



**HAL**  
open science

# Les scores hanche et genou en orthopédie : validation d'un score, définition d'une différence minimale cliniquement importante et méthodes de Mapping

Sophie Putman

## ► To cite this version:

Sophie Putman. Les scores hanche et genou en orthopédie : validation d'un score, définition d'une différence minimale cliniquement importante et méthodes de Mapping. Médecine humaine et pathologie. Université de Lille, 2020. Français. NNT : 2020LILUS014 . tel-03267674

**HAL Id: tel-03267674**

**<https://theses.hal.science/tel-03267674>**

Submitted on 22 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE DE LILLE  
**ECOLE DOCTORALE BIOLOGIE SANTE**

THESE D'UNIVERSITE

Recherche clinique, innovation technologique, santé  
publique

**Les scores hanche et genou en orthopédie**

Validation d'un score, définition d'une différence minimale  
cliniquement importante et méthodes de Mapping

Présentée et soutenue publiquement le 12 novembre 2020

**Par Sophie Putman**

---

**JURY**

**Rapporteur et Président du jury : Pr Stéphane Boisgard**

**Rapporteur : Pr Matthieu Ollivier**

**Invité : Pr Cristian Preda**

**Co-directeur de thèse : Pr Alain Duhamel**

**Co-directeur de thèse : Pr Henri Migaud**

---

## Remerciements

Je tiens à remercier tous les membres du jury pour leur participation.

Merci au Pr Boisgard d'avoir accepté de juger ce travail. Merci pour votre disponibilité, votre participation au comité de suivi. Vous aviez trouvé le sujet intéressant lors de ces comités. Certaines publications de votre équipe correspondent au sujet. J'espère que le résultat final vous conviendra. Merci pour votre soutien dans ce parcours universitaire.

Merci au Pr Ollivier d'avoir accepté de juger ce travail. Vos dernières publications ont en commun certains sujets de cette thèse et j'espère que vous trouverez dans ce travail des éléments prêtant à explorer de nouvelles pistes.

Merci au Pr Preda que j'ai rencontré pendant les comités de suivi, qui m'a aidée à réfléchir sur cette thèse. Merci pour votre accompagnement et votre disponibilité. Merci beaucoup de juger ce travail.

Merci au Pr Duhamel pour toutes ses remarques constructives, permettant le plus souvent d'éclaircir les points qui me paraissaient logiques. Merci de m'avoir encadrée du master 2 à cette thèse, merci pour votre disponibilité.

Merci au Pr Migaud pour son aide quotidienne. Si je suis là, c'est grâce ou à cause de vous car vous m'avez fait découvrir la chirurgie orthopédique lors de mon stage d'externe.

Merci également au Pr Pasquier, qui m'a toujours soutenue. Dans les moments de doute, nos conversations m'ont permis de continuer.

Merci au Pr Girard, qui m'a permis d'avancer dans ce travail. Merci pour ton soutien.

Merci aux chefs de clinique et assistants du service.

Merci aux équipes de l'unité de statistiques, ainsi qu'aux équipes d'orthopédie du CHU de Lille.

## RESUME EN FRANÇAIS

Les scores d'autoévaluation (PROM = Patient Reported Outcome Measurement) font partie intégrante de l'évaluation moderne des patients. Ils permettent par exemple en orthopédie de définir un statut de base pour ces patients et ensuite après un traitement chirurgical ou médical, de définir une amélioration. Ces scores nécessitent des étapes de validation permettant leur utilisation (validité interne et externe, effet plancher et plafond, répétabilité, sensibilité au changement, traduction dans différentes langues,). Une fois ces étapes effectuées, la différence minimale cliniquement importante (DMCI ou MCID pour Minimal Clinical Important Difference) pour le patient peut être définie afin de fixer un niveau minimum d'amélioration. La DMCI, qui est une notion variable d'un pays à l'autre, peut être calculée par différentes méthodes notamment la méthode des ancres nécessitant une question supplémentaire afin de définir si le patient se juge comme amélioré ou non. Malgré ces données, le manque d'harmonisation dans le choix des scores rend difficile la comparaison de ces scores notamment pour les méta-analyses ou les revues de la littérature. Ainsi pour cette thèse, une cohorte prospective de patients souffrant d'une arthrose de hanche et du genou a été créée afin : 1) de valider différents scores, 2) de définir la DMCI pour ces scores en France, 3) de permettre d'obtenir par un processus statistique, un moyen de connaître le résultat d'un score en n'ayant à sa disposition que le résultat d'un autre score (mapping ou crosswalk). A l'occasion de cette thèse, les score HOOS-12 et KOOS-12, version réduite à 12 items du score HOOS (40 items) et du score KOOS (42 items) ont bénéficié d'une validation en français. La DMCI pour les scores HOOS et Oxford-12 hanche a pu être définie. Cette thèse a permis d'explorer une option de calcul de la DMCI en utilisant un item du score lui-même permettant d'envisager l'usage de la méthode des ancres en l'absence d'une question supplémentaire. Différentes méthodes de mapping ont été testées sur le score HOOS et tous ses dérivés afin de définir la méthode la plus optimale ainsi que le score le plus adapté pour obtenir un mapping avec le score Oxford-12. La méthode equipercentile a permis de réaliser une table de conversion entre le score Oxford-12 Hanche et le score HOOS-12.

Mots clés : hanche, genou, DMCI, Mapping, score, arthroplastie

## RESUME EN ANGLAIS

Patient reported outcome measurement scores (PROMs) are part of modern patient's assessment. For example, in orthopedics they are used to define a baseline level and after a surgical or medical treatment, to define an improvement. These scores require a validation procedure (internal and external validity, Responsiveness, reliability, translation in different languages, assessment of floor and ceiling effect...). Patient's satisfaction after a surgical procedure can be measured through MCID (Minimal Clinical Important Difference). The MCID represents the minimum clinically relevant difference for the patient but is different between countries. MCID can be calculated by various methods, including the anchor method requiring an additional question to define the improvement. Despite these data, the lack of harmonization in the choice of scores makes difficult to compare these scores, especially for meta-analyzes or systematic reviews. In this thesis, a prospective hip and knee cohort was created to: 1) validate different scores, 2) define the MCID of these scores in France, 3) define a statistical process to deduct a score result from another score (Mapping or crosswalk). The HOOS-12 and KOOS-12 scores (abridged versions to 12 items of the HOOS score (40 items) and the KOOS score (42 items)) were validated in French. The MCID for HOOS and hip Oxford scores has been defined for the French population. An option to calculate the MCID using a score item itself was explored, which allowed us to consider using the anchor method in the absence of additional questions. Different methods have been tested to define the most optimal method and the most optimal score for mapping between HOOS and Oxford-12. A conversion table was created between the Oxford-12 hip score and HOOS-12 by the equipercetile method.

Key words: hip, knee, PROM, MCID, Mapping, arthroplasty

## **Valorisation du travail de thèse**

### **1) Articles**

-Quel questionnaire choisir en hanche et genou prothétique ? Comment évaluer la réponse à ces questionnaires ? Qu'est-ce qu'une MCID ou un PASS? Dans la catégorie MISE AU POINT. Maitrise orthopédique numéro 293 avril 2020.

-Do language changes modify the properties of a shortened score and already validated in a complete version? French validation of the HOOS-12 and KOOS-12 scores in primary knee and hip arthroplasty. Article accepté par la revue Orthopaedics and Traumatology: Surgery and Research (OTSR)

-The Minimal Clinically Important Difference in a French speaking population of primary hip replacement: Is it possible to determine the MCID using one item of the PROM with the Anchor strategy? Article accepté par la revue OTSR

-Mapping and crosswalk of the Oxford-12 hip score and different versions of the HOOS score: Is there a pathway for translation? Article soumis au CORR (Clinical Orthopaedics and Related Research®)

### **2) Présentations orales**

-Satisfaction des PTG : notion de MCID, Controverses en orthopédie, Nantes, 2018

-Instabilité fémoro-patellaire. Quel score choisir? Symposium SOFCOT 2019 sous la direction de Franck Remy et Roger Badet. SOFCOT Paris

-Au-delà de la p-value, comment correctement comprendre les statistiques d'un article scientifique et les résultats des études traitant d'intelligence artificielle : comment interpréter un score A venir, conférences d'enseignement Société Française d'Arthroscopie Genève 2020

## **TABLE DES MATIERES**

1. Introduction.....	8
1.1 La chirurgie prothétique de hanche et de genou .....	8
1.1.1 Les arthroplasties ou prothèses .....	8
a. Les arthroplasties de hanche.....	9
b. Les arthroplasties de genou .....	11
1.1.2 Les pathologies traitées .....	13
1.1.3 Objectifs chirurgicaux.....	14
1.2 Les scores d'évaluation.....	14
1.3 Objectifs de la thèse.....	16
2. Validation du score HOOS-12 et KOOS-12.....	18
2.1 Présentation.....	18
2.2 Article.....	19
2.3 Mise en perspective .....	30
3. Différence minimale cliniquement importante.....	33
3.1 Présentation .....	33
3.2 Article.....	34
3.3 Mise en perspective .....	51
4. Mapping.....	54
4.1 Présentation.....	54
4.2 Article.....	55
4.3 Mise en perspective .....	73
5. Synthèse.....	75
5.1 Points forts de la thèse.....	75
5.2 Limites de la thèse .....	76
5.2 Perspectives de la thèse .....	76
6. Conclusion.....	77
7. Bibliographie.....	78
8. Annexes.....	85

8.1 Score HOOS .....	85
8.2 Score Oxford de hanche .....	91
8.3 Score KOOS .....	94
8.4 Score Oxford de genou .....	100
8.5 Avis du comité pour la protection des personnes .....	103



# 1.Introduction

Nicolas Andry (1658-1742), auteur d'un ouvrage consacré aux « difformités » du corps des enfants a donné son nom à cette spécialité « orthopédie » (vient du grec « Orthos » qui signifie droit et « Paidion », qui signifie enfant) (1).

La chirurgie orthopédique et traumatologique s'intéresse à la prise en charge des pathologies ostéo-articulaires.

Spécialité initialement intégrée dans la chirurgie générale, celle-ci s'est progressivement individualisée après la guerre de 1939-1945 dans l'organisation des soins (1).

La chirurgie orthopédique est une chirurgie qui a vu son essor au XX ème siècle, notamment avec le développement des chirurgies prothétiques, l'amélioration des techniques chirurgicales et des matériaux à disposition (1).

Celle-ci reste en développement constant en raison du vieillissement de la population, de la demande fonctionnelle de plus en plus importante des patients.

## 1.1 La chirurgie prothétique de hanche et de genou

### ***1.1.1 Les arthroplasties ou prothèses***

Les arthroplasties ou prothèses sont des composants mis en place pour remplacer une articulation atteinte. Le terme de prothèse totale est utilisé en cas de remplacement complet de l'articulation et de prothèse partielle en cas de changement d'une partie de l'articulation (Figure 1).

On emploie le terme de chirurgie primaire lors de la première pose d'une prothèse chez un patient et de chirurgie de reprise ou de révision en cas de changement prothétique.



**Figure 1 exemple radiographique d'une prothèse totale de hanche et de genou (A1 : radiographie de hanche avant l'intervention A2 : radiographie avec la hanche prothétique, B1 : radiographie de genou avant l'intervention chirurgicale, B2 : radiographie de genou avec le genou prothétique)**

### ***a. Les arthroplasties de hanche***

La première chirurgie de hanche pour raideur articulaire a eu lieu en 1826 par John Barton (2). Le développement de nouveau d'une raideur fit penser certains à interposer du matériel.

C'est à Marius Nygaard Smith Petersen que l'on doit l'invention de la véritable première prothèse de hanche en 1923 avec l'implantation d'une première prothèse en verre (2). Malheureusement, toutes ces prothèses furent cassées. L'histoire veut qu'en se rendant en 1937 chez son dentiste, celui-ci utilisant le Vitallium, suggéra l'idée à Smith Petersen d'utiliser ce composant. Cela marque le passage en 1938 d'un simple matériel d'interposition à un matériel définitif avec plus de 1000 prothèses implantées entre 1938 et 1952 (2).

En France, les frères Judet développent dès 1936 l'utilisation de prothèses avec l'ajout d'une structure métallique centrale (Figure 2) (2).

**Figure 2 Prothèse de Judet**

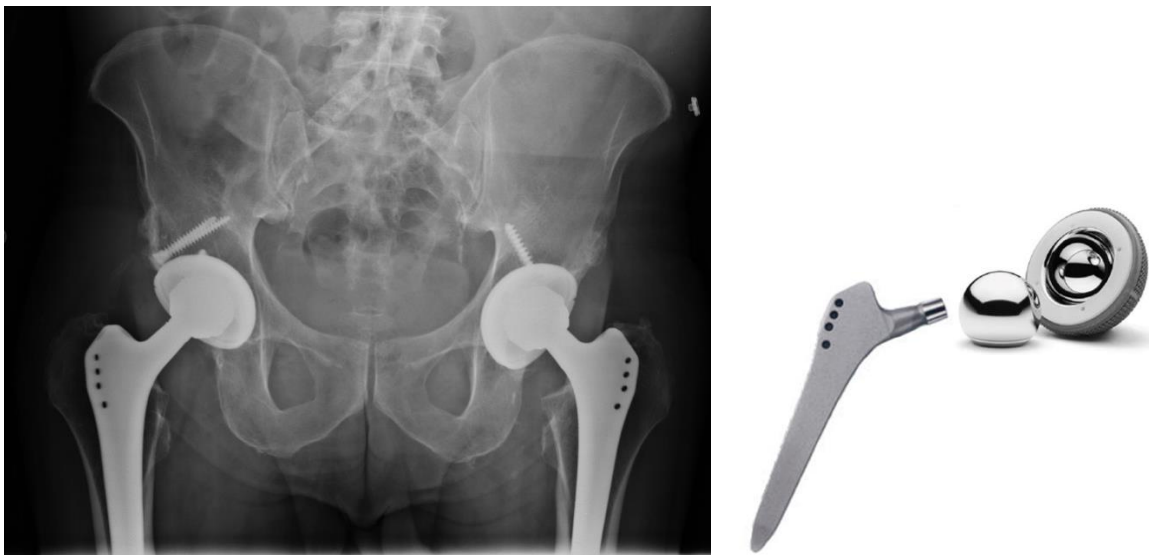
En Angleterre Sir John Charnley a introduit le polyéthylène pour la pièce acétabulaire (appelée



aussi cupule), l'utilisation du poly-méthyle méthacrylate de méthyle (appelé ciment dans le langage courant) pour fixer les prothèses dans l'os, là aussi en utilisant un procédé de dentisterie (2).

La deuxième partie du XX<sup>ème</sup> siècle a permis le développement de meilleurs matériaux, d'une meilleure fixation osseuse.

On développa des fixations sans ciment, de nouveaux couples de frottement (Figure 3).



**Figure 3 Exemple d'une prothèse totale de hanche à couple de frottement métal-métal**

Dans un article de 2018, Ferguson et al. (3) estime que plus d'un million de prothèses sont posées par an avec une durée de vie (pas de changement d'implant prothétique) de ces arthroplasties à plus de 25 ans entre 77 % et 58 % (4).

La moyenne d'âge des patients prise en charge pour une prothèse est de 69 ans [61-76] au Royaume uni (3), par exemple.

En 2018 en France, il ya eu 148 965 actes de prothèse primaires dont 124 251 prothèses totales (5).

Entre 2018 et 2050, Erivan et al, ont prévu une augmentation entre 41,9 et 114,3 % pour les prothèses de hanche primaires(5).

### ***b. Les arthroplasties de genou***

Des essais ont débutés au XIX ème siècle, mais le développement des prothèses totales de genou (PTG) n'a véritablement commencé qu'à partir des années 1950 avec l'introduction d'une prothèse charnière (6). La formation du groupe GUEPAR en France en 1968 (Groupe pour l'Utilisation et l'étude des prothèses articulaires ) a permis également le développement de cette technologie , toujours à partir d'une PTG charnière ( Figure 4) (6) .

Ces prothèses charnières fonctionnent sans ligament. Les deux pièces prothétiques sont reliées l'une à l'autre par un lien mécanique.

La fixation osseuse des PTG a utilisé et utilise encore majoritairement le ciment, introduit par Charnley pour les prothèses de hanche.



**Figure 4 Prothèse charnière**

L'évolution du concept prothétique s'est fait vers des prothèses plus conservatrices des structures ligamentaires et osseuses (6) sans moyen d'union entre les pièces fémorales et tibiales, on parle aussi de prothèse à glissement (Figure 5).



**Figure 5 Exemple de prothèse totale de genou**

Les prothèses totales de genou sont désormais plus fréquentes que les PTH. Il y a plus de 700000 interventions aux USA par an contre 370000 PTH (7).

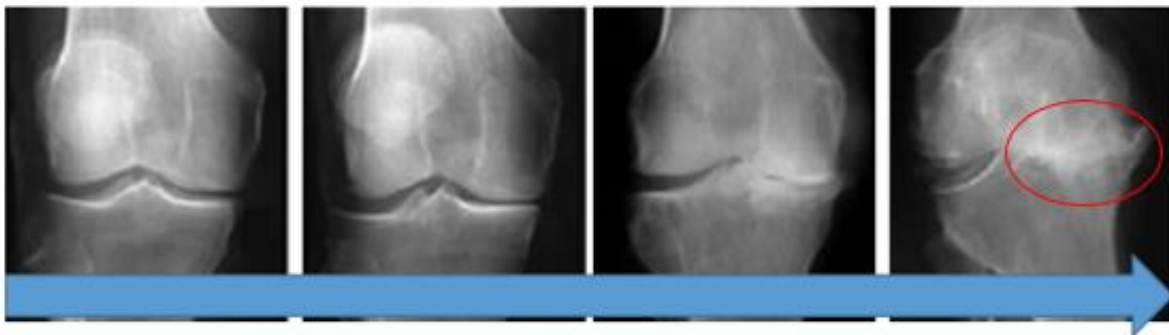
Selon Evans et al. (8), la moyenne d'âge des patients est de 65 ans, avec une durée de vie de ces implants à 25 ans de 82% (8).

En 2018, la chirurgie prothétique de première intention représente en France plus de 113 600 procédures (9). Les prothèses de première intention augmentaient de 32,2 % entre 2012 et 2018 en France(9).

Entre 2018 et 2050, les scénarios prévoient une augmentation de 30,8 % à 152,8 %selon Erivan et al. (9).

### **1.1.2 Les pathologies traitées**

L'arthrose de hanche ou coxarthrose et l'arthrose de genou ou gonarthrose sont les principales causes de chirurgie prothétique pouvant correspondre à plus de 90% des indications selon les pays (3,7). L'arthrose correspond à la destruction progressive du cartilage articulaire (Figure 6) (3,7). Celles-ci sont appelées primaires lors du vieillissement normal de l'articulation. Elles sont appelées secondaires lors de la destruction plus rapide de l'articulation en raison d'une pathologie préexistante comme par exemple un traumatisme (arthrose post-traumatique).



**Figure 6 Exemple d'une dégradation arthrosique d'un genou avec l'apparition progressive d'un pincement de l'articulation à la partie interne du genou (entouré en rouge)**

Les autres étiologies conduisant à une arthroplastie sont les maladies inflammatoires (Polyarthrite rhumatoïde, spondylarthrite ankylosante..), l'ostéonécrose, les fractures, les pathologies tumorales ou les infections (3,7).

### **1.1.3 Objectifs chirurgicaux**

Les objectifs initiaux de ces chirurgies prothétiques étaient de restaurer une articulation plus mobile, moins douloureuse, objectif le plus souvent atteint.

Un des objectifs chirurgicaux souvent retrouvé est une durée de vie de ces implants la plus longue possible, permettant d'éviter une réintervention pour usure des composants prothétiques chez ces patients, faisant initialement privilégier ces interventions chez des patients plus âgés.

Cet objectif est presque atteint avec des durées de vie de ces implants à 25 ans tout à fait acceptables (4,8) permettant d'élargir les possibilités thérapeutiques.

Les objectifs ont donc évolué avec l'augmentation de l'usage de ces arthroplasties, notamment en raison des demandes fonctionnelles plus importantes des patients.

Désormais un des objectifs est une articulation dite « oubliée », à savoir un patient n'ayant aucune douleur et ne faisant pas la différence entre son articulation native et son articulation prothétique (10-13).

Ceci n'est actuellement pas le cas pour les genoux où l'on retrouve 15 à 20% de patients insatisfait du résultat (7).

Ces objectifs nécessitent donc de mieux comprendre le ressenti du patient, avant son intervention pour comprendre sa demande et ensuite après l'intervention afin d'améliorer nos pratiques.

## **1.2 Les scores d'évaluation**

Ces chirurgies comme toute procédure nécessitent une évaluation afin de savoir si elles doivent être continuées, si elles apportent un réel confort pour le patient. L'évaluation des patients après une chirurgie par exemple ou un traitement sont marquée par l'usage de scores permettant de quantifier un résultat (10).

On considère 3 types de scores ou questionnaires à l'usage des patients : les questionnaires faits par un tiers évaluateur, les auto-questionnaires ou PROM (Personal Reported Outcome Measurement), les questionnaires hybrides ou mixtes correspondant à des questionnaires avec une partie faite par un tiers évaluateur et une partie faite par le patient lui-même.

Les questionnaires faits par un tiers étaient les premiers questionnaires utilisés mais ces questionnaires introduisent un biais car le patient pouvait être influencé par le tiers examinateur. Certains de ces questionnaires plus anciens s'intéressaient également plus au résultat qu'à la qualité de vie.

Prenons par exemple le score Postel Merle d'Aubigné (PMA) (initialement créé sur 12 points puis modifié en 1970 sur 18 points) (Figure 7). Celui-ci est retrouvé dans les chirurgies de hanche, allant de 0 à 18 points en 3 catégories de 6 points : marche, douleur, mobilité (14).

DOULEUR		MOBILITE			MARCHE-STABILITE
		Pas d'attitude vicieuse	Attitude vicieuse en flexion ou RE	Attitude vicieuse en ABD, ADD ou RI	
6	Aucune	Flexion $\geq 90^\circ$	Aucune	Aucune	Normale ou illimitée
5	Rare Légère	$70^\circ$ à $90^\circ$	Aucune	Aucune	Limitée ou légère boiterie si prolongée Canne pour longues distances Pas d'instabilité
4	Après 30mn à 1h de marche	$50^\circ$ à $70^\circ$	Abaisser la note de 1 point	Abaisser la note de 2 points	Canne pour sortir Boiterie nette Légère instabilité
3	Après 10 à 20mn de marche	$30^\circ$ à $50^\circ$			Canne en permanence Instabilité
2	Avant 10mn de marche	$< 30^\circ$			2 cannes
1	Immédiatement à La marche				Béquilles
0					Marche impossible

**Figure 7 Score de Postel Merle d'Aubigné**

En même temps que l'évolution de la société, l'évolution des objectifs des patients et des chirurgiens a introduit de plus en plus la notion de qualité de vie dans ces questionnaires, ce que nous ne retrouvons pas par exemple dans ce score PMA (14).

A ce jour, l'évaluation de patients repose donc plus fréquemment sur les PROM permettant d'éviter le biais du tiers évaluateur et sur des questionnaires évaluant la qualité de vie (10,11,12,13,15).

C'est le cas par exemple du questionnaire HOOS (16,17), avec des questions comme « Avez-vous modifié votre façon de vivre pour éviter les activités qui pourraient aggraver votre problème de hanche ? »



### 1.3 Objectifs de la thèse

Les scores d'autoévaluation (PROM : Patient Reported Outcome Measurement) font donc partie désormais de l'arsenal quotidien d'évaluation des patients en orthopédie (10,15) tant en préopératoire pour évaluer un état de base, qu'en postopératoire pour évaluer l'efficacité d'un traitement.

