

MINISTERE DE LA SANTE
REGION GRAND EST
INSTITUT LORRAIN DE FORMATION DE MASSO-KINESITHERAPIE DE NANCY

**EFFICACITE DES TESTS CLINIQUES DE LA RHIZARTHROSE : UNE REVUE
SYSTEMATIQUE DE LA LITTERATURE AVEC INITIATION A LA META-ANALYSE
STATISTIQUE**

SOUS LA DIRECTION DE M. Thomas CLEMENT

Mémoire présenté par **Arnaud SCHWARZ**,
étudiant en 5ème année de masso-kinésithérapie,
en vue de valider l'UE28 dans le cadre de la
formation initiale du Diplôme d'État de
Masseur-Kinésithérapeute
Promotion 2022-2026.



Avertissement

Ce document est le fruit d'un long travail et a été validé par l'auteur et son directeur de mémoire en vue de l'obtention de l'UE 28, Unité d'Enseignement intégrée à la formation initiale de masseur-kinésithérapeute.

L'IFMK de Nancy n'est pas garant du contenu de ce mémoire mais le met à disposition de la communauté scientifique élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : secretariat@kine-nancy.eu

Liens utiles

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<https://www.service-public.fr/professionnels-entreprises/vosdroits/F23431>



UE28

Mémoire

DÉCLARATION SUR L'HONNEUR CONTRE LE PLAGIAT

Je soussigné SCHWARZ Arnaud,

étudiant en K5 et réalisant un mémoire d'initiation à la recherche et sous la direction de M. CLEMENT Thomas l'IFMK de Nancy,

certifie que le manuscrit présenté en vue de la soutenance est le fruit d'un travail original et que toutes les sources utilisées ont été clairement indiquées.

Je certifie, de surcroît, que je n'ai ni copié, ni utilisé des idées ou des formulations tirées d'un ouvrage, article ou mémoire, en version imprimée ou électronique, sans mentionner précisément leur origine et que les citations sont expressément signalées entre guillemets (ou par une autre disposition graphique sans ambiguïté).

Conformément à la loi, le non-respect de ces dispositions me rend passible de poursuites devant la commission disciplinaire et les tribunaux de la République française pour plagiat universitaire.

Fait à Nancy, le 11/01/2026

Signature



UE28

Mémoire

Attestation d'utilisation éthique d'une Intelligence Artificielle (IA)

1. Informations générales

Je soussigné SCHWARZ Arnaud,

Étudiant(e) en K5 à l'IFMK de Nancy réalisant un mémoire d'initiation à la recherche,
sous la direction de M. CLEMENT Thomas.

2. Utilisation de l'IA (indiquer votre autoévaluation avec l'échelle numérique dans la dernière colonne)

Aide à la compréhension de concepts / recherche bibliographique	0 (pas du tout) ----- ☐ 10 (beaucoup)	0
Traduction	0 (pas du tout) ----- ☐ 10 (beaucoup)	2
Reformulation / amélioration du style	0 (pas du tout) ----- ☐ 10 (beaucoup)	0
Planification ou structure du manuscrit	0 (pas du tout) ----- ☐ 10 (beaucoup)	0
Génération de contenu (texte, figures, images...)	0 (pas du tout) ----- ☐ 10 (beaucoup)	0
Traitement des résultats et analyse des données	0 (pas du tout) ----- ☐ 10 (beaucoup)	2

Autres utilisations à préciser :

NA

Outils d'IA utilisés (ChatGPT, Gemini, Perplexity, Mistral...) :

Rayyan ; DeepL ; ChatGPT

3. Engagement

- Je m'engage à ce que cette déclaration reflète fidèlement mon usage de l'IA dans le respect des règles des textes réglementaires en vigueur (règlement intérieur, charte sur les IA, règlement des examens...).

Je déclare également avoir :

- Vérifié la pertinence des informations fournies par l'intelligence artificielle,
- Cité les sources de mes emprunts,
- Apporté une contribution personnelle suffisante à l'élaboration du devoir.

Fait à Nancy, le 11/01/2026

Signature

Remerciements

A M. Gouilly, pour m'avoir donné ma chance de pouvoir changer d'orientation professionnelle et d'exercer ce si beau métier. Le culot a payé !

A M. Clément, pour sa prise en charge dans la réussite de ce travail de Mémoire. Nous avons sans doute perdu quelques cheveux dans la bataille... Mais cela valait le coup !

A M. Ferring, pour son humanité, sa volubilité unique, et pour avoir été un référent pédagogique sur lequel j'ai toujours pu compter pendant la formation.

Aux différents professeurs de l'Institut, qui ont permis de nourrir notre kiné et nous faire comprendre que le compagnonnage est sans doute l'aspect le plus important du métier.

Au personnel administratif, qui sait se montrer aux petits soins au quotidien : ce sont des héros de l'Institut à leur échelle (Mme Biediger, Mme Heintz, Mme Serdet, M. Jacquot, une pensée aussi pour Mme Mangin qui a été mon premier contact de l'IFMK).

A Mme Ferry, qui a validé initialement ma démarche de reconversion au sein du CHRU de Nancy, lorsque je n'étais encore qu'un petit manip radio intéressé par la kiné hospitalière.

A mes blessures répétées, qui m'ont amené à fréquenter le cabinet de ceux qui deviendront d'ici quelques temps des confrères, et qui ont fait quelque part germer la petite graine.

A Papa et Maman, merci de m'avoir supporté, moi et mon caractère pas toujours facile, depuis toutes ces années. Nous allons pouvoir nous retrouver pleinement, après ces moments de « révisions » sans fin.

A Alexis, j'espère que le temps n'aura pas d'emprise sur notre relation et que nous saurons toujours être là l'un pour l'autre, comme depuis le premier jour de notre existence. TDB.

A Mamie Thérèse, tu es le pilier de la famille et j'espère pouvoir encore profiter de ton expérience de la vie le plus tard possible.

A tous les amis de ma vie d'avant, qui m'ont soutenu dans ma démarche et ne m'ont pas oublié pendant ces quatre ans où je n'ai pas été un bon copain.

A mes amis de l'IFMK, ceux qui sont tombés en cours de route, ceux qui sont allés au bout : la compréhension, le soutien et l'entraide ont compté énormément pour mener à bien notre projet commun (je ne nomme personne, les vrai(e)s savent). Merci également à mes relecteurs fraîchement diplômés, Marion et Alexis.

A Didier, Bernard, Pascal, et tant d'autres...

Aux patients, sans qui nous ne sommes rien.

A Arnold, Choa, Gelberman, Merritt, Model, Ribak et Sela, sans qui ce travail n'existerait pas. Et surtout à Manon, soutien indéfectible et pierre angulaire de ma vie depuis six ans. Tout ceci n'aurait jamais été possible sans toi, sache-le. Nous allons enfin pouvoir reprendre sereinement le cours de notre belle histoire, et tu n'auras plus besoin de parler en mon nom pour dire aux autres comment ça se passe en kiné ☺

Résumé

Réalisé avec le support de la liste de contrôle PRISMA 2020 for Abstracts (Preferred Reporting Items for Systematic reviews and Meta-Analyses), disponible à l'adresse : <https://www.prisma-statement.org/abstracts>.

