalence.* CMAJ, 2013. 185(11): p. E537-44.
263. Skou, S.T., E.M. Roos, and M.B. Laursen, *A Randomized, Controlled Trial of Total Knee Replacement.* N Engl J Med, 2016. 374(7): p. 692.
264. Zou., H., Hastie., *Regularization and variable selection via the elastic net.* Journal of the Royal Statistical Society, 2005. Series B (Statistical Methodology)(67(2)): p. p. 301-320.
265. Dionne, C.E., et al., *A clinical return-to-work rule for patients with back pain.* CMAJ, 2005. 172(12): p. 1559-67.

266. Harrell, F.E., Jr., et al., *Regression models for prognostic prediction: advantages, problems, and suggested solutions.* Cancer Treat Rep, 1985. 69(10): p. 1071-77.
267. Breinman, L., *Random forest.* Machine Learning, 2001. 45:5–32.
268. Kotti, M., et al., *Detecting knee osteoarthritis and its discriminating parameters using random forests.* Med Eng Phys, 2017. 43: p. 19-29.
269. Mezghani, N., et al., *Mechanical biomarkers of medial compartment knee osteoarthritis diagnosis and severity grading: Discovery phase.* J Biomech, 2017. 52: p. 106-112.
270. Bordini, B.J., A. Stephany, and R. Kliegman, *Overcoming Diagnostic Errors in Medical Practice.* J Pediatr, 2017.
271. Dagan, N., et al., *External validation and comparison of three prediction tools for risk of osteoporotic fractures using data from population based electronic health records: retrospective cohort study.* BMJ, 2017. 356: p. i6755.
272. Riddle, D.L., et al., *External Validation of a Prognostic Model for Predicting Nonresponse Following Knee Arthroplasty.* J Arthroplasty, 2017. 32(4): p. 1153-1158 e1.
273. Stiell, I.G. and G.A. Wells, *Methodologic standards for the development of clinical decision rules in emergency medicine.* Ann Emerg Med, 1999. 33(4): p. 437-47.
274. Stanton, T.R., et al., *Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions.* Phys Ther, 2010. 90(6): p. 843-54.
275. Mintken, P.E., et al., *Examination of the Validity of a Clinical Prediction Rule to Identify Patients With Shoulder Pain Likely to Benefit From Cervicothoracic Manipulation.* J Orthop Sports Phys Ther, 2017. 47(4): p. 252-260.
276. Tsien, C.L., et al., *Using classification tree and logistic regression methods to diagnose myocardial infarction.* Stud Health Technol Inform, 1998. 52 Pt 1: p. 493-7.
277. Selfe, J., et al., *Are there three main subgroups within the patellofemoral pain population? A detailed characterisation study of 127 patients to help develop targeted intervention (TIPPs).* Br J Sports Med, 2016. 50(14): p. 873-80.
278. Roshanov, P.S., et al., *Can computerized clinical decision support systems improve practitioners' diagnostic test ordering behavior? A decision-maker-researcher partnership systematic review.* Implement Sci, 2011. 6: p. 88.
279. Whiting, P., et al., *A review identifies and classifies reasons for ordering diagnostic tests.* J Clin Epidemiol, 2007. 60(10): p. 981-9.
280. Graham, I.D., et al., *Lost in knowledge translation: time for a map?* J Contin Educ Health Prof, 2006. 26(1): p. 13-24.
281. Hoffmann, T.C., V.M. Montori, and C. Del Mar, *The connection between evidence-based medicine and shared decision making.* JAMA, 2014. 312(13): p. 1295-6.
282. Dogba MJ, M.M., Stacey D, Brière N, Légaré F., *The Evolution of an Interprofessional Shared Decision-Making Research Program: Reflective Case Study of an Emerging Paradigm.* International Journal of Integrated Care., 2016. 6(3):4. DOI: http://doi.org/10.5334/ijic.2212.

283. Legare, F., et al., *Validating a conceptual model for an inter-professional approach to shared decision making: a mixed methods study.* J Eval Clin Pract, 2011. 17(4): p. 554-64.
284. Thompson-Leduc, P., et al., *Shared decision-making behaviours in health professionals: a systematic review of studies based on the Theory of Planned Behaviour.* Health Expect, 2014.
285. Stacey, D., et al., *Decision aids for people facing health treatment or screening decisions.* Cochrane Database Syst Rev, 2014(1): p. CD001431.
286. Stacey, D., et al., *Impact of patient decision aids on appropriate and timely access to hip or knee arthroplasty for osteoarthritis: a randomized controlled trial.* Osteoarthritis Cartilage, 2016. 24(1): p. 99-107.
287. Decary, S., et al., *Decisional conflict screening for a diversity of primary care decisions. Are we SURE yet?* J Clin Epidemiol, 2017.
288. Tousignant-Laflamme, Y., et al., *Does shared decision making results in better health related outcomes for individuals with painful musculoskeletal disorders? A systematic review.* Journal of Manual & Manipulative Therapy, 2017: p. 1-7.
289. Traeger, A.C., R.N. Moynihan, and C.G. Maher, *Wise choices: making physiotherapy care more valuable.* J Physiother, 2017. 63(2): p. 63-65.
290. Born, K.B., et al., *Engaging patients and the public in Choosing Wisely.* BMJ Qual Saf, 2017. 26(8): p. 687-691.

Annexes

1. Detailed search strategy with keywords and descriptors for article 1

PUBMED		
#	Search details	Results
1	(((((reliab* OR reproduc* OR validity OR accuracy OR variability OR "predictive value")) OR (interobserver* OR intertester* OR interrater* OR interexamin* OR intraobserver* OR intratester* OR intrarater* OR intraexamin* OR "observer variation" [Mesh])) OR ((inter OR intra) AND (rater* OR examin* OR tester* OR observer*)))) AND ((((patholog* OR lesion* OR ruptur$ OR torn OR tear* OR trauma OR traumas OR effusion* OR instabilit* OR laxity OR injur* OR disorder* OR syndrome OR pain OR osteoarthr* OR alignment)) AND (knee* OR "anterior cruciate ligament*" OR "posterior cruciate ligament*" OR "medial collateral ligament*" OR menisc* OR patell* OR "knee" [Mesh] OR "knee joint" [Mesh] OR "ligaments, articular" [Mesh] OR "patella" [Mesh])) AND ("physical examination" OR "physical examination" [Mesh] OR "clinical assessment" OR test OR tests OR manual* OR manoeuv* OR maneuv* OR palpation OR examiner*))	5345

MEDLINE		
#	Search details	Results
1	(patholog$ or lesion$ or ruptur$ or torn or tear$ or trauma$ or effusion$ or instabilit$ or laxity or injur$ or disorder$ or syndrome or pain or osteoarthr$ or alignment). mp.	4 840 034
2	(knee$ or "anterior cruciate ligament$" or "posterior cruciate ligament*" or "medial collateral ligament*" or menisc$ or patell$). mp. or knee/ or exp knee joint/ or exp ligaments, articular/ or patella/	144 958
3	("physical examination" or "clinical assessment" or examiner$ or test$ or manual$ or man?euv$ or palpation). mp. or exp physical examination/	4 110 216
4	(reliab$ or reproduc$ or validity or accuracy or variability). mp.	1 312 329
5	((inter or intra) adj (rater$ or examin$ or tester$ or observer$)). mp.	12 210
6	(interobserver$ or intertester$ or interrater$ or interexamin$ or intraobserver$ or intratester$ or intrarater$ or intraexamin$). mp. or observer variation/	46 037
7	4 or 5 or 6	1 325 926
8	1 and 2 and 3 and 7	3935

	CINAHL	
#	Search details	Results
1	pathолog* or lesion* or ruptur* or torn or tear* or trauma$ or effusion* or instabilit* or laxity or injur* or disorder* or syndrome or pain or osteoarthr* or alignment	965 537
2	knee or knees or “anterior cruciate ligament*” or “posterior cruciate ligament*” or “medial collateral ligament*” or menisc$ or patell* or MH “Knee” or MH “Knee Joint+” or MH “Patella” OR MH “Patellar Ligament” or MH “Anterior Cruciate Ligament” or MH “Medial Collateral Ligament, knee” or MH “Posterior Cruciate Ligament”	44 735
3	“physical examination” or “clinical assessment” or examiner* or test* or manual* or maneuv* or manoeuv* or palpation or MH “Physical Examination”	700 651
4	reliab* or reproduc* or validity or accuracy or variability	201 057
5	MH “Reliability” or MH “Reliability and Validity” or MH “Interrater reliability” or MH “Intrarater reliability” or “Test-Retest Reliability”	48 676
6	((inter or intra) N (rater* or examin* or tester* or observer*	292 870
7	interobserver* or intertester* or interrater* or interexamin* or intraobserver* or intratester* or intrarater* or intraexamin*	24 654
8	S4 OR S5 OR S6 OR S7	452 480
9	S1 AND S2 AND S3 AND S8	3 783

	EMBASE	
#	Search details	Results
1	(pathолog$ or lesion$ or ruptur$ or torn or tear$ or trauma$ or effusion$ or instabilit$ or laxity or injur$ or disorder$ or syndrome or pain or osteoarthr$ or alignment). mp.	6 587 026
2	(knee or knees or “ anterior cruciate ligament$ ” or “ posterior cruciate ligament$ ” or “ medial collateral ligament$ ” or menisc$ or patell$). mp. or knee/ or exp knee ligament/ or patella ligament/ or patella/ or knee meniscus/	176 966
3	(“ physical examination ” or “ clinical assessment ” or examiner$ or test$ or manual$ or man?euv$ or palpation). tw, kw. Or exp physical examination/	3 319 371