Riddle et al. (15) en 2008 sur l'usage des scores d'évaluation en hanche et genou retrouvaient en ce sens plus de 20 questionnaires utilisés en pathologie de la hanche et plus de 14 pour les pathologies du genou.

Ce travail de thèse initié en 2016, s'est intéressé aux scores d'évaluation en chirurgie prothétique de la hanche et du genou.

En 2016 et 2017, les publications sur Pubmed concernant les prothèses de hanche ou de genou hanche et genou sont nombreuses. On retrouve en effet plus de 600 publications en 2016 et plus de 900 publications en 2017 sur l'arthroplastie de hanche. Ces publications utilisent des questionnaires d'évaluation avec un mélange d'auto-questionnaire et de score d'évaluation par le praticien. Si l'on regarde uniquement les auto-questionnaires utilisés, 2 scores apparaissent régulièrement utilisés, le score HOOS (16,17) (Annexe 1) (50 citations en 2016, 45 citations en 2017) et le score Oxford-12 (18,19) (Annexe 2) (61 citations en 2016, 49 citations en 2017).

De même, on retrouve en effet plus de 600 publications en 2016 et plus de 1100 publications en 2017 sur l'arthroplastie de genou. Ces publications utilisent des questionnaires d'évaluation avec un mélange d'auto-questionnaire et de score d'évaluation par le praticien. Si l'on regarde uniquement les auto-questionnaires utilisés, 2 scores apparaissent régulièrement utilisés, le score KOOS (20,21) (Annexe 3) (51 citations en 2016, 52 citations en 2017) et le score Oxford-12 (22,23) (Annexe 4) (87 citations en 2016, 89 citations en 2017).

Ces scores apparaissent donc comme de scores largement utilisés. C'est la raison pour laquelle, nous les avons choisis pour ce travail de thèse débuté en 2016.

Ce travail de thèse s'est articulé autour de trois problématiques : comment valider un score ? Comment prendre en compte le résultat d'un score ? Comment reconstruire un score à partir d'un autre score (méthode de mapping) ?

Afin de répondre à ces différentes questions, une cohorte prospective a été créée avec la rédaction d'un protocole de recherche spécifique permettant d'obtenir l'accord du comité

pour la protection des personnes (Annexe 5). L'accord a été obtenu en 2018 permettant de débiter un recueil prospectif.

Cette thèse est ainsi composée de trois parties : un article de validation d'un score raccourci pour le scores HOOS-12 et KOOS 12, un article sur la différence minimale cliniquement importante en France pour les scores HOOS et Oxford-12 hanche et un article de Mapping entre le score Oxford-12 hanche et le score HOOS-12.

## 2. Validation du score HOOS-12 et KOOS-12

### 2.1 Présentation

L'usage d'un score est soumis à des recommandations de validation à la fois lors de la création d'un score mais également une validation transculturelle lors de son passage dans une autre langue que sa langue d'origine.

Ces étapes de validation sont bien codifiées (24), notamment avec les recommandations COSMIN (25).

L'usage de plus en plus régulier de ces scores notamment en orthopédie pour l'évaluation des patients a mis en évidence des difficultés de remplissage pour les patients pour les scores comprenant beaucoup d'items (26,27). Ainsi par exemple, les scores HOOS (16,17) et KOOS (20,21) autorisent des données manquantes pour simplifier ce problème.

Les équipes à l'origine de ces scores, ont exploré une autre possibilité que le recours à des méthodes ad hoc en cas de données manquantes, la réduction du nombre d'items notamment en utilisant des procédés statistiques permettant d'exclure les items considérés comme redondants (28-30).

Le score HOOS et KOOS avec une forme à 12 items est apparue en 2018 (28-30) Ce nouveau format a été conçu par les auteurs afin d'être plus exhaustif que les autres formats raccourcis (format PS (Physical function Short-form) (31-33), format JR (Joint Replacement) (34,35).

Ce format à 12 items est apparu intéressant en raison de son format proche du score Oxford-12 (18,19,22,23) mais ce nouveau format nécessitait une validation transculturelle afin de pouvoir l'utiliser. En effet le changement de langage peut modifier l'intérêt de ce nouveau score.

Aussi nous avons mené une étude prospective afin de déterminer si : 1) cette nouvelle version à 12 items possède les mêmes qualités en langue française que le score initial à 40 et 42 items 2) valider l'utilisation du score HOOS-12 et KOOS-12 en langue française sur une population d'arthroplastie primaire de hanche et de genou (validité, fiabilité, sensibilité au changement).

Cet article a été accepté par la revue Orthopaedics and Traumatology: Surgery and Research (OTSR).

## 2.2 Article

Does change in language change the properties of a shortened score previously validated in its complete version? Validation of the French versions of the HOOS-12 and KOOS-12 scores in primary knee and hip arthroplasty.

S. Putman<sup>a,b,c\*</sup>, H. Migaud<sup>a,b</sup>, G Pasquier<sup>a,b</sup>, J Girard<sup>a,b</sup>, Preda C<sup>e,f</sup>, A Duhamel<sup>c,d</sup>

a Univ. Lille, CHU Lille, ULR 4490, Hôpital Salengro, F-59000 Lille, France

b Département Universitaire de Chirurgie Orthopédique et Traumatologique, Hôpital Roger Salengro, CHU de Lille, Place de Verdun, 59037 Lille, France

c Univ. Lille, CHU Lille, EA 2694 - Metrics: Evaluation des Technologies de Santé et des Pratiques Médicales, F-59000 Lille, France

d CHU Lille, Unité de Méthodologie et Biostatistiques, F-59000 Lille, France

e Laboratory of Mathematics Paul Pain Levé, UMR CNRS 8524, University of Lille, France

f Lille Catholic Hospitals, Biostatistics Department Delegation for Clinical Research and Innovation, Lille Catholic University, Lille, France

\* Corresponding author: Sophie Putman, Département Universitaire de Chirurgie Orthopédique et Traumatologique, Hôpital Roger Salengro, CHU de Lille, Place de Verdun, 59037 Lille, France  
Email: [sophie.putman@wanadoo.fr](mailto:sophie.putman@wanadoo.fr)

### Abstract

### Background

The HOOS and KOOS scoring questionnaires comprise respectively 40 and 42 items; a shorter 12-item version was recently developed, but remains to be validated in a French-speaking population. We therefore conducted a prospective study 1) to determine whether the new 12-item versions in French are equivalent to the longer HOOS and KOOS versions, and 2) to validate the French-language HOOS-12 and KOOS-12 patient-reported outcome measures in a population of primary total hip and knee arthroplasty: validity, reliability, and responsiveness.

### Hypothesis

The change in language in a score already validated in its long version does not alter its properties in the short version.

### Material and method

One hundred patients (59 male, 41 female) undergoing primary total hip arthroplasty and 100 patients (43 male, 57 female) undergoing primary total knee arthroplasty were prospectively

included. They filled out the original HOOS or KOOS questionnaires, their simplified versions (PS: Physical function Short form; JR: Joint Replacement) and the short HOOS-12 and KOOS-12 versions, and also the Oxford-12 score assessing the affected joint, preoperatively, then at 6-12 months.

## **Results**

The 100% response rate confirmed ease of use. There were no redundant items. There were strong correlations between the 12-item and longer versions ( $>0.9$ ). The HOOS-12 and KOOS-12 scores were reliable and valid: 1) there were no ceiling or floor effects for pre- or post-operative KOOS-12 scores, although a ceiling effect was found for HOOS-12 postoperatively (20% of patients having maximum scores of 100); 2) internal consistency was confirmed, with Cronbach alpha  $>0.8$ ; 3) external consistency between Oxford-12 and HOOS-12/KOOS-12 was excellent, with Pearson correlation coefficient  $> 0.8$ . Sensitivity to pre-/post-operative change was confirmed, with effect size  $> 0.8$ .

## **Discussion**

The present study confirmed the usefulness of this new 12-item form for HOOS and KOOS. Properties were identical between the French- and English-language versions, authorizing everyday use of these simpler versions.

Keywords: PROM, Hip, Knee, KOOS, HOOS, Total hip arthroplasty, Total knee arthroplasty

Level of evidence: IV, prospective study without control group

## **1. Introduction**

Scoring systems, and patient-reported outcome measures (PROM) in particular, are recommended for patient assessment, and notably for postoperative quality of life [1]. The Hip Osteoarthritis Outcome Score (HOOS) and Knee Osteoarthritis Outcome Score (KOOS) were respectively developed in 2003 and 1998 [2,3]. They comprise respectively 40 and 42 items, in 5 domains: pain (P), symptoms (S), function and activities of daily living (F), sport and leisure activities (SP), and quality of life (Q). Each domain is scored out of 100 points (0 to 100), and results are presented for the 5 domains, without global score.

Simplified versions were developed: PS (Physical function Short form) and JR (Joint Replacement) [4-7]. In 2018, the original team developed 12-item versions, HOOS-12 and KOOS-12 [8-10], as the short PS and JR versions [4-7] were felt not to provide as full an assessment as the original versions. Redundant items, determined on statistical analysis, were deleted.

Ornetti et al. [11,12] validated French-language versions of HOOS and KOOS and of the short PS version [13]. Item formulations were unchanged in developing the 12-item versions, so no new translation was needed, having already been validated. Nevertheless, transition to French may affect certain properties such as validity, reliability and responsiveness, with potential saturation: i.e., prematurely hitting a ceiling or floor beyond which patients can no longer be differentiated. To our knowledge, these possible effects have not been assessed, and we therefore conducted a prospective study 1) to determine whether the new 12-item versions in French are equivalent to the longer HOOS and KOOS versions, and 2) to validate the French-language HOOS-12 and KOOS-12 questionnaires in a population of primary total hip and knee arthroplasty: validity, reliability, and responsiveness. The study hypothesis was that the change in language in a score already validated in its long version does not alter its properties in the short version, and that the 12-item HOOS and KOOS questionnaires were utilizable in their French versions.

## **2. Material and methods**

### **2.1 Patients**

A prospective study was conducted in the University Hospital of Lille, France (registration n° AC 2017-A01911-52, Clinical Trials NCT04057651). One hundred patients (59 male, 41 female; mean age, 61 ± 11 years [range, 30-85 years]) undergoing primary total hip arthroplasty and 100 patients (43 male, 57 female; mean age, 67±11 years [range, 48-89 years]) undergoing primary total knee arthroplasty were included. All patients were adult (>18 years), with osteoarthritis or osteonecrosis, and provided written consent. Exclusion criteria comprised: lack of national health insurance cover, age <18 years, incapacity to provide consent, pregnancy, and body-mass index > 35 or <18.

### **2.2 Methods**

The HOOS-12 [9] and KOOS-12 [10] self-administered questionnaires, comprising 12 items scored 0-4, were distributed by mail prospectively in paper version on eve of surgery and at a

minimum 6 months' follow-up. The Oxford-12 [14-16] questionnaire was administered at the same time points, and also the long versions of HOOS [2] and KOOS [3].

HOOS-12 and KOOS-12 comprise 3 domains: pain (P) (4 items), function and activities of daily living (F) (4 items), and quality of life (Q) (4 items) (Appendices 1 and 2). Scores per domain were calculated out of 100 (0-100), as in the original versions, with the mean of the 3 as final HOOS-12/KOOS-12 score. A "global" HOOS KOOS long-version score was likewise calculated out of 100 as the mean of the 5 domains (P+S+F+SP+Q).

## **2.3 Assessment**

Statistical analyses used SPSS software. All tests were 2-tailed, with 5% first-order risk.

Main endpoint assessment was based on the criteria defined by Bouletreau et al. [17] and the COSMIN group [18] : validity, reliability, and responsiveness.

\*Validity concerns relevance to the phenomenon under study, and assesses item quality in terms of redundancy. Principal component analysis (PCA) was performed to determine whether items were distributed along a single dimension or as subgroups [19]. Validity also concerns external consistency, assessed by the correlation with Oxford-12 scores (also comprising 12 items) and with the PS and JR scores derived from the HOOS and KOOS.

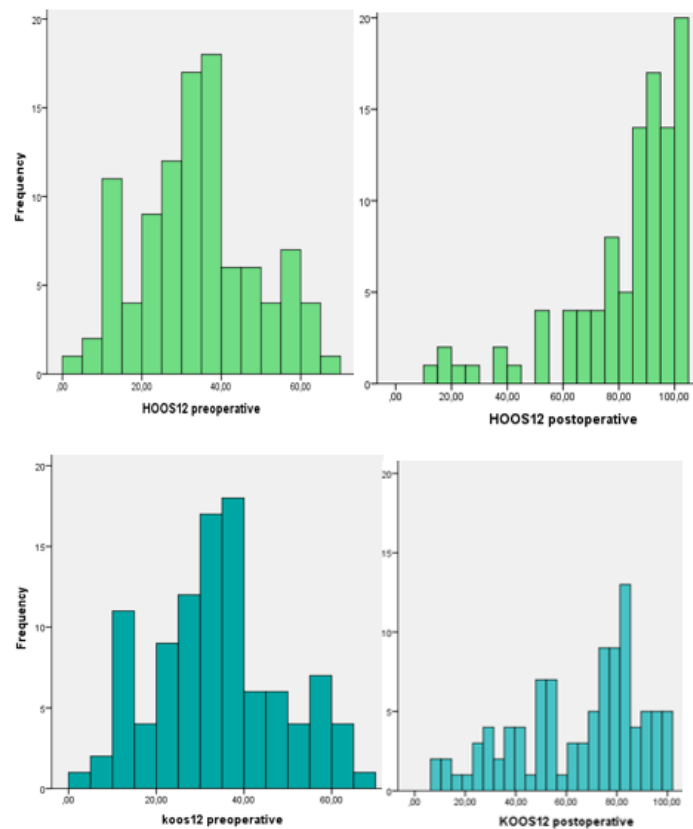
\* Reliability was assessed as internal consistency on Cronbach alpha, considered good if >0.7 [20].

\* Responsiveness was assessed by the difference between preoperative and 6-month scores on Cohen's kappa, with 0.2 as "weak" effect, 0.5 "moderate", and 0.8 "strong" [21].

## **3. Results**

### **3.1 Distribution of results**

Response rate was 100%. There were no floor or ceiling effects preoperatively, but HOOS-12 showed a ceiling effect at follow-up (20% of patients scoring the maximum, 100 points) (Figure 1).



**Figure 1** Distribution of HOOS-12 and KOOS-12 scores in 100 patients before and after surgery; ceiling effect for postoperative HOOS-12.

### 3.2 Validity

PCA revealed a single dimension for both HOOS-12 [9] and KOOS-12 [10], with respectively 50% and 51% of variance explained by the first principal component. There were no redundant items.

External consistency assessment found good correlation for HOOS-12 [9] and KOOS-12 [10] with Oxford-12 score, both pre- and post-operatively (Pearson  $r$ , > 0.5). The 12-item versions correlated more strongly with the “global” scores on the HOOS and KOOS long versions, both pre- and post-operatively, than with the PS and JR derivatives. All correlations were significant ( $p < 0.05$ ) (Tables 1 and 2).



Correlation of HOOS-12 [9] with:	Preoperative	Postoperative
Oxford-12[15]	0.821	0.906
HOOS global (5 HOOS domains) [2]	0.942	0.979
HOOS JR [6]	0.803	0.950
HOOS PS [4]	0.866	0.905

Table 1: Correlation between HOOS-12 [9] and Oxford-12 [14], HOOS global score and JR (Joint Replacement) and PS (Physical function Short form) forms, pre- and post-operatively (6 months).

Correlation of KOOS-12 [10] with:	Preoperative	Postoperative
Oxford-12 [16]	0.828	0.886
KOOS global (5 KOOS domains) [3]	0.953	0.974
KOOS JR [7]	0.920	0.937
KOOS PS [5]	0.933	0.893

Table 2: Correlation between KOOS-12 [10] and Oxford-12 [16], KOOS global score [3] and JR (Joint Replacement) and PS (Physical function Short form) forms, pre- and post-operatively (6 months).

### 3.2 Reliability

Cronbach alpha was good (>0.7), both pre- and post-operatively (Table 3).

Coefficient de Cronbach	Preoperative	Postoperative
HOOS-12 [9]	0.893	0.952
HOOS-12 Pain [9]	0.754	0.897
Hoos-12 Function [9]	0.802	0.913
HOOS-12 Quality of life [9]	0.827	0.905
KOOS-12 [10]	0.909	0.948
KOOS-12 Pain [10]	0.772	0.906
KOOS-12 Function [10]	0.813	0.864
KOOS-12 Quality of life [10]	0.782	0.891

Table 3: Cronbach alpha [18] for pre- and post-operative (6 months) scores; good reliability if alpha > 0.7

### 3.3 Responsiveness

Pre- and post-operative scores were significantly different. The effect size was strong (>0.8) (Table 4): i.e., HOOS-12 and KOOS-12 were responsive to change.

	Preoperative (*)	Postoperative (*)	P (+)	Effect size (**)
<b>HOOS-12 [9]</b>	33 +/-14 [4-68]	82 +/-20 [14-100]	<0.001	3
<b>HOOS-12 Pain [9]</b>	38+/-16[0-75]	86+/-20[18-100]	<0.001	2.9
<b>Hoos-12 Function [9]</b>	37+/-16[0-75]	83+/-21[12-100]	<0.001	2.7
<b>HOOS-12 Quality of life [9]</b>	24+/-17[0-75]	77+/-26 [0-100]	<0.001	3
<b>KOOS-12 [10]</b>	32+/-16[0-75]	65+/-24[9-100]	<0.001	2 .06
<b>KOOS-12 Pain [10]</b>	36+/-17[0-85]	71+/-25 [18-100]	<0.001	2.05
<b>KOOS-12 Function [10]</b>	35+/-19[0-85]	65+/-25[0-100]	<0.001	1.57
<b>KOOS-12 Quality of life [10]</b>	25+/-18 [0-81]	59+/-29[0-100]	<0.001	1.88

Table 4: Responsiveness\*: mean +/- standard deviation [range]; +: matched Student test; \*\*:on Cohen criteria, effect size 0.2 is “weak”, 0.5 “moderate” and 0.8 “strong” [19].

#### 4. Discussion

The present study demonstrated that the simplified 12-item HOOS and KOOS PROMs were utilizable in their French-language versions. Score properties were unaffected by the translation into French. There were no redundant items: i.e., each item analyzed one aspect of the hip or knee. Principal components analysis revealed a single dimension: i.e., the scores assessed the hip or knee globally. The items selected for the transition from the long to the 12-item versions showed strong correlation with global long-version scores: i.e., the authors’ intention to respect the spirit of the original scores in the new short versions was conserved in the French translation. Marot et al. [22] showed that KOOS scores varied with age, gender and body-mass index. It was therefore necessary to validate the new French version, as the translation might betray the intention of the authors of the 12-item versions, despite the initial work by Ornetti et al.[11-13].

Perez et al. [23] did not show the number of items to affect understanding, but nevertheless suggested that fatigue and loss of concentration might affect results quality. Deleting redundant items from the 12-item questionnaires, without affecting the initial properties, makes for greater ease of use. The original authors estimated long-form administration time to average 10 minutes [2,3]; Gudbergsen et al. [24] reported 7 minutes. The 12-item versions are still exhaustive, but save time and avoid having missing data.

The scores proved reliable, valid and sensitive to change. In the present series, data for floor and ceiling effects were comparable for the KOOS-12 postoperatively [10]. In contrast, the ceiling effect rate for HOOS-12 at 1 year was 12-14% for Gandek et al. [9], which was not strictly a ceiling effect as in the present study, but nevertheless not far from the present rate of 20%. This rate was lower than for the other scores derived from HOOS (27-36% for HOOS JR), testifying to the interest of the 12-item version in differentiating patient progression profiles [6,9]. Bourlez et al. [25] showed that the KOOS daily living and quality of life domains showed ceiling and floor effects. The new 12-item score avoids this limitation of the 5-domain long version.

The present Cronbach alpha results were similar to those of Gandek et al. [9,10], testifying to item homogeneity in the construction of the score and absence of change according to type of population (French- versus English-speaking). Like in the present series, Cronbach's alpha was weaker in the pain (P) questionnaires (close to 0.7) [9,10].

Again like for Gandek et al. [9,10], effect size was greater for the hip than the knee, but still strong (>0.8), testifying to responsiveness. The difference between French- and English-speaking populations did not alter good differentiation according to treatment.

The present study had certain limitations. 1) Reproducibility was not assessed. For PROMs, only intra-observer reproducibility applies, and even this is a question of debate [26]. It was tested on the cross-cultural validation of HOOS and KOOS by Ornetti et al. [11-12]. 2) Sample size is open to discussion, as suggested by Kayaalp et al. [26] in their Letter to the Editor about the French-language validation of the IKS score. Rouquette and Falissard [27] recommended a minimum 300 subjects, but that was for creation and validation rather than cross-cultural validation of a score. The present series was equivalent to or greater than various other reports of score translation into French [15,16,28,29].

## **5. Conclusion**

The simplified 12-item versions of HOOS and KOOS (HOOS-12 and KOOS-12) proved reliable, valid and responsive to change in their French versions. The authors had sought to conserve the philosophy of the original HOOS and KOOS, and translation into French did not affect this. The 12-item versions are easier to use, and can be used in French.

## References

- 1) Canovas F, Dagneaux L. Quality of life after total knee arthroplasty. *OrthopTraumatolSurg Res* 2018;104(1S):S41–S46.
- 2) Nilsson AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS) - validity and responsiveness in total hip replacement. *BMC MusculoskeletDisord* 2003;4:10.
- 3) Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee injury and osteoarthritis outcome score (KOOS). Development of a self-administered outcome measure. *J Orthop Sports PhysTher* 1998;28:88-96.
- 4) Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan, PG. An OARSI/OMERACT initiative: the development of a short measure of physical function for hip OA. HOOS-Physical Function Shortform (HOOS-PS). *Osteoarthritis Cartilage* 2008;16:551-9.
- 5) Perruccio AV, Lohmander LS, Canizares M, Tennant A, Hawker GA, Conaghan PA. The development of a short measure of physical function for knee OA. KOOS-Physical Function Short-form (KOOS-PS) – an OARSI/OMERACT initiative. *Osteoarthritis Cartilage* 2008;16:542-50.
- 6) Lyman S, Lee YY, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin OrthopRelatRes* 2016;474:1472–82.
- 7) Lyman S, Lee YY, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: A Short-form Knee Arthroplasty Outcomes Survey. *ClinOrthopRelat Res* 2016;474:1461-71.
- 8) Gandek B, Roos EM, Franklin PD, Ware JE Jr. Item selection for 12-item short forms of the Knee injury and Osteoarthritis Outcome Score (KOOS-12) and Hip disability and Osteoarthritis Outcome Score (HOOS-12). *Osteoarthritis Cartilage* 2019;27:746-753.
- 9) Gandek B, Roos EM, Franklin PD, Ware JE Jr. A 12-item short form of the Hip disability and Osteoarthritis Outcome Score (HOOS-12): tests of reliability, validity and responsiveness. *Osteoarthritis Cartilage*.2019;27:754-761.
- 10) Gandek B, Roos EM, Franklin PD, Ware JE, Jr. A 12-item short form of the Knee injury and Osteoarthritis Outcome Score (KOOS-12): Tests of reliability, validity and responsiveness. *Osteoarthritis Cartilage* 2019;27:762-770.
- 11) Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, Roos EM, Guillemin F, Maillefert JF Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. *Osteoarthritis Cartilage* 2010;18:522-529.
- 12) Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, Roos EM, Guillemin F, Maillefert JF. Cross-cultural adaptation and validation of the French version of the Knee injury

and Osteoarthritis Outcome Score (KOOS) in knee osteoarthritis patients. *Osteoarthritis Cartilage* 2008;16:423-8.

13) Ornetti P, Perruccio AV, Roos EM, Lohmander LS, Davis AM, Maillefert JF Psychometric properties of the French translation of the reduced KOOS and HOOS (KOOS-PS and HOOS-PS). *Osteoarthritis Cartilage* 2009;17:1604-8.

14) Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996;78:185-90.

15) Delaunay C, Epinette JA, Dawson J, Murray D, Jolles BM. Cross-cultural adaptations of the Oxford-12 HIP score to the French speaking population *Orthop Traumatol Surg Res* 2009;95:89-99.

16) Jenny JY, Diesinger Y. Validation of a French version of the Oxford knee questionnaire. *Orthop Traumatol Surg Res* 2011;97:267-71.

17) Bouletreau A, Chouanière D, Wild P, Fontana JM. Construire, traduire et valider un questionnaire propos d'un exemple EUROQUEST, service épidémiologie INRS, 1999;178:ISSN 0397.4529. <http://www.inrs.fr/dms/inrs/Publication/B-5-2-040-3566-01/ns178.pdf>.

18) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-45.

19) Fleiss JL. *The design and analysis of clinical experiments* New York: John Wiley & Sons, 1986.

20) Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.

21) Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition. Hillsdale: Lawrence Erlbaum, New York 1988.

22) Marot V, Murgier J, Carrozzo A, Reina N, Monaco E, Chiron P, Berard E, Cavaignac E. Determination of normal KOOS and WOMAC values in a healthy population. *Knee Surg Sports Traumatol Arthrosc* 2019;27:541–548.

23) Perez JL, Mosher ZA, Watson SL, Sheppard ED, Brabston EW, et al. Readability of Orthopaedic Patient-reported Outcome Measures: Is There Fundamental Failure to Communicate? *Clin Orthop Relat Res* 2017;475:1936-1947.

24) Gudbergson H, Bartels EM, Krusager P, Wæhrens EE, Christensen R, et al. Test-retest of computerized health status questionnaires frequently used in the monitoring of knee osteoarthritis: a randomized crossover trial. *BMC Musculoskelet Disord.* 2011;12:190. doi:10.1186/1471-2474-12-190.

- 25) Bourlez J, Canovas F, Duflos C, Dagneaux L. Are modern knee outcomes scores appropriate for evaluating patellofemoral degeneration in osteoarthritis? Evaluation of the ceiling and floor effects in knee outcomes scores. *OrthopTraumatolSurg Res* 2019;105:599–603.
- 26) Kayaalp ME. Comment on: "French adaptation of the new Knee Society Scoring System for total knee arthroplasty" by Debette C., Parratte S., Maucort-Boulch D., Blanc G., Pauly V., Lustig S., Servien E., Neyret P., Argenson JN. published in *OrthopTraumatolSurg Res* 2014;100(5):531-4. *OrthopTraumatolSurg Res* 2018;104:733–734.
- 27) Rouquette A, Falissard B. Sample size requirements for the internal validation of psychiatric scales. *Int J MethodsPsychiatrRes* 2011;20:235-49.
- 28) Debette C, Parratte S, Maucort-Boulch D, Blanc G, Pauly V, et al. French adaptation of the new Knee Society Scoring System for total knee arthroplasty. *OrthopTraumatolSurg Res* 2014;100:531-4.
- 29) Klouche S, Giesinger JM, Sariali EH. Translation, cross-cultural adaption and validation of the French version of the Forgotten Joint Score in total hip arthroplasty. *Orthop TraumatolSurgRes* 2018;104:657-661.

## 2.3 Mise en perspective

La validation d'un score est actuellement une étape bien codifiée, nécessitant de respecter les critères fixés par le Groupe COSMIN (25) : validité (validity), fiabilité (reliability), sensibilité au changement (responsiveness) (figure 8).



Figure 8 Validation d'un score selon Mokkink et al. COSMIN (Consensus-based Standard for the selection of health Measurement Instruments) (25)

Détaillons ces étapes en quelques mots.

La validité est composée de plusieurs paramètres : validité d'apparence (face validity), validité contre critère (criterion validity), validité de contenu (content validity), validité du construit ou validité de structure (construct validity).

La validité d'apparence résulte du jugement subjectif d'experts du domaine, experts le plus souvent à l'origine de la création du score. On considère en effet que si les créateurs du score sont des experts du domaine, la validité d'apparence est respectée.

L'étude de sa validité contre critère, à savoir sa comparaison par rapport à une référence appelée « gold standard » est parfois difficile à définir en raison de la multiplicité des scores.

La validité de structure s'intéresse à la pertinence du questionnaire vis-à-vis du phénomène étudié (qualité des items, score discriminant, absence de redondance)

Un score est fiable s'il donne des résultats reproductibles dans des situations identiques : reproductibilité de la mesure par différents observateurs (reproductibilité inter-observateur), reproductibilité de la mesure dans le temps (reproductibilité intra-observateur) si les conditions sont identiques. Pour la fiabilité, le consensus d'expert du groupe COSMIN, a ajouté à cette notion de reproductibilité, la notion de consistance interne(25).La consistance interne est une notion qui découle de la notion d'homogénéité mais réfère plus précisément selon le groupe COSMIN à la notion d'interrelation entre les items (« the inter relatedness among the items ») et est appréciée par le calcul du coefficient alpha de Cronbach (36).

La Sensibilité au changement quant à elle correspond à la capacité du score à se modifier par exemple après un traitement.

Ces étapes doivent être respectées avant de pouvoir utiliser un score.

Malgré cela certaines étapes restent soumises à discussion, notamment la reproductibilité d'un score.

En cas d'auto-questionnaire, celle-ci peut être qu'intra-observateur car le sujet est son propre témoin.

Le problème posé par cette étape est le délai nécessaire afin de refaire faire le questionnaire. Il faut que celui-ci soit suffisamment long pour que le patient puisse oublier ses réponses mais il faut qu'il soit suffisamment court pour ne pas créer de changement. Les auteurs utilisent souvent un délai d'un à deux semaines mais sans véritable consensus (37,38).

Jenny et al. (23) pour la validation transculturelle du score Oxford-12 genou privilégient la sensibilité au changement de ce questionnaire plutôt que sa reproductibilité en estimant que l'on se sert plus volontiers du fait que le questionnaire doit permettre de juger d'une amélioration ou d'une détérioration après traitement permettant de comparer le patient à son état avant traitement.



Dans notre article sur la validation du score HOOS-12 et KOOS-12, la fiabilité du score a été testée par le coefficient de Cronbach (36) mais nous n'avons pas testé la reproductibilité du questionnaire. La reproductibilité avait déjà été testée par les auteurs ainsi que par Ornetti et al. (17,21) lors de la traduction de la version longue des questionnaires HOOS et KOOS. Quah et al. (39) ont montré que ces scores restent très variables avant traitement sur une population préopératoire avec le score Oxford-12 mais la sensibilité au changement après traitement est néanmoins retrouvée, montrant l'intérêt de celle-ci.

La sensibilité au changement est donc un critère de choix dans l'évaluation de patients permettant de connaître l'effet de leur traitement. Pour autant comme pour la reproductibilité, l'évaluation à distance n'est pas codifiée en termes de délai minimum et peut varier selon les auteurs. Hunq et al (40) utilisent un délai de 3 et 6 mois. Gandek et al. (29,30) sur la validation du score HOOS-12 et KOOS-12 utilisent un délai de 6 et 12 mois. Dans notre article sur la validation du HOOS-12 et KOOS-12 ; un délai proche de 1 an a permis de voir cette sensibilité au changement. Pour autant certains auteurs recommandent un délai plus long, notamment sur les arthroplasties totales de genou. Lyman et al. (35) réévaluent les patients à 2 ans. Browne et al. (41) recommandent une évaluation à 12 mois après les arthroplasties de hanche et le genou. Un consensus sur cette période semble donc nécessaire.

Le nombre de patients à inclure pour ce type d'investigation reste discuté également. Dans leur étude Rouquette et Falissard (42) sur un score en psychiatrie ont montré la nécessité d'un minimum de 300 patients. Terwee et al. (43) jugeaient comme acceptable un minimum de 100 patients pour évaluer les propriétés d'un score, ce qui est le plus souvent utilisé dans les validations transculturelles (37,38,44). Ce nombre reste donc soumis à discussion et doit certainement être modifié selon qu'il s'agisse d'une validation transculturelle d'un score déjà validé ou de la validation d'un score nouvellement créé. En effet notre article sur la validation transculturelle d'un score a pu s'appuyer sur un travail préalable pour la création du ce score ce qui a permis une comparaison avec les données de la langue originelle permettant de limiter le nombre de sujet.

Les étapes de validité intègrent dans certains cas l'existence d'un effet plafond ou d'un effet plancher correspondant à une saturation du score à plus de 15% (45). Ces notions sont intéressantes pour évaluer ces scores. Bourlez et al (46) ont utilisé cette notion pour évaluer le côté discriminant des scores de genou pour l'évaluation de l'articulation fémoro-patellaire. Ces notions d'effet plancher ou plafond devraient être plus présents dans la présentation des résultats des scores.

## 3. Différence minimale cliniquement importante

### 3.1 Présentation

En dépit de leur utilisation en routine, il n'existe pas de seuil consensuel permettant de définir une réelle amélioration (ou détérioration) de la fonction du patient. La question de l'évaluation de ces PROM s'est ensuite posée. Fallait-il l'évaluer comme une différence significative simple avec la recherche d'un  $p < 0,05$  ou introduire dans cette évaluation la notion d'amélioration clinique ?

Jaeschke et al. (47) ont ainsi introduit ainsi la notion de différence minimale cliniquement importante (DMCI) (Minimal Clinically Important Difference – MCID) correspondant à la plus petite différence que les patients considèrent comme importante.

Cette notion de différence clinique est désormais utilisée régulièrement mais pose le problème de son calcul (48). Trois méthodes se distinguent (49-51) : la méthode des ancres basée sur une évaluation par les patients de leur propre réponse à partir d'une question, la méthode basée sur la distribution des résultats pour un critère (méthode statistique déterminant la modification nécessaire pour déterminer que la réponse soit supérieure à celle pouvant être liée au hasard), la méthode Delphi correspondant à un consensus d'expert. Il est recommandé d'utiliser deux méthodes (49), le plus souvent la méthode des ancres et la méthode de la distribution.

Les DMCI représentent un instrument de mesure essentiel pour évaluer une nouvelle procédure ou l'effet d'une procédure dans une population donnée (arthrose versus arthrite inflammatoire, chirurgie primaire et de reprise) mais reste variable selon la population étudiée.

Lorsque cette thèse a été débutée il n'y avait pas à notre connaissance de DMCI en France pour les scores choisis (HOOS (16,17), KOOS (20,21), Oxford-12 hanche et genou (18,19,22,23), objet du deuxième article de cette thèse. Ce travail a aussi exploré une option à savoir utiliser une question du score comme ancre, permettant d'envisager un usage plus répandu de la DMCI dans les études, notamment rétrospectives.

Cet article a été accepté par la revue OTSR.

## 3.2 Article

### Can Minimal Clinically Important Difference be determined in a French-speaking population of primary hip replacement using one PROM item with the Anchor strategy?

S. Putman <sup>a,b,c,d\*</sup>, J Dartus <sup>a,b</sup>, H. Migaud <sup>a,b</sup>, G. Pasquier <sup>a,b</sup>, J. Girard <sup>a,b</sup>, C. Preda <sup>b,e,f</sup>, A. Duhamel <sup>b,c,d</sup>

a CHU Lille, Service d'Orthopédie, Hôpital Salengro, Place de Verdun, F-59000 Lille, France

b Université Lille, 59000, France

c Univ. Lille, CHU Lille, ULR2694 –METRICS : Evaluation des Technologies de Santé et des Pratiques Médicales, F-59000 Lille, France

d CHU Lille, Department of Biostatistics, F-59000 Lille, France

e Laboratory of Mathematics Paul Painlevé, UMR CNRS 8524, University of Lille, France

f Lille Catholic Hospitals, Biostatistics Department, Delegation for Clinical Research and Innovation, Lille Catholic University, Lille, France

\* Corresponding author: Sophie Putman, CHU Lille, Service d'Orthopédie, Hôpital Salengro, Place de Verdun, F-59000 Lille, France

E-mail: [sophie.putman@wanadoo.fr](mailto:sophie.putman@wanadoo.fr)

#### **Abstract**

##### **Background**

The impact of surgery on the patient is classically assessed on pre- and post-treatment scores; it is, however, increasingly recommended to weight these results according to “MCID”, Minimal Clinically Important Difference, using either the data distribution method or the anchor method, which consists in an extra question specifically targeting the patient's improvement. MCIDs vary between populations and to our knowledge there have been no investigations in France regarding total hip replacement (THR). We therefore conducted a prospective study in a THR population to determine: 1) whether MCIDs for scores in France were comparable to the international literature data; 2) whether a general item taken from a different score could serve as anchor; and 3) whether an item from the actual questionnaire itself could serve as anchor.

##### **Hypothesis**

When pre- and post-treatment scores are available, an item from the questionnaire itself can serve as anchor for MCID.

##### **Material and method**

In a prospective observational study, 123 primary THR patients (69 male, 54 female), out of 150 initially included, filled out the 5 domains of the HOOS questionnaire and the Oxford-12 questionnaire, preoperatively and at 6-12 months. MCID was calculated on the distribution-

based and anchor-based methods. Two Oxford items (questions 1 and 2) and 2 HOOS items (questions S1 and Q4) were used as anchors, as well as a supplementary question on improvement and the Forgotten Joint Score (FJS).

### **Results:**

At a mean  $10.12 \pm 1.2$  months' follow-up [range, 6.5-11.9 months], Oxford-12 score increased from  $19 \pm 8$  [3-35] to  $40 \pm 10$  [8-48] ( $p < 0.001$ ), all HOOS components showed improvement, and FJS at last follow-up was  $71 \pm 29$  [0-100]. The general items (Oxford question 1 and HOOS question Q4) were more discriminating than the joint-specific items (Oxford question 2 and HOOS question S1). On the 3 anchors (improvement rated 1 to 5, Oxford question 1 and HOOS question Q4), 3 to 5 patients showed deterioration, 5 to 6 were unchanged, 30 to 40 were slightly improved, and 73 to 80 were improved by THR.

Mean MCID on both distribution and anchor methods was 9 [5.5-12] for 'Oxford-12, 20 [12-27] for HOOS Symptoms, 26 [10-36] for HOOS Pain, 22 [11.5-28] for HOOS Function, 26 [13-34] for HOOS Sport and 22 [14-28] for HOOS Quality of Life.

### **Discussion**

MCID for the Oxford-12 and HOOS scores in a French population was comparable to literature data. Using a score item as anchor to define improvement is possible, but only if a general item is used.

Keywords: PROM, Hip, HOOS, Total hip replacement

Level of evidence: IV, prospective study without control group

Clinical Trials registration n°: NCT04057651

## 1. Introduction

Patient-Reported Outcome Measures (PROM) are increasingly used in postoperative assessment, and themselves need to be assessed [1]. They provide scores for patient status to be evaluated. With their increasing use, Jaeschke et al. [1] introduced the concept of Minimal Clinically Important Difference (MCID), being the smallest difference patients consider important. Although MCID differs between populations and according to the method by which it is determined, it is increasingly used in the context of questionnaire assessment [2,3] and is important for interpreting the results of surgery and differences between procedures. In trials using PROMs as main endpoint, MCID determines the expected difference in superiority studies or the non-inferiority threshold in non-inferiority studies.

Two methods [4,5] are mainly used for calculating MCID, one based on score distribution, and the other using an extra question known as an “anchor” to categorize results as “improvement/no change/worsening”. In a systematic review of the literature, Celik et al. [6] analyzed this idea in hip, knee and ankle questionnaires and found a variety of methods, notably for choosing an anchor. Some scores, such as the Hip Osteoarthritis Outcome Score (HOOS) [7,8], include a general item: “Finally, are you bothered by your hip?”, with 5 possible responses. In patient assessment, these scores are available pre- and post-treatment, including the answers to these general questions that help in interpreting the results.

MCID varies between populations [9] and, to our knowledge, there have been no French investigations in total hip replacement (THR). We therefore conducted a prospective study in a primary THR population, using the HOOS [7-8] and Oxford-12 [10-11] scores, to determine: 1) whether MCIDs for these scores in France were comparable to the international literature data; 2) whether a general item taken from a different score could serve as anchor; and 3) whether an item from the actual questionnaire itself could serve as anchor. The study hypothesis was that, when pre- and post-treatment scores are available, an item from the questionnaire itself can serve as anchor for MCID.

## 2. Material and methods

### 2.1 Patients

A prospective study was conducted between 2018 and 2019 in the University Hospital Center of Lille, France (local registration n° 2017-A01911-52; Clinical Trials registration n° NCT04057651) with approval from the “Est IV” data protection committee. All patients aged

≥18 years undergoing THR for osteoarthritis or osteonecrosis were included on signature of informed consent.

Exclusion criteria comprised: lack of French national health insurance cover, age <18 years, inability to provide consent, pregnancy, and body-mass index > 35 or <18. Thus, 150 patients were included: 90 male, 60 female; mean age 64 ± 13 years (range, 34-89 years).

## **2.2 Methods**

Patients filled out the hip-specific HOOS and Oxford-12 questionnaires [7, 8,10,11] on eve of surgery, and then by mail at 6-12 months, when the FJS Forgotten Joint Score was also associated [12].

They also answered a question rating improvement on a 1-5 scale: “Compared to before surgery, how do you rate the present state of your hip? (1 = worse, 2 = slightly worse, 3 = the same, 4 = slightly better, 5 = better or much better)”. 123 of the 150 patients (82%: 69 male, 54 female) responded at ≥6 months.

## **2.3 Assessment**

Two methods were used to determine MCID [4-6,13]:

### **1) Distribution-based**

This is a statistical method based on the distribution of the variable over the population as a whole, determining the minimal change required for the response to be better than chance. In the overall population, we used the mean pre- to post-treatment change,  $\Delta$ , with standard deviation (SD), such that  $SD_{\Delta} * 0.5 = MCID$  [14].

We also used the Minimal Detectable Change (MDC) at 95% confidence level:  $MDC_{95} = 1.96\sqrt{2SEM}$  (Standard Error of Measurement) [15].

### **2) Anchor-based**

Several anchors were tested, with the 5-point improvement question as reference.

For the HOOS, the anchor was question 4: “In general, how much difficulty do you have with your hip?” with 5 possible responses (Figure 1).

Q4. In general, how much difficulty do you have with your hip?

- none       Mild       Moderate       severe       Extreme  
 0 point      1 point      2 points      3 points      4 points

Possible difference between response before surgery and response after surgery

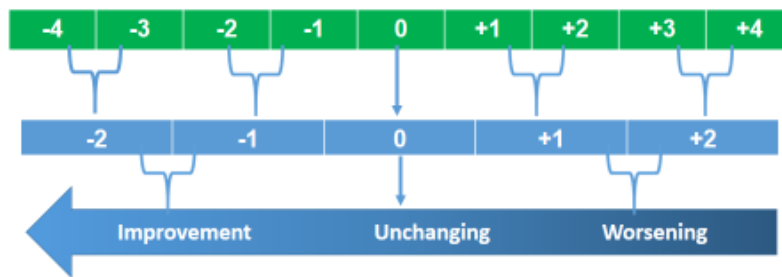


Figure 1: Definition of improvement on a HOOS item. Possible change ranges from -4 to +4. Results are then grouped to have a score from -2 to +2, where 0 represents no change, -1 and -2 improvement, and +1 and +2 worsening.

For Oxford-12, the anchor was question 1: “How would you describe the pain you usually have in your hip?”, again with 5 responses (Figure 2)

How would you describe the pain you usually have in your hip?

- None       Very Mild       Mild       Moderate       Severe  
 4 points      3 points      2 points      1 point      0 points

Possible difference between response before surgery and response after surgery

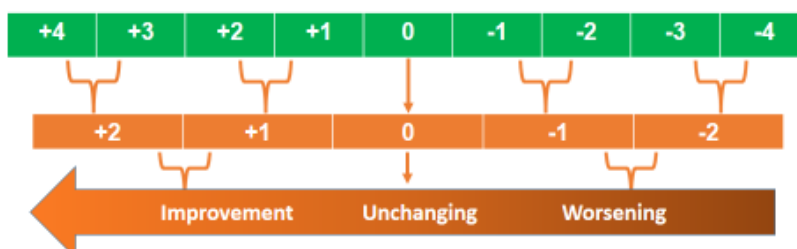


Figure 2 : Definition of improvement on an Oxford-12 item. Possible change ranges from -4 to +4. Results are then grouped to have a score from -2 to +2, where 0 represents no change, -1 and -2 improvement, and +1 and +2 worsening

To assess discrimination, we applied the same procedure on 2 items unrelated to clinical improvement:

- Oxford item 2: "Have you had any trouble with washing and drying yourself (all over) because of your hip?"

- HOOS item S1: "Do you feel grinding, hear clicking or any other kind of noise from your hip?"

For these anchors, results were considered unchanged if pre- and post-operative responses were the same (0), and improved for an improvement of  $\geq 1$  point (Figures 1 and 2). Improvement of 1 point was taken as equivalent to the response "slightly better" and improvement of 2 points to "better or much better".

The ROC was used to assess the discriminatory power of pre- to post-operative difference in scores in 2 sub-populations: with and without improvement. The group without improvement comprised patients with worsened, unchanged and only slightly improved results; the group with improvement comprised patients responding "better or much better".

Areas Under the ROC Curve (AUC) were calculated. Score sensitivity corresponded to the percentage of patients showing improvement in agreement with the anchor, with score differential exceeding the MCID. Specificity corresponded to the percentage without improvement in agreement with the anchor, with score differential less than the MCID. Thus, the largest possible AUC was sought:  $> 0.9$  was considered exceptional,  $0.8-0.9$  excellent,  $0.7-0.8$  acceptable and  $0.5$  non-discriminating [16]. MCID was calculated for anchors that were more discriminating than the reference anchor (the 1-5 point improvement question).

Once populations had been determined for each anchor considered as discriminating, the Youden index was used to determine the MCID threshold: i.e., point on the ROC curve where [sensitivity + specificity -1] was optimal [17].

## **2.4 Statistics**

Statistical analysis used SPSS software (IBM, Bois-Colombes, France). Quantitative variables were reported as mean and standard deviation for normal distribution or else median and interquartile range. Distribution normality was checked graphically and assessed on Shapiro-Wilk test. Qualitative variables were reported as number and percentage, and were compared between groups on Student test, after checking normal distribution. The significance



threshold was set at  $p=0.05$ . Effect size was calculated on Cohen's criteria: 0.2, weak; 0.5, moderate; and 0.8, strong [18].

### 3. Results

#### 3.1 Score distributions

123 of the initial 150 patients (69 male, 54 female) responded, at a mean  $10.12 \pm 1.2$  months (range, 6.5-11.9 months). Table 1 shows preoperative and follow-up scores. There was systematic significant overall improvement ( $p < 0.001$ ).

	N	Preoperative	6-12 months' FU	$\Delta^*$	Effect size	$p^{**}$
<b>Oxford-12</b>	123	19 [3-35] +/-8	40,8 [8-48] +/-10	21 [-12, +43] +/-11	1,92	<0.001
<b>HOOS Symptom</b>	123	41 [0-85] +/-17	83 [15-100] +/-19	42 [-25, +90] +/-24	1,76	<0.001
<b>HOOS Pain</b>	123	40 [2-75] +/-14	86 [17-100] +/-19	46 [-15, +85] +/-20	2,26	<0.001
<b>HOOS Function</b>	123	41 [3-88] +/-16	85 [19-100] +/-19	44 [-30, +88] +/-23	1,93	<0.001
<b>HOOS Sport</b>	123	23 [0-81] +/-17	75 [0-100] +/-25	51 [-18, +100] +/-26	1,92	<0.001
<b>HOOS Quality of Life</b>	123	24 [0-86] +/-18	78 [0-100] +/-24	54 [-50, +100] +/-28	1,93	<0.001
<b>Forgotten joint score (FJS)</b>	123		71 [0-100] +/-29			

**Table 1: Distribution of pre- and post-operative scores. Results reported as mean [range] +/- standard deviation; \*  $\Delta$ : mean difference; \*\*: pre- to post-operative p-value (matched Student test). On Cohen's criteria, effect size 0.2 is "weak", 0.5 "moderate" and 0.8 "strong".**

### 3.2 Choice of anchors

Table 2 shows AUCs.