EFFICACITE DES TESTS CLINIQUES DE LA RHIZARTHROSE : UNE REVUE SYSTEMATIQUE DE LA LITTERATURE AVEC INITIATION A LA META-ANALYSE STATISTIQUE		
Introduction	<i>Généralités</i>	La rhizarthrose, maladie articulaire dégénérative, est diagnostiquée selon des paramètres cliniques et para-cliniques convergents, mais non harmonisés.
	<i>Objectifs</i>	Après une collecte de données scientifiques, nous souhaitons proposer un <i>cluster</i> de tests, facilitant le triage kinésithérapique.
Matériel et méthode	<i>Critères d'éligibilité</i>	Les articles devaient traiter du diagnostic clinique de la pathologie, incluant des données statistiques, et les tests décrits précisément. L'intervention devait être réalisée par tout intervenant formé et habilité. Les articles devaient inclure des patients non opérés, sans limite d'âge et rédigés en français ou anglais.
	<i>Sources d'informations</i>	Nous avons exploré les bases de données PubMed, Embase, LILACS et PEDro entre le 23/07/2025 et le 05/01/2026.
	<i>Risque de biais</i>	La recherche fut supportée par les grilles QUADAS.
	<i>Synthèse des résultats</i>	La méta-analyse fut justifiée par le <i>Cochrane Library Handbook</i> dédié.
Résultats	<i>Études incluses</i>	Sept études furent incluses, regroupant 10 tests.
	<i>Synthèse de résultats</i>	Les <i>MP extension test</i> , <i>MC extension test</i> , <i>MC flexion test</i> et <i>Traction-shift test</i> furent simplement décrits. Six tests furent méta-analysés. Le <i>Grind test</i> présentait une sensibilité de 42% (IC95 31-53%) et une spécificité de 91% (IC95 84-95%) ; le <i>Palpation test</i> donnait une sensibilité de 90% (IC95 83-94%) et une spécificité de 83% (IC95 72-90%) ; le <i>Lever test</i> accordait une sensibilité de 90% (IC95 52-99%) et une spécificité de 85% (IC95 59-96%) ; l' <i>Adduction stress test</i> montrait une sensibilité de 76% (IC95 14-98%) et une spécificité de 89% (IC95 67-97%) ; l' <i>Extension stress test</i> recevait une sensibilité de 83% (IC95 37-98%) et une spécificité de 91% (IC95 75-97%) ; quant au <i>Seesaw test</i> , une sensibilité de 57% (IC95 28-81%) et une spécificité de 84% (IC95 74-91%).
Discussion	<i>Interprétation</i>	Le <i>cluster</i> théorique regroupe les <i>Grind test</i> , <i>Palpation test</i> , <i>Lever test</i> , <i>Extension stress test</i> et <i>Adduction stress test</i> .
Mots-clés : rhizarthrose ; tests cliniques ; méta-analyse statistique ; raisonnement clinique		

EFFECTIVENESS OF CLINICAL TESTS FOR RHIZARTHROSIS : A SYSTEMATIC REVIEW WITH AN INTRODUCTION TO STATISTICAL META-ANALYSIS		
Background	<i>Rationale</i>	Rhizarthrosis, a degenerative joint disease, is diagnosed according to convergent but non-harmonized clinical and para-clinical parameters.
	<i>Objectives</i>	After collecting scientific data, we wanted to propose a cluster of tests, facilitating physiotherapy triage.
Methods	<i>Eligibility criteria</i>	The articles had to deal with the clinical diagnosis of the pathology, including statistical data, and the tests precisely described. The intervention had to be carried out by a trained and qualified practitioner. The articles had to include unoperated patients, without age limit and written in French or English.
	<i>Information sources</i>	We explored the PubMed, Embase, LILACS and PEDro databases between July 23, 2025 and January 5, 2026.
	<i>Risk of bias</i>	The research was justified by QUADAS grids.
	<i>Synthesis methods</i>	The meta-analysis was justified by the dedicated Cochrane Library Handbook.
Results	<i>Study selection</i>	Seven studies were included, involving 10 tests.
	<i>Results of syntheses</i>	The MP extension test, MC extension test, MC flexion test and Traction-shift test were simply described. Six tests were meta-analyzed. The Grind test had a sensitivity of 42% (CI95 31-53%) and a specificity of 91% (CI95 84-95%) ; the Palpation test gave a sensitivity of 90% (CI95 83-94%) and a specificity of 83% (CI95 72-90%) ; the Lever test granted a sensitivity of 90% (CI95 52-99%) and a specificity of 85% (CI95 59-96%) ; the Adduction stress test presented a sensitivity of 76% (CI95 14-98%) and a specificity of 89% (CI95 67-97%) ; the Extension stress test showed a sensitivity of 83% (CI95 37-98%) and a specificity of 91% (CI95 75-97%) and the Seesaw test revealed a sensitivity of 57% (CI95 28-81%) and a specificity of 84% (CI95 74-91%).
Discussion	<i>Interpretation</i>	The theoretical cluster includes Grind test, Palpation test, Lever test, Extension stress test and Adduction stress test.
Keywords : rhizarthrosis ; clinical tests ; statistical meta-analysis ; clinical reasoning		

Table des matières

1. INTRODUCTION	1
1.1. Contexte.....	1
1.2. Environnement anatomique et biomécanique.....	1
1.2.1. La base de la colonne du pouce	1
1.2.1.1. Structures osseuses et articulaires	2
1.2.1.2. Structures capsulo-ligamentaires	2
1.2.1.3. Structures musculaires.....	2
1.2.1.4. Structures vasculaires.....	3
1.2.1.5. Structures nerveuses	3
1.3.2. Cinésiologie de l'articulation trapézo-métacarpienne.....	3
1.4. La rhizarthrose	5
1.4.1. Épidémiologie.....	5
1.4.2. Etio-physiopathologie	5
1.4.3. Symptômes, déformations cliniques et classification radiologique	6
1.5. Architecture de l'examen clinique kinésithérapique général	7
1.5.1. Anamnèse.....	7
1.5.2. Inspection	8
1.5.3. Palpation	8
1.5.4. Les bilans cliniques.....	8
1.5.5. Synthèse issue de l'examen	8
1.6. Place de l'examen clinique de la rhizarthrose	9
1.7. Lien avec les <i>clusters</i> de tests cliniques existants	9
1.8. Problématisation.....	10
1.8.1. Questions de recherche	11
1.8.2. Hypothèses.....	11
2. MATERIEL ET METHODES.....	12
2.1. Revue systématique de la littérature	12
2.1.1. Création de l'équation de recherche.....	12
2.1.2. Stratégie de recherche bibliographique.....	12
2.1.3. Sélection des articles.....	13
2.1.3.1. Suppression des doublons	13
2.1.3.2. Sélection selon les critères d'éligibilité.....	13
2.1.3.3. Lecture complète et réalisation de fiches de lecture.....	14
2.1.4. Extraction des données.....	14
2.1.5. Établissement de la qualité méthodologique, du risque de biais et de la préoccupation concernant l'applicabilité des articles inclus.....	15
2.2. Méta-analyse statistique	15
2.2.1. Extraction des données.....	16
2.2.2. Analyse des données	16
2.2.3. Évaluation de l'hétérogénéité.....	17
2.2.4. Estimation du biais de publication	17
3. RESULTATS.....	19
3.1. Revue systématique de la littérature	19
3.1.1. Diagramme de flux.....	19
3.1.2. Présentation des articles inclus	20
3.1.2.1. Méthodes des études.....	20