4	(reliab$ or reproduc$ or validity or accuracy or variability). mp.	1 846 828
5	((inter or intra) adj (rater$ or examin$ or tester$ or observer$)). mp.	17 804
6	(interobserver$ or intertester$ or interrater$ or interexamin$ or intraobserver$ or intratester$ or intrarater$ or intraexamin$). mp. or observer variation/	41 503
7	intrarater reliability/ or interrater reliability/ or test retest reliability/ or reliability/	112 517
8	4 or 5 or 6 or 7	1 856 756
9	1 and 2 and 3 and 8	3660

2. Detailed search strategy with keywords and descriptors for article 2

PUBMED

#	Search details	Results
1	(((((reliab* OR reproduc* OR validity OR accuracy OR variability OR "predictive value")) OR (interobserver* OR intertester* OR interrater* OR interexamin* OR intraobserver* OR intratester* OR intrarater* OR intraexamin* OR "observer variation" [Mesh])) OR ((inter OR intra) AND (rater* OR examin* OR tester* OR observer*)))) AND ((((patholog* OR lesion* OR ruptur$ OR torn OR tear* OR trauma OR traumas OR effusion* OR instabilit* OR laxity OR injur* OR disorder* OR syndrome OR pain OR osteoarthr* OR alignment)) AND (knee* OR "anterior cruciate ligament*" OR "posterior cruciate ligament*" OR "medial collateral ligament*" OR menisc* OR patell* OR "knee" [Mesh] OR "knee joint" [Mesh] OR "ligaments, articular" [Mesh] OR "patella" [Mesh])) AND ("physical examination" OR "physical examination" [Mesh] OR "clinical assessment" OR test OR tests OR manual* OR manoeuv* OR maneuv* OR palpation OR examiner*))	5345

MEDLINE

#	Search details	Results
1	(patholog$ or lesion$ or ruptur$ or torn or tear$ or trauma$ or effusion$ or instabilit$ or laxity or injur$ or disorder$ or syndrome or pain or osteoarthr$ or alignment). mp.	4 840 034
2	(knee$ or "anterior cruciate ligament$" or "posterior cruciate ligament*" or "medial collateral ligament*" or menisc$ or patell$). mp. or knee/ or exp knee joint/ or exp ligaments, articular/ or patella/	144 958
3	("physical examination" or "clinical assessment" or examiner$ or test$ or manual$ or man?euv$ or palpation). mp. or exp physical examination/	4 110 216
4	(reliab$ or reproduc$ or validity or accuracy or variability). mp.	1 312 329
5	((inter or intra) adj (rater$ or examin$ or tester$ or observer$)). mp.	12 210
6	(interobserver$ or intertester$ or interrater$ or interexamin$ or intraobserver$ or intratester$ or intrarater$ or intraexamin$). mp. or observer variation/	46 037
7	4 or 5 or 6	1 325 926
8	1 and 2 and 3 and 7	3935

CINAHL

#	Search details	Results
1	patholog* or lesion* or ruptur* or torn or tear* or trauma$ or effusion* or instabilit* or laxity or injur* or disorder* or syndrome or pain or osteoarthr* or alignment	965 537
2	knee or knees or "anterior cruciate ligament*" or "posterior cruciate ligament*" or "medial collateral ligament*" or menisc$ or patell* or MH "Knee" or MH "Knee Joint+" or MH "Patella" OR MH "Patellar Ligament" or MH "Anterior Cruciate Ligament" or MH "Medial Collateral Ligament, knee" or MH "Posterior Cruciate Ligament"	44 735

3	"physical examination" or "clinical assessment" or examiner* or test* or manual* or maneuv* or manoeuv* or palpation or MH "Physical Examination"	700 651
4	reliab* or reproduc* or validity or accuracy or variability	201 057
5	MH "Reliability" or MH "Reliability and Validity" or MH "Interrater reliability" or MH "Intrarater reliability" or "Test-Retest Reliability"	48 676
6	((inter or intra) N (rater* or examin* or tester* or observer*	292 870
7	interobserver* or intertester* or interrater* or interexamin* or intraobserver* or intratester* or intrarater* or intraexamin*	24 654
8	S4 OR S5 OR S6 OR S7	452 480
9	S1 AND S2 AND S3 AND S8	3 783

EMBASE

#	Search details	Results
1	(patholog$ or lesion$ or ruptur$ or torn or tear$ or trauma$ or effusion$ or instabilit$ or laxity or injur$ or disorder$ or syndrome or pain or osteoarthr$ or alignment). mp.	6 587 026
2	(knee or knees or " anterior cruciate ligament$ " or " posterior cruciate ligament$ " or " medial collateral ligament$ " or menisc$ or patell$). mp. or knee/ or exp knee ligament/ or patella ligament/ or patella/ or knee meniscus/	176 966
3	(" physical examination " or " clinical assessment " or examiner$ or test$ or manual$ or man?euv$ or palpation). tw, kw. Or exp physical examination/	3 319 371
4	(reliab$ or reproduc$ or validity or accuracy or variability). mp.	1 846 828
5	((inter or intra) adj (rater$ or examin$ or tester$ or observer$)). mp.	17 804
6	(interobserver$ or intertester$ or interrater$ or interexamin$ or intraobserver$ or intratester$ or intrarater$ or intraexamin$). mp. or observer variation/	41 503
7	intrarater reliability/ or interrater reliability/ or test retest reliability/ or reliability/	112 517
8	4 or 5 or 6 or 7	1 856 756
9	1 and 2 and 3 and 8	3660

3. Liste des variables de l'anamnèse

Patients' socio-demographic caracteristics
Gender
Age
Marital Status
Household living status
Education level
Employment status
Work physical demands (seated, light/moderate/hard)
Personal income
Family income
Absenteeism from work (number of days)
Use and type of walking aids
Comorbidities
History taking elements and clinical measures
Affected side (left/right/bilateral)
First time consultation (yes/no/uncertain)
Traumatic or progressive onset of symptoms
Duration of knee symptoms (weeks/months/more than a year)
Knee pain location (anterior/posterior/left/right/diffuse)
Sign of neuropathy (stabbing or shooting pain?, tingling in feet?)
Duration of walking (minutes to more than an hour)
Mechanism of injury (if needed, not mutually exclusive): falling on knee (weight-bearing), direct external force/blow to the knee, trauma following a jump landing, pivoting trauma, foot/leg stuck/blocked on the ground or "popping" sensation during trauma.
Immediate or late/delayed pain apparition after trauma (if needed) (no/yes immediately/yes some time after but before 24 h/yes after 24h/uncertain)
Ability/inability to pursue activities after trauma (if needed) (yes/no/uncertain)
Immediate or late/delayed knee effusion after trauma (if needed) (no/yes less than 2 hours after trauma/yes more than 2h after trauma/uncertain).
Weight (in kilograms)
Height (in meters)
Passive and active knee range of motion in flexion or extension (in degrees)

4. Formulaire de l’anamnèse

Évaluation Subjective

1. Introduction

Nous vous remercions d'avoir accepté de participer au projet de recherche intitulé "Validité et fidélité des éléments de l'histoire et des tests physiques pour le diagnostic des pathologies communes au genou". L'objectif du projet est de déterminer les éléments de l'examen clinique les plus pertinents pour le médecin et les combiner pour qu'il puisse éventuellement être en mesure de faire un diagnostic valide sans avoir recours à l'imagerie médicale. Ceci pourrait améliorer l'accessibilité aux soins pour une problématique au genou et diminuer les coûts de santé.

Ce projet a été préalablement accepté par le comité d'éthique de la recherche. Toutes les données que vous fournirez dans le présent formulaire demeureront confidentielles et sécurisées sur les serveurs de l'équipe de recherche de la clinique orthopédique de l'hôpital Maisonneuve-Rosemont.

Le présent formulaire contient 4 sections et devrait prendre environ 30 minutes à compléter. N'hésitez pas à demander de l'aide au professionnel de recherche au besoin.

1. Date:

JJ MM AAAA

Indiquer la date de l'évaluation ☐ / ☐ / ☐

2. Numéro du participant:

Indiquez le numéro du participant. ☐

2. Données Socio-Démographiques

Les questions suivantes portent sur vous et votre problème de santé.

Voici quelques questions d’ordre général qui nous permettront de comparer votre état de santé à celui d’autres personnes ayant des caractéristiques semblables aux vôtres.