Area under the ROC curve	Improvement anchor, 1-5 points	Oxford-12 question 1	HOOS question Q4 (quality of life)	Oxford-12 question 2	HOOS question S1 (symptom)
Oxford-12	0.825 [0.763-0.907]	0.869 [0.806- 0.932]	0.869 [0.806-0.932]	0.531 [0.402-0.661]	0.648 [0.546-0.720]
HOOS Symptom	0.819 [0.738-0.900]	0.839 [0.767-0.912]	0.827 [0.736-0.891]	0.427 [0.311-0.643]	0.693 [0.592-0.783]
HOOS Pain	0.761 [0.664-0.859]	0.877 [0.811-0.944]	0.837 [0.751-0.903]	0.479 [0.357-0.602]	0.688 [0.580-0.796]
HOOS Function	0.788 [0.702-0.873]	0.830 [0.757-0.904]	0.813 [0.762-0.909]	0.514 [0.393-0.630]	0.631 [0.518-0.744]
HOOS Sport	0.808 [0.726-0.890]	0.879 [0.819-0.940]	0.850 [0.782-0.917]	0.420 [0.310-0.529]	0.606 [0.501-0.711]
HOOS Quality of life	0.822 [0.736-0.907]	0.859 [0.789-0.924]	0.920 [0.900- 0.979]	0.497 [3.679-0.616]	0.645 [0.544-0.744]

**Table 2: Area under the curve for various anchors with 95% confidence intervals. The greater the AUC, the more discriminating the anchor; AUC 0.5 is non-discriminating.**

They lay between 0.7 and 0.9 for the 3 “general” anchors: improvement question, Oxford question 1 and HOOS question Q4. The AUCs for Oxford question 1 and HOOS question Q4 were slightly greater than for the usual 1-5 point improvement question, and could thus be considered discriminating and be used for calculating MCID. In contrast, the anchors with little relevance to improvement (Oxford question 2 and HOOS question S1) were less discriminating than the general items (Oxford question 1 and HOOS question Q4), with smaller AUCs, close to 0.5, indicating that they could not be used for calculating MCID (Figure 3).

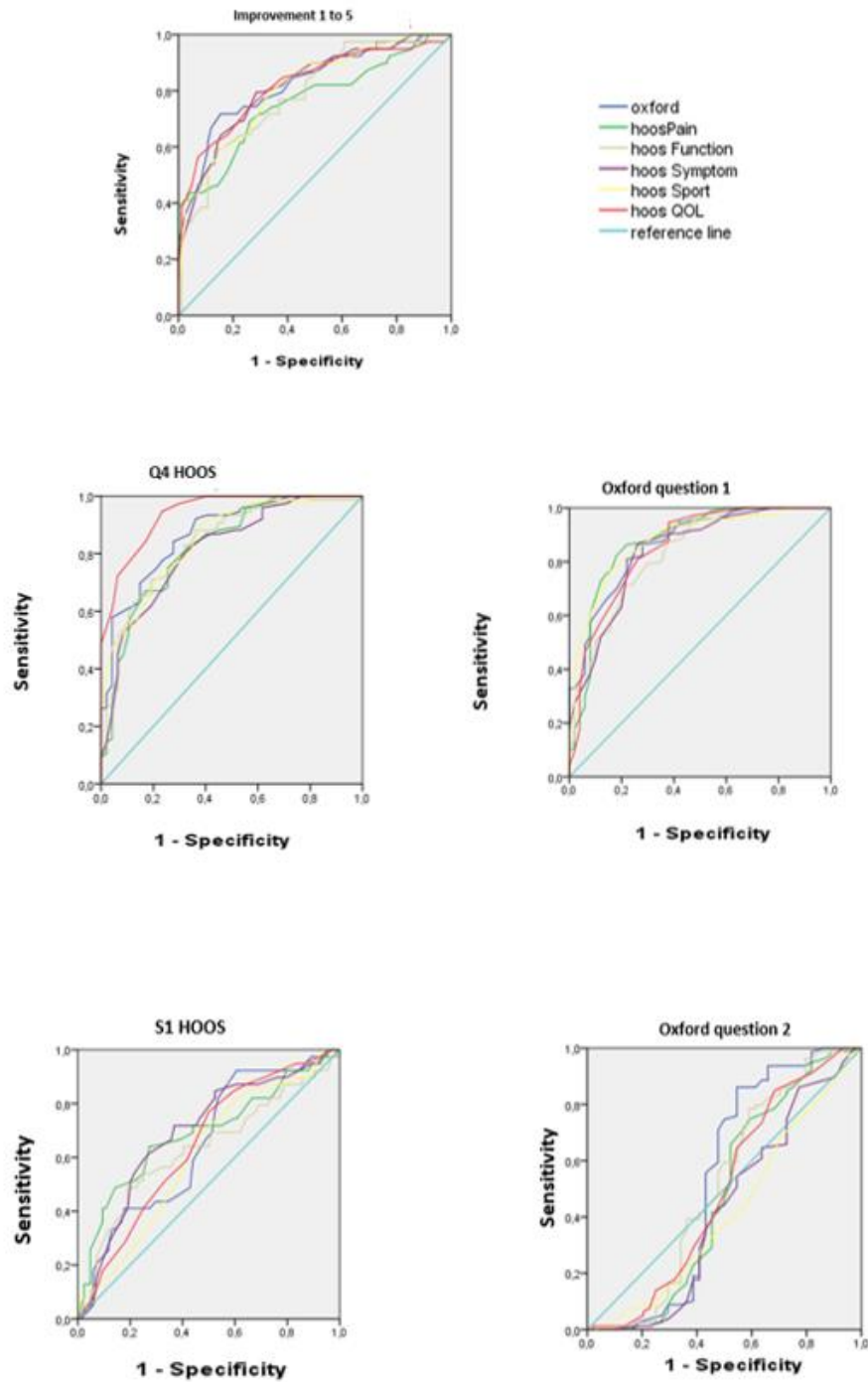


Figure 3: ROC curves of anchors used for all scores. Y-axis sensitivity; X-axis, 1- specificity. The 45° line corresponds to AUC 0.5 (non-discriminating).

### 3.3 Score distributions and calculation of MCID according to anchor

On the 3 anchors (improvement question, Oxford question 1 and HOOS question Q4), there were 3-5 patients with deterioration, 5-6 unchanged, 30-40 slightly improved, and 73-80 improved by surgery (Table 3).

Score	Improvement anchor, 1-5		$\Delta M$ [range] +/- standard deviation	Oxford question 1 anchor		$\Delta M$ [range]	HOOS question Q4 anchor		$\Delta M$ [range]
		N			N			N	
Oxford-12	1	5	-3.8 [-13. +6] +/- 8	-2	0		+2		
	2	0		-1	5	-5 [-14. +2] +/-6	+1	3	-1 [-12. +10] +/-15
	3	4	9 [-3. +10] +/-10	0	5	8 [0. +14] +/-6	0	6	1 [-10. +10] +/-7
	4	30	15 [-6. +36] +/-9	+1	40	15 [-6. +31] +/-8	-1	38	14[-6. +33] +/-8
	5	84	25 [7. +43] +/-8	+2	73	27 [9. +43] +/-8	-2	76	26 [8. +43] +/-8
HOOS Symptom	1	5	-9 [-25. +5] +/-12	-2	0		+2		
	2	0		-1	5	-10 [-25. +0] +/-10	+1	3	-7.5 [-25. +10] +/-24
	3	4	13 [0. +40] +/-19	0	5	16 [5. 30] +/-11	0	6	0 [-15. +25] +/-20
	4	30	30 [-10. +70] +/-20	+1	40	31 [-10. +70] +/-20	-1	38	31 [0. +75] +/-14
	5	84	51 [5. 90] +/-19	+2	73	54 [10. 90] +/-14	-2	76	53 [10.90] +/-18
HOOS Pain	1	5	-2 [-15. +15] +/-14	-2	0		+2		
	2	0		-1	5	-5 [-55. +12] +/-12	+1	3	-3 [-15. +10] +/-17
	3	4	+15 [-15. +40] +/-22	0	5	24 [-3. +50] +/-23	0	6	4 [-12. +12] +/-20
	4	30	39 [3. +72] +/-20	+1	40	36 [+3. +72] +/-15	-1	38	38 [-4. +72] +/-17
	5	84	53 [25. +85] +/-14	+2	73	57 [32. +85] +/-12	-2	76	55 [25. +85] +/-13
HOOS Function	1	5	-4 [-19. +16] +/-16	-2	0		+2		
	2	0		-1	5	-9 [-19. 0] +/-8	+1	3	-6 [-16. +30] +/-15
	3	4	18 [-1. +32] +/-16	0	5	29 [5. +52] +/-15	0	6	2 [-20. 26] +/-15
	4	30	35 [0. +75] +/-19	+1	40	34 [-5. +75] +/-20	-1	38	34 [-1. +75] +/-18
	5	84	51 [30. +88] +/-18	+2	73	54 [18. +88] +/-15	-2	76	53 [30. 88] +/-18
HOOS Sport	1	5	-6 [-18. +12] +/-12	-2	0		+2		
	2	0		-1	5	-3 [-18. +18] +/-14	+1	3	0 [-19. +19] +/-26
	3	4	5 [0. +12] +/-6	0	5	18 [-12. +20] +/-30	0	6	9 [-12. +37] +/-20
	4	30	40 [0. +100] +/-21	+1	40	36 [-12. 81] +/-19	-1	38	36 [-12. +75] +/-22
	5	84	61 [0. +100] +/-20	+2	73	65 [6. +100] +/-19	-2	76	64 [6. +100] +/-19
HOOS Quality of life	1	5	-9 [-50. +3] +/-23	-2			+2		21[-12. +43] +/-11
	2	0		-1		-9 [-50. +25] +/-23	+1	3	-25 [-50. 0] +/-30
	3	4	29 [-10. +56] +/-23	0		28 [0. +56] +/-27	0	6	3 [-6. +18] +/-8
	4	30	39 [-6. +100] +/-27	+1		39 [-6. +93] +/-24	-1	38	35 [0. +69] +/-18
	5	84	64 [6. +100] +/-19	+2		68 [25. +100] +/-17	-2	76	70 [37. +100] +/-16

**Table 3: Distribution of scores according to anchor. N= number of subjects;  $\Delta M$  = pre- to post-operative deference in mean. Blue= "worse of much worse"; Green = "slightly worse"; Yellow = "no change"; Orange = "slightly better"; Red = "better or much better"**

There was thus overall improvement except in 9-10 patients, who were dissatisfied or unchanged, this number being comparable between anchors. MCID was calculated according to the predefined distribution-based method and the method based on the most discriminating anchors (Table 4).

	Distribution-based		Improvement anchor, 1-5	Oxford question 1 anchor	HOOS question Q4 anchor	Mean MCID for distribution-based method	Mean MCID for anchor-based method	Mean MCID for both methods
	SD <sub>Δ</sub> *0,5	MD C95 %						
<b>Oxford-12</b>	5,5	7	11	12	12	6	12	9
<b>HOOS Symptom</b>	12	15	20	26	27	13	24	20
<b>HOOS Pain</b>	10	16	32	35	36	13	34	26
<b>HOOS Function</b>	11,5	12	29	30	28	11	29	22
<b>HOOS Sport</b>	13	22	31	32	34	17	32	26
<b>HOOS Quality of life</b>	14	17	26	28	26	15	27	22

**Table 4 : MCID per score and per method. Mean MCID calculated per method, then for both methods.**

## 4. Discussion

### 4.1 Results for the series

The present study showed that MCID in a French-speaking population was close to reports for other populations (Tables 5 and 6).

	Number of patients	Country	Procedure	Method	MCID
Beard et al. [19]	82415	USA	Arthroplasty	Anchor	7.5
Fernandez et al. [20]	361	Spain	Arthroplasty	Anchor	7
Impelizari et al. [21]	102	Switzerland	Arthroscopy	Anchor	8-10
Present series	123	France	Arthroplasty	Anchor and distribution	9 [5.5-12]

**Table 5: Oxford-12 hip MCID in the literature, according to country and method**

**Table 6: MCID for HOOS in the literature, according to country and method**

	Number of patients	Country	Procedure	Method	MCID
Nwachukwu et al. [22]	49	USA	Arthroscopy	Distribution	Sport 13 Function 7,9
Paulsen et al. [23]	1288	Denmark	Arthroplasty	Anchor and distribution	Pain 24 Quality of life 17
Lyman et al. [3]	2323	USA	Arthroplasty	Anchor	Pain 36 Symptoms 20 Function 14 Quality of life 13
Goodman et al. [24]	4801	USA	Arthroplasty	Anchor	Pain 22 Function 18
Kuo et al. [25]	271	USA	Arthroplasty	Anchor	Pain 25 Symptoms 20 Function 19 Quality of life 7 Sport 16
Present series	123	France	Arthroplasty	Anchor and distribution	Pain 26 [10-36] Symptoms 20 [12-27] Function 22 [11-28] Quality of life 23 [14-28] Sport 26 [13-34]

These findings were borne out by an FJS-12 of  $71 \pm 29$  points, close to the mean value of  $70.9 \pm 33$  reported by Giesinger et al. in a general US population [26].

Even so, although the study defined threshold values for a French population for the first time, the results varied according to method and are only indicative. Results vary between countries [23]: Lyman et al. [3] suggested that environmental factors could explain the difference found in Denmark [23], with better quality of life on international rankings; results may also differ between populations: Kuo et al. [25] found specific results in a population of veterans. MCID may also vary with pathology: values after hip arthroscopy varied slightly, notably for the HOOS (Table 6) [21].

Results further vary according to method. Kuo et al. [25] showed that, like in the present study, the distribution-based method gave smaller MCIDs than the anchor-based method. For the distribution-based method, Copay et al. [5] recommend using the MDC95, to have a result independent of sample size. ROC curve analysis is also debatable, as it depends on the choice of index: some teams use the Youden index, others use specificity at 0.8 [27,28].

#### **4.2 Can an item from one score serve as anchor for another?**

For an item from one score to serve as anchor for another, it has to be general enough to assess progression: less general items were less sensitive to improvement than items focusing on quality of life or pain. Question 1 of the Oxford-12 score, for example, was interesting, as Goodman et al. [24] had shown preoperative pain to be related to improvement. Danoff et al. [29] defined an MCID by pain on visual analogue scale. Other studies also used items from a different score: Lyman et al. [3] used the Satisfaction Survey of the HSS score, assessing improvement on 6 levels.

Other studies used a different score in its entirety: Lee et al. [30] used the whole Oxford-12 score to define MCID for the Knee Society rating system, taking advantage of the known Oxford-12 MCID of 5 points' improvement. Kuo et al [25] used a 100-point 4-question score for which there was consensus: e.g., that 25 points corresponds to severe dissatisfaction. Van der Wees et al. [31] used a different consensual breakdown of the Oxford-12 score (>41, excellent; 34-41, good; 27-33, moderate; <27 poor), defining MCID as improvement by 1 category. Using a different score in its entirety requires interpretation to be known, which usually means an expert consensus. Whether one item or the whole score is used, the score needs to be based on several response levels, to enable distinctions to be made in the

population, like with anchors having between 3 and 15 possible responses [32]. Responses often need to be grouped according to presence/lack of improvement.

#### **4.3 Can an item serve as anchor for its own score?**

Using an item to serve as anchor for its own score seemed interesting, but needs assessing in other studies as we were not able to find any other hip series using this method. The interest is that no extra question is needed, so that the score assesses itself, with a gain in objectivity inasmuch as the patient does not realize that the assessment is being made. However, several anchors, or at least several methods, are often necessary [4].

#### **4.4 Study limitations**

The present study had several limitations;

- 1) Choice of anchor was subjective; however, asking at least 2 questions enhances the strength of the MCID value.
- 2) Categorizing based on item results is debatable. The choice was supported by the number of patients per subgroup, which was close to the number for the reference anchor (1-5 point improvement). To calculate MCID, responses often need to be grouped. In a study of MCID for the HOOS joint reconstruction score, Hung et al. [33] used a 7-question anchor, classifying patients from -3 to +3, then grouped patients reporting slight improvement (+1) together with slight deterioration and no change.
- 3) Sample size was small. However, Terwee et al. [34] consider 100 patients enough to assess score properties. In a review of the literature, Copay et al. [5] advocated using MDC95 to limit the impact of sample size.
- 4) The present strict population selection was a limitation; however, MCID varies between populations, and our criteria limited bias, as scores vary widely, as Marot et al. [35] showed for the KOOS, with variation according to age, gender and body-mass index.

### **5. Conclusion**

Minimal Clinically Important Difference (MCID) was comparable in a French THR to other reported series. Using a score item to determine MCD appears to be possible if the item concerns general assessment.



## References

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control ClinTrials*1989;10:407-15.
2. Singh JA, Schleck C, Harmsen S, Lewallen D. Clinically important improvement thresholds for Harris Hip Score and its ability to predict revision risk after primary total hip arthroplasty. *BMC MusculoskeletDisord*2016;17:256.
3. Lyman S, Lee YY, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clin Orthop Relat Res* 2018;476:2432-41.
4. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102-9.
5. Copay AG, Eyberg B, Chung AS, Zurcher KS, Chutkan N, Spangehl MJ. Minimum Clinically Important Difference: Current Trends in the Orthopaedic Literature, Part II: Lower Extremity: A Systematic Review. *JBJs Rev*2018;6:e2.
6. Celik D, Çoban Ö, Kılıçoğlu Ö. Minimal clinically important difference of commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a systematic review. *J Clin Epidemiol* 2019;113:44-57.
7. Nilsson AK, Lohmander LS, KlässboM, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord* 2003;4:10.
8. Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, et al. Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. *Osteoarthritis Cartilage* 2010;18:522-529.
9. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010;63:524-34.
10. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996;78:185-90.
11. Delaunay C, Epinette JA, Dawson J, Murray D, Jolles BM. Cross-cultural adaptations of the Oxford-12 Hip score to the French speaking population *Orthop Traumatol Surg Res* 2009;95:89-99.
12. Klouche S, Giesinger JM, Sariali EH. Translation, cross-cultural adaptation and validation of the French version of the Forgotten Joint Score in total hip arthroplasty. *Orthop Traumatol Surg Res* 2018;104:657-661.

13. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7:541-6.
14. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;41:582-592.
15. Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? *J Biopharm Stat* 2004;14:97-110.
16. Hosmer DW, Lemeshow S. *Applied logistic regression*, 2<sup>nd</sup> Ed. Chapter 5, John Wiley and Sons, New York (2000), pp 160-164.
17. Youden WJ. Index for rating diagnostic test. *Cancer* 1950;3:32-5.
18. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition. Hillsdale: Lawrence Erlbaum, New York 1988.
19. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol* 2015;68:73-9.
20. Martín-Fernández J, Gray-Laymón P, Molina-Siguero A, Martínez-Martín J, García-Maroto R, et al. Cross-cultural adaptation and validation of the Spanish version of the Oxford Hip Score in patients with hip osteoarthritis. *BMC Musculoskelet Disord* 2017;18:205.
21. Impellizzeri FM, Mannion AF, Naal FD, Hersche O, Leunig M. The early outcome of surgical treatment for femoroacetabular impingement: success depends on how you measure it. *Osteoarthritis Cartilage* 2012;20:638-45
22. Nwachukwu BU, Chang B, Rotter B, Kelly BT, Ranawat AS, Nawabi DH. Minimal clinically important difference and substantial clinical benefit after revision hip arthroscopy. *Arthroscopy* 2018;34:1862-1868
23. Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. *Acta Orthop* 2014;85:39-48.
24. Goodman SM, Mehta BY, Mandl LA, Szymonifka JD, Finik J, et al. Validation of the Hip Disability and Osteoarthritis Outcome Score and Knee Injury and Osteoarthritis Outcome Score pain and function subscales for use in total hip replacement and total knee replacement clinical trials. *J Arthroplasty* 2019;35: In Press. doi: 10.1016/j.arth.2019.12.038.
25. Kuo AC, Giori NJ, Bowe TR, Manfredi L, Lalani NF, Nordin DA, Harris AHS. Comparing methods to determine the Minimal Clinically Important Differences in Patient-Reported Outcome Measures for veterans undergoing elective total hip or knee arthroplasty in veterans health administration hospitals. *JAMA Surg* 2020.

26. Giesinger JM, Behrend H, Hamilton DF, Kuster MS, Giesinger K. Normative values for the Forgotten Joint Score-12 for the US general population. *J Arthroplasty* 2019;34:650-655.
27. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006;163:670-5.
28. Aletaha D, Funovits J, Ward MM, Smolen JS, Kvien TK. Perception of improvement in patients with rheumatoid arthritis varies with disease activity levels at baseline. *Arthritis Rheum* 2009;61:313-20.
29. Danoff JR, Goel R, Sutton R, Maltenfort MG, Austin MS. How much pain is significant? Defining the Minimal Clinically Important Difference for the VisualAnalog Scale for pain after total joint arthroplasty. *J Arthroplasty* 2018;33(Suppl7):S71-S75.
30. Lee WC, Kwan YH, Chong HC, Yeo SJ. The minimal clinically important difference for Knee Society Clinical Rating System after total knee arthroplasty for primary osteoarthritis. *Knee Surg Sports Traumatol Arthrosc* 2017;25:3354-3359.
31. Van der Wees PJ, Wammes JJ, Akkermans RP, Koetsenruijter J, Westert GP, et al. Patient-reported health outcomes after total hip and knee surgery in a Dutch University Hospital Setting: Results of twenty years clinical registry. *BMC Musculoskeletal Disord* 2017;18:97. doi: 10.1186/s12891-017-1455-y.
32. Sedaghat AR. Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures. *Otolaryngol Head Neck Surg* 2019;161:551-560.
33. Hung M, Bounsanga J, Voss MW, Saltzman CL. Establishing minimum clinically important difference values for the Patient-Reported Outcomes Measurement Information System Physical Function, hip disability and osteoarthritis outcome score for joint reconstruction, and knee injury and osteoarthritis outcome score for joint reconstruction in orthopaedics. *World J Orthop* 2018;9:41-49.
34. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651-7.
35. Marot V, Murgier J, Carrozzo A, Reina N, Monaco E, Chiron P, Berard E, Cavaignac E. Determination of normal KOOS and WOMAC values in a healthy population. *Knee Surg Sports Traumatol Arthrosc* 2019;27:541-5.

### 3.3 Mise en perspective

La variation avant et après traitement est une option couramment utilisée mais n'introduit pas le résultat ressenti par le patient. La notion de DMCI introduit cette notion de ressenti avec toute sa subjectivité.

Pour autant, on peut se perdre dans la littérature avec cette notion pouvant regrouper différents termes, comme l'ont souligné Maredupaka et al. (52). D'autres notions apparaissent dans la littérature pouvant se rapprocher de la notion de DMCI. Nous retrouvons en effet la notion de MID (Minimally important Difference), cette notion est le plus souvent confondue avec la notion de MCID (49,50),

La notion de MDC (Minimum Detectable Change) (50) correspond quant à elle à la mesure statistique de la méthode de la distribution, correspondant au MCID.

Si les notions de MCID, MCD et MID se rapprochent, d'autres notions différentes sont utilisées dans la littérature : Nous retrouvons ainsi la notion de « retour à la normale », estimé de façon arbitraire par un score meilleur de 2 SD par rapport à la moyenne du score initial score >2 SD préopératoire (53).