3.1.2.2.	Populations étudiées	21
3.1.2.3.	Tests comparés	22
3.1.2.4.	Tests de référence utilisés	23
3.1.2.5.	Flux et temporalité des études	23
3.1.2.6.	Tableau détaillé des tests cliniques	24
3.1.2.7.	Statistiques en tableaux 2x2.....	24
3.1.2.8.	Diagrammes détaillés des études	25
3.1.3.	Risque de biais et applicabilité.....	26
3.2.	Méta-analyse statistique	28
3.2.1.	Synthèse décisionnelle	28
3.2.2.	Forest plots.....	28
3.2.3.	SROC.....	31
3.2.4.	« Pseudo-ROC ».....	33
3.2.5.	Recherche de biais de publication	33
4.	<i>DISCUSSION</i>	37
4.1.	Forces et faiblesses des articles inclus	37
4.2.	Forces et faiblesses des tests analysés	37
4.3.	Intérêt et limites de la démarche méthodologique présentée	38
4.4.	Intérêt et limites des résultats pour la pratique et la profession.....	39
4.5.	Intérêt et limites des résultats par rapport à l'état des lieux de la littérature	40
4.6.	Réflexion autour des questions de recherche	40
5.	<i>LIMITES DE L'ETUDE</i>.....	45
6.	<i>PERSPECTIVES DE RECHERCHE</i>	46
7.	<i>CONCLUSION</i>	46
8.	<i>FINANCEMENT</i>	47
9.	<i>DECLARATION DE CONFLIT D'INTERET</i>.....	47
10.	<i>BIBLIOGRAPHIE</i>.....	48
ANNEXES		54

Tables des illustrations

Liste des figures

FIGURE 1 : CONFORMATION OSSEUSE DE L'ARTICULATION TRAPEZO-METACARPIENNE (SOURCE : BIOMECANIQUE SELON KAPANDJI).....	2
FIGURE 2 : LES AXES DE MOUVEMENT DU POUCE (SOURCE : BIOMECANIQUE SELON DUFOUR).....	3
FIGURE 3 : A. BILANS DE L'OPPOSITION ET DE LA CONTRE-OPPOSITION DU POUCE (SOURCE : BIOMECANIQUE SELON KAPANDJI) ET B. PRISES FONCTIONNELLES AU QUOTIDIEN DE LA MAIN ET DU POUCE (SOURCE : BIOMECANIQUE SELON DUFOUR).....	4
FIGURE 4 : FORCE DE PRESSION EXPERIMENTALE ENGENDREE SUR L'ARTICULATION TRAPEZO-METACARPIENNE (SOURCE : COONEY & CHAO, 1977)	5
FIGURE 5 : PROCESSUS DEGENERATIF GENERAL DE L'ARTHROSE (SOURCE : LIGUES-RHUMATISME.CH/RHUMATISMES-DE-A-A-Z/ARTHROSE)	6
FIGURE 6 : A. DEFORMATION VISIBLE A UN STADE PRECOCE (SOURCE : CHEVALIER, 2012) ET B. A UN STADE CLINIQUE AVANCE (SOURCE : COLONNA, 2024).....	7
FIGURE 7 : CLASSIFICATION RADIOLOGIQUE SELON EATON-LITTLER (SOURCE : TRADUIT DE COLONNA, 2024)	7
FIGURE 8 : MECANISME DE RAISONNEMENT KINESITHERAPIQUE (SOURCE : DESJARDINS, 2024).....	11
FIGURE 9 : DIAGRAMME DE FLUX DU TRAVAIL DE RECHERCHE SYNTHETISE SELON LA METHODE PRISMA.....	19
FIGURE 10 : RECAPITULATIFS DU RISQUE DE BIAIS.....	27
FIGURE 11 : META-ANALYSE STATISTIQUE DU GRIND TEST	29
FIGURE 12 : META-ANALYSE STATISTIQUE DU PALPATION TEST	29
FIGURE 13 : META-ANALYSE STATISTIQUE DU LEVER TEST	29
FIGURE 14 : META-ANALYSE STATISTIQUE DU SEESAW TEST	30
FIGURE 15 : META-ANALYSE STATISTIQUE DU ADDUCTION STRESS TEST.....	30
FIGURE 16 : META-ANALYSE STATISTIQUE DU EXTENSION STRESS TEST.....	30
FIGURE 17 : COURBES RECAPITULATIVES SROC DE LA META-ANALYSE STATISTIQUE.....	32
FIGURE 18 : COURBES "PSEUDO-ROC" DES TESTS UNIQUES	33
FIGURE 19 : RECHERCHE DE BIAIS DE PUBLICATION GLOBAL.....	34
FIGURE 20 : RECAPITULATIFS DE L'ANALYSE EN SOUS-GROUPE.....	35
FIGURE 21 : DEMARCHE DANS LE CADRE DE LA CLINIMETRIE RAISONNEE EN KINESITHERAPIE (SOURCE : PIETTE, 2021 CORRIGE).....	41
FIGURE 22 : NOMOGRAMME DE FAGAN PROJETANT UN EXEMPLE A PARTIR DE LA PREVALENCE ET DES RV+ ET RV- (SOURCE : CARAGUEL CGB, 2013).....	43

Liste des tableaux

TABEAU I : LA QUESTION DE RECHERCHE SOUS LE FORMAT PIT+R (ADAPTE DE PICO).....	12
TABEAU II : CRITERES D'ELIGIBILITE APPLIQUES A LA REVUE SYSTEMATIQUE.....	13
TABEAU III : FORMULES STATISTIQUES UTILES AU TRAVAIL DE MEMOIRE (SOURCE : COCHRANE LIBRARY).....	14
TABEAU IV : INFORMATIONS DE BASE SUR LA METHODE UTILISEE POUR LES ARTICLES	20
TABEAU V : DESCRIPTION DES POPULATIONS ETUDIEES.....	21
TABEAU VI : DESCRIPTION DES TESTS COMPARES.....	22
TABEAU VII : DESCRIPTION DES REFERENCES UTILISEES	23
TABEAU VIII : DESCRIPTION DU FLUX ET TIMING DES ETUDES.....	24
TABEAU IX : STATISTIQUES COMPLETES DES TESTS INDIVIDUELS COLLECTES DANS LES ETUDES..	25
TABEAU X : RECAPITULATIF DECISIONNEL DE LA META-ANALYSE	28
TABEAU XI : GRILLES QUADAS VALORISEES ET MOYENNEES	42
TABEAU XII : TABLEAU DES RAPPORTS DE VRAISEMBLANCE DES TESTS.....	43
TABEAU XIII : RECAPITULATIF GENERAL PAR TESTS.....	44

Liste des annexes

ANNEXE I : EQUATION DE RECHERCHE DANS LES BASES DE DONNEES BIBLIOGRAPHIQUES.....	54
ANNEXE II : LIGNE DIRECTRICE D'ECRITURE STARD 2015 EN FRANÇAIS	55
ANNEXE III : ECHELLE METHODOLOGIQUE QUADAS-C TRADUITE EN FRANÇAIS	56
ANNEXE IV : TABLEAU DETAILLE DES TESTS CLINIQUES IDENTIFIES.....	61
ANNEXE V : DIAGRAMMES DETAILLES DES ETUDES	64
ANNEXE VI : PACKAGES ET APPELS SUR R.....	71