1. Quel est votre sexe?

◯ Homme ◯ Femme

2. Quel est votre âge?

☐

3. Quel est votre état civil?

◯ Célibataire/ jamais marié(e) ◯ Marié(e) / union de fait ◯ Divorcé(e)/ séparé(e) ◯ Veuf /veuve

Page 1

Évaluation Subjective

4. Présentement combien de personnes habitent avec vous ?
(incluant votre conjoint le cas échéant)

5. Quel est le plus haut niveau de scolarité que vous avez complété?

- ◯ Partie du primaire
- ◯ Primaire complété
- ◯ Partie du secondaire
- ◯ Secondaire complété
- ◯ Partie du collégial (incluant : école de métiers, collège commercial privé, institut technique, école de sciences infirmière, école normale)
- ◯ Collégial terminé (diplôme ou certificat obtenu) (incluant : école de métiers, collège commercial privé, institut technique, école de sciences infirmière, école normale)
- ◯ Études partielles à l'université
- ◯ Universitaire (1er cycle, Baccalauréat acquis)
- ◯ Universitaire (2e et 3e cycle en cours ou terminé)

6. Présentement quelle est votre situation d'emploi ?

- ◯ J'ai un ou des emplois
- ◯ Je suis en congé de maladie
- ◯ Je suis retraité
- ◯ Je n'ai pas d'emploi

7. Choisissez la catégorie représentant le mieux les exigences physiques de votre travail (profession ou travail à la maison):

- ◯ Travail en position assise (ex. : bureau)
- ◯ Travail physique léger nécessitant de la marche légère continue ou intermittente
- ◯ Travail physique lourd (ex : construction, marche modérée à importante)
- ◯ Je ne travaille pas

8. Quel était approximativement votre revenu personnel total l'AN DERNIER avant déductions d'impôts?
Cette question n'est pas obligatoire, sélectionnez "Ne désire pas répondre" au besoin.

Évaluation Subjective

9. Quel était approximativement le revenu familial total l'AN DERNIER avant déductions d'impôts?
Cette question n'est pas obligatoire, sélectionnez "Ne désire pas répondre" au besoin.

10. Combien de jours de travail avez-vous manqué à cause de votre douleur durant les 12 derniers mois?

11. Quel type d'aide à la marche utilisez-vous présentement?

- ◯ Canne
- ◯ Marchette
- ◯ Déambulateur
- ◯ Béquille
- ◯ Fauteuil roulant
- ◯ Quadriporteur
- ◯ Je n'utilise aucune aide à la marche

12. Au cours de votre vie, votre médecin vous a-t-il déjà dit que vous aviez:

	Oui/Non/Incertain	En quelle année ce problème de santé a-t-il été diagnostiqué pour la 1ère fois	Présentement, ce problème de santé vous limite-t-il dans l'une ou l'autre de vos activités régulières?
Arthrose ou arthrite			
Maladies du coeur			
Tension artérielle élevée (haute pression)			
Diabète			

Autre (veuillez préciser)

Merci pour vos réponses.
Veuillez poursuivre la 2e section à la page suivante.

3. Évaluation subjective auto-administrée

La prochaine série de question vise à décrire plus précisément les raisons de votre consultation.

Évaluation Subjective

1. À quel genou avez-vous de la douleur ou éprouvez-vous de la difficulté?

◯ Droit ◯ Gauche ◯ Droit et Gauche

2. Est-ce la première fois que vous consultez pour un problème au(x) genou(x)?

◯ Oui ◯ Non ◯ Ne sait pas

3. Le problème pour lequel vous consultez est-il relié à un évènement traumatique (accident) ou à une apparition progressive?

◯ Traumatique (accident)

◯ Apparition progressive

4. Depuis quand avez-vous de la douleur ou de la difficulté avec votre/vos genou(x)?

5. Pouvez-vous situer la douleur?

◯ À l'intérieur du genou

◯ À l'extérieur du genou

◯ À l'avant du genou (rotule)

◯ À l'arrière du genou

◯ Diffuse (ne peut être ciblée, à plusieurs endroits)

◯ Autre

6. Ressentez-vous une douleur de type élancement ou brulement dans la jambe?

◯ Oui ◯ Non ◯ Incertain

7. Ressentez-vous des picotements dans la jambe ou les orteils?

◯ Oui ◯ Non ◯ Incertain

8. Combien de temps êtes-vous capable de marcher sans douleur?

◯ Je ne suis pas capable de marcher sans douleur

◯ Moins de 15 minutes

◯ 15-30 minutes

◯ 30-45 minutes

◯ 45-60 minutes

◯ Plus de 60 minutes

◯ Incertain

Évaluation Subjective

Les dernières questions portent sur le mécanisme de blessure si vous avez eu un traumatisme (accident) à votre genou. Si votre problématique est d'apparition progressive, s'il vous plait sélectionnez "Ma problématique n'est pas survenue suite à un accident" pour les dernières questions.

9. Êtes-vous tombé sur votre genou lors de l'accident?

◯ Oui ◯ Non ◯ Incertain ◯ Ma problématique n'est pas survenue suite à un accident

10. Avez-vous frappé votre genou lors de l'accident?

◯ Oui ◯ Non ◯ Incertain ◯ Ma problématique n'est pas survenue suite à un accident

11. Est-ce que la douleur est survenue suite à un saut?

◯ Oui ◯ Non ◯ Incertain ◯ Ma problématique n'est pas survenue suite à un accident

12. Est-ce que votre genou s'est tordu (rotation) lors de l'accident?

◯ Oui ◯ Non ◯ Incertain ◯ Ma problématique n'est pas survenue suite à un accident

13. Est-ce que votre pied ou votre jambe est resté coincé au sol lors de l'accident?

◯ Oui ◯ Non ◯ Incertain ◯ Ma problématique n'est pas survenue suite à un accident

14. Avez-vous entendu un "pop" ou "clic" au genou lors de l'accident?

◯ Oui ◯ Non ◯ Incertain ◯ Ma problématique n'est pas survenue suite à un accident

Évaluation Subjective

15. Y a-t-il eu de la douleur au moment de l'accident?

- ◯ Non
- ◯ Oui, immédiatement
- ◯ Oui, quelques heures après
- ◯ Oui, avant 24h
- ◯ Oui, après 24h
- ◯ Incertain
- ◯ Ma problématique n'est pas survenue suite à un accident

16. Étiez-vous en mesure de poursuivre vos activités suite à l'accident?

◯ Oui ◯ Non ◯ Incertain ◯ Ma problématique n'est pas survenue suite à un accident

17. Est-ce qu'un gonflement est survenu suite à l'accident?

- ◯ Non
- ◯ Oui, moins de 2 h suivant l'incident
- ◯ Oui, plus de 2 h suivant l'incident, mais avant 24h
- ◯ Ma problématique n'est pas survenue suite à un accident

Merci pour vos réponses.
Veuillez poursuivre la 3e section à la page suivante.

4. Questionnaire du genou KOOS

INSTRUCTIONS
Cette section vous demande votre opinion sur votre genou. Il nous permettra de mieux connaître ce que vous ressentez et ce que vous êtes capable de faire dans vos activités de tous les jours.

Répondez à chaque question. Veuillez cocher une seule case par question. En cas de doute, cochez la case qui vous semble la plus adaptée à votre cas.

Ces questions concernent vos symptômes au cours des huit derniers jours.

1. Est-ce que votre genou gonfle?

◯ Jamais ◯ Rarement ◯ Parfois ◯ Souvent ◯ Tout le temps

5. Questionnaire Knee Injury and Osteoarthritis Outcome Score (KOOS)

QUESTIONNAIRE DE GENOU KOOS

DATE: _____ DATE DE NAISSANCE: _____

NOM: _____

INSTRUCTIONS

Ce questionnaire vous demande votre opinion sur votre genou. Il nous permettra de mieux connaître ce que vous ressentez et ce que vous êtes capable de faire dans votre activité de tous les jours.

Répondez à chaque question. Veuillez cocher une seule case par question. En cas de doute, cochez la case qui vous semble la plus adaptée à votre cas.

Symptomes

Ces questions concernent vos symptômes au cours des **huit derniers jours.**

S1. Est-ce que votre genou gonfle?

Jamais	Rarement	Parfois	Souvent	Tout le temps
☐	☐	☐	☐	☐

S2. Ressentez-vous des ou entendez-vous des craquements ou n'importe quel autre type de bruit en bougeant le genou?

Jamais	Rarement	Parfois	Souvent	Toujours
☐	☐	☐	☐	☐

S3. Est-ce que votre genou accroche ou se bloque en bougeant?

Jamais	Rarement	Parfois	Souvent	Toujours
☐	☐	☐	☐	☐

S4. Pouvez-vous étendre votre genou complètement?

Toujours	Souvent	Parfois	Rarement	Jamais
☐	☐	☐	☐	☐

S5. Pouvez-vous plier votre genou complètement?

Toujours	Souvent	Parfois	Rarement	Jamais
☐	☐	☐	☐	☐

Raideur

Ces questions concernent la raideur de votre genou au cours des **huit derniers jours**. La raideur est la sensation d'avoir du mal à bouger le genou.

S6. Le matin au réveil, la raideur de votre genou est:

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

S7. Après être resté(e) assis(e), couché(e), ou au repos pendant la journée, la raideur de votre genou est:

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

Douleur

P1. Avez-vous souvent mal au genou?

Jamais	Une fois par mois	Une fois par semaine	Tous les jours	Tout le temps
☐	☐	☐	☐	☐

Au cours des **huit derniers jours,** quelle a été l'importance de votre douleur du genou en faisant les activités suivantes?

P2. En tournant, pivotant sur votre jambe

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

P3. En étendant complètement le genou

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

P4. En pliant complètement le genou

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

P5. En marchant sur un terrain plat

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

P6. En montant ou en descendant les escaliers

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

P7. Au lit la nuit

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

P8. En restant assis(e) ou couché(e)

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

P9. En restant debout

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

Fonction, vie quotidienne

Les questions suivantes concernent ce que vous êtes capable de faire. Au cours des **huit derniers jours**, quelle a été votre difficulté pour chacune des activités suivantes?