Sayers et al. (53) individualisent 4 méthodes : MID, MCID, retour à la normale mais également la méthode OMERACT OARSI (OO) qui définit une amélioration soit >50% ou 20 points en valeur absolue dans une échelle ou >20% ou <10 points en valeur dans au moins deux échelles. Cette méthode a été établie par des experts et se rapproche donc de la méthode Delphi. Sayers et al. (53) se proposent d'unifier toutes les notions retrouvées.

Ces différentes notions découlent de l'hétérogénéité des méthodes de calculs de cette DMCI.

Une uniformisation des méthodes est nécessaire afin de comparer les résultats de la MCID.

Les méthodes de la distribution restent une base concrète statistique. Néanmoins cette méthode de la distribution comporte de nombreuses façons d'être calculée. On peut retrouver des méthodes utilisant 0,5\*déviation standard (54). Cette méthode peut utiliser soit la population initiale ou alors la déviation standard du delta entre avant et après le traitement (54). Dans notre article sur la DMCI, nous avons privilégié le delta montrant selon nous un véritable changement. Il faut donc être très vigilant sur la technique utilisée. Copay et al. (50) recommandent pour la méthode de la distribution, l'utilisation de la SEM (Standard Error of Measurement) moins soumise à la variation des échantillons, permettant de calculer le MDC (Minimal Detectable Change). Cette notion de SEM ne doit pas être confondue avec l'erreur standard de la moyenne. Ce MDC peut être calculé à 80 %, 90%, 95 % avec à 95 % soit  $MDC_{95} = 1,96\sqrt{2SEM}$ . Beard

et al. (55) pour le score oxford utilise le MDC 90%. Copay et al. (50) recommandent l'usage du MDC à 95 %.

De même on retrouve plusieurs méthodes de calcul à partir de la méthode des ancres permettant de diviser la population en différentes sous-groupes (pire, identique, légèrement amélioré...). Jaeschke et al. (47) puis Redelmeier et Lorig (56) ont proposé des méthodes à partir des données obtenues à partir des réponses obtenues à cette question ancre. Pour Redelmeier et Lorig (56), on définit le MCID par la moyenne des différences entre les catégories identique et légèrement améliorée. Actuellement, on retrouve surtout la méthode ROC pour le calcul de ce DMCI (50,54,57). Pour autant comme nous l'avons souligné dans notre article, certains auteurs prennent 80% de la sensibilité, d'autres l'indice de Youden (58-60). Cet indice est le plus souvent utilisé mais un consensus serait nécessaire. Il apparaît donc important dans ces méthodes de fixer un consensus afin de comparer des résultats selon ces mêmes techniques.

L'usage de 2 méthodes (ancre et distribution) reste l'option la plus utilisée pour définir cette notion de DMCI (49,61). Il est souvent difficile en rétrospectif par exemple d'utiliser la méthode des ancres si celle-ci n'a pas été initialement décidée.

Notre article a permis d'identifier une option à savoir la possibilité d'utiliser une partie d'un questionnaire comme ancre. Cette idée permettra plus facilement aux auteurs si elle est confirmée par d'autres séries d'utiliser plus facilement la méthode des ancres et ainsi d'avoir de façon systématique, deux méthodes d'évaluation de la DMCI (méthode des ancres et distribution).

Se pose ensuite la question de la présentation des résultats, certaines équipes différencient chaque résultat de DMCI (48,61). Dans notre article, nous proposons une moyenne avec l'intervalle entre le minimum et le maximum car comme l'a montré Kuo et al. (61) sur l'évaluation de la DMCI chez des vétérans, le résultat est très variable selon l'ancre choisie ou la méthode. En effet comme dans notre article, dans l'article de Kuo et al. (61) la méthode de la distribution donne des résultats plus faibles que la méthode des ancres. Il apparaît donc intéressant de donner l'écart entre le minimum et le maximum, car cette DMCI est avant tout le reflet d'un ressenti du patient.

Cette notion de DMCI est complétée dans les maladies chroniques par la notion de PASS (patient acceptable symptom state), notion introduit par Tubach et al. (62). Cette notion apparaît en complément de la notion de DMCI (63,64). Le PASS a été défini comme le niveau de symptômes, pour lequel les patients se considèrent comme allant bien. La référence externe la plus couramment utilisée pour déterminer le seuil du PASS est : « En prenant en compte toutes les activités que vous avez dans votre vie quotidienne, votre niveau de douleur et aussi

votre handicap fonctionnel, si vous deviez rester dans votre état pendant les prochains mois à venir, considérez-vous votre état actuel comme étant satisfaisant ? »

Ce PASS est plus couramment utilisé dans les maladies chroniques comme par exemple les maladies rhumatologiques, contrairement à la DMCI qui juge plus de l'amélioration mais pourrait aussi être développé en orthopédie en complément de la DMCI.

# 4. Mapping

## 4.1 Présentation

Malgré la complexité de ces étapes de validation et l'identification des DMCI, de nombreux scores sont à notre disposition et, en l'absence de consensus, chaque équipe opte en fonction de sa pratique pour l'un ou l'autre. Lan et al. en 2020 (65) soulignent cette absence de consensus sur l'utilisation de scores de genou après arthroplastie et mettent en évidence une modification des usages de certains scores selon les années. Cette absence de standardisation rend difficile la comparaison des séries entre les équipes ainsi que la comparaison des résultats publiés notamment lors des revues de la littérature ou lors de la réalisation de méta analyses.

Des méthodes statistiques (score mapping ou crosswalk) ont été développées pour passer d'un score à l'autre (66), en créant un modèle statistique propre à ces scores, notamment pour obtenir le EQ-5D partir d'autres scores (66). Ces méthodes statistiques peuvent permettre de prédire les valeurs des différentes dimensions d'un score à partir des réponses connues pour un autre score, tout en garantissant ses propriétés métrologiques. Ghomrawi et al. (67) puis Polascik et al. (68) ont utilisé la méthode equipercentile (69,70) en orthopédie afin d'obtenir une table de conversion entre différents scores. Dakin et al. (66) ainsi que Fayers et al. (71) ont montré que différentes méthodes existent et peuvent être testées.

Ce troisième article s'intéresse à la possibilité d'un mapping entre les scores choisis, en choisissant le meilleur score adapté à cette procédure entre le score Oxford-12 hanche (18,19) et les différentes formes du HOOS (16,17,28,29,31,33), puis en choisissant la méthode apparaissant comme la meilleure.

Cet article a été soumis au CORR (Clinical Orthopaedics and Related Research®).

## 4.2 Article

### Mapping and crosswalk of the Oxford Hip Score and different versions of the HOOS

S. Putman<sup>a,b,c,d\*</sup>, Preda C<sup>ef</sup>, J Girard<sup>a,b</sup>, H. Migaud<sup>a,b</sup>, A Duhamel<sup>c,d</sup>

a CHU Lille, Orthopedics Department, Hôpital Salengro, F-59000 Lille, France

b University of Lille, 59000, France

c Univ. Lille, CHU Lille, ULR2694 – METRICS: évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France

d CHU Lille, Department of Biostatistics, F-59000 Lille, France

e Laboratory of Mathematics Paul Painlevé, UMR CNRS 8524, University of Lille, France

f Lille Catholic Hospitals, Biostatistics Department, Delegation for Clinical Research and Innovation, Lille Catholic University, Lille, France

\* Corresponding author: Sophie Putman

Email: sophie.putman@wanadoo.fr

Ethics committee approval IDRCB2017-A01911-52, Clinical Trials NCT04057651

#### Abstract

##### Introduction:

Patient reported outcome measures (PROMs) such as the Oxford-12 Hip Score or the HOOS are some of the tools used in daily orthopedics practice to evaluate patients, although there is no consensus about which PROM is the best, making comparisons between studies more difficult. Various mapping methods can be used to develop crosswalk tables to go from Oxford-12 scores to HOOS scores (and its derivatives HOOS PS, HOOS JR, HOOS-12) and vice-versa. This led us to carry out a prospective study with the Oxford-12 and HOOS and its derivatives with the objectives of 1) determining which version of the HOOS had the best mapping with the Oxford-12, 2) defining the most appropriate mapping method using selected indicators, 3) generating crosswalk tables between these two PROMs.

##### Materials and Methods:

The study enrolled 500 adult patients (294 men, 206 women) with hip osteoarthritis or avascular necrosis of the femoral head who completed the HOOS and Oxford-12 before undergoing primary total hip replacement. Various mapping methods (linear regression, tobit



regression, quantile regression) and equating methods (linear equating, equipercntile method) were used along with cross-validation to determine the best model and best PROM for mapping using the mean absolute error (MAE),  $R^2$ , Kolmogorov Smirnov distance as criteria.

### Results:

The Oxford-12 and HOOS were strongly correlated ( $> 0.5$ ), both in the HOOS subcategories and the HOOS PS, HOOS JR and HOOS-12. The correlation was the strongest with the HOOS-12 ( $r = 0.825$ ) and according to the MAE and  $R^2$  criteria, when comparing the various methods for the HOOS and its derivatives. This meant the HOOS-12 was best suited for mapping. To map the HOOS-12 and Oxford-12, regression methods were the best according to the MAE (5.97 regressions vs 6.2 equating) while the equipercntile method was the best according to the Kolmogorov Smirnov distance (0.04 for equipercntile method vs 0.11 for regressions). The  $R^2$  for all the methods was 0.64 to 0.67. In the end, the equipercntile method was chosen to generate crosswalk tables between the HOOS-12 and Oxford-12.

### Conclusion:

Mapping is possible between the HOOS, its derivatives and the Oxford-12 score. The resulting crosswalk tables can be used to compare clinical studies that did not include both outcome scores, especially meta-analyses and systematic reviews.

**Level of evidence:** V – Prospective study

Keywords: outcome score, PROM, mapping, equating

## **Introduction**

Patient-reported outcome measures (PROMs) are one of the tools used in daily orthopedics practice to evaluate patients [1]. However, the lack of standardization when it comes to PROMs makes it difficult compare studies between teams and published results and to do meta-analyses. In 2008, Riddle et al. [28] found that more than 20 different PROMs were being used in hip surgery.

Statistical methods (score mapping or crosswalk) have been developed to go from one PROM to another [1, 11, 20, 25]. These statistical methods can be used to predict the values of one PROM based on the responses in another PROM, while preserving its metrological properties. These have been most widely used with quality of life scores [20]. In 2017, Ghomrawi et al. [14] did a conversion between the UCLA score and Lower Extremity Activity Scales using

the equipercntile method. Polascik et al [26] used this same method to develop crosswalk tables between the HOOS JR and the Oxford Hip Score.

However, there is no consensus on which mapping method should be used to go from one PROM to another when using only the data from the first PROM. This pertains to orthopedics when we want to go from the Oxford-12 score to the HOOS [21,23] or one of its derivatives (HOOS PS [Physical Function Shortform], HOOS JR [Joint Replacement], HOOS-12) [4,12,13,19]. This led us to carry out a prospective study with the Oxford-12 and HOOS and its derivatives with the objectives of 1) determining which version of the HOOS had the best mapping with the Oxford-12, 2) defining the most appropriate mapping method using selected indicators, 3) generating crosswalk tables between these two PROMs.

## **Materials and Methods**

This was a prospective cohort study done at the Lille University Hospital between May 2018 and September 2019 (IRB approval AC 2017-A01911-52, Clinical Trials NCT04057651).

### ***Outcome scores***

The HOOS [21] has 40 items with five possible responses, graded from 0 to 4. This score has five subcategories: pain (P) (10 items), symptoms and stiffness (S) (5 items), activities of daily living (ADL) (17 items), function in sports and recreational activities (SP) (4 items), quality of life (Q) (4 items). The score in each subcategory is calculated out of 100 points (minimum 0; maximum 100). Other versions of the HOOS have been developed based on the original 40-question questionnaire, such as the HOOS PS (5 items) [4]; HOOS JR (6 items) [19]. In 2019, the team that had developed the original HOOS simplified it to 12 items, calling it the HOOS-12 [12, 13]. In our study, we also analyzed the total HOOS as the mean of the five HOOS subcategories.

The Oxford Hip Score (Oxford-12) [5] consists of 12 questions with five possible responses, graded from 0 to 4. These scores are added up to calculate the final score out of 48 points (0 = worse possible score, 48 = best score with no symptoms). Both PROMs have been validated and have undergone trans-cultural adaptation for the French-speaking population [6, 23, 24].

### ***Patient cohort***

Between May 2018 and September 2019 (IRB approval AC 2017-A01911-52, Clinical Trials), every patient over 18 years of age with hip osteoarthritis or avascular necrosis of the femoral head was invited to participate in the study. After providing written informed consent, they filled out the HOOS and Oxford-12 questionnaires before undergoing total hip arthroplasty. Patients

were excluded if they were not covered by the national health insurance scheme, were under 18 years of age, were unable to consent, were pregnant or had a body mass index  $> 35$  or  $< 18$ . The patients filled out the same questionnaires again after a minimum of 6 months postoperative.

### ***Statistical methods***

There is no accepted method to define how many subjects are needed for a mapping study. Thus, we determined the number of subjects required based on the study by Rouquette and Falissard [29] who recommended having at least 300 subjects to validate a PROM. The statistical analysis was carried out by the Biostatistics Department of the Lille University Hospital using the software, R (R Foundation for Statistical Computing, Vienna, Austria).

Except when specified, the analysis was done on the preoperative dataset. The Pearson correlation coefficient was calculated between the Oxford-12 and the HOOS. Dorans [7, 8] recommend having a correlation coefficient of at least 0.3 between two PROMs before starting the mapping procedure.

Two broad types of methods can be used: equating methods and regression methods. We used five of the most common methods: linear regression (LR), quantile regression (QR), Tobit regression (Tobit), linear equating method (LinEQ), equipercetile method (EQUI) [3, 16, 17, 18, 27]. These methods were evaluated based on three criteria: MAE (mean absolute error),  $R^2$  and the Kolmogorov Smirnov distance (KS) [2,31]. The latter calculates the distance between a PROM's observed distribution and the one reconstructed by mapping. The goal is minimize the MAE, minimize the KS distance and have the  $R^2$  be as close as possible to 1.

### ***Selecting the HOOS score that maps the best to the Oxford-12***

For a given method and PROM, the single value calculated based on the study sample provides no information about the uncertainty in the criterion's estimate and thereby cannot be used to directly compare the quality of the mapping using the different HOOS. To estimate the distribution of the criteria, we used the bootstrap method with replacement [9]. If  $N$  is the size of the total sample, we did  $k = 100$  bootstrap iterations with  $n \leq 500$  individuals randomly drawn (with replacement) among the 500 individuals in study cohort. To determine the variations in the distribution of the criteria due to sample size, we repeated the analysis by changing  $n$  from 50 to 500 in increments of 50. We would expect the dispersion in the criterion's distribution to be reduced and then to stabilize when  $n$  increases. For each bootstrap iteration, mapping of the nine HOOS versions was done with the five methods; the three mapping quality criteria were calculated for each of these methods. The criterion calculated with the sample on which the method was developed is affected by optimism bias (same sample used to predict a score's

value and to judge the quality of this prediction). To account for this bias, we estimated the criteria for each bootstrap iteration using leave-one-out cross-validation [15].

Lastly, we did a graphical representation of each method (LR, QR, Tobit, LinEQ, EQUI) and one given criterion (MAE,  $R^2$ , KS) to visualize the criterion's distribution for the nine HOOS versions relative to the size  $n$  of the sample (mean values and 95% confidence intervals). This resulted in 15 charts being used to identify which HOOS version was best suited to mapping with the Oxford-12.

### ***Selecting the best mapping method***

For the chosen HOOS, the procedure described above (100 bootstrap iterations of size  $n$  [ $n=50$  to 500 by increments of 50] and cross-validation on each iteration) was used to compare the mapping methods. We did a graphical representation of how effective the mapping was for the chosen HOOS using all five methods with fixed criterion. The best method was then selected based on the three criteria (MAE,  $R^2$  and KS). The observed density and the estimated density after mapping of the chosen HOOS were determined on the full sample with each method using a non-parametric procedure involving a Gaussian kernel [30].

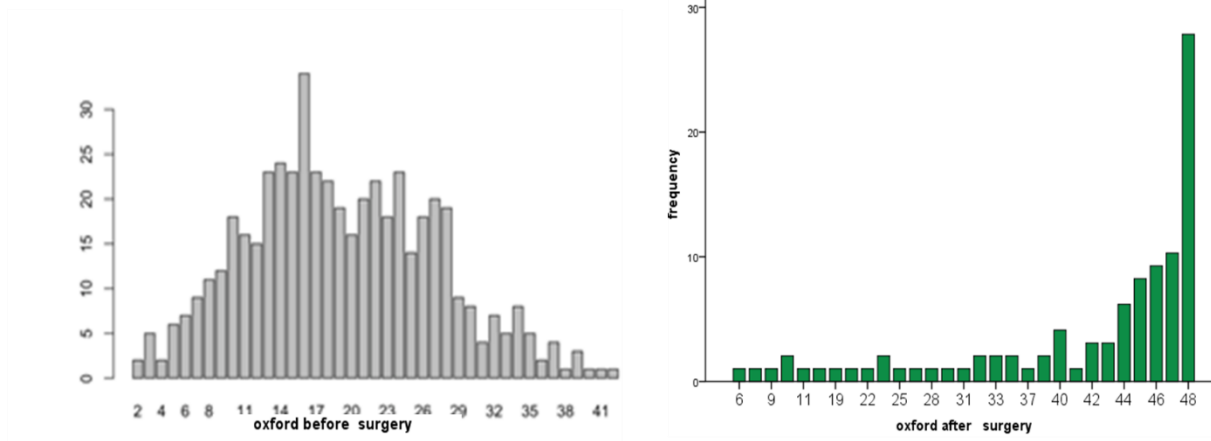
### ***Creating the crosswalk tables***

To generate a larger distribution of the minimum values of the PROM towards the maximum values, the crosswalk table combines the preoperative and postoperative datasets as described by Ghomrawi et al. [14].

## **Results**

### ***Patient cohort***

Five hundred patients (294 men, 206 women) were enrolled having a mean age of  $61 \pm 13.7$  years (range, 22-89) before the total hip replacement. These patients were evaluated again postoperatively after a minimum of 6 months (range, 6-12). There was no missing data. There was no floor or ceiling effect in the Oxford-12 and HOOS or its derivatives during the preoperative period (Figure 1). A ceiling effect (>15% maximum scores) was found during the postoperative measurement for the Oxford-12 (Figure 1).



**Figure 1: Oxford-12 score before and after surgery. There is a ceiling effect in the Oxford-12 with saturation of the maximum scores**

Given this ceiling effect, the preoperative data set was used to analyze and select the best suited HOOS variant, along with best method, as it had the largest variation.

***HOOS score that best maps to the Oxford-12***

The correlation between the Oxford-12 and the HOOS versions was above the 0.3 threshold defined by Dorans and ranged from 0.59 for the HOOS QOL to 0.84 for the HOOS global (Table 1).

	HOOS symptom s	HOOS pain	HOOS ADL	HOOS sport	HOOS QOL	HOOS global	HOOS PS	HOOS JR	HOOS -12
Oxford -12	0.649	0.787	0.835	0.661	0.586	0.842	0.775	0.795	0.825

Abbreviations: ADL: activity daily; QOL: quality of life; JR: Joint Replacement; PS: Physical function Short form

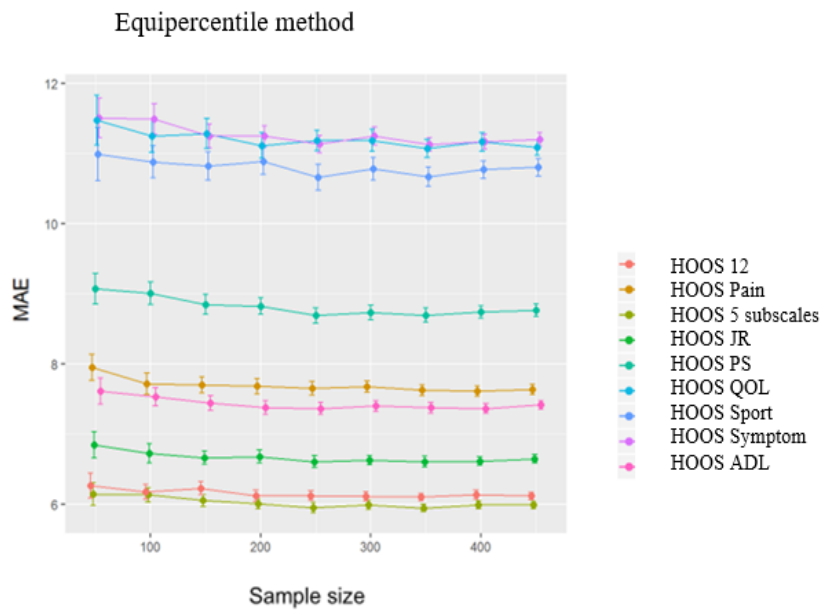
**Table 1: Pearson correlation coefficient between HOOS and Oxford-12**

Based on the 15 charts used to compare the mapping of the nine HOOS versions by method (LR, QR, Tobit, linEQ, EQUI) and criterion (MAE, R<sup>2</sup>, KS), the HOOS-12 and the HOOS global appeared to have the best mapping. Table 2 summarizes our analysis of the 15 graphs.

	Linear regression	Quantile regression	Tobit regression	Linear equating	Equipercetile method
MAE	HOOS-12 / HOOS global	HOOS-12 / HOOS global	HOOS-12 / HOOS global	HOOS-12 / HOOS global	HOOS-12 / HOOS global
KS	HOOS ADL/ HOOS global	HOOS ADL/ HOOS global	HOOS ADL/HOOS Global	HOOS global/ HOOS ADL	HOOS PS / HOOS JR / HOOS Sport / HOOS QOL
R <sup>2</sup>	HOOS 12 / HOOS global/ HOOS ADL	HOOS 12 / HOOS global/HOOS ADL	HOOS 12 / HOOS global/HOOS ADL	HOOS 12 / HOOS global/HOOS ADL	HOOS 12 / HOOS global/HOOS ADL

**Table 2: HOOS scores that provide the best mapping with the Oxford-12 based on different criteria (MAE, R<sup>2</sup> and Kolmogorov Smirnov distance) and different methods (linear regression, quantile regression, Tobit regression, linear equating, equipercetile method).**

Figure 2 is an example of the HOOS comparison using the MAE criterion and equipercetile method.



**Figure 2: Graphical representation of the HOOS based on the MAE criterion; the points are the mean value calculated on 100 bootstrap iterations of size n (n = 50 up to 500 by increasements of 50). The bars correspond to 95% confidence intervals.**

The HOOS-12 was created by the authors who created the HOOS to be used as an overall score. They do not recommend using the HOOS global as an overall score. Thus, the HOOS-12 was used for mapping purposes to generate the crosswalk tables.

***Best mapping method***

Figure 3 shows how the five methods compare based on the KS criterion by using the bootstrap estimates (mean and 95% CI) as a function of sample size for the HOOS-12 mapping.