Abréviations

AMSTAR : Assessing the Methodological quality of SysTemAtic Reviews
AUC : Area Under the Curve
CRAN : Comprehensive R Archive Network
DOR : Diagnostic Odd Ratio
DTA : Diagnostic Test Accuracy
EULAR : EUropean aLLiance of Associations for Rheumatology
FN : Faux Négatif
FP : Faux Positif
GRADE : Grading of Recommendations, Assessment, Development and Evaluations
HAS : Haute Autorité de Santé
HDLM : Histoire De La Maladie
HSROC : Hierarchical Summary Receiver Operating Characteristic
IA : Intra Articulaire
IC95 : Indice de Confiance à 95%
M : Malade
MC : MetaCarpal
MK : Masseur – Kinésithérapeute
MP : Metacarpo-Phalangien
NM : Non Malade
OR : Odd Ratio
PICO : Patient, Intervention, Comparaison, Outcome
PRISMA : Preferred Reporting Items for Systematic reviews and Meta-Analyses
QUADAS : QUality Assessment of Diagnostic Accuracy Studies
RV : Rapport de Vraisemblance
S : Sujet
SD : Standard Deviation
SDRC : Syndrome Douloureux Régional Complexe
STARD : STAndards for Reporting of Diagnostic accuracy studies
TMJ : Trapezo-Metacarpal Joint
VN : Vrai Négatif
VP : Vrai Positif
VPN : Valeur Prédictive Négative
VPP : Valeur Prédictive Positive

1. INTRODUCTION

1.1. Contexte

Les tests cliniques sont des outils mis à la disposition du kinésithérapeute qui permettent théoriquement de faciliter l'évaluation diagnostique et le suivi d'un patient. L'émergence dans la littérature internationale depuis une vingtaine d'années de groupes de tests, communément nommés *clusters*, autorisent à conforter encore davantage une hypothèse clinique grâce à un ajustement des paramètres statistiques comme la sensibilité, la spécificité, la valeur prédictive ou encore le rapport de vraisemblance. Ces démarches technologiques doivent bien sûr être réalisées en fonction de la symptomatologie du patient et de manière pertinente, avec comme but principal de proposer un traitement kinésithérapique optimal ou, le cas échéant, une réorientation médicale.

D'autre part, l'arthrose est l'affection rhumatologique la plus répandue chez l'être humain. Elle est caractérisée par une évolution inflammatoire à bas bruit, marquée par la possibilité de phases de poussées et touchant l'environnement articulaire dans sa globalité (1). Cette pathologie complexe, reliée à un certain nombre de facteurs de risques, ne peut se limiter uniquement à une usure articulaire de surcharge ou de mésusage. Pour preuve, la rhizarthrose, centrée autour de l'articulation trapézo-métacarpienne, touche essentiellement les femmes ménopausées après cinquante ans et de manière bilatérale (un lien étant cependant établi avec une gestuelle professionnelle répétitive et contraignante) (2). Il est par ailleurs aisé de comprendre que cette localisation à la base de la colonne du pouce engendre des incapacités et des restrictions de participation, et en fait un problème de Santé Publique important devant des notions de chronicité et d'évolutivité.

1.2. Environnement anatomique et biomécanique

1.2.1. La base de la colonne du pouce

L'emploi volontaire d'un terme générique en titre de cette partie vise à cerner la région anatomique qui constituera la base de la réflexion menée dans ce travail de Mémoire : nous nous situons au niveau de la loge thénar à la face antérieure de la main, et globalement en regard de la tabatière anatomique à sa face postérieure. Les éléments anatomiques présentés ci-après sont issus du tome 2 d'Anatomie de l'appareil locomoteur selon Dufour, référence d'apprentissage en la matière (3).

1.2.1.1. Structures osseuses et articulaires

Au niveau des structures osseuses en présence, nous retrouvons une continuité fonctionnelle proximo-distale débutant par le scaphoïde, se poursuivant par le trapèze puis continuant avec les os propres du pouce (1^{er} métacarpien puis 1^{ère} et 2^{ème} phalange du pouce). De manière spécifique, l'articulation trapézo-métacarpienne est une articulation en selle non congruente qui permet l'orientation du pouce dans les trois plans de l'espace. C'est donc cette articulation qui est à l'origine de la grande mobilité du pouce (**Fig. 1**).

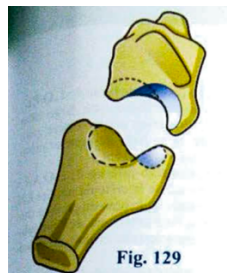


Figure 1 : conformation osseuse de l'articulation trapézo-métacarpienne (source : biomécanique selon Kapandji)

1.2.1.2. Structures capsulo-ligamentaires

La capsule trapézo-métacarpienne, qui est une partie individualisée de l'interligne carpo-métacarpienne, est une structure lâche qui permet selon Dufour les composantes rotatoires (4). Au niveau trapézo-métacarpien peuvent être cités les ligaments obliques antérieur, postéro-latéral et postéro-médial, qui assurent le renfort capsulaire. Par ailleurs, au niveau métacarpo-phalangien, les ligaments collatéraux - surtout le ligament oblique postéro-médial - assurent la stabilité du pouce dans les mouvements d'opposition comme lors des mouvements de pince pollici-digitale par exemple.

1.2.1.3. Structures musculaires

Nous pouvons dénombrer dix muscles participant à la mobilité du pouce, dont quatre extrinsèques et six intrinsèques. La loge thénar comporte l'ensemble des muscles intrinsèques : le court abducteur du I, l'opposant du I, le court fléchisseur du I et l'adducteur du I, auxquels nous pouvons ajouter les 1ers interosseux dorsaux et palmaires pour compléter l'énumération. Par ailleurs, avec des corps musculaires à distance, nous retrouvons dans cette zone de la base du pouce le long fléchisseur du I, le court et le long extenseur du I ainsi que le long abducteur du I par leurs tendons. Nous noterons également à ce niveau la présence de gaines, entourant et favorisant les coulissements tendineux normaux.

1.2.1.4. Structures vasculaires

La tabatière anatomique contient principalement l'artère radiale, qui y devient superficielle et donc palpable avant de plonger vers l'arcade palmaire profonde. Elle est encadrée par le tendon du long extenseur du I en médial et ceux du court extenseur et du long abducteur du I en latéral. Le réseau veineux possède les mêmes caractéristiques en arcades mais ne présente pas de particularité fonctionnelle ou anatomique notable dans le cadre de notre travail.

1.2.1.5. Structures nerveuses

Au niveau de la sensibilité cutanée, la base du pouce est innervée principalement par la racine C6. Sur le plan tronculaire, la sensibilité cutanée de cette région est assurée par le nerf radial à la face dorsale et par le nerf médian à la face palmaire. Sur le plan de la motricité, nous retrouvons une innervation par le nerf médian pour le long fléchisseur du I, le court abducteur du I, l'opposant du I et pour le faisceau superficiel du court fléchisseur du I. Le nerf ulnaire innerve le faisceau profond du court fléchisseur du I, l'adducteur du I et tous les interosseux de la main. Le nerf radial innerve le long abducteur du I, ainsi que les court et long extenseurs du I.

1.3.2. Cinésiologie de l'articulation trapézo-métacarpienne

Tout d'abord, il convient de dire que le pouce n'est pas dans l'axe des doigts longs : sa position de neutralité est définie avec environ 40° dans chacun des trois plans de l'espace orthonormé. Les trois degrés de liberté évoqués précédemment sont dus à la forme de selle de « cheval scoliotique » d'après Kapandji (5). Les mouvements présents sont donc : Flexion/Extension ($25/0/35$ selon Debrunner), Abduction/Adduction ($30/0/20$) et Pronation/Supination. Ce dernier mouvement, évalué à 90° dans son amplitude totale, est automatique : sont associées flexion et pronation, ainsi qu'extension et supination de manière simultanée (4) (**Fig. 2**).