A1. Descendre les escaliers

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A2. Monter les escaliers

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A3. Vous relever d'une position assise

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A4. Rester debout

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A5. Vous pencher en avant pour ramasser un objet

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A6. Marcher sur un terrain plat

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A7. Monter ou descendre de voiture

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A8. Faire vos courses

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A9. Mettre vos chaussettes ou vos collants

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A10. Sortir du lit

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A11. Enlever vos chaussettes ou vos collants

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A12. Vous retourner ou garder le genou dans la même position en étant couché(e)

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A13. Entrer ou sortir d'une baignoire

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A14. Rester assis(e)

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A15. Vous asseoir ou vous relever des toilettes

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A16. Faire de gros travaux ménagers (déplacer des objets lourds, récurer les sols,…)

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

A17. Faire des petits travaux ménagers (faire la cuisine, faire la poussière,…).

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

Activités, sport et loisirs

Les questions suivantes concernent ce que vous êtes capable de faire au cours d'autres activités. Au cours des **huit derniers jours**, quelle a été votre difficulté pour les activités suivantes?

SP1. Rester accroupi(e)

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

SP2. Courir

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

SP3. Sauter

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

SP4. Tourner, pivoter sur votre jambe

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

SP5. Rester à genoux

Absente	Légère	Modérée	Forte	Extrême
☐	☐	☐	☐	☐

Qualité de vie

Q1. Pensez-vous souvent à votre problème de genou?

Jamais	Une fois par mois	Une fois par semaine	Tous les jours	Tout le temps
☐	☐	☐	☐	☐

Q2. Avez-vous modifié votre façon de vivre pour éviter les activités qui pourraient aggraver votre problème de genou?

Pas du tout	Un peu	Modérément	Beaucoup	Totalement
☐	☐	☐	☐	☐

Q3. Est-ce qu'un manque de confiance dans votre genou vous gêne?

Pas du tout	Un peu	Modérément	Beaucoup	Totalement
☐	☐	☐	☐	☐

Q4. Finalement, êtes-vous gêné(e) par votre genou?

Pas du tout	Un peu	Modérément	Beaucoup	Extrêmement
☐	☐	☐	☐	☐

******Merci beaucoup d'avoir répondu à ce questionnaire******

6. Questionnaire Kessler-6

Ne rien écrire dans la marge

K6+

Questionnaire remis par :_____
Identifiant :
|__|__|__|__|__|__|__|__|__|__|

Date de réponse __ __ / __ __ / __ __ __ __

Étiquette si disponible	Identifiant patient ou client :
Nom :	
Prénom(s) :	
Date de naissance : __ __ / __ __ / __ __ __ __	Sexe : M ☐1 F ☐2
Adresse :	

SANTÉ MENTALE

Les questions suivantes portent sur la façon dont vous vous êtes senti(e) au cours des **30 derniers jours**. Pour chaque question, veuillez entourer le numéro de la réponse correspondant le mieux au nombre de fois où vous avez éprouvé ce sentiment.

Q1.	Au cours des 30 derniers jours, combien de fois avez-vous eu le sentiment...	Tout le temps	La plupart du temps	Quel-quefois	Rarement	Jamais
a)	…d'être nerveux/nerveuse ?	1	2	3	4	5
b)	…d'être désespéré(e) ?	1	2	3	4	5
c)	…d'être agité(e) ou incapable de tenir en place ?	1	2	3	4	5
d)	…d'être tellement déprimé(e) que rien ne pouvait vous remonter le moral ?	1	2	3	4	5
e)	…que tout vous demandait un effort ?	1	2	3	4	5
f)	…de n'être bon(ne) à rien ?	1	2	3	4	5

SANTÉ MENTALE

QUESTIONNAIRE D'AUTO-ÉVALUATION K6+(1/2)

TSVP

K6-SA - France/French - Version of 24 Oct 11 - Mapi Institute.
ID6310 / K6-SA_AU1.0_fre-FR.doc

Q2. Les questions précédentes portaient sur des sentiments que vous avez pu éprouver au cours des 30 derniers jours. Dans l'ensemble, au cours des 30 derniers jours, avez-vous éprouvé ces sentiments plus souvent, à peu près aussi souvent ou moins souvent que d'habitude ? (Si vous n'éprouvez jamais ces sentiments, entourez « 4 ».)

Plus souvent que d'habitude				**Moins souvent que d'habitude**		
Beaucoup plus souvent	**Plutôt plus souvent**	**Un peu plus souvent**	**Aussi souvent que d'habitude**	**Un peu moins souvent**	**Plutôt moins souvent**	**Beaucoup moins souvent**
1	**2**	**3**	**4**	**5**	**6**	**7**

Les questions suivantes portent sur les conséquences que ces sentiments ont pu avoir sur vous au cours des 30 derniers jours. Vous n'avez pas besoin d'y répondre si vous avez répondu « Jamais » **aux six questions** de la Q1.

Ne rien écrire dans la marge

Q3. Au cours des 30 derniers jours, pendant combien de jours avez-vous été totalement incapable de travailler ou d'accomplir vos activités habituelles en raison de ces sentiments ?

_____ **(Nombre de jours)**

Q4. **Sans compter les jours indiqués dans la réponse à la question précédente**, pendant combien de jours au cours des 30 derniers n'avez-vous pu faire que la moitié, voire moins, des choses que vous auriez normalement été capable de faire si vous n'aviez pas éprouvé ces sentiments ?

_____ **(Nombre de jours)**

Q5. Au cours des 30 derniers jours, combien de fois avez-vous consulté un médecin ou un autre professionnel de santé à propos de ces sentiments ?

_____ **(Nombre de fois)**

	Tout le temps	**La plupart du temps**	**Quel-quefois**	**Rare-ment**	**Jamais**
Q6. Au cours des 30 derniers jours, est-il arrivé que ces sentiments soient principalement provoqués par des problèmes de santé physique ?	**1**	**2**	**3**	**4**	**5**

QUESTIONNAIRE D'AUTO-ÉVALUATION K6+(2/2)

Merci d'avoir répondu à ce questionnaire.

7. Guide de standardisation de l'examen physique [77, 78, 113]

Tests	Description des tests et interprétation des résultats
Examen palpatoire	
Examen palpatoire des différentes structures du genou.	Effectuez un examen palpatoire des structures suivantes : tendon quadricipital, tendon patellaire, interlignes articulaire médiales et latérales, apex patellaire, facettes patellaires médiales et latérales, tête du péroné, patte d'oie, creux poplité, tendons sembraneux, semi-tendineux et biceps fémoral. Comparez entre le genou affecté et sain afin de vérifier que la pression excessive ne cause pas un faux positif. Pour la palpation de l'interligne articulaire, s'assurer que le genou est à 90 degrés de flexion pour standardiser ce test. **Positif :** le participant dit éprouver de la douleur/inconfort lors de la pression sur la structure. La douleur/inconfort doit reproduire celle pour laquelle le participant a consulté (à différencier de la douleur due à la pression, mais que le patient ne reconnait pas). Un spasme est aussi interprété comme un test positif. **Négatif :** le participant dit ne ressentir aucune douleur lors de la pression sur la structure. L'évocation d'un inconfort non représentatif de leur Sy est considérée comme étant négative.
Présence de chaleur à la palpation (OA + inflammation)	Lors de l'examen palpatoire, déterminez s'il y a présence de chaleur au niveau de l'articulation. Comparez au genou sain. Comparez avec la température au niveau de la cuisse. **Positif :** lors de la palpation, vous ressentez une chaleur au niveau du genou de façon globale et celle-ci est différente du genou sain. **Négatif :** vous ne ressentez aucune différence de température entre les deux genoux.
Présence de crépitements	Placer votre main sur le genou afin de couvrir la patella et le plus de surface possible. Tenir l'interligne articulaire entre le 1er et le 5e métacarpe. Comparer les deux genoux afin d'éviter un faux positif. Vous pouvez déterminez la présence de crépitements dans quatre situations: lors de la mobilisation de la patella (signe du rabot), lors de la flexion/extension active ou lors de la flexion/extension passive, lorsque le patient se lève (mise en charge). **Positif:** vous ressentez ou entendez des vibrations caractéristiques des crépitements. **Négatif:** vous ne ressentez aucun crépitement.
Observations générales	
Limitation des amplitudes de mouvement en flexion et extension passive.	Demandez au participant d'effectuer une flexion complète suivie d'une extension complète du genou avec le membre inférieur sain suivi du genou affecté. Le participant doit revenir en extension