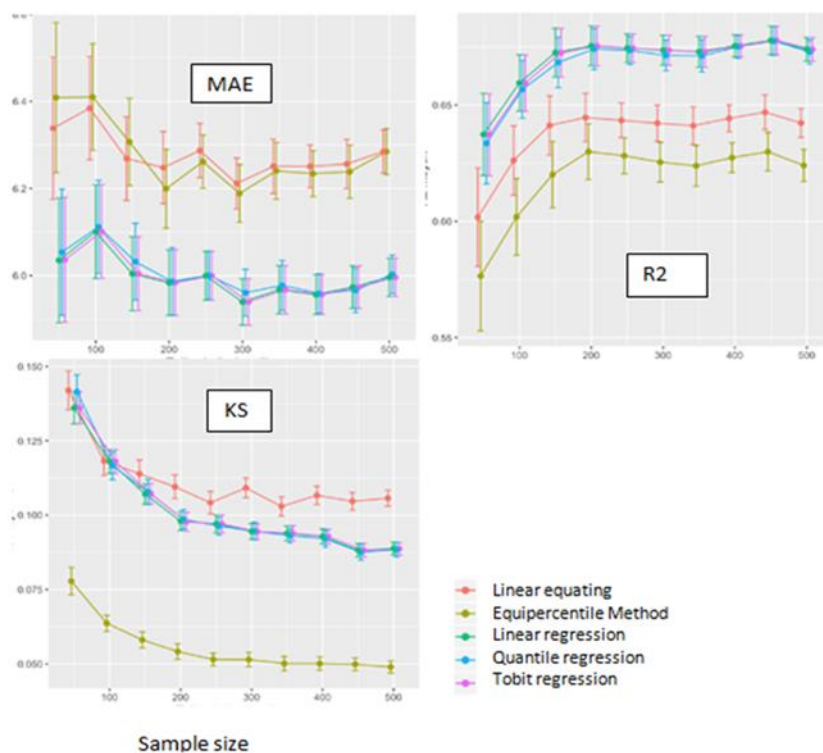


Figure 3: Graphical representation of the KS criterion for the HOOS-12 with the five methods; the points are the mean value calculated on 100 bootstrap iterations of size n (n = 50 up to 500 by increasements of 50). The bars correspond to 95% confidence

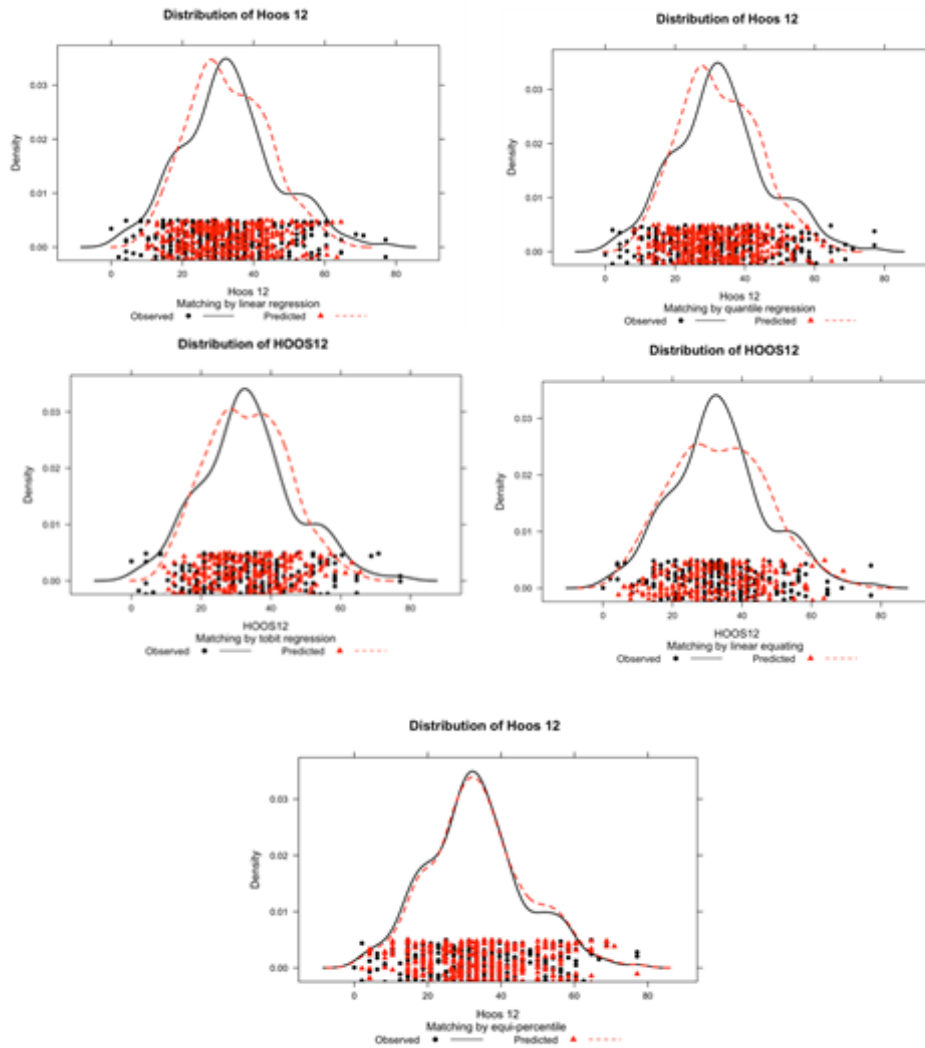
Table 3 lists the unique values for the criteria calculated on the full study cohort (N = 500).

	MAE	R <sup>2</sup>	KS
<b>Linear regression</b>	5.97	0.67	0.07
<b>Tobit regression</b>	5.97	0.67	0.07
<b>Quantile regression</b>	5.95	0.67	0.07
<b>Linear equating</b>	6.26	0.64	0.11
<b>Equipercentile method</b>	6.28	0.64	0.04

Table 3: Results of the different methods based on the predefined criteria (MAE, R<sup>2</sup>, Kolmogorov Smirnov) for mapping between the HOOS-12 and Oxford-12. The MAE was best with the regression methods (linear, Tobit and quantile) while the Kolmogorov Smirnov distance was best with the equating methods (linear equating, equipercentile method)



The methods generated similar results for the  $R^2$  (ranging from 0.64 to 0.67) and the MAE (ranging from 5.95 to 6.28). For the KS criterion, the equipercentile method was the best by far, with distance reduced by 43% relative to the regression methods and by a factor of 2.5 relative to the linear equating method. Only the equipercentile method resulted in very similar observed and estimated densities (Figure 4).



**Figure 4: Observed density of the HOOS-12 with the entire patient cohort (N = 500) and the density of the HOOS-12 after mapping from the Oxford-12. The densities are estimated for each method using the Gaussian kernel method**

Consequently, we selected the equipercentile method to generate the crosswalk tables. When we looked at how the HOOS-12 mapping performed for the five methods based on the three selection criteria, as a function of sample size, the three criteria became stable starting at 300 individuals (Figure 3).

**Crosswalk tables**

We used the preoperative and postoperative (minimum follow up 6 months) data sets (N = 695) for this step. Crosswalk tables were generated in both directions: Oxford-12 to HOOS-12 and HOOS-12 toward Oxford-12 (Tables 4, 5).

Oxford →	HOOS12		Oxford →	HOOS12
0	0		25	42
2	2		26	44
3	4		27	48
4	6		28	52
5	8		29	52
6	10		30	54
7	15		31	56
8	15		32	58
9	17		33	60
10	19		34	65
11	21		35	67
12	21		36	67
13	25		37	71
14	27		38	73
15	29		39	75
16	29		40	77
17	31		41	79
18	31		42	81
19	33		43	83
20	35		44	86
21	35		45	90
22	38		46	92
23	40		47	96
24	40		48	100

**Table 4: Crosswalk table for the Oxford-12 to the HOOS-12.**

HOOS12→	Oxford	HOOS12→	Oxford	HOOS12→	Oxford	HOOS12→	Oxford
0	2	26	14	51	28	76	39
1	2	27	14	52	28	77	39
2	2	28	15	53	30	78	41
3	3	29	15	54	30	79	41
4	3	30	17	55	31	80	42
5	5	31	17	56	31	81	42
6	5	32	18	57	32	82	43
7	5	33	18	58	32	83	43
8	5	34	20	59	33	84	44
9	6	35	20	60	33	85	44
10	6	36	22	61	34	86	44
11	7	37	23	62	34	87	44
12	7	38	23	63	34	88	45
13	7	39	23	64	34	89	45
14	7	40	25	65	35	90	45
15	9	41	25	66	35	91	45
16	9	42	26	67	36	92	47
17	10	43	26	68	36	93	47
18	10	44	27	69	37	94	47
19	11	45	27	70	37	95	48
20	11	46	27	71	37	96	48
21	13	47	27	72	37	97	48
22	13	48	28	73	39	98	48
23	13	49	28	74	39	99	48
24	13	50	28	75	39	100	48
25	14						

**Table 5: Crosswalk table for the HOOS-12 to the Oxford-12.**

## Discussion

Mapping between the HOOS and Oxford-12 is feasible using the simplified version of the HOOS (preferably the HOOS-12). The resulting crosswalk tables can be used in meta-analysis studies when a study only has one of these two PROMs. The equipercentile method appears to be the best method according to the KS distance, which represents the difference between the observed distribution and the one estimated by mapping. Starting at 300 subjects, increasing the sample size further does not improve the mapping any further.

Our study has certain limitations. 1) While the size of our sample (500 subjects) may seem small, 300 patients appear to be sufficient to validate a subjective evaluation scale as suggested by Rouquette and Falissard [29]. We used the leave-one-out cross-validation method [15] that allows each patient to be excluded from the model when it is the subject of the mapping, thereby allowing us to refine the model with a sample of 500 patients.

The graphical evaluation of the number of subjects showed stabilization of the methods, with 300 patients being sufficient, although this is not definite proof. Dakin [3] reviewed mapping studies and found that the number of subjects varied greatly, going from 48 to more than 30,000. 2) The single-center nature of our study is also a limitation, although it provided consistent primary hip replacement indications in a contemporary population. 3) While restricting the sample to patients who were scheduled for total hip replacement could be perceived as a limitation, it provided us with a population that has similar indications. 4) We had to use postoperative data to generate the crosswalk tables. Ghomrawi et al. [14] also used postoperative data to generate their crosswalk tables since using the preoperative or postoperative data sets by themselves did not result in a complete score distribution. The fact that we used the preoperative data set to validate may be questionable; however the absence of ceiling or floor effects [32] in this dataset compared to the postoperative dataset validates this data for the selection step for the method and score.

*(Question 1)* Mapping between the Oxford-12 and the HOOS or one of its derivatives is feasible. Verification of the prerequisites as recommended by Oppe et al. [22] is vital before mapping. For equating methods, Dorans [7] set a threshold of 0.3, although it is a relatively low value for correlations. We believe it is important to evaluate clinical relevance along with the degree of correlation. If the PROMs of interest are not widely used by healthcare professionals in their practice, the mapping process is useless. Polascik et al [26] used the HOOS JR, although the correlation was not as good as the HOOS-12 in our study. Because the HOOS-12 was created recently by the original authors of the HOOS and resembles the HOOS global [21], it allowed us to achieve satisfactory mapping from the Oxford-12. The 12-question format made the mapping simpler and made it possible to generate a crosswalk table. Certain HOOS

subcategories (especially the SP and QOL) were less correlated to the Oxford-12 (Table 1). The ADL subcategory had the highest but not the best correlation to the Oxford-12. With these two PROMs available, we felt it was justified to select the HOOS-12 from a clinical point of view as this better captures the various concepts of the HOOS.

*(Question 2)* Various mapping methods exist. When Dakin [3] did a review of economics literature in 2018, they found that regression methods were most used to develop health economics studies, without having identified the reference method (linear, Tobit, quantile). Equating methods were mainly developed in the context of classical test theory. Fayers and Hays in 2014 [10] and then Thompson et al. [33] in 2018 proposed supplementing regression methods with equating method for economic models. In fact, regression methods – in particular the linear regression often used for mapping – has the drawback of using the mean value, which does not always correspond to the score's distribution. Consequently, Tobit regression, quantile regression or mixed models are preferred [28]. Equating methods were used by Ghomrawi et al. [14], and by Polascik et al [26] when they performed mapping between the UCLA and the Lower Extremity Activity Scales. Lee et al. [17] demonstrated that each method has limitations when used for mapping. We were able to perfect our model by comparing these various methods. The MAE criterion is mainly used with regression methods and appears to be the best for these methods. The less specific KS distance [31] evaluating the distance between the predicted model and the observed model produces better results with equating methods. It is still relevant when comparing methods that do not use regression, like the equipercentile method, which we observed graphically.

*(Question 3)* The equipercentile method was used to generate crosswalk tables between the Oxford-12 and HOOS-12. This is the same method used by Ghomrawi et al [14] to generate crosswalk tables. This table is simple to use clinically, making it easy to go from one PROM score to another. Ghomrawi et al. [14] had proposed using this type of table so that healthcare professionals did not have to go through all the statistical steps. Since principles defining equating methods were followed, it is possible to go from one PROM to another, and vice-versa.

## **Conclusion**

Mapping is possible between the HOOS and the Oxford-12; the best mapping is achieved by pairing the HOOS-12 with the Oxford-12. The resulting crosswalk tables can be used in meta-analyses to compare clinical studies that did not include both outcome scores, or to compare the results of one study to another, or to simplify the clinical assessment since the patients would not need to complete both PROMs.

## References

1. Askew RL, Kim J, Chung H, Cook KF, Johnson KL, Amtmann D. Development of a crosswalk for pain interference measured by the BPI and PROMIS pain interference short form. *Qual Life Res.* 2013;22:2769–27
2. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference-based measures of health to generic preference-based measures. *Eur J Health Econ.* 2010 Apr;11(2):215-2
3. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health Qual Life Outcomes.* 2013;11:151.
4. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage.* 2008;16:551-559.1482.
5. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br.* 1996;78:185-9
6. Delaunay C, Epinette JA, Dawson J, Murray D, Jolles BM. Cross-cultural adaptation of the Oxford-12 HIP score to the French speaking population. *Orthop Traumatol Surg Res.* 2009;95:89-99.
7. Dorans NJ. Equating methods and sampling designs. *Appl Meas Educ.* 1990;3:3-17. ht
8. Dorans NJ. Linking scores from multiple health outcome instruments. *Qual Life Res.* 2007;16 Suppl 1:85-94.
9. Efron B. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika.* 1981;68:589–599.
10. Fayers PM, Hays RD. Should linking replace regression when mapping from profile-based measures to preference-based measures? *Value Health.* 2014;17:261-5.
11. Kiadaliri A, Alava MH, Roos EM, Englund M. Mapping EQ-5D-3L from the Knee Injury and Osteoarthritis Outcome Score (KOOS). *Qual Life Res.* 2020;29:265-274.
12. Gandek B, Roos EM, Franklin PD, Ware JE Jr. Item selection for 12-item short forms of the Knee injury and Osteoarthritis Outcome Score (KOOS-12) and Hip disability and Osteoarthritis Outcome Score (HOOS-12). *Osteoarthritis Cartilage.* 2019;27:746-753.
13. Gandek B, Roos EM, Franklin PD, Ware JE Jr. A 12-item short form of the Hip disability and Osteoarthritis Outcome Score (HOOS-12): tests of reliability, validity and responsiveness. *Osteoarthritis Cartilage.* 2019;27:754-761.
14. Ghomrawi HM, Lee YY, Herrero C, Joseph A, Padgett D, Westrich G, Parks M, Lyman S. A Crosswalk between UCLA and Lower Extremity Activity Scales *Clin Orthop Relat Res.* 2017;475:542-548.
15. Hastie T, Tibshirani R. and Friedman J. (2009) *The Elements of Statistical Learning* (2nd edition), Springer-Verlag.

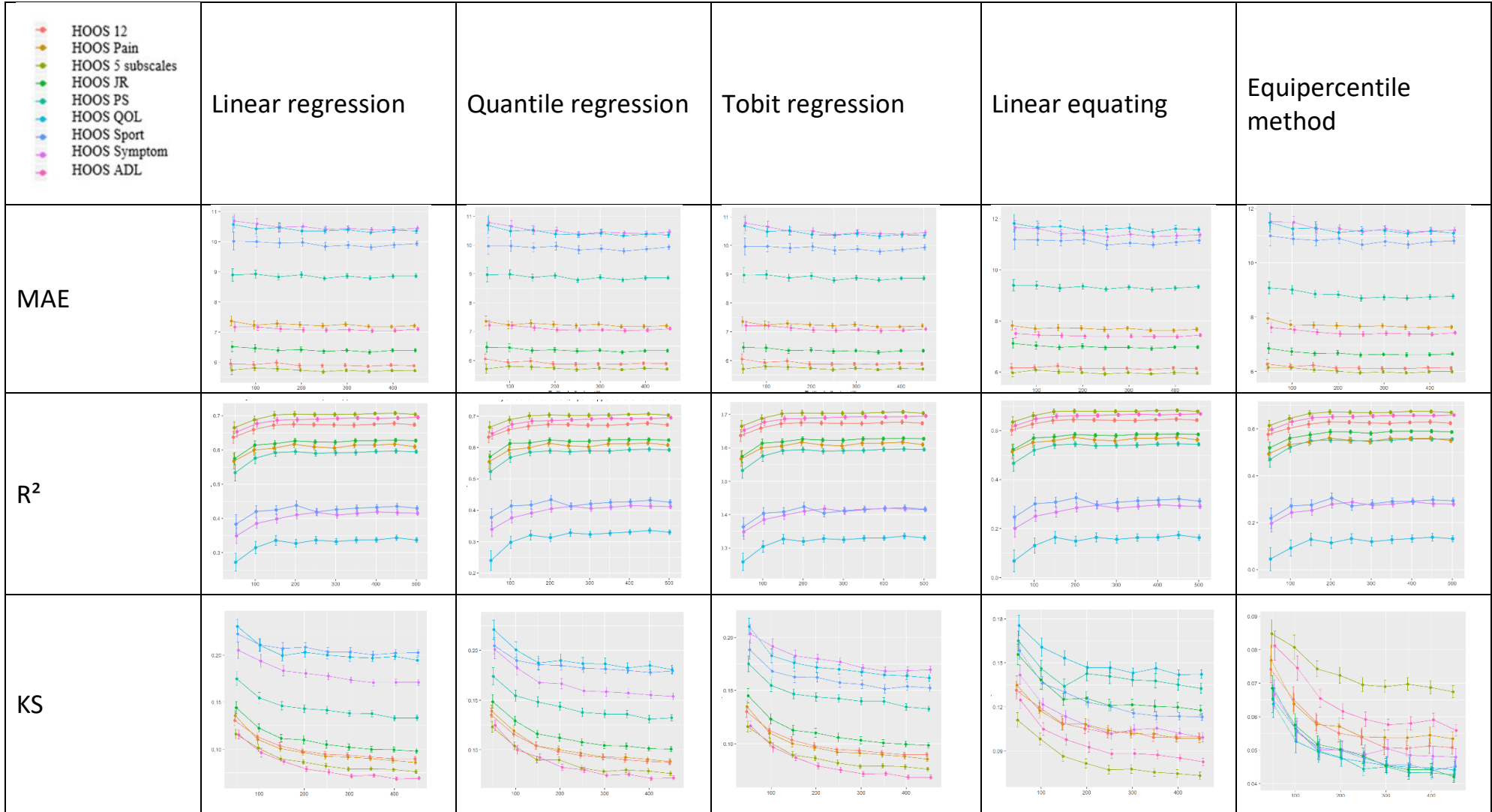
16. Hogan TP. *Psychological Testing: a practical introduction* (2e éd.). New-Jersey Wiley 2007:
17. Lee CF, Ng R, Luo N, Cheung YB. Development of Conversion Functions Mapping the FACT-B Total Score to the EQ-5D-5L Utility Value by Three Linking Methods and Comparison with the Ordinary Least Square Method. *Appl Health Econ Health Policy*. 2018;16:685-695.
18. Livingston SA. *Equating Test Scores (Without IRT)*. Princeton, NJ: Educational Testing Services; 2014.
19. Lyman S, Lee YY, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin Orthop Relat Res*. 2016;474:1472.
20. Nemes S, Garellick G, Salomonsson R, Rolfson O. Crosswalk algorithms for the conversion of mean EQ-5D indices calculated with different value sets. *Scand J Public Health*. 2016;44:455–461.
21. Nilsson AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS) validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord*. 2003 30;4:10.
22. Oppe M, Devlin N, Black N. Comparison of the underlying constructs of the EQ-5D and Oxford Hip Score: implications for mapping. *Value Health*. 2011;14:884-91.
23. Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, Roos EM, Guillemin F, Maillefert JF. Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. *Osteoarthritis Cartilage*. 2010;18:522-9.
24. Ornetti P, Perruccio AV, Roos EM, Lohmander LS, Davis AM, Maillefert JF Psychometric properties of the French translation of the reduced KOOS and HOOS (KOOS-PS and HOOS-PS). *Osteoarthritis Cartilage*. 2009;17:1604-8.
25. Oude Voshaar MA, Ten Klooster PM, Taal E, Wolfe F, Vonkeman H, Glas CA, Van De Laar MA. Linking physical function outcomes in rheumatology: performance of a crosswalk for converting Health Assessment Questionnaire scores to Short Form 36 physical functioning scale scores. *Arthritis Care Res(Hoboken)*. 2014;66:1754–1758.
26. Polascik BA, Hidaka C, Thompson MC, et al. Crosswalks Between Knee and Hip Arthroplasty Short Forms: HOOS/KOOS JR and Oxford. *J Bone Joint Surg Am*. 2020;102(11):983-990. doi:10.2106/JBJS.19.00916
27. Price LR, Lurie A, Wilkins C. EQUIPERCENT: A SAS Program for Calculating Equivalent Scores Using the Equipercentile Method. *Appl Psychol Measurement*. 2001;25:332–341.
28. Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: a systematic review. *Arthritis Rheum*. 2008;59:876-83.

29. Rouquette A, Falissard B. Sample size requirements for the internal validation of psychiatric scales. *Int J Methods Psychiatr Res.* 2011;20:235-49.
30. Sheather JS. Density estimation. *Statistical Science* 2004; 19(4): 588–597
31. Stephens MA. Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables. *J Royal Stat Soc Series B*, 32:115–122, 1970.
32. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
33. Thompson NR, Lapin BR, Katzan IL. Mapping PROMIS Total Health Items to EuroQol (EQ-5D) Utility Scores Using Linear and Equipercetile Equating. *Pharmacoeconomics.* 2017;35:1167-1176.

### **Supplementary Data**

*15 charts used to compare the mapping of the nine HOOS versions by method (LR, QR, Tobit, linEQ, EQUI) and criterion (MAE, R<sup>2</sup>, KS)*





## 4.3 Mise en perspective

Le mapping permet de comparer des résultats entre différentes séries et a pour but l'amélioration des données des méta-analyses.

Différentes techniques de Mapping existent. Dakin et al. (66) dans une revue de la littérature économique en 2018 retrouvaient essentiellement des méthodes de régression sans pour autant retrouver de méthode de référence. Parmi les méthodes de régression, la régression linéaire souvent utilisée dans les mapping optimise la prédiction moyenne et ne prend donc pas en compte la distribution du score. D'autres méthodes, comme la régression Tobit (permettant d'utiliser des valeurs cibles dans le modèle) ou alors la régression quantile (régression par la médiane ou les quantiles) peuvent être préférées. Les méthodes de régression cherchent à prédire une variable appelée variable « à expliquer » en créant un modèle prenant en compte une variable appelée variable « explicative ».

Fayers et Hays (71) en 2014 puis Thompson et al. (72) en 2018 proposaient l'utilisation des méthodes d'equating en complément des méthodes de régression dans les modèles économiques. Les méthodes d'equating ont été développées initialement dans les sciences humaines, puis ont été utilisées dans les sciences économiques en complément des méthodes de régression. Ces méthodes ne sont pas des méthodes de prédiction. Le score A permet d'obtenir le score B et le score B permet de retrouver le score A sans inadéquation. Les méthodes d'equating sont dites symétriques.

Ces méthodes de mapping posent certains problèmes. Lesquelles choisir et comment les évaluer ? Combien, de patients sont nécessaires ?