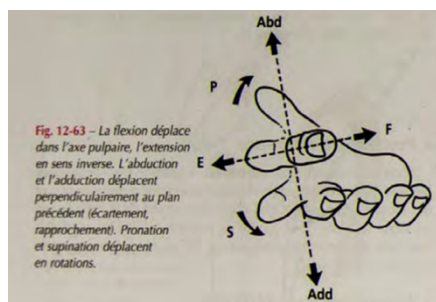


Figure 2 : les axes de mouvement du pouce (source : biomécanique selon Dufour)

D'après Dufour encore, la conformation des pièces osseuses mises en jeu au niveau articulaire propose un roulement-glissement en sens inverse dans le mouvement d'abduction/adduction, puisque la base du 1^{er} métacarpien est convexe et la cavité articulaire du trapèze en regard est concave. Lors du mouvement de flexion/extension, il s'agit d'un roulement-glissement dans le même sens (avec la justification anatomique opposée).

Autre concept fonctionnel, toujours selon Kapandji (5), le mouvement d'opposition correspond au mouvement de prise et la contre-opposition au lâchage, caractéristiques de notre fonction de bipèdes évolués au regard de la phylogénèse. La qualité de ces mouvements peut être objectivée par des bilans fonctionnels (**Fig. 3a**).

Pour le mouvement d'opposition, il existe une division conceptuelle avec la notion de petite course et celle de grande course. Cette différence est notamment objectivée dans le cadre d'une lésion du nerf médian, avec un mouvement d'adduction simple, en petite course, pour réaliser le mouvement dans le plan de la main et sans composante rotatoire.

En pratique, il existe de très nombreuses possibilités fonctionnelles d'utilisation du pouce en opposition selon les circonstances exigées (**Fig. 3b**). Une grande majorité des activités de la vie courante réalisées avec les mains requière en pratique l'utilisation de cette opposition. Ces possibilités demandent une variation du niveau de force et de précision engagé, et donc une amplitude importante des contraintes engendrées au niveau de l'articulation trapézo-métacarpienne.

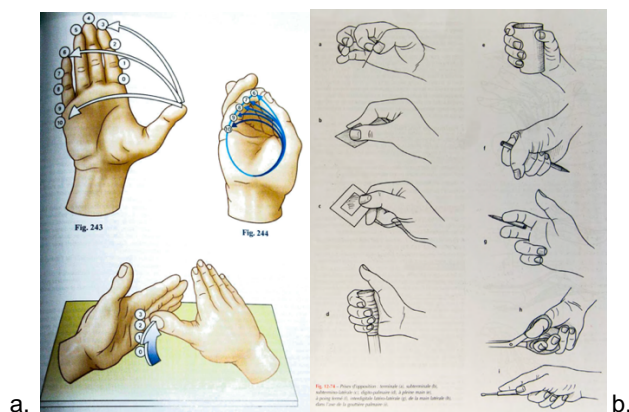


Figure 3 : a. bilans de l'opposition et de la contre-opposition du pouce (source : biomécanique selon Kapandji) et b. prises fonctionnelles au quotidien de la main et du pouce (source : biomécanique selon Dufour)

Nous pouvons citer, parmi les travaux historiques qui ont permis de mettre en avant les contraintes biomécaniques en compression de l'articulation trapézo-métacarpienne, ceux de Cooney et Chao en 1977 (6) (**Fig. 4**). Ces chercheurs ont établi qu'une pression de 12kg sur l'articulation trapézo-métacarpienne était appliquée à partir d'un mouvement de serrage de 1kg en prise subtermino-latérale par le jeu des forces musculaires et des bras de levier. Cette notion alimente donc l'usure potentielle au long cours de l'articulation trapézo-métacarpienne, dans le cadre d'un surmenage articulaire professionnel par exemple.

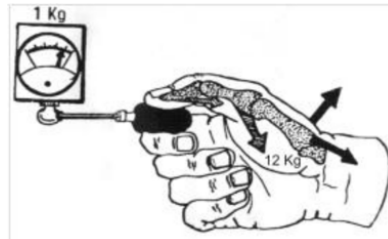


Figure 4 : force de pression expérimentale engendrée sur l'articulation trapézo-métacarpienne (source : Cooney & Chao, 1977)

De plus, d'autres travaux réalisés par Boutan ont mis en évidence l'importance d'un couple musculaire 1^{er} Interosseux dorsal – Opposant optimal dans l'équilibre de l'articulation trapézo-métacarpienne, ce duo permettant une stabilisation active de l'articulation dans tous les mouvements et tous les plans (7). Nous en déduisons donc qu'un déséquilibre musculaire peut aussi être à l'origine ou à la conclusion d'un défaut de cinématique articulaire trapézo-métacarpienne, et potentiellement générer une usure articulaire anormale.

1.4. La rhizarthrose

1.4.1. Épidémiologie

Comme énoncé dans le contexte de ce Mémoire, la pathologie touche davantage les femmes que les hommes (80% vs. 20% en moyenne), l'atteinte peut être bilatérale et évolue en général après cinquante ans et l'apparition de la ménopause (8). Il est difficile de quantifier de manière exacte l'incidence de la pathologie sur la population générale, et ce pour plusieurs raisons : peu de statistiques disponibles, et beaucoup de patients ne consultant pas malgré déformation, douleurs ou impotence. La rhizarthrose est donc traditionnellement une pathologie sous-diagnostiquée, ou bien diagnostiquée tardivement et/ou fortuitement. Dans tous les cas, des études historiques, comme celle réalisée par Cushnagan et al. sur une population arthrosique, montrent que la main (30%) fait partie du trio de tête des arthroses avec celles de la hanche (19%) et du genou (41,2%) (9). Sur une population âgée de soixante-dix ans et plus, lors de l'étude épidémiologique Framingham Study réalisée sur 1041 sujets incluant 36% d'hommes et 64% de femmes, la prévalence d'une arthrose de main symptomatique, et donc douloureuse, est de 26,2% chez les femmes et 13,4% chez les hommes (10).

1.4.2. Etio-physiopathologie

La rhizarthrose - du grec *rhiza* : racine, sous-entendu de la main - est un processus dégénératif (**Fig. 5**) inflammatoire à bas bruit du cartilage de l'articulation trapézo-

métacarpienne aboutissant à sa destruction progressive. Les signes généraux de l'arthrose sont marqués par une dégénérescence du cartilage articulaire, une sclérose de l'os sous-chondral, la formation d'ostéophytes, et parfois une synovite réactionnelle (11). Pour rappel, dans le contexte biomécanique, sa forme en selle la soumet également à des contraintes en cisaillement théoriquement délétères (12).

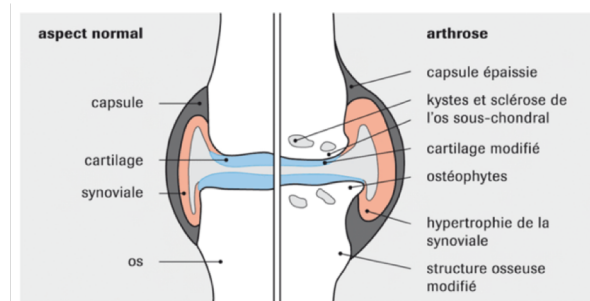


Figure 5 : processus dégénératif général de l'arthrose (source : liques-rhumatisme.ch/rhumatismes-de-a-a-z/arthrose)

Au niveau étiologique, les facteurs de risque principaux décrits par l'EULAR sont donc le sexe féminin, la senescence, une hyperlaxité ligamentaire, une gestuelle professionnelle contraignante ou répétitive, une fracture mal traitée (notamment de Bennett ou de Rolando), une obésité, des antécédents familiaux, ou encore un déséquilibre musculaire (13). Des sources spécialisées concordantes notent qu'une atteinte régionale peut aussi être évoquée avec l'arthrose de type scapho-trapézo-trapézoïdien (STT), bien qu'elle soit rare et touche surtout la femme âgée sur un terrain de chondrocalcinose (8,14).