	entre les deux mouvements pour éviter une pression au niveau du dos. Ensuite, effectuez de façon passive une flexion complète suivie d'une extension complète du genou avec le membre inférieur sain suivi du genou affecté. **Positif :** vous observez que le participant n'est pas en mesure de faire une flexion/extension équivalente entre le membre inférieur sain et affecté. Vous observez que l'amplitude passive en flexion/extension n'est pas équivalente entre le membre inférieur sain et affecté. **Négatif :** vous observez une amplitude semblable entre les deux membres inférieurs pour la flexion/extension active et passive.
Douleur à l'extension résistée isométrique à 90° et 30° (SFP) de flexion	Le participant est en position assise. Placez le membre inférieur atteint en position de flexion à 90°. Placez votre main au niveau de la cheville du participant et demandez-lui d'essayer de faire une extension du genou en forçant contre vous. Résistez afin que le membre ne bouge pas. **Positif :** le participant dit ressentir une douleur/inconfort représentatif des Sy habituels au genou au moment de la contraction résistée. **Négatif :** le participant ne ressent aucune douleur au moment de la contraction résistée. Le participant dit ressentir un inconfort non représentatif des Sy habituels.
Évaluation visuelle de la morphologie des membres inférieurs	Le participant est debout en position les pieds collés. Observez la morphologie de ses membres inférieurs. Sélectionner les variations morphologiques observées parmi celles listées.
Évaluation visuelle de la morphologie patellaire	Le participant est debout en position confortable. Observez la morphologie de ses patellas. Sélectionner les variations morphologiques observées parmi celles listées.
Tests spécifiques pour l'inflammation du genou	
Test du flot (4-8 ml extra fluide)	Le participant est en décubitus dorsal. Avec une main placée en médial du genou, effectuez 2-3 mouvements de « brosse » vers le haut, pour ensuite effectuer un mouvement de brosse vers le bas avec l'autre main placée en externe du genou. Attendre deux secondes pour s'assurer du déplacement des fluides. Vous pouvez effectuez la manoeuvre jusqu'à 2-3 fois consécutives avant d'interpréter le test. Le test est considéré positif lorsque vous observez un léger gonflement qui apparait sous l'aspect médial de la patella. Un gonflement autour du tendon rotulien (médial ou latéral) est aussi considéré comme un test positif.

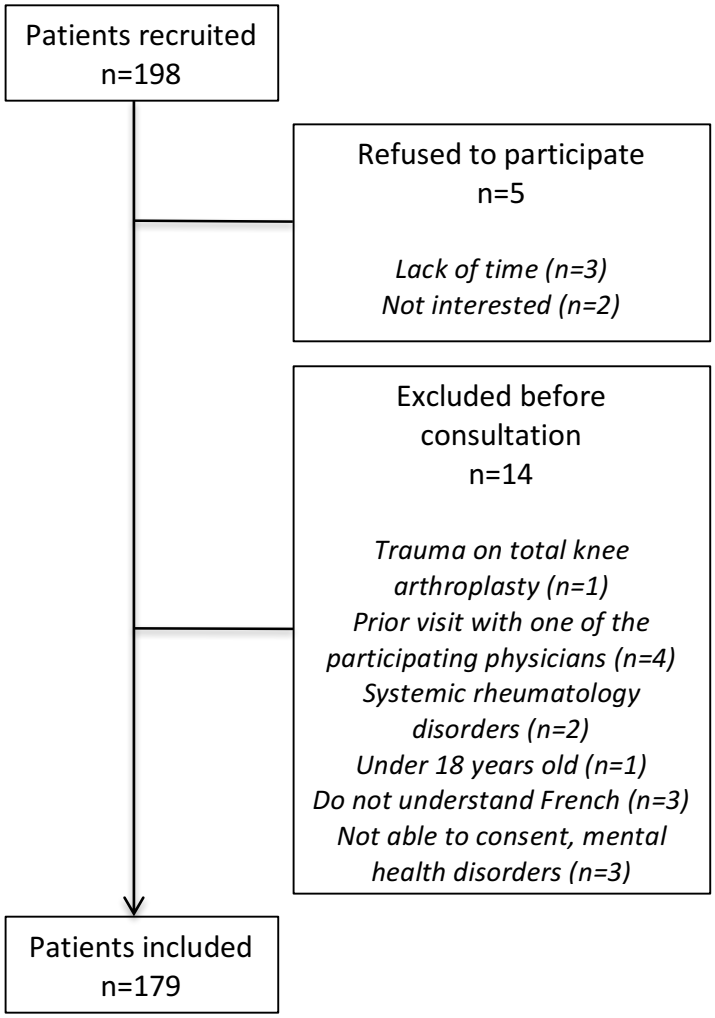
Test du glaçon (40-50 ml extra fluides)	Le participant est en décubitus dorsal. Vous placez une main supérieure au genou et l'autre main inférieure au genou. Vous ramenez les vos deux mains vers le genou. À l'aide de votre pouce, poussez la patella vers la trochlée et observez le retour de la patella dans sa position originale. Le test est considéré positif lorsque vous observez que la patella semble « flotter » vers sa position originale.
Tests spécifiques pour blessures ligamentaires	
Lachman	Le participant est en décubitus dorsal. Pliez le genou entre 15 et 30 degrés de flexion. Stabilisez le fémur distal avec une main, alors que votre autre main agrippe le tibia proximal par l'arrière. Appliquez une force antérieure pour tirer le tibia vers l'avant. La force doit être produite de façon rapide et de courte durée. Le test est considéré comme positif si le tibia produit un déplacement antérieur plus grand du côté affecté que du côté sain.
Test du tiroir antérieur	Le participant est en décubitus dorsal. Pliez le genou à 90 degrés, assoyez vous sur le pied du participant et agrippez avec vos deux mains le tibia proximal avec vos pouces sur les plateaux tibiaux en antérieur et les index palpant les tendons des ischio-jambiers en postérieur. Appliquez une force antérieure pour tirer le tibia vers l'avant. La force doit être produite de façon lente. Le test est considéré comme positif si le tibia produit un déplacement antérieur plus grand du côté affecté que du côté sain.
Test du Pivot Shift	Le participant est en décubitus dorsal. Placez le genou en extension complète pour ensuite demander au participant de complètement relâcher sa musculature (le test est difficilement interprétable si le participant produit de la co-contraction). Entourez fermement le tibia en proximal et en distal et effectuez une rotation interne du membre inférieur. En tenant cette position de rotation interne, amorcer lentement une flexion passive du genou. Le test est considéré positif si vous observez une subluxation externe du plateau tibial dès le début de la flexion ou palpez un son/vibration mat ou un « clic ». Vous pouvez effectuer la manoeuvre quelques fois selon la tolérance du participant avant de statuer sur le résultat du test.
Test du tiroir postérieur	Le participant est en décubitus dorsal. Pliez le genou à 90 degrés, assoyez vous sur le pied sur participant et agrippez avec vos deux mains le tibia proximal avec vos pouces sur les plateaux tibiaux en antérieur et les index palpant les tendons des ischio-jambiers en

	postérieur. Appliquez une force postérieure pour pousser le tibia vers l'arrière. La force doit être produite de façon lente. Le test est considéré comme positif si le tibia produit un déplacement postérieur plus grand du côté affecté que du côté sain.
Test du ligament collatéral interne à 0 et 30 degrés	Le participant est en décubitus dorsal. Fléchissez le genou à 0 degrés (extension complète) et saisissez fermement le tibia entre votre bras et le côté de votre corps, vous êtes placés en externe du membre inférieur du participant. Placez l'autre main en distal du fémur, sur la face externe du genou. La main en externe du genou applique une légère force médiale alors que l'autre bras effectue une légère rotation externe du tibia simultanément. Refaite le test à 30 degrés. Le test est considéré positif si vous ressentez une ouverture médiale excessive en comparaison avec le côté sain **et/ou** que le participant se plaint de douleur/inconfort représentative des Sy lors du test.
Test du ligament collatéral externe à 0 et 30 degrés	Le participant est en décubitus dorsal. Fléchissez le genou à 0 degrés (extension complète) et saisissez fermement le tibia entre votre bras et le côté de votre corps, vous êtes placés en interne du membre inférieur du participant. Placez l'autre main en distal du fémur, sur la face interne du genou. La main en interne du genou applique une légère force latérale alors que l'autre bras effectue une légère rotation externe du tibia simultanément. Refaite le test à 30 degrés. Le test est considéré positif si vous ressentez une ouverture latérale excessive en comparaison avec le côté sain **et/ou** que le participant se plaint de douleur/inconfort représentative des Sy lors du test.
Tests spécifiques pour blessures méniscales	
Test de McMurray	Le participant est en décubitus dorsal. Prenez le membre inférieur au niveau du talon alors que l'autre main saisit le genou entre le pouce et l'index afin de palper l'interligne articulaire médiale et latérale. Effectuez passivement une flexion complète du genou. Pour tester le ménisque médial, effectuez à l'aide de la main inférieure une rotation externe du tibia pour ensuite étendre lentement le genou. Pour tester le ménisque latéral, effectuez une rotation interne du tibia pour ensuite étendre lentement le genou. Vous pouvez effectuer un mouvement de "scooping" en flexion complète afin d'essayer de faire ressortir le "clic" pour ensuite enchainer avec l'extension. Vous devez pouvoir faire une flexion au moins presque complète pour interpréter ce test.