Le choix d'une méthode reste difficile. Le groupe ISPOR (73) a établi des recommandations pour les études de Mapping dans le cadre d'étude medico-économiques avec pour but d'obtenir des QALYs (Quality Adjusted Life Years). Ils ont établi des recommandations sur l'organisation des études mais ne recommandent pas une méthode ou une autre, ni un critère d'évaluation, tout étant dépendant des données. Ils recommandent néanmoins de tester différentes méthodes. Dans notre article nous avons testé plusieurs méthodes. Les critères d'évaluation comme le MAE (mean absolute error), le  $R^2$  nous ont aidés à faire notre choix mais ces critères sont plus spécifiques aux méthodes de régression. Thompson et al. (72) dans leur étude sur le mapping par les méthodes d'equating utilisent aussi le MAE ainsi que le MSE (Mean Square Error). Nous avons utilisé la distance de Kolmogorov Smirnov (74) permettant de comparer ces différentes méthodes (régression et equating) puis une évaluation graphique. Le choix de ces indicateurs reste à discuter selon les méthodes choisies. Plusieurs nous semblent nécessaires.

Le  $R^2$  est moins utilisé car moins informatif mais il a l'avantage d'être entre 0 et 1 donc plus facile d'interprétation pour les non statisticiens.

Le nombre de patients nécessaires n'est pas codifié pouvant aller de 48 à plus de 30000 dans la revue de Dakin et al. (66). Cela étant dépendant des scores et de la population étudiée (66). Une évaluation graphique a permis dans notre article de montrer une certaine stabilité vers 200 à 300 patients. L'usage d'une technique de Bootstrap a permis de conforter ce chiffre. Des études complémentaires sur un nombre minimal de patient sont nécessaires afin d'harmoniser les résultats dans la littérature.

Les méthodes de mapping est un domaine de recherche en évolution. Certains auteurs préconisent désormais des modèles de réponse à l'item. La théorie des réponses aux items (TRI ou IRT dans la littérature anglophone) s'efforce de produire une estimation des propriétés de l'item qui soit indépendante d'un groupe particulier d'individus(70). Le modèle le plus connu est le modèle de Rasch (75). Ce modèle n'est possible que si la pondération des items est identique et nécessite une unidimensionnalité du score, non réalisable dans notre travail car le score HOOS global reste un score à plusieurs dimensions. Ghomrawi et al. (67) puis Polascik et al. (68) avaient éliminé ce modèle de leur étude de mapping, notamment à cause de la nécessité de l'unidimensionnalité. Néanmoins ces modèles permettent de créer un modèle pour chaque item puis un modèle complet et leur usage pour des scores à plusieurs dimensions doit être étudié. Ils sont d'ailleurs déjà utilisés pour réduire les items d'un score (28).

## 5. Synthèse

L'usage des PROM a permis de remettre en avant le ressenti du patient. L'objectif de ceux-ci est de connaître le résultat dans notre cas d'une chirurgie fréquente et en augmentation avec pour but « une articulation oubliée » ?

Cette thèse s'est intéressée aux scores d'évaluation en orthopédie en explorant 3 facettes des scores : comment valider un score ? Comment prendre en compte le résultat d'un score ? Comment reconstruire un score à partir d'un autre score (méthode de mapping)

Chaque article présenté dans cette thèse a mis en évidence des difficultés pour harmoniser les publications sur ces scores et des perspectives sur chaque article afin d'harmoniser les résultats.

### 5.1 Points forts de la thèse

Cette thèse sur les scores en orthopédie a permis :

1. De valider un score plus facile d'utilisation que le score initial. Nous avons pu montrer que ce score à 12 items est utilisable en langue française et que celui-ci se rapproche plus d'un score d'évaluation globale. Cette validation, nous a permis de montrer sa possibilité de mapping avec le score Oxford 12(18,19).
2. De montrer une option de calcul de la DMCI à partir d'un item d'un score si celui-ci est « général ». Cette option est intéressante dans les séries rétrospectives ne permettant pas d'avoir accès au DMCI autrement que par la méthode de la distribution. Cette option qui reste encore à explorer permettra de calculer plus souvent la DMCI et ainsi d'avoir des résultats plus comparables.
3. De montrer qu'un mapping entre deux scores est possible permettant d'avoir des résultats comparables. Cet article sur le mapping s'est intéressé aux différentes méthodes pouvant y aboutir. Nous avons exploré différentes méthodes avant de choisir la méthode equipercentile choisie d'emblée par Ghomwari (67) et al. et Polascik (68) at

al. Nous avons montré que la méthode equipercentile est possible, mais que d'autres méthodes peuvent se discuter et doivent être explorées selon les cas.

## 5.2 Limites de la thèse

Ce travail comporte certaines limites :

-L'article sur la validation du score HOOS-12et KOOS-12 n'a intégré que 100 patients,. Cependant ce nombre est jugé comme un nombre suffisant dans la plupart des publications transculturelles (37,38, 44). Cette étude a été réalisée lors de l'apparition dans la littérature du score à 12 items en 2018 (article en ligne en 2018 et publié en 2019 dans la revue) (28-30). Une partie de ce travail était de s'assurer que ces scores étaient bien validés selon les recommandations actuelles.

-Les articles sur la DMCI ainsi que le mapping s'intéressent à la hanche. La création de la cohorte de hanche a été plus simple que le genou. En effet dans le protocole, nous avons exclu les chirurgies de reprise afin d'avoir une population plus homogène selon l'avis du CPP. La chirurgie de reprise de genou est une part importante de notre activité, l'inclusion a été plus longue. La cohorte genou a permis d'intégrer 300 patients mais les dernières inclusions ont plus tardives. Un article sera proposé sur le mapping de genou pour ces scores ainsi qu'un article sur la DMCI.

- La cohorte utilisée est monocentrique, créant possiblement un biais dans l'évaluation. La mise en place de ce travail de thèse ainsi que la création de la cohorte a été longue car le sujet balaye un sujet large que sont les scores d'évaluation. Ces articles peuvent servir de base à de futures recherches multicentriques.

## 5.2 Perspectives de la thèse

Ce travail a permis la création d'une cohorte prospective et permet d'envisager d'autres publications sur le sujet :

-Mapping entre les score oxford genou et le score KOOS ou ses dérivés

-Définition de la DMCI pour les scores de genou et validation de l'usage d'un item du score comme ancre

-Introduction de la notion de PASS pour ces scores

## 6. Conclusion

La chirurgie orthopédique a beaucoup évolué depuis son origine notamment avec la création des arthroplasties permettant de corriger l'usure liée au vieillissement, permettant de redonner aux patients une meilleure qualité de vie.

Mais l'amélioration des techniques nécessitent également une meilleure évaluation de celles-ci, afin de comparer des résultats et de connaître l'efficacité d'un traitement.

Les scores remplis par le patient ont permis d'améliorer cette évaluation en introduisant la notion de ressenti du patient.

Ces scores sont néanmoins encore en évolution afin de trouver des formats adaptés à la population étudiée. Une uniformisation des pratiques a été faite pour la validation de ces scores mais il reste encore des points à éclaircir. Ainsi, un consensus sur la méthode de définition de la DMCI semble nécessaire afin d'harmoniser les publications et un usage plus régulier des méthodes de mapping, permettant d'avoir des données comparables si les scores dans les articles sont différents, semble également nécessaire afin d'améliorer la qualité des Meta analyses.

## 7. Bibliographie

- 1) Lecerf G. Historique de la société française de chirurgie orthopédique et traumatologique. Un siècle d'innovation française en chirurgie orthopédique et traumatologique. ISBN978-2-7598-23-2.
- 2) Hernigou P, Poignard A, Manicom O. Historique de la prothèse de hanche. Prothèse totale de hanche. Les choix. Cahiers d'enseignement de la SOFCOT. ISBN2-84299-706-9.
- 3) Ferguson RJ, Palmer AJ, Taylor A, Porter ML, Malchau H, Glyn-Jones S. Hip replacement. *Lancet*. 2018;392(10158):1662-1671.
- 4) Evans JT, Evans JP, Walker RW, Blom AW, Whitehouse MR, Sayers A. How long does a hip replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *Lancet*. 2019;393(10172):647-654.
- 5) Erivan R, Villatte G, Dartus J, Reina N, Descamps S, Boisgard S. Progression and projection for hip surgery in France, 2008-2070: Epidemiologic study with trend and projection analysis. *OTSR*. 2019;105(7):1227-1235. doi:10.1016/j.otrs.2019.07.021
- 6) Ollivier M, Paratte S, Argenson JN. Dessin des implants et cinématique femorotibiale . Prothèses totales de genou. Cahier d'enseignement de la SOFCOT. ISBN978-2-294-74916-2.
- 7) Price AJ, Alvand A, Troelsen A, et al. Knee replacement. *Lancet*. 2018;392(10158):1672-1682.
- 8) Evans JT, Walker RW, Evans JP, Blom AW, Sayers A, Whitehouse MR. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up *Lancet*. 2019;393(10172):655-663.
- 9) Erivan R, Tardieu A, Villatte G, et al. Knee surgery trends and projections in France from 2008 to 2070. *OTSR*. 2020;106(5):893-902.
- 10) Canovas F, Dagneaux L. Quality of life after total knee arthroplasty. *OTSR* 2018;104(1S):S41–S46.
- 11) Behrend H, Giesinger K, Giesinger JM, Kuster MS. The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. *J Arthroplasty*. 2012;27(3):430-436.e1.
- 12) Puliero B, Blakeney WG, Beaulieu Y, Vendittoli PA. Joint Perception After Total Hip Arthroplasty and the Forgotten Joint. *J Arthroplasty*. 2019;34(1):65-70.
- 13) Eichler D, Beaulieu Y, Barry J, Massé V, Vendittoli PA. Perception of a Natural Joint After Total Knee Arthroplasty. *J Arthroplasty*. 2020;35(2):358-363.

- 14) D'Aubigné RM, Postel M: Functional results of hip arthroplasty with acrylic prosthesis. *J Bone Joint Surg Am.* 1954;36A:451-475.
- 15) Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: a systematic review. *Arthritis Rheum* 2008;59:876-83.
- 16) Nilsson AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS) e validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord* 2003;4:10.
- 17) Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, Roos EM, Guillemin F, Maillefert JF Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. *Osteoarthritis Cartilage* 2010;18:522-529.
- 18) Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996;78:185-90.
- 19) Delaunay C, Epinette JA, Dawson J, Murray D, Jolles BM. Cross-cultural adaptations of the Oxford-12 HIP score to the French speaking population *Orthop Traumatol Surg Res* 2009;95:89-99.
- 20) Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee injury and osteoarthritis outcome score (KOOS). Development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 1998;28:88-96.
- 21) Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, Roos EM, Guillemin F, Maillefert JF. Cross-cultural adaptation and validation of the French version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in knee osteoarthritis patients. *Osteoarthritis Cartilage* 2008;16:423-8.
- 22) Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg [Br]* 1998;80-B:63-9
- 23) Jenny JY, Diesinger Y. Validation of a French version of the Oxford knee questionnaire. *Orthop Traumatol Surg Res* 2011;97:267-71.
- 24) Bouletreau A, Chouanière D, Wild P, Fontana JM. Construire, traduire et valider un questionnaire propos d'un exemple EUROQUEST, service épidémiologie INRS, 1999;178:ISSN 0397.4529. <http://www.inrs.fr/dms/inrs/Publication/B-5-2-040-3566-01/ns178.pdf>.
- 25) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-45.



- 26) Perez JL, Mosher ZA, Watson SL, Sheppard ED, Brabston EW, et al. Readability of Orthopaedic Patient-reported Outcome Measures: Is There Fundamental Failure to Communicate? *Clin Orthop Relat Res* 2017;475:1936-1947.
- 27) Gudbergesen H, Bartels EM, Krusager P, Wæhrens EE, Christensen R, et al. Test-retest of computerized health status questionnaires frequently used in the monitoring of knee osteoarthritis: a randomized crossover trial. *BMC Musculoskelet Disord.* 2011;12:190..
- 28) Gandek B, Roos EM, Franklin PD, Ware JE Jr. Item selection for 12-item short forms of the Knee injury and Osteoarthritis Outcome Score (KOOS-12) and Hip disability and Osteoarthritis Outcome Score (HOOS-12). *Osteoarthritis Cartilage* 2019;27:746-753.
- 29) Gandek B, Roos EM, Franklin PD, Ware JE Jr. A 12-item short form of the Hip disability and Osteoarthritis Outcome Score (HOOS-12): tests of reliability, validity and responsiveness. *Osteoarthritis Cartilage.*2019;27:754-761.
- 30) Gandek B, Roos EM, Franklin PD, Ware JE, Jr. A 12-item short form of the Knee injury and Osteoarthritis Outcome Score (KOOS-12): Tests of reliability, validity and responsiveness. *Osteoarthritis Cartilage* 2019;27:762-770.
- 31) Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan, PG. An OARSI/OMERACT initiative: the development of a short measure of physical function for hip OA. HOOS-Physical Function Shortform (HOOS-PS). *Osteoarthritis Cartilage* 2008;16:551-9.
- 32) Perruccio AV, Lohmander LS, Canizares M, Tennant A, Hawker GA, Conaghan PA. The development of a short measure of physical function for knee OA. KOOS-Physical Function Short-form (KOOS-PS) – an OARSI/OMERACT initiative. *Osteoarthritis Cartilage* 2008;16:542-50.
- 33) Ornetti P, Perruccio AV, Roos EM, Lohmander LS, Davis AM, Maillefert JF Psychometric properties of the French translation of the reduced KOOS and HOOS (KOOS-PS and HOOS-PS). *Osteoarthritis Cartilage* 2009;17:1604-8.
- 34) Lyman S, Lee YY, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin Orthop Relat Res* 2016;474:1472–82.
- 35) Lyman S, Lee YY, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: A Short-form Knee Arthroplasty Outcomes Survey. *Clin Orthop Relat Res* 2016;474:1461-71.
- 36) Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
- 37) Debette C, Parratte S, Maucort-Boulch D, Blanc G, Pauly V, et al. French adaptation of the new Knee Society Scoring System for total knee arthroplasty. *OTSR* 2014;100:531-4.

- 38) Klouche S, Giesinger JM, Sariali EH. Translation, cross-cultural adaption and validation of the French version of the Forgotten Joint Score in total hip arthroplasty. *OTSR* 2018;104:657-661.
- 39) Quah C, Holmes D, Khan T, Cockshott S, Lewis J, Stephen A. The variability in Oxford hip and knee scores in the preoperative period: is there an ideal time to score?. *Ann R Coll Surg Engl* 2018;100(1):16-20.
- 40) Hung M, Saltzman CL, Greene T, et al. Evaluating instrument responsiveness in joint function: The HOOS JR, the KOOS JR, and the PROMIS PF CAT. *J Orthop Res*. 2018;36(4):1178-1184.
- 41) Browne JP, Bastaki H, Dawson J. What is the optimal time point to assess patient-reported recovery after hip and knee replacement? A systematic review and analysis of routinely reported outcome data from the English patient-reported outcome measures programme. *Health Qual Life Outcomes*. 2013;11:128.
- 42) Rouquette A, Falissard B. Sample size requirements for the internal validation of psychiatric scales. *Int J Methods Psychiatr Res* 2011;20:235-49.
- 43) Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651-7.
- 44) Erivan R, Villatte, G, Chaput, T, Mulliez, A, Ollivier M., Descamps, S., Boisgard, S. French translation and cultural adaptation of a questionnaire for patients with hip or knee prosthesis. *OTSR*, 105(3), 435–440.
- 45) Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- 46) Bourlez J, Canovas F, Duflos C, Dagneaux L. Are modern knee outcomes scores appropriate for evaluating patellofemoral degeneration in osteoarthritis? Evaluation of the ceiling and floor effects in knee outcomes scores. *Orthop Traumatol Surg Res*. 2019;105(4):599-603.
- 47) Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control ClinTrials*1989;10:407-15.
- 48) Lyman S, Lee YY, Mc Lawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clin Orthop Relat Res* 2018;476:2432-41.
- 49) Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102-9.

- 50) Copay AG, Eyberg B, Chung AS, Zurcher KS, Chutkan N, Spangehl MJ. Minimum Clinically Important Difference: Current Trends in the Orthopaedic Literature, Part II: Lower Extremity: A Systematic Review. *JBJS Rev* 2018;6:e2.
- 51) Celik D, Çoban Ö, Kılıçoğlu Ö. Minimal clinically important difference of commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a systematic review. *J Clin Epidemiol* 2019;113:44-57.
- 52) Maredupaka S, Meshram P, Chatte M, Kim WH, Kim TK. Minimal clinically important difference of commonly used patient-reported outcome measures in total knee arthroplasty: review of terminologies, methods and proposed values. *Knee Surg Relat Res.* 2020;32(1):19. Published 2020 Apr 9.
- 53) Sayers A, Wylde V, Lenguerrand E, et al. A unified multi-level model approach to assessing patient responsiveness including; return to normal, minimally important differences and minimal clinically important improvement for patient reported outcome measures. *BMJ Open.* 2017;7(7):e014041.
- 54) Sedaghat AR. Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures. *Otolaryngol Head Neck Surg* 2019;161:551-560.
- 55) Beard, D. J., Harris, K., Dawson, J., Doll, H., Murray, D. W., Carr, A. J., & Price, A. J. (2015). Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *Journal of clinical epidemiology*, 68(1), 73–79.
- 56) Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements: an illustration in rheumatology. *Arch InternMed* 1993;153:1337e42.
- 57) Jacquet C, Pioger C, Khakha R, Steltzlen C, Kley K, Pujol N, Ollivier M. Evaluation of the "Minimal Clinically Important Difference" (MCID) of the KOOS, KSS and SF-12 scores after open-wedge high tibial osteotomy *Knee Surg Sports Traumatol Arthrosc.* 2020.
- 58) Youden WJ. Index for rating diagnostic test. *Cancer* 1950;3:32-5.
- 59) Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006;163:670-5.
- 60) Aletaha D, Funovits J, Ward MM, Smolen JS, Kvien TK. Perception of improvement in patients with rheumatoid arthritis varies with disease activity levels at baseline. *Arthritis Rheum* 2009;61:313-20.
- 61) Kuo AC, Giori NJ, Bowe TR, Manfredi L, Lalani NF, Nordin DA, Harris AHS. Comparing methods to determine the Minimal Clinically Important Differences in Patient-Reported Outcome Measures for veterans undergoing elective total hip or knee arthroplasty in veterans health administration hospitals. *JAMA Surg* 2020.

- 62) Tubach F, Ravaud P, Baron G .Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Annals of the Rheumatic Diseases* 2005;64:34-37.
- 63) Levy DM, Kuhns BD, Chahal J, Philippon MJ, Kelly BT, Nho SJ. Hip Arthroscopy Outcomes With Respect to Patient Acceptable Symptomatic State and Minimal Clinically Important Difference. *Arthroscopy*. 2016;32(9):1877-1886.
- 64) Tashjian RZ, Deloach J, Porucznik CA, Powell AP. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *J Shoulder Elbow Surg*. 2009;18(6):927-932.
- 65) Lan RH, Bell JW, Samuel LT, Kamath AF. Evolving Outcome Measures in Total Knee Arthroplasty: Trends and Utilization Rates Over the Past 15 Years. *J Arthroplasty*. 2020;S0883-5403(20)30680-X
- 66) Dakin H, Abel L, Burns R, Yang Y. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health Qual Life Outcomes*. 2018;16(1):31. doi:10.1186/s12955-018-0857-3
- 67) Ghomrawi HM, Lee YY, Herrero C, Joseph A, Padgett D, Westrich G, Parks M, Lyman S. A Crosswalk between UCLA and Lower Extremity Activity Scales Clin Orthop Relat Res 2017;475:542-548.
- 68) Polascik BA, Hidaka C, Thompson MC, et al. Crosswalks Between Knee and Hip Arthroplasty Short Forms: HOOS/KOOS JR and Oxford. *J Bone Joint Surg Am*. 2020;102(11):983-990.
- 69) Dorans NJ. Equating methods and sampling designs. *Appl Meas Educ* 1990;3:3-17.
- 70) Dorans NJ. Linking scores from multiple health outcome instruments. *Qual Life Res*. 2007;16 Suppl 1:85-94.
- 71) Fayers PM, Hays RD. Should linking replace regression when mapping from profile-based measures to preference-based measures? *Value Health*. 2014;17:261-5.2106/JBJS.19.00916
- 72) Thompson NR, Lapin BR, Katzan IL. Mapping PROMIS Total Health Items to EuroQol (EQ-5D) Utility Scores Using Linear and Equipercentile Equating. *Pharmacoeconomics* 2017;35:1167-1176.
- 73) Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to Estimate Health-State Utility from Non-Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report [published correction appears in *Value Health*. 2017 Sep;20(8):1226]. *Value Health* 2017;20(1):18-27.

- 74) Stephens MA. Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables. *J Royal Stat Soc Series B*, 32:115–122, 1970.
- 75) Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Denmarks Paedagogiske Institut.

## 8. Annexes

### 8.1 Score HOOS

Le score HOOS a été créé en 2003 à partir du WOMAC. Ce questionnaire contient 40 items avec 5 choix de réponse, cotées de 0 à 4. Zéro représentant l'absence de symptômes et 4 les symptômes les plus forts.

En cas de deux cases cochées, il est recommandé de choisir le symptôme le plus fort.

Ce score est divisé en 5 catégories : douleur (D)(10 items), symptômes (S) (5 items), fonction et activités de la vie quotidienne (F) (17 items), activités sport et loisir (SP) (4 items), qualité de vie (Q) (4 items).

Les scores sont ensuite calculés sur 100 en utilisant la formule suivante :

$$\text{HOOS Douleur} = 100 - \text{MOYENNE}(D1-D10) / 4 * 100$$

$$\text{HOOS Symptômes} = 100 - \text{MOYENNE}(S1-S5) / 4 * 100$$

$$\text{HOOS Activités quotidiennes} = 100 - \text{MOYENNE}(F1-F17) / 4 * 100$$

$$\text{HOOS Sport} = 100 - \text{MOYENNE}(SP1-SP4) / 4 * 100$$

$$\text{HOOS Qualité de vie} = 100 - \text{MOYENNE}(Q1-Q4) / 4 * 100$$

Des items manquants sont possibles. Avant 2013, 2 items pouvant être manquants dans chaque sous-catégorie. Depuis 2013, des nouvelles règles de gestion des items manquants ont été introduits autorisant 50 % d'items manquants.

	Nombre d'items nécessaires avant 2013	Nombre d'items nécessaires après 2013
Douleur	8	5
Symptômes	3	3
Activité de la vie quotidienne	15	9
Sport	2	2
Qualité de vie	2	2

#### Nombre d'items manquants acceptés pour le score HOOS

Ce score a également été modifié sous différentes versions afin de réduire son nombre d'items.

Le HOOS PS se compose de 5 items correspondant à l'activité physique. Ce score est calculé par la somme des items puis est mis sur 100 grâce à une table de conversion. Initialement, 100 correspondait au plus mauvais score puis par analogie avec le score HOOS, une table de conversion a été réalisée afin de transformer le score pour obtenir un score allant de 0 à 100 avec 100 correspondant au meilleur score possible.

Le HOOS JR se compose de 6 items issus d'une diminution du nombre d'items à l'aide d'un modèle de Rasch. Ce score est également calculé en faisant la somme des items puis mis sur 100 par l'intermédiaire d'une table de conversion.

En 2019, l'équipe à l'origine du score a développé une version de 12 items le HOOS-12 considérant que les versions plus courtes n'étaient pas exhaustives.