1.4.3. Symptômes, déformations cliniques et classification radiologique

D'après Athlani, des douleurs mécaniques apparaissent tout d'abord sur le mouvement type de pince pollici-digitale, en opposition de pouce (15). Cliniquement, nous retrouvons donc une perte progressive de la capacité d'extension et d'abduction de la colonne du pouce, base de ce mouvement fonctionnel d'opposition. Il existe un mécanisme progressif de subluxation de l'articulation trapézo-métacarpienne, visible en fonction de l'évolution anatomique de la déformation (**Fig. 6a**). Nous pouvons retrouver une déformation en « Z » ou *pouce adductus* à un stade avancé, comprenant un enraidissement de l'articulation trapézo-métacarpienne, une diminution du premier espace inter-métacarpien (avec une subluxation de M1 qui se place en adduction par rapport au plan de la main), une hyper extension métacarpo-phalangienne et une flexion inter-phalangienne (**Fig. 6b**). Cette déformation est donc associée à une amyotrophie globale de l'éminence thénar, et une néo-pince pollici-digitale par compensation fonctionnelle est utilisée par le patient. En règle générale, la déformation en « Z » apparaît alors que les douleurs disparaissent (8).



Figure 6 : a. déformation visible à un stade précoce (source : Chevalier, 2012) et b. à un stade clinique avancé (source : Colonna, 2024)

Au niveau radiologique, nous citerons la classification Eaton-Littler permettant d'établir quatre stades de la rhizarthrose (16)(**Fig. 7**). Il est à noter qu'il n'y a pas de lien entre le stade radiologique et le stade clinique selon Batra, la radiologie permettant simplement de définir un cadre conceptuel de prise en charge médicale en fonction du niveau de plainte éventuel du patient (17). Il est utile de préciser que la littérature décrit déjà des douleurs chez les femmes jeunes, de moins de cinquante ans, aux deux premiers stades de classification (16).

Stade I : stade de la synovite		Stade II : stade de la laxité capsulaire	
• Les contours articulaires sont normaux • Élargissement possible de l'articulation trapézo-métacarpienne suggérant un épanchement articulaire ou une laxité ligamentaire • Aucune formation d'ostéophytes		• Rétrécissement de l'espace articulaire trapézo-métacarpien • Petite formation d'ostéophytes sur la face ulnaire de la surface articulaire trapéziennne distale (- de 2mm de diamètre) • Absence de subluxation articulaire ou 1/3 de l'ATM	
Stade III : destruction articulaire importante		Stade IV : arthrose pantrapéziennne	
• Rétrécissement supplémentaire de l'espace articulaire avec modifications kystiques et ossification sclérotique • Ostéophytes proéminents au niveau du bord ulnaire du trapèze distal (+ de 2mm de diamètre) • Subluxation modérée dans le sens radial et dorsal à la base du premier métacarpien • Arthrose légère de l'articulation scapho-trapéziennne		• Subluxation importante de l'articulation • Rétrécissement de l'espace articulaire comme au stade III • Modifications kystiques et sclérotiques de l'os sous-chondral • Erosion et destruction importantes de l'articulation scapho-trapéziennne	

Figure 7 : classification radiologique selon Eaton-Littler (source : traduit de Colonna, 2024)

1.5. Architecture de l'examen clinique kinésithérapique général

Le modèle est largement décrit dans la littérature kinésithérapique. Pour la région de la main, celui-ci est hérité de l'ouvrage de démarche clinique selon Dufour (18).

1.5.1. Anamnèse

Cet interrogatoire est fondamental : il permet de balayer l'historique personnel, familial, la profession ou encore les habitudes de vie du patient. L'histoire de la maladie, avec une

notion de recherche de traumatisme ou d'apparition des symptômes est également importante. Un bilan de la douleur exhaustif doit être mené aussi à cette étape pour faire le tri entre les douleurs mécaniques et inflammatoires, mais aussi pour s'assurer de pouvoir réaliser un bilan complet sereinement.

1.5.2. Inspection

Cette étape consiste en une recherche de déformations visibles, de troubles vasomoteurs, d'évaluation cutanée et des phanères essentiellement.

1.5.3. Palpation

Nous cherchons à évaluer avec nos capteurs digitaux la nature des tissus évalués, mais aussi une éventuelle douleur rapportée à la palpation : nociception, allodynie, hyperesthésie ou encore anesthésie peuvent être recherchées. Ces éléments seront à prendre en compte pour mener à bien notre diagnostic kinésithérapique.

1.5.4. Les bilans cliniques

Selon la recherche à mener sur le patient, différents bilans peuvent être engagés :

- neurologique : réflexes ostéo-articulaires, sensibilité cutanée, évaluation de la force musculaire, neurodynamique
- orthopédique : mobilisation passive/active (permet une évaluation des amplitudes articulaires et des capacités musculaires en rapport)
- fonctionnel : tests spécifiques, tests de provocation (reproduction ou diminution des symptômes), recherche de ressaut, craquement...

1.5.5. Synthèse issue de l'examen

A cette étape, le clinicien peut réaliser une formulation des hypothèses diagnostiques. Un point important à signaler est que le kinésithérapeute ne pose pas de diagnostic médical, il réalise un raisonnement dans le cadre du bilan diagnostic kinésithérapique. Deux embranchements sont possibles ici en finalité, avec soit l'établissement d'un plan de traitement kinésithérapique adapté, soit une réorientation médicale car découverte de drapeaux rouges.

1.6. Place de l'examen clinique de la rhizarthrose

Dans le cadre de nos recherches préalables, un grand nombre d'articles dans la littérature scientifique supporte l'idée que pour mener à bien un diagnostic solide issu des hypothèses de prise en charge, il convient à tout clinicien de réaliser un examen rigoureux à la recherche de preuves en s'aidant d'un arsenal clinique fiable, ainsi que de chercher et d'éliminer les diagnostics différentiels pouvant dégrader l'interprétation par des tests spécifiques et/ou des examens para-cliniques. Ces besoins cliniques sont équivalents en médecine générale (19), en chirurgie spécialisée (20) ou en kinésithérapie (18).

Le modèle "*gold standard*" communément admis, bien que perfectible, de la prise en charge de la rhizarthrose regroupe trois éléments majeurs utilisés ensemble :

- l'anamnèse : incluant l'histoire de la maladie, la prise en compte des facteurs de risque et la plainte rapportant en général une notion de douleur chronique, à tendance mécanique, localisée, sans traumatisme ;
- l'examen clinique spécifique : non harmonisé et non standard, opérateur dépendant et pouvant inclure des tests retrouvés dans de nombreux articles comme le *Grind test* (qui correspond à une mise en compression de l'articulation trapézo-métacarpienne accompagnée d'une rotation axiale à la recherche de reproduction du symptôme douloureux) (15,21–23) ;
- les examens para-cliniques : la radiographie étant notamment un étalon pour l'état des lieux anatomique et l'élimination de certains diagnostics différentiels, avec une nuance concernant les spécificités vues précédemment rendant la technique seule non pertinente (voir partie 1.4.3.).