	Le test est considéré positif si vous entendez ou palpez un son/vibration mat ou de type « clic ». Comme pour le Thessaly, nous ajoutons qu'une douleur représentative des Sy du patient lors de la manoeuvre est considérée comme étant positive (douleur en torsion). Par contre la douleur à la surpression au scooping en flexion complète ne représente pas un McMurray positif. Bien faire la distinction avec la sensation de crépitements émanant de la rotule, souvent présente chez les patients ayant de l'OA fémoropatellaire.
Test de Thessaly à 20 degrés	Le participant est en position debout. Demandez au participant de se tenir sur une seule jambe avec le genou légèrement plié (squat partiel, 20 degrés). Prenez les mains du participant. Effectuez quelques pas de côtés afin de guider le participant vers une rotation de son corps. Commencer vers la gauche, puis vers la droite. Reproduire 3 fois sans pause de chaque côté. Le pied du participant doit rester en place au sol. Le test est considéré positif si le participant se plaint d'une douleur/inconfort à l'interligne articulaire ou d'une sensation de blocage ou d'agrippement (accrochage). Bien faire la distinction avec une douleur antérieure qui est souvent présente chez les patients ayant une pathologie fémoropatellaire.
Tests spécifiques pour syndrome fémoro-patellaire et/ou instabilité rotulienne	
Douleur à la compression/mobilisation de la patella (Signe du Rabot)	Le participant est en décubitus dorsal. Poussez la patella directement contre la trochlée. Effectuez des translations latérales et inféro-supérieures (mobilisations). Le test est considéré comme étant positif s'il reproduit la douleur de consultation.
Signe de Clark (Patellar Grind Test)	Le participant est en décubitus dorsal. Entourez l'aspect supérieur de la patella avec votre main de sorte à faire une résistance à la patella vers le haut. Vous pouvez stabiliser le membre inférieur avec votre autre main. Demandez au participant de contracter le quadriceps. Le test est considéré positif si le participant se plaint de douleur/inconfort représentatif des Sy. Ce test créé beaucoup de faux positif dans la population générale. Exécuter sur le genou sain afin de s'assurer qu'il s'agit d'un vrai positif (douleur supérieure du côté affecté).
Test d'appréhension patellaire (30° fx)	Le participant est en décubitus dorsal. Le membre inférieur est placé passivement à 30 degrés de flexion. Exercez une force latérale sur l'aspect médial de la patella. Le test est considéré comme étant positif si le participant démontre de l'appréhension en résistant la force latérale ou si la

	douleur/inconfort de consultation est reproduite. Comparez avec l'autre genou pour s'assurer de la validité de l'observation. Une douleur à la facette patellaire externe ne représente pas un test d'appréhension positif.
Désalignement (signe en J) de la patella lors de la contraction quadricipitale	Le participant est en décubitus dorsal. Vous pouvez stabiliser le membre inférieur avec votre autre main. Demandez au participant de faire une contraction du quadriceps. Observez la direction de la patella lors de la contraction et/ou palpez légèrement l'aspect supérieur de la patella afin de sentir la direction de la patella lors de la contraction. Vous pouvez observer ce signe lors de l'extension active du genou en position assise, ou lors de la montée et descente de la marche. Le test est considéré comme étant positif si vous observez ou palpez une direction patellaire à la contraction latérale plutôt que supérieure. Comparez avec l'autre genou pour s'assurer de la validité de l'observation d'un point de vue de son utilité clinique.
Évaluation fonctionnelle	
Analyse du squat (Dlr et force fonctionnelle)	Demander au participant d'effectuer un squat le plus complet possible. Assurez-vous qu'il est dans un environnement sécuritaire. Notez si le participant se plaint de douleur lors de l'exécution et/ou s'il démontre un désalignement (ex: valgus, instabilité) ou une limitation de l'amplitude de mouvement, signe d'une diminution du contrôle moteur. Il s'agit d'un test visant à évaluer la force fonctionnelle et globale au MI.
Analyse de la montée et de la descente d'une marche (Dlr et force fonctionnelle)	Demandez au participant de monter et de descendre une marche d'escalier standard (environ 25 cm). Assurez-vous qu'il est dans un environnement sécuritaire. Notez si le participant se plaint de douleur lors de l'exécution et/ou s'il démontre un désalignement (ex: valgus, instabilité) ou une limitation de l'amplitude de mouvement, signe d'une diminution du contrôle moteur. Comparez avec l'autre genou pour s'assurer de la validité de l'observation. Il s'agit d'un test visant à évaluer la force fonctionnelle et globale au MI.

8. Flow chart of patients' recruitment in article 3



9. 2x2 contingency tables to assess the validity and inter-rater reliability of the complete musculoskeletal examination for each knee disorder in article 3

ACL injury		Medical expert		
		Yes	No	Total
Physiotherapist	Yes	8	1	9
	No	0	170	170
	Total	8	171	179

Meniscal injury		Medical expert		
		Yes	No	Total
Physiotherapist	Yes	35	6	41
	No	1	137	138
	Total	36	143	179

Patellofemoral pain		Medical expert		
		Yes	No	Total
Physiotherapist	Yes	41	4	45
	No	4	130	134
	Total	45	134	179

Osteoarthritis		Medical expert		
		Yes	No	Total
Physiotherapist	Yes	72	3	75
	No	7	97	104
	Total	79	100	179

Others knee disorder*		Medical expert		
		Yes	No	Total
Physiotherapist	Yes	9	0	9
	No	2	168	170
	Total	11	168	179

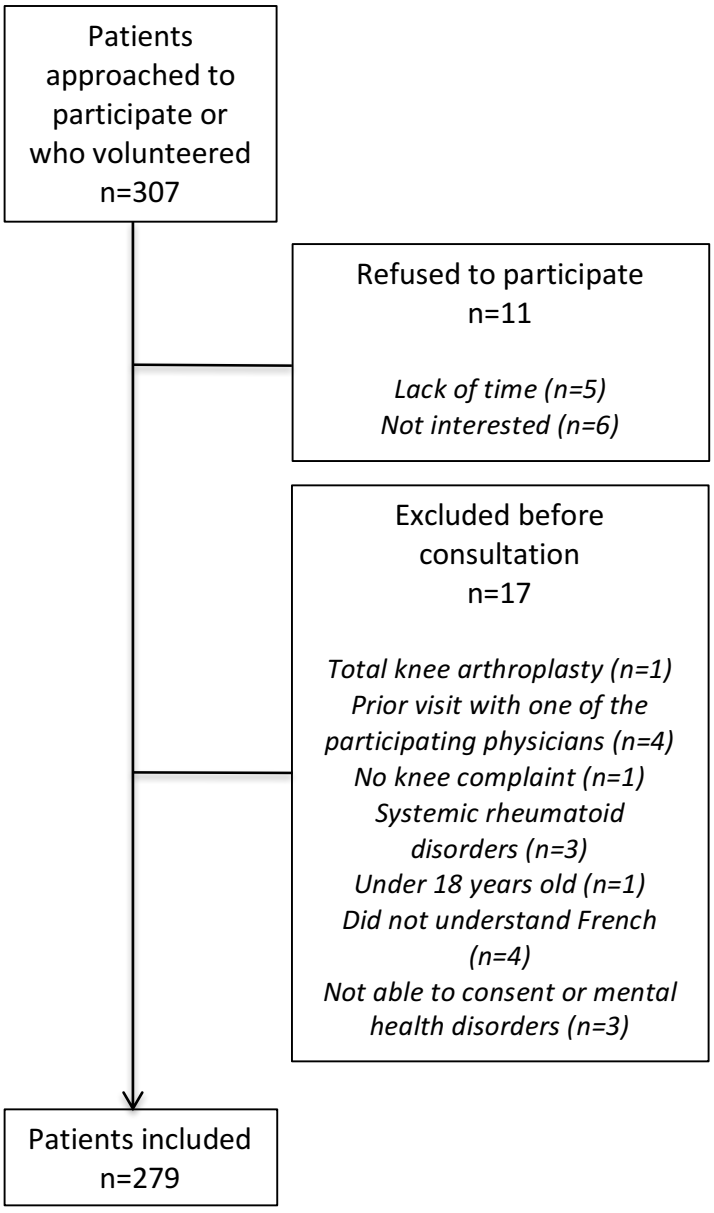
**Other diagnosis included: contusion of the tibial plateau (n=2), PCL tear (n=1), soleus tear (n=1), psychosomatic origin (n=1), muscular spasms linked to multiple sclerosis (n=1), hamstring tendinopathy (n=3), medial collateral ligament injury (n=1), functional instability in the absence of ACL or meniscal injury (n=1).*

10. 3x3 contingency table to assess the inter-rater reliability of the triage of surgical candidates and conservative care in article 3

		Medical expert			
		Yes	No	Uncertain	Total
Physiotherapist	Yes	21	11	0	32
	No	2	139	2	143
	Uncertain	0	0	4	4
	Total	23	150	6	179

"Yes" are surgical candidates, "No" are conservative care candidates, "Uncertain" needed more information to confirm treatment options.