HOOS	HOOS PS	HOOS JR	HOOS 12
<b>SYMPTOMES</b>			
S1. Bruit ?			
S2. écarter largement les jambes			
S3. marcher à grands pas			
<b>RAIDEUR</b>			
S4. Le matin au réveil			
S5. Après être resté(e) assis(e), couché(e)			
<b>DOULEUR</b>			
D1. souvent mal à la hanche ?			✓
D2. En étendant complètement la hanche			
D3. En pliant complètement la hanche			
D4. En marchant sur un terrain plat			✓
D5. En montant ou en descendant les escaliers		✓	✓
D6. Au lit la nuit :			
D7. En restant assis(e) ou couché(e) :			✓
D8. En restant debout :			
D9. En marchant sur une surface dure (asphalte, béton) :			
D10. En marchant sur une surface irrégulière :		✓	
<b>FONCTION, VIE QUOTIDIENNE</b>			
F1. Descendre les escaliers	✓		
F2. Monter les escaliers			
F3. Relever d'une position assise		✓	✓
F4. Rester debout :			✓
F5. pencher en avant pour ramasser un objet		✓	
F6. Marcher sur un terrain plat			
F7. Monter ou descendre de voiture			✓
F8. Faire vos courses			
F9. Mettre vos chaussettes ou vos collants			
F10. Sortir du lit			
F11. Enlever vos chaussettes ou vos collants			
F12. retourner ou garder la hanche dans la même position en étant couché		✓	
F13. Entrer ou sortir d'une baignoire	✓		
F14. Rester assis(e)	✓	✓	
F15. Vous asseoir ou vous relever des toilettes			
F16. Faire de gros travaux ménagers			
F17. Faire des petits travaux ménagers			
<b>ACTIVITES SPORT ET LOISIRS</b>			
SP1. Rester accroupi :			
SP2. Courir :	✓		
SP3. Tourner, pivoter sur votre jambe :	✓		
SP4. Marcher sur une surface irrégulière :			✓
<b>QUALITE DE VIE</b>			
Q1. Pensez-vous souvent à votre problème de hanche ?			✓
Q2. modifié votre façon de vivre			✓
Q3. manque de confiance			✓
Q4 êtes-vous gêné(e) par votre hanche ?			✓
Nombre items 40	5	6	12

Items correspondant au score HOOS, HOOS PS, HOOS JR et HOOS -12



## Questionnaire de Hanche HOOS

DATE: \_\_\_\_\_ DATE DE NAISSANCE: \_\_\_\_\_

Ce questionnaire vous demande votre opinion sur votre hanche. 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. Ressentez-vous des ou entendez-vous des craquements ou n'importe quel autre type de bruit provenant de votre hanche ?

- Jamais       Rarement       Parfois       Souvent       Toujours

S2. Quelle est votre difficulté pour écarter largement les jambes ?

- Absente       Légère       Modérée       Forte       Extrême

S3. Quelle est votre difficulté pour marcher à grands pas ?

- Absente       Légère       Modérée       Forte       Extrême

### RAIDEUR

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

S4. Le matin au réveil, la raideur de votre hanche est :

- Absente       Légère       Modérée       Forte       Extrême

S5. Après être resté(e) assis(e), couché(e), ou au repos pendant la journée, la raideur de votre hanche est :

- Absente       Légère       Modérée       Forte       Forte Extrême

### DOULEUR

D1. Avez-vous souvent mal à la hanche ?

- Jamais       Une fois/mois       Une fois/semaine       Tous les jours       Tout le temps

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

D2. En étendant complètement la hanche :

- Absente       Légère       Modérée       Forte       Extrême

D3. En pliant complètement la hanche :

- Absente       Légère       Modérée       Forte       Extrême

D4. En marchant sur un terrain plat :

- Absente       Légère       Modérée       Forte       Extrême

D5. En montant ou en descendant les escaliers :

- Absente       Légère       Modérée       Forte       Extrême

D6. Au lit la nuit :

- Absente       Légère       Modérée       Forte       Extrême

D7. En restant assis(e) ou couché(e) :

Absente       Légère       Modérée       Forte       Extrême

D8. En restant debout :

Absente       Légère       Modérée       Forte       Extrême

D9. En marchant sur une surface dure (asphalte, béton) :

Absente       Légère       Modérée       Forte       Extrême

D10. En marchant sur une surface irrégulière :

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 ?

F1. Descendre les escaliers :

Absente       Légère       Modérée       Forte       Extrême

F2. Monter les escaliers :

Absente       Légère       Modérée       Forte       Extrême

F3. Vous relever d'une position assise :

Absente       Légère       Modérée       Forte       Extrême

F4. Rester debout :

Absente       Légère       Modérée       Forte       Extrême

F5. Vous pencher en avant pour ramasser un objet :

Absente       Légère       Modérée       Forte       Extrême

F6. Marcher sur un terrain plat :

Absente       Légère       Modérée       Forte       Extrême

F7. Monter ou descendre de voiture :

Absente       Légère       Modérée       Forte       Extrême

F8. Faire vos courses :

Absente       Légère       Modérée       Forte       Extrême

F9. Mettre vos chaussettes ou vos collants :

Absente       Légère       Modérée       Forte       Extrême

F10. Sortir du lit :

Absente       Légère       Modérée       Forte       Extrême

F11. Enlever vos chaussettes ou vos collants :

- Absente       Légère       Modérée       Forte       Extrême

F12. Vous retourner ou garder la hanche dans la même position en étant couché :

- Absente       Légère       Modérée       Forte       Extrême

F13. Entrer ou sortir d'une baignoire :

- Absente       Légère       Modérée       Forte       Extrême

F14. Rester assis(e) :

- Absente       Légère       Modérée       Forte       Extrême

F15. Vous asseoir ou vous relever des toilettes :

- Absente       Légère       Modérée       Forte       Extrême

F16. Faire de gros travaux ménagers (déplacer des objets lourds, récurer les sols,...) :

- Absente       Légère       Modérée       Forte       Extrême

F17. Faire des petits travaux ménagers (faire la cuisine, faire la poussière,...) :

- Absente       Légère       Modérée       Forte       Extrême

### ACTIVITES 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 :

- Absente       Légère       Modérée       Forte       Extrême

SP2. Courir :

- Absente       Légère       Modérée       Forte       Extrême

SP3. Tourner, pivoter sur votre jambe :

- Absente       Légère       Modérée       Forte       Extrême

SP4. Marcher sur une surface irrégulière :

- Absente       Légère       Modérée       Forte       Extrême

### QUALITE DE VIE

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

- Jamais       Une fois/mois       Une fois/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 hanche ?

- Pas du tout       Un peu       Modérément       Beaucoup       Totalement

Q3. Etes-ce qu'un manque de confiance dans votre hanche vous gêne ?

Pas du tout       Un peu       Modérément       Beaucoup       Totalement

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

Pas du tout       Un peu       Modérément       Beaucoup       Extrêmement

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

## 8.2 Score Oxford de hanche

Ce questionnaire créé en 1996 contient 12 items avec 5 choix de réponse.

Initialement crée sur de 12 (meilleur score) à 60 points (plus mauvais score). Les auteurs recommandent de le coter de 0 (plus mauvais score) à 48 points (meilleur score) avec des items, cotés de 0 à 4. Quatre représentants l'absence de symptômes et 0 les symptômes les plus forts.

(Murray, D. W., Fitzpatrick, R., Rogers, K., Pandit, H., Beard, D. J., Carr, A. J., & Dawson, J. (2007). The use of the Oxford hip and knee scores. *The Journal of bone and joint surgery. British volume*, 89(8), 1010–1014.)

Ces items sont ensuite additionnés afin de calculer un score de 0 à 48.

0 = plus mauvais score

48 = meilleur score, absence de symptômes

Il peut également apparaitre des présentations de résultats sous forme d'un score sur 60. Dans ce cas, la présentation est inversée : 60 correspond au plus mauvais score et 12 au meilleur score possible.

Il est facile de passer d'un score sur 48 à un score sur 60 en utilisant la formule :  $60 - (\text{résultat du score sur 48}) = \text{Résultat du patient sur 60}$

## Questionnaire Oxford Prothèse de Hanche

DATE: \_\_\_\_\_

Coté Opéré :  Droit  Gauche

*Le but de cette échelle est d'aider à évaluer l'impact que votre douleur à la hanche a eu sur votre vie quotidienne au cours des quatre dernières semaines. Le résultat sera l'un des facteurs que votre médecin prendra en considération avant de discuter avec vous les prochaines étapes dans la gestion de votre problème.*

*Cocher une seule case pour chaque question.*

### **Durant ces 4 dernières semaines**

1. Comment décririez-vous la douleur que vous habituellement ressentie dans votre hanche ?

- Aucune  Minime  Légère  Modérée  Sévère

2. Avez-vous eu des difficultés pour vous laver et vous sécher le corps vous-même (des pieds à la tête) à cause de votre hanche ?

- Aucune difficulté  Difficultés minimes  Difficultés modérées  Difficultés majeures  Impossible à réaliser

3. Avez-vous eu des difficultés à cause de votre hanche pour entrer ou sortir d'une voiture ou pour utiliser les transports en commun (*quelque soit le mode de transport utilisé*)?

- Aucune difficulté  Difficultés minimes  Difficultés modérées  Difficultés majeures  Impossible à réaliser

4. Avez-vous été capable de mettre seul(e) vos bas, collants ou chaussettes ?

- Oui, facilement  Avec très peu de difficultés  Avec quelques difficultés  Avec beaucoup de difficultés  Non, impossible

5. Avez-vous pu faire tout(e) seul(e) des courses pour la maison ?

- Oui, facilement  Avec très peu de difficultés  Avec quelques difficultés  Avec beaucoup de difficultés  Non, impossible

6. Combien de temps pouviez-vous marcher (sans vous arrêter) avant que la douleur dans votre hanche ne devienne très importante ? (avec ou sans canne) ?

- Pas de douleur ou plus de 30  De 16 à 30 minutes  De 5 à 15 minutes  Autour de la maison seulement  Pas du tout ou douleur sévère à la

7. Avez-vous pu monter un étage par les escaliers ?

- Oui, facilement  Avec très peu de difficultés  Avec quelques difficultés  Avec beaucoup de difficultés  Non, impossible

8. Après être resté assis (pour un repas par exemple), quel degré de douleur avez-vous ressenti en vous levant de la chaise à cause de votre hanche ?

- Pas douloureux du tout  Légèrement douloureux  Modérément douloureux  Très douloureux  Insupportable

9. Avez-vous boité en marchant, à cause de votre hanche ?

- Rarement ou jamais  Quelquefois, ou juste au début  Souvent, pas seulement au début  La plupart du temps  Tout le temps

10. Avez-vous ressenti au niveau de votre hanche malade (ou opérée) une douleur soudaine, vive et intense (en coup de poignard, spasme, en vrille, etc...) ?

- Jamais  Seulement 1 ou 2 jours  Quelques jours  La plupart des jours  Chaque jour

11. La douleur de votre hanche vous a-t-elle gêné(e) dans votre travail ou vos activités habituelles (tâches ménagères comprises) ?

- Pas du tout  Un peu  Modérément  Fortement  Tout le temps

12. Avez-vous souffert de douleurs de votre hanche au lit la nuit ?

- Jamais  Seulement 1 ou 2 nuits  Quelques nuits  La plupart des nuits  Toutes les nuits

### 8.3 Score KOOS

Ce questionnaire crée en 1998 contient 42 items avec 5 choix de réponse, cotées de 0 à 4. Zéro représentant l'absence de symptômes et 4 les symptômes les plus forts.

En cas de deux cases cochées, il est recommandé de choisir le symptôme le plus fort.

Ce score est divisé en 5 dimensions : douleur (P) (9 items), symptômes (S) (7 items) fonction et activités de la vie quotidienne (A) (17 items), activités sport et loisir (SP) (5 items), qualité de vie (Q) (4 items).

Les scores sont ensuite calculés sur 100 en utilisant la formule suivante :

KOOS Douleur:  $=100 - \text{MOYENNE}(P1-P9)/4 \times 100$

KOOS Symptomes:  $=100 - \text{MOYENNE}(S1-S7)/4 \times 100$

KOOS Activités quotidiennes :  $=100 - \text{MOYENNE}(A1-A17)/4 \times 100$

KOOS Sport:  $=100 - \text{MOYENNE}(SP1-SP5)/4 \times 100$

KOOS Qualité de vie :  $=100 - \text{MOYENNE}(Q1-Q4)/4 \times 100$

Des items manquants sont possibles. Avant 2012, 2 items pouvant être manquants dans chaque sous-catégorie. Depuis 2012, des nouvelles règles de gestion des items manquants ont été introduits autorisant 50 % d'items manquants.

	Nombre d'items nécessaires avant 2012	Nombre d'items nécessaires après 2012
Douleur	7	5
Symptômes	5	4
Activité de la vie quotidienne	15	9
Sport	3	3
Qualité de vie	2	2

#### Nombre d'items manquants autorisés pour le score KOOS

Ce score a également été modifié sous différentes versions afin de réduire son nombre d'items

Le KOOS PS se compose de 7 items correspondant à l'activité physique. Initialement, 100 correspondait au plus mauvais score puis par analogie avec le score HOOS, une table de conversion a été réalisé afin de transformer le score pour obtenir un score allant de 0 à 100 avec 100 correspondant au meilleur score possible.

Le KOOS JR se compose de 7 items issus d'une diminution du nombre d'items par la réalisation d'un modèle de Rasch. Ce score est également calculé en faisant la somme des items puis mis sur 100 par l'intermédiaire d'une table de conversion.

En 2018, l'équipe à l'origine du score a développé une version de 12 items le KOOS 12 considérant que les versions plus courtes n'étaient pas exhaustives.



<b>KOOS</b>	<b>KOOS PS</b>	<b>KOOS JR</b>	<b>KOOS 12</b>
<b>SYMPTOMES</b>			
S1. gonflement ?			
S2. Bruits			
S3. Blocage			
S4. Extension complète			
S5. Flexion complète			
<b>RAIDEUR</b>			
S6. Le matin au réveil		✓	
S7. Après être resté(e) assis(e), couché(e)			
<b>DOULEUR</b>			
P1. souvent mal au genou			✓
P2. En tournant, pivotant		✓	
P3. En étendant complètement		✓	
P4. En pliant complètement			
P5. En marchant sur un terrain plat			✓
P6. En montant ou en descendant les escaliers		✓	✓
P7. Au lit la nuit			
P8. En restant assis(e) ou couché(e) :			✓
P9. En restant debout :		✓	
<b>FONCTION. VIE QUOTIDIENNE</b>			
A1. Descendre les escaliers			
A2. Monter les escaliers			
A3. Relever d'une position assise	✓	✓	✓
A4. Rester debout			✓
A5. Pencher en avant pour ramasser un objet	✓	✓	
A6. Marcher sur un terrain plat			
A7. Monter ou descendre de voiture			✓
A8. Faire vos courses			
A9. Mettre vos chaussettes ou vos collants	✓		
A10. Sortir du lit	✓		
A11. Enlever vos chaussettes ou vos collants			
A12. Retourner en étant couché			
A13. Entrer ou sortir d'une baignoire			
A14. Rester assis(e)			
A15. Vous asseoir ou vous relever des toilettes			
A16. Faire de gros travaux ménagers			
A17. Faire des petits travaux ménagers			
<b>ACTIVITES SPORT ET LOISIRS</b>			
SP1. Rester accroupi(e)	✓		
SP2. Courir			
SP3. Sauter			
SP4. Tourner, pivoter sur votre jambe	✓		✓
SP5. Rester à genoux	✓		
<b>QUALITE DE VIE</b>			
Q1. Pensez-vous souvent à votre problème de genou			✓
Q2. Modifier votre façon de vivre			✓
Q3. Manque de confiance dans votre genou			✓
Q4. êtes-vous gêné(e) par votre genou			✓
<b>Nombre items 42</b>	<b>7</b>	<b>7</b>	<b>12</b>

**Items correspondants au score KOOS, KOOS PS, KOOS JR et KOOS-12**

## Questionnaire de Genou KOOS

DATE: \_\_\_\_\_ DATE DE NAISSANCE: \_\_\_\_\_

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 provenant de 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/mois       Une fois/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 la hanche dans la même position en étant couché :

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

## ACTIVITES 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

#### QUALITE DE VIE

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

Jamais       Une fois/mois       Une fois/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. Etes-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\*\*\***

#### **8.4 Score Oxford de genou**

Ce questionnaire crée en 1998 contient 12 items avec 5 choix de réponse, cotées de 0 à 4. Quatre représentant l'absence de symptômes et 0 les symptômes les plus forts.

Ces items sont ensuite additionnés afin de calculer un score de 0 à 48.

0 = plus mauvais score

48 = meilleur score, absence de symptômes

## Questionnaire Oxford 12 pour le Genou

DATE: \_\_\_\_\_

Coté Opéré :  Droit  Gauche

*Le but de cette échelle est d'aider à évaluer l'impact que votre douleur au genou a eu sur votre vie quotidienne au cours des quatre dernières semaines. Le résultat sera l'un des facteurs que votre médecin prendra en considération avant de discuter avec vous les prochaines étapes dans la gestion de votre problème.*

*Cocher une seule case pour chaque question.*

### **Durant ces 4 dernières semaines**

1. Comment décririez-vous la douleur que vous habituellement ressentie dans votre genou ?

- Nulle  Très légère  Légère  Modérée  Sévère

2. Avez-vous eu des difficultés pendant la toilette et l'habillage à cause de votre genou ?

- Aucune  Très légères  Légères  Importantes  Activité impossible

3. Avez-vous eu des difficultés à rentrer ou sortir d'une voiture ou à utiliser les transports en commun à cause de votre genou?

- Aucune  Très légères  Légères  Importantes  Activité impossible

4. Pendant combien de temps pouvez-vous marcher sans ressentir des douleurs importantes du genou?

- Plus de 30 minutes  De 16 à 30 minutes  De 5 à 15 minutes  Moins de 5 minutes  Les douleurs apparaissent immédiatement

5. Après être resté assis longtemps, avez-vous des douleurs du genou lorsque vous vous relevez ?

- Aucune  Très légères  Légères  Importantes  Insupportables

6. Est-ce que vous boitez à cause de votre genou ?

- Jamais rarement  Parfois  Souvent  Le plus souvent  Toujours

7. Pouvez-vous vous accroupir et vous relever?

- Facilement  Avec des petites difficultés  Avec des difficultés moyennes  Avec beaucoup de difficultés  C'est impossible

8. Avez-vous des douleurs du genou pendant la nuit?

- Jamais  Parfois  Souvent  Le plus souvent  Toujours

9. Est-ce que les douleurs de votre genou gênent vos activités habituelles ?

- Jamais       Un peu       Modérément       Beaucoup       Les activités sont impossibles

10. Avez-vous l'impression que votre genou est instable ?

- Jamais       Parfois       Souvent       Le plus souvent       Toujours

11. Pouvez-vous faire vos achats vous-même ?

- Facilement       Avec des petites difficultés       Avec des difficultés moyennes       Avec beaucoup de difficultés       C'est impossible

12. Pouvez-vous descendre les escaliers ?

- Facilement       Avec des petites difficultés       Avec des difficultés moyennes       Avec beaucoup de difficultés       C'est impossible

## 8.5 Avis du comité pour la protection des personnes

### COMITÉ DE PROTECTION DES PERSONNES « EST IV »

1, place de l'Hôpital  
67091 STRASBOURG Cedex  
☎ : 03.88.11.60.03  
Fax : 03.88.11.63.48  
E-mail : cpp.est4@chru-strasbourg.fr

Strasbourg, le 6 avril 2018

**Le Président**

CHRU DE LILLE  
DIRECTION DE LA RECHERCHE EN SANTÉ

18 AVR. 2018

**Madame Amélie HERBAUT LECOCQ**  
Département de la Recherche en Santé  
CHRU Lille  
2 avenue Oscar Lambret  
59037 LILLE Cedex

<b>Référence CPP</b>	17/ 59
<b>Titre de la recherche</b>	Etude des auto-questionnaires d'évaluation utilisés en chirurgie de la hanche et du genou
<b>Identification</b>	IDRCB 2017-A01911-52 / Référence 2016_72
<b>Promoteur</b>	CHRU Lille 2 avenue Oscar Lambret 59037 LILLE Cedex
<b>Investigateur principal</b>	Docteur Sophie PUTMAN Hôpital Roger Salengro Service d'Orthopédie Place de Verdun 59037 Lille Cedex

Madame,

Je vous remercie pour votre réponse du 26 décembre 2017 complétée par mail du 29 mars 2018, faisant suite à mon courrier du 13 octobre 2017 concernant le projet de recherche référencé ci-dessus.

Les précisions apportées par le promoteur répondent aux attentes du Comité et j'ai donc le plaisir de vous adresser l'avis favorable que le CPP Est IV a émis lors de sa séance du 10 octobre 2017.

Je vous prie de croire, Madame, en l'assurance de mes salutations les meilleures.

Professeur Philippe WOLF



Les scores d'autoévaluation (PROM = Patient Reported Outcome Measurement) font partie intégrante de l'évaluation moderne des patients. Ils permettent par exemple en orthopédie de définir un statut de base pour ces patients et ensuite après un traitement chirurgical ou médical, de définir une amélioration. Ces scores nécessitent des étapes de validation permettant leur utilisation (validité interne et externe, effet plancher et plafond, répétabilité, sensibilité au changement, traduction dans différentes langues.). Une fois ces étapes effectuées, la différence minimale cliniquement importante (DMCI ou MCID pour Minimal Clinical Important Difference) pour le patient peut être définie afin de fixer un niveau minimum d'amélioration. La DMCI, qui est une notion variable d'un pays à l'autre, peut être calculée par différentes méthodes notamment la méthode des ancres nécessitant une question supplémentaire afin de définir si le patient se juge comme amélioré ou non. Malgré ces données, le manque d'harmonisation dans le choix des scores rend difficile la comparaison de ces scores notamment pour les méta-analyses ou les revues de la littérature. Ainsi pour cette thèse, une cohorte prospective de patients souffrant d'une arthrose de hanche et du genou a été créée afin : 1) de valider différents scores, 2) de définir la DMCI pour ces scores en France, 3) de permettre d'obtenir par un processus statistique, un moyen de connaître le résultat d'un score en n'ayant à sa disposition que le résultat d'un autre score (mapping ou crosswalk). A l'occasion de cette thèse, les score HOOS-12 et KOOS-12, version réduite à 12 items du score HOOS (40 items) et du score KOOS (42 items) ont bénéficié d'une validation en français. La DMCI pour les scores HOOS et Oxford-12 hanche a pu être définie. Cette thèse a permis d'explorer une option de calcul de la DMCI en utilisant un item du score lui-même permettant d'envisager l'usage de la méthode des ancres en l'absence d'une question supplémentaire. Différentes méthodes de mapping ont été testées sur le score HOOS et tous ses dérivés afin de définir la méthode la plus optimale ainsi que le score le plus adapté pour obtenir un mapping avec le score Oxford-12. La méthode equipercentile a permis de réaliser une table de conversion entre le score Oxford-12 Hanche et le score HOOS-12.

Patient reported outcome measurement scores (PROMs) are part of modern patient's assessment. For example, in orthopedics they are used to define a baseline level and after a surgical or medical treatment, to define an improvement. These scores require a validation procedure (internal and external validity, Responsiveness, reliability, translation in different languages, assessment of floor and ceiling effect...). Patient's satisfaction after a surgical procedure can be measured through MCID (Minimal Clinical Important Difference). The MCID represents the minimum clinically relevant difference for the patient but is different between countries. MCID can be calculated by various methods, including the anchor method requiring an additional question to define the improvement. Despite these data, the lack of harmonization in the choice of scores makes difficult to compare these scores, especially for meta-analyses or systematic reviews. In this thesis, a prospective hip and knee cohort was created to: 1) validate different scores, 2) define the MCID of these scores in France, 3) define a statistical process to deduct a score result from another score (Mapping or crosswalk). The HOOS-12 and KOOS-12 scores (abridged versions to 12 items of the HOOS score (40 items) and the KOOS score (42 items)) were validated in French. The MCID for HOOS and hip Oxford scores has been defined for the French population. An option to calculate the MCID using a score item itself was explored, which allowed us to consider using the anchor method in the absence of additional questions. Different methods have been tested to define the most optimal method and the most optimal score for mapping between HOOS and Oxford-12. A conversion table was created between the Oxford-12 hip score and HOOS-12 by the equipercentile method.