L'adage commun « *treat patients, not x-rays* », repris par Steven Glickel, un spécialiste international du sujet de la rhizarthrose, prend donc son sens ici (24).

En se plaçant donc de notre point de vue de professionnels non médicaux, une anamnèse bien menée associée à la présence de tests cliniques spécifiques efficaces pour la détection de la maladie pourraient donc largement faire converger la supposition clinique. L'examen clinique spécifique est donc un pan entier de la prise en charge accessible à notre niveau de compétences, et son exploration pourrait être améliorée et uniformisée. Cet état de fait correspond aussi à une demande des cliniciens, avec des besoins urgents de classification d'une pathologie peu traitée en recherche scientifique (12).

1.7. Lien avec les *clusters* de tests cliniques existants

La sensibilité évalue l'aptitude d'un test clinique à identifier correctement un individu malade, tandis que la spécificité évalue l'aptitude de ce même test à identifier correctement

un individu sain (25). Ces deux paramètres représentent la qualité propre d'un test clinique, et de nombreux cas de figure rendent parfois l'utilisation d'un test isolé non concluant (26). Nous détaillons l'exemple du *cluster* de Wainner publié en 2003 pour expliciter notre propos (27). Cette étude prospective a été réalisée sous une forme aveugle chez une population suspectée de troubles neurologiques cervicaux, et a permis de déterminer les tests spécifiques qui reproduisaient ou diminuaient les symptômes parmi une batterie complète de tests cliniques. Conjointement à cela, un questionnaire était fourni pour collecter les symptômes et l'historique du patient. Un examen para-clinique de type électromyogramme était aussi réalisé pour mettre en évidence une lésion neurologique périphérique. Des analyses statistiques ont permis d'objectiver les résultats et d'identifier un *cluster* robuste de tests cliniques capable de confirmer l'hypothèse d'une radiculopathie cervicale dans un contexte clinique évocateur, alors que l'utilisation isolée de ces tests ne permettait qu'une objectivation modérée du diagnostic recherché.

1.8. Problématisation

Des revues de la littérature et des études de la portée récentes ont été rédigées concernant l'examen clinique de la rhizarthrose, et ces articles correspondent à un état des lieux large mais peu spécifique des pratiques dans le domaine (21,22). Ayant étudié des *clusters* diagnostics dans le cadre de notre formation, nous nous sommes demandé si le cheminement scientifique aboutissant à la création de ces groupes de tests pouvait être reproduit avec la rhizarthrose. C'est donc ici que nous souhaitons positionner notre travail de recherche à travers différents objectifs, en collectant tout d'abord les comparaisons de tests cliniques présents dans la littérature et en les analysant, puis en proposant de regrouper les plus pertinents statistiquement pour répondre à la question. Nous poursuivrons en tentant de statuer sur l'intérêt de constituer un *cluster* lié à cette pathologie chronique dans un contexte grandissant de triage et d'accès direct aux soins, et donc de l'autonomisation d'un kinésithérapeute « augmenté » devant la réorganisation médicale grandissante (28) (**Fig. 8**). Un autre objectif opérationnel de ce travail sera d'évaluer l'apport conjoint du raisonnement et des outils cliniques au contact d'une population dont le vieillissement, et donc les pathologies dégénératives non urgentes mais à fort pouvoir limitant, augmente.

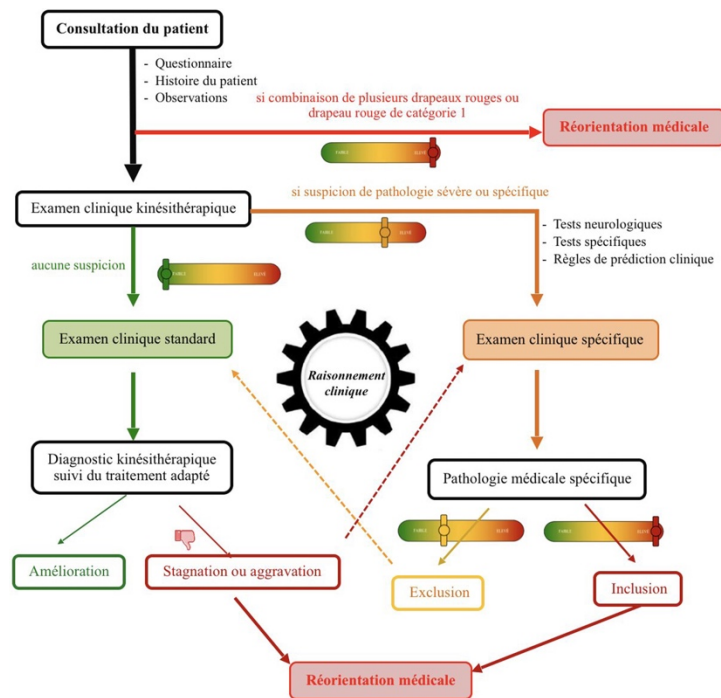


Figure 8 : mécanisme de raisonnement kinésithérapique (source : Desjardins, 2024)

1.8.1. Questions de recherche

En se rapportant à la problématisation et aux objectifs opérationnels qui découlent de notre réflexion, il convient donc de poser la question de recherche principale liée à ce travail. **Quels tests cliniques existants sont statistiquement fiables pour détecter une rhizarthrose au sein d'une population à risque ?**

Des sous-questions alimentent naturellement la réflexion, en fonction de l'orientation que nous souhaitons donner au contenu. La constitution d'un *cluster* de confirmation statistiquement significatif est-elle possible lorsqu'un doute clinique apparaît ? Les kinésithérapeutes, à travers l'exemple de cet outil, pourront-ils prendre part, dans un avenir proche, à des actes de prévention dans le cadre de l'augmentation des compétences lié au triage des patients ?

1.8.2. Hypothèses

Hypothèse 1 : plusieurs tests cliniques, bien décrits et reproductibles, sont statistiquement efficaces pour mettre en évidence une rhizarthrose.

Hypothèse 2 : le *cluster* de tests proposé est statistiquement significatif et permet de conforter un doute clinique.

Hypothèse 3 : l'outil envisagé correspond à un besoin clinique grandissant de la part des kinésithérapeutes.

2. MATERIEL ET METHODES

Un guide méthodologique de la littérature scientifique coréenne issu de nos recherches préalables, ainsi que celui édité par la *Cochrane Library*, nous ont permis de pouvoir faire reposer cette revue spécifique associée à une méta-analyse statistique comparant l'efficacité de tests diagnostiques sur des fondations solides (29–31).

2.1. Revue systématique de la littérature

Cette revue systématique de la littérature a été rédigée en s'appuyant sur la liste de contrôle PRISMA 2020, dernière version actualisée et possédant une traduction validée en Français (disponible à l'adresse : <https://www.prisma-statement.org/translations>).

2.1.1. Création de l'équation de recherche

L'équation de recherche a été formulée à partir des items du modèle PICO (Patient/ Intervention/ Comparaison/ Outcome), en s'adaptant la spécificité statistique de notre sujet avec la formulation PIT (Population/ Index test/ Target condition)(31). A ce modèle a été ajouté un item de Référence test pour inclure des comparaisons statistiques dans la démarche (**Tab. I**). Les synonymes trouvés pour chaque item ont été connectés par l'opérateur booléen « OR », et les différents items ont été connectés par l'opérateur booléen « AND ».

Tableau I : la question de recherche sous le format PIT+R (adapté de PICO)

Item de la question	Mots clés associés
Population (Patients)	thumb OR basal joint OR trapeziometacarpal joint OR carpometacarpal joint
Index test (Intervention)	test OR assessment OR evaluation OR examination
Référence test (Comparateur)	sensitivity OR specificity OR diagnosis OR accuracy
Target condition (Outcome)	osteoarthritis OR arthritis OR rhizarthrosis

Étant donné la variabilité propre de chacune de bases de données interrogées, les différentes versions de l'équation ont été mises à disposition du lecteur (**Annexe I**).

2.1.2. Stratégie de recherche bibliographique

Comme conseillé par l'article de l'équipe de travail coréenne, l'équation de recherche a été soumise aux trois bases de données citées (PubMed, Embase et LILACS), ainsi que la base de données PEDro (30). La recherche s'est déroulée du 23/07/2025 au 02/10/2025, et une veille documentaire a été mise en place jusqu'au 05/01/2026 en cas d'article répondant aux critères définis. Les résultats ont été exportés sous format .ris vers les logiciels Rayyan, Zotero et Apple Numbers.

2.1.3. Sélection des articles

2.1.3.1. Suppression des doublons

Le logiciel Rayyan a été utilisé pour une recherche de doublons automatisée, puis une confrontation manuelle a été réalisée sur la base des informations fondamentales des articles (titre de l'article, date de publication et auteurs). Ce travail a pu être mis en œuvre par l'intermédiaire de la liste transférée entre les logiciels Apple Numbers et Zotero, dans le but d'éliminer chaque doublon potentiel.

2.1.3.2. Sélection selon les critères d'éligibilité

Après élimination des doublons, la liste obtenue a été soumise à lecture des titres et résumés afin d'exclure ou d'inclure les articles dans la revue systématique, selon les critères établis (**Tab. II**).

Tableau II : critères d'éligibilité appliqués à la revue systématique

	Critères d'inclusion	Critères d'exclusion
Population (Patients)	Patients non opérés au niveau de la région étudiée ; Hommes et femmes adultes sans distinction ; Pas de limite d'âge supérieur	Population ayant subi une chirurgie de la région concernée ; Recherches sur enfants et adolescents, animaux ou cadavériques
Index test (Intervention)	Tests cliniques décrits précisément (avec texte et dessins/photos) ; Tout intervenant formé et habilité aux tests cliniques (kinésithérapeute, médecins, pratiques avancées)	Tests non décrits de manière précise ; Tout autre support d'évaluation que les tests cliniques (questionnaire par exemple) ; Tests réalisés par des non soignants
Référence test (Compareur)	Articles avec données statistiques comparatives chiffrées	Articles sans collecte de données statistiques exploitables ; Retrait des articles uniquement narratifs
Target condition (Outcome)	Articles traitant du diagnostic clinique et paraclinique de la pathologie	Articles incluant une intervention thérapeutique pouvant fausser le résultat des tests ; Exclusion des pathologies inflammatoires
Période de temps	Toutes dates	X
Type de publication	Toute étude comparative susceptible d'apporter des informations utiles au sujet, soumise à comité de lecture	Littérature grise (thèse, mémoire) ou article non accessible ou étude de cas ou rapport de congrès
Langage	Langue française ou anglaise	Toute autre langue empêchant la compréhension technique de l'article

Une échelle de surlignage en couleurs a été définie sur le logiciel Apple Numbers pour réaliser un tri efficient (couleur rouge pour un article exclu à la lecture du titre ; couleur jaune pour un article douteux dont le résumé a dû être lu avant décision ; couleur verte pour un article correspondant aux critères d'inclusion initiaux).

2.1.3.3. Lecture complète et réalisation de fiches de lecture

Les articles sélectionnés en liste secondaire (les articles douteux finalement conservés après lecture du résumé et ceux correspondant aux critères d'inclusion initiaux) ont été lus dans leur intégralité. Les articles finalement exclus l'ont été s'ils ne répondaient pas à tous les critères d'éligibilité fixés. Des fiches ont été réalisées en se basant sur la méthodologie d'écriture pour les études portant sur la précision des tests diagnostiques STARD 2015 (*Standards for Reporting of Diagnostic Accuracy Studies*), traduite en français québécois (32)(**Annexe II**), en se confrontant à la version de 2003, traduite par M. Gedda en 2014 (33).

2.1.4. Extraction des données

Les données ont été traitées sous forme de tableaux regroupant les catégories pertinentes pour l'analyse : la méthode des études (type d'étude, pays de l'étude, mode de recrutement, conflits d'intérêt, financement, passage en commission d'éthique), la population étudiée (nombre de patients, âge moyen, répartition des groupes, but de la consultation, lieu de l'étude), les tests comparés (nom de la technique, instructions détaillées, photographie/image, classification claire, évaluateur, synchronisation entre les différents tests, randomisation, mise en aveugle, mode d'évaluation), la référence utilisée (composition, *gold standard* optimal, classification claire, mise en aveugle, commentaires), le flux et temporalité de l'étude (données manquantes, patients perdus de vue, synchronisation entre les différents tests et la référence, les patients ont tous reçu les tests et la référence) et les statistiques en tableau 2x2 avec vrais positifs (VP), faux positifs (FP), vrais négatifs (VN) et faux négatifs (FN) auquel nous ajouterons les rapports de vraisemblance positifs (RV+) et négatifs (RV-), ainsi que l'*odd ratio* (OR). Devant l'hétérogénéité des articles en terme d'écriture statistique, des formules mathématiques furent utilisées pour uniformiser l'écriture des tableaux (**Tab. III**).

Tableau III : formules statistiques utiles au travail de Mémoire (source : Cochrane Library)

Informations générales d'un tableau 2x2				
		Maladie (issu du Test référence)		
		OUI	NON	
Test index comparé	Positif	VP	FP	Total Tests + (T+)
	Négatif	FN	VN	Total Tests - (T-)
		Malades M (M = S - NM) (M = VP + FN)	Non malades NM (NM = S - M) (NM = FP + VN)	Sujets S (S = M + NM)
Sensibilité (Sn) = VP / M = VP / (VP+FN) Spécificité (Sp) = VN / NM = VN / (FP+VN)				
Calculs des VP, VN, FP, FN, RV+, RV- et OR				
VP = Sn x M = M - FN VN = Sp x NM = NM - FP FP = NM - VN FN = M - VP RV+ = Sn / (1-Sp) = (VP / [VP + FN]) / (FP / [FP + VN]) RV- = (1-Sn) / Sp = (FN / [VP + FN]) / (VN / [FP + VN]) OR = RV+ / RV- = (Sn x Sp) / (1-Sn) x (1-Sp) = (VP x VN) / (FP x FN)				

2.1.5. Établissement de la qualité méthodologique, du risque de biais et de la préoccupation concernant l'applicabilité des articles inclus

La qualité méthodologique des articles sélectionnés pour l'étude a été réalisée à l'aide des outils QUADAS-2 et QUADAS-C (*Quality Assessment of Diagnostic Accuracy Studies*). QUADAS-C (C pour *Comparative*) a été traduit de l'Anglais au Français avec l'application DeepL à partir du modèle site web Open Science Framework du groupe de travail QUADAS-C (disponible sur le site internet suivant : <https://osf.io/hq8mf/files/>), puis confronté à la version française de QUADAS-2 traduite par Rostagno et Pallot, et corrigé par nos soins si besoin (34). L'outil QUADAS-C a été utilisé pour les études comportant plusieurs tests à comparer à la référence, tandis que QUADAS-2 était prévu lorsque le test comparé à la référence est unique. Il est également à noter qu'une section QUADAS-2 préalable est intégrée dans l'outil QUADAS-C, ce dernier