11. Flow chart of patients' recruitment in article 4, 5, 6 and 7



12. 2x2 contingency tables for the diagnostic validity of history elements and physical examination tests when individually performed for partial or complete ACL tears in article 4

Pivoting traumatic mechanism

		Reference standard		
		Yes	No	Total
Classification	Yes	34	37	71
	No	9	199	208
	Total	43	236	279

Popping sensation during trauma

		Reference standard		
		Yes	No	Total
Classification	Yes	25	18	43
	No	18	218	236
	Total	43	236	279

Lachman

		Reference standard		
		Yes	No	Total
Classification	Yes	35	5	40
	No	8	231	239
	Total	43	236	279

Pivot Shift

		Reference standard		
		Yes	No	Total
Classification	Yes	30	4	34
	No	9	191	200
	Total	39	195	234

13. 2x2 contingency tables for diagnostic validity of history elements and physical examination tests when individually performed for complete ACL tears only in article 4

Pivoting traumatic mechanism

		Reference standard		
		Yes	No	Total
Classification	Yes	21	50	71
	No	1	207	208
	Total	22	257	279

Immediate effusion

		Reference standard		
		Yes	No	Total
Classification	Yes	15	16	31
	No	7	241	248
	Total	22	257	279

Lachman

		Reference standard		
		Yes	No	Total
Classification	Yes	18	22	40
	No	4	235	239
	Total	22	257	279

Pivot Shift

		Reference standard		
		Yes	No	Total
Classification	Yes	16	18	34
	No	4	196	200
	Total	20	214	234

14. 2x2 contingency tables of diagnostic clusters using recursive partitioning for partial or complete ACL tears in article 4

Combination of history with pivot and popping sensation during trauma

		Reference standard		
		Yes	No	Total
Classification	Yes	25	14	39
	No	18	222	240
	Total	43	236	279

Combination of negative history with pivot or popping sensation during trauma and negative Lachman or pivot shift

		Reference standard		
		Yes	No	Total
Classification	Yes	40	30	70
	No	3	206	209
	Total	43	236	279

15. 2x2 contingency tables of diagnostic clusters using recursive partitioning for complete ACL tears only in article 4

Combination of history of pivot during trauma and immediate effusion after trauma and a positive Lachman test

		Reference standard		
		Yes	No	Total
Classification	Yes	18	12	30
	No	4	245	249
	Total	22	257	279

16. 2x2 contingency tables of diagnostic clusters using recursive partitioning for traumatic and degenerative symptomatic meniscal tears in article 5

High specificity and sensitivity diagnostic clusters for traumatic SMT

		Reference standard		
		Yes	No	Total
Classification	Yes	32	25	57
	No	3	219	222
	Total	35	244	279

High specificity diagnostic clusters for degenerative SMT

		Reference standard		
		Yes	No	Total
Classification	Yes	26	21	47
	No	19	213	232
	Total	45	234	279

High sensitivity diagnostic clusters for degenerative SMT

		Reference standard		
		Yes	No	Total
Classification	Yes	42	81	123
	No	3	153	156
	Total	45	234	279

17. Description of other physical examination tests techniques for symptomatic degenerative meniscal tears in article 5

Tests	Description of technique and positive outcomes.
Visual assessment of passive and active knee flexion and extension range of motion	The patient is in supine position. The clinician asks the patient to actively move his/her healthy knee to its maximum flexion and then back to its maximal extension. The patient then repeat the movement with his injured knee. The clinician then repeats the same movement by passively flexing and extending the patient's knee. A positive test occurs if the active or passive range of motion is restricted between the healthy and injured knee. Restricted range of motion may both be both of painful or of mechanical origin.
Visual assessment of knee and lower limb alignment	The patient is in standing position with his/her feet close together. The clinician visually assesses the knee and lower limb alignment. The knees and lower limbs are considered misaligned if a varus, valgus or recurvatum is present.
Pain during squat or stairs descending	The clinician asks the patient to complete a squat to 90° of knee flexion and to descend a 20-centimetres step. A positive test occurs if the patient reports pain to his knee during the movements.
Alignment during squat or stairs descending	The clinician asks the patient to complete a squat to 90° of knee flexion and to descend a 20-centimetres step. A positive test occurs if the clinician observes loss of control or asymmetry of the lower limb alignment during the movement between the healthy and injured knee.

18. 2x2 contingency tables of diagnostic clusters using recursive partitioning for symptomatic knee osteoarthritis in article 6

High specificity clusters for SOA

		Reference standard		
		Yes	No	Total
Classification	Yes	82	7	89
	No	47	143	190
	Total	129	150	279

High sensitivity clusters for SOA

		Reference standard		
		Yes	No	Total
Classification	Yes	119	44	163
	No	10	106	116
	Total	129	150	279

19. 2x2 contingency tables of diagnostic clusters using recursive partitioning for patellofemoral pain in article 7

High specificity diagnostic clusters for patellofemoral pain

		Reference standard		
		Yes	No	Total
Classification	Yes	48	15	63
	No	27	189	216
	Total	75	204	279

High sensitivity diagnostic clusters for patellofemoral pain

		Reference standard		
		Yes	No	Total
Classification	Yes	69	72	141
	No	6	132	138
	Total	75	204	279

20. Certificat d'éthique

HMR Hôpital Maisonneuve-Rosemont
Centre affilié à l'Université de Montréal

Pour vous, pour la vie

Certificat éthique

Le Comité d'éthique de la recherche de l'Hôpital Maisonneuve-Rosemont a approuvé et assurera le suivi du projet de recherche intitulé:

Validité et fidélité des éléments de l'histoire et des tests physiques pour le diagnostic des pathologies communes au genou. (Réf. CÉR : 14061)

présenté par ***Monsieur François Desmeules.*** Cette étude est conforme aux normes éthiques actuelles.

Ce certificat est valide pour la période du ***04 novembre 2014*** au ***04 novembre 2015.***

COMITÉ D'ÉTHIQUE DE LA RECHERCHE HMR

Dr Peter Vavassis
Président
Comité d'éthique de la recherche
Hôpital Maisonneuve-Rosemont

PV/cl

21. Formulaire de consentement

Hôpital Maisonneuve-Rosemont
Centre affilié à l'Université de Montréal

Pour vous, pour la vie

FORMULAIRE D'INFORMATION ET DE CONSENTEMENT (FIC)

Titre de l'étude : Validité et fidélité des éléments de l'histoire et des tests physiques pour le diagnostic des pathologies communes au genou.

Chercheur principal : François Desmeules, pht. Ph. D.

Co-chercheur(s) : Simon Décary, pht, candidat au doctorat en sciences de la réadaptation
Pascal-André Vendittoli, MD, MSc., FRCS
Debbie Feldman, pht, Ph. D.
Marie-Pierre Sylvestre, Ph. D.
Marie-Pierre Pelletier, Ph. D.
Jean-Pierre Pelletier, MD, CSPQ, FRCP(C)

Organisme subventionnaire :

1- Fonds de recherche en santé du Québec (FRQ-S): bourse doctorale pour professionnel de la santé (M. Simon Décary).
2- Fonds de démarrage Centre de Recherche de L'Hôpital Maisonneuve-Rosemont (M. François Desmeules).

Veuillez noter que le terme « FIC » est utilisé pour désigner « Formulaire d'information et de consentement » dans ce document.

INTRODUCTION

Nous vous invitons à participer à un projet de recherche. Cependant, avant d'accepter de participer à ce projet et de signer ce FIC, veuillez prendre le temps de lire, de comprendre et de considérer attentivement les renseignements suivants. Vous pouvez apporter une copie non signée du FIC à la maison pour y réfléchir et en discuter avec votre famille ou vos amis avant de prendre une décision.

Ce formulaire peut contenir des mots que vous ne comprenez pas. Nous vous invitons à poser toutes vos questions au chercheur responsable du projet et aux membres de l'équipe de recherche, afin qu'ils vous expliquent tout mot ou renseignement qui n'est pas clair.

NATURE ET OBJECTIFS DE L'ÉTUDE

Vous avez un problème avec un ou vos genoux. Cette problématique est survenue lors d'un incident ou sans cause définie. Suite à une consultation médicale en première ligne (ex. : clinique sans rendez-vous, urgence, médecin de famille, etc.), vous avez été dirigé à la clinique orthopédique ou à l'Unité de Médecine Familiale (UMF) de l'Hôpital Maisonneuve-Rosemont à Montréal afin de consulter un orthopédiste ou un médecin de famille ayant une expertise en pathologies du système locomoteur pour des examens plus approfondis. Le but de cette étude est de vérifier si le diagnostic des pathologies du genou basé uniquement sur les questions que posent le médecin et des tests physiques est valide sans avoir recours à des tests d'imagerie médicale comme le rayon X ou la résonnance magnétique.

En effet, comme vous, plusieurs individus doivent être mis sur une liste d'attente pour obtenir un diagnostic médical spécialisé ou un examen d'imagerie médicale suite à une première consultation avec un médecin. La prise en charge adéquate est donc souvent retardée, ce qui peut causer de la douleur et une diminution de la qualité de vie. Bien que la littérature scientifique démontre que l'imagerie est souvent non nécessaire pour obtenir un diagnostic pour une pathologie commune au genou, plusieurs cliniciens hésitent à amorcer une prise en charge rapide à cause d'un manque de confiance en la validité de l'examen clinique basé uniquement sur l'histoire et les résultats des tests physiques. La littérature démontre aussi que le diagnostic clinique d'un orthopédiste, d'un physiothérapeute ou d'un médecin ayant une expertise en pathologies du système locomoteur est valide.

L'objectif de cette étude est donc de trouver les meilleurs éléments de l'histoire et les meilleurs tests physiques pour les combiner et créer des règles de prédiction clinique visant à aider les cliniciens à faire un diagnostic valide plus rapidement. Puisque la plupart des pathologies communes au genou peuvent être traitées efficacement lorsqu'elles sont prises rapidement, nous croyons qu'un diagnostic valide et rapide permettra d'améliorer la prise en charge des individus comme vous.

DÉROULEMENT DE L'ÉTUDE

La consultation médicale avec un médecin dans le cadre de ce projet de recherche est différente de la procédure habituelle. Dans la procédure habituelle, lors du rendez-vous avec le médecin, l'individu ayant une problématique au genou est évalué directement par un le médecin sénior. Suite à l'évaluation, le médecin sénior peut demander un examen en imagerie au besoin pour ensuite proposer une prise en charge adéquate. Un examen normal est d'une durée d'environ deux à trois heures à cause des temps d'attente.

Vous devrez remplir des questionnaires. Le temps maximal requis pour remplir tous ces questionnaires est estimation à environ trente (30) minutes. Vous répondrez à ces questionnaires à la clinique. Par la suite, deux évaluateurs, un physiothérapeute suivi d'un orthopédiste ou un médecin de famille ayant une expertise en pathologies du système locomoteur viendront vous évaluer à tour de rôle dans le bureau de consultation. Vous ne devrez pas informer le médecin expert des tests faits par le physiothérapeute. Le médecin expert vous dévoilera les résultats de votre évaluation et votre diagnostic final à la fin de la

consultation. Ainsi, en comparaison avec une consultation standard, le fait d'être évalué par deux (2) évaluateurs pourrait prolonger le temps de consultation de 20 minutes.

Le projet de recherche se terminera après l'évaluation par les deux évaluateurs. À ce moment, le médecin sénior qui vous informera de votre diagnostic et poursuivra votre prise en charge normale. Il décidera à ce moment si une imagerie est nécessaire et vous informera de l'approche de traitement requise pour votre condition. Au total, le projet de recherche requiert environ 60 minutes de votre temps. Aucune autre présence ne sera requise de votre part pour le projet de recherche et vous poursuivrez votre prise en charge avec votre médecin.

RISQUES / INCONVÉNIENTS / INCONFORTS

Inconforts associés aux procédures

Il est possible que vous ressentiez de la douleur au genou suite à l'évaluation par un professionnel. Si cela devait se produire, vous serez exclus du projet de recherche et rencontrerez directement l'orthopédiste afin de poursuivre la consultation normale.

Inconforts associés aux questionnaires

Remplir les questionnaires pourrait générer de l'anxiété ou un certain inconfort. Des pauses sont donc prévues tout au long des questionnaires. Si tel est le cas, n'hésitez pas à en parler avec le personnel de l'étude qui saura vous référer au besoin.

AVANTAGES

Vous ne retirerez aucun bénéfice personnel de votre participation à ce projet de recherche. Cependant, les résultats obtenus contribueront à l'avancement des connaissances dans le domaine de l'orthopédie et du diagnostic des pathologies communes au genou.

COMPENSATION FINANCIÈRE

Vous ne recevrez aucune compensation financière pour votre participation à cette étude.

INDEMNISATION EN CAS DE PRÉJUDICE

Si vous deviez subir quelque préjudice que ce soit suite à l'administration des procédures reliées à ce projet de recherche, vous recevrez tous les soins et services requis par votre état de santé, sans frais de votre part.

En acceptant de participer à ce projet, vous ne renoncez à aucun de vos droits ni ne libérez les chercheurs, le commanditaire et l'établissement de leur responsabilité civile et professionnelle.

PARTICIPATION VOLONTAIRE ET DROIT DE RETRAIT DE L'ÉTUDE

Votre participation à ce projet de recherche est volontaire. Vous êtes donc libre de refuser d'y participer. Vous pouvez également vous retirer de ce projet à n'importe quel moment, sans avoir à donner de raison, en faisant connaître votre décision au chercheur responsable du projet ou à un membre de l'équipe de recherche.

Votre décision de ne pas participer à ce projet de recherche ou de vous en retirer n'aura aucune conséquence sur la qualité des soins et des services auxquels vous avez droit, ni sur votre relation avec le chercheur responsable du projet et les autres intervenants.

Le chercheur responsable et le comité d'éthique de la recherche de l'Hôpital Maisonneuve-Rosemont (HMR) peuvent mettre fin à votre participation sans votre consentement, si de nouvelles découvertes ou informations indiquent que votre participation au projet n'est plus dans votre intérêt, si vous ne respectez pas les consignes du projet de recherche ou s'il existe des raisons administratives d'abandonner le projet.

Si vous vous retirez ou êtes retiré du projet, l'information déjà obtenue dans le cadre de ce projet sera conservée aussi longtemps que nécessaire pour assurer votre sécurité et celle des autres participants de recherche et ainsi, se conformer aux exigences réglementaires. Toutefois, aucune nouvelle donnée ne sera recueillie de vos dossiers.

Toute nouvelle connaissance acquise au cours du projet pouvant affecter votre décision de continuer d'y participer vous sera communiquée sans délai verbalement et par écrit.

CONFIDENTIALITÉ

Durant votre participation à ce projet, le chercheur responsable ainsi que son personnel recueilleront et consigneront dans un dossier de recherche les renseignements vous concernant. Seuls les renseignements nécessaires pour répondre aux objectifs scientifiques de ce projet seront recueillis.

Ces renseignements peuvent inclure les informations contenues dans vos dossiers médicaux concernant votre état de santé passé et présent, vos habitudes de vie ainsi que les résultats de tous les tests, examens et procédures que vous aurez à subir durant ce projet. Votre dossier peut aussi contenir d'autres renseignements tels que votre nom, votre sexe, votre date de naissance et votre origine ethnique.

Tous les renseignements recueillis demeureront strictement confidentiels dans les limites prévues par la loi. Afin de préserver votre identité et la confidentialité de vos renseignements, vous ne serez identifié que par un numéro de code. La clé du code reliant votre nom à votre dossier de recherche sera conservée par le chercheur responsable.

Également, les données du projet pourraient servir pour d'autres analyses de données reliées au projet, ou pour l'élaboration de projets de recherche futurs. Les données pourront être publiées dans des revues spécialisées ou faire l'objet de discussions scientifiques, mais il sera

impossible de vous identifier. Ces données seront conservées pendant 7 ans par le chercheur responsable.

À des fins de surveillance et de contrôle, votre dossier de recherche ainsi que vos dossiers médicaux pourront être consultés par une personne mandatée par le comité d'éthique de la recherche de l'HMR, par l'établissement ou par des organismes publics autorisés. Toutes ces personnes et ces organismes adhèrent à une politique de confidentialité.

À des fins de protection, notamment afin de pouvoir communiquer avec vous rapidement, vos nom et prénom, vos coordonnées et la date de début et de fin de votre participation au projet seront conservés pendant 1 an après la fin du projet, dans un répertoire à part maintenu par le chercheur responsable.

Vous avez le droit de consulter votre dossier de recherche pour vérifier les renseignements recueillis et les faire rectifier au besoin, et ce, aussi longtemps que le chercheur responsable du projet ou l'établissement détiennent ces informations. Cependant, afin de préserver l'intégrité scientifique du projet, vous pourriez n'avoir accès à certaines de ces informations qu'une fois votre participation terminée.

Les résultats de la recherche pourront vous être fournis sur demande lorsque l'étude sera complétée et que les résultats seront publiés.

FINANCEMENT

Le chercheur principal du projet ainsi que l'établissement ont reçu un financement d'organismes subventionnaires publics pour mener à bien ce projet de recherche**.**

PERSONNES-RESSOURCES

Si vous avez des questions ou éprouvez un problème relié à ce projet de recherche, veuillez communiquer avec :

M. François Desmeules, pht. Ph. D., chercheur principal
ou
M. Simon Décary, pht, candidat au doctorat en sciences de la réadaptation

SURVEILLANCE DES ASPECTS ÉTHIQUES
Le comité d'éthique de la recherche de l'HMR a approuvé ce projet de recherche et en assure le suivi. De plus, il approuvera au préalable toute révision et toute modification apportée au FIC ainsi qu'au protocole de recherche. Pour toute question concernant vos droits à titre de participant à ce projet de recherche ou si vous avez une plainte ou un commentaire à formuler, vous pouvez contacter le commissaire local aux plaintes et à la qualité des services de HMR.

CONSENTEMENT ET SIGNATURES

Titre de l'étude : Validité et fidélité des éléments de l'histoire et des tests physiques pour le diagnostic des pathologies communes au genou.

Participant de recherche

J'ai pris connaissance du présent formulaire d'information et de consentement. Je reconnais qu'on m'a expliqué le projet, qu'on a répondu à mes questions et qu'on m'a laissé le temps voulu pour prendre une décision.

☐ _____ (initiales) J'autorise le chercheur à informer mon médecin traitant de ma participation à ce projet :
Nom et adresse du médecin : _____

Je consens à participer à ce projet de recherche aux conditions qui y sont énoncées. Une copie signée et datée du présent FIC me sera remise. Il n'y aura pas de copie supplémentaire de ce document déposée à mon dossier médical; seul l'original sera conservé dans mon dossier de la clinique de recherche.

_____	_____	_____
Nom (lettres moulées)	Signature	Date

Personne qui obtient le consentement

J'ai expliqué au participant de recherche les termes du présent FIC et j'ai répondu aux questions qu'il m'a posées.

_____	_____	_____
Nom (Lettres moulées)	Signature	Date

Engagement du chercheur principal

Je certifie qu'on a expliqué au participant les termes du présent FIC, répondu à ses questions et clairement indiqué qu'il demeure libre de se retirer de l'étude en tout temps, et ce, sans préjudice. Je m'engage avec l'équipe de recherche à respecter ce qui a été convenu au FIC et à remettre une copie signée au participant.

_____	_____	_____
Nom (lettres moulées)	Signature	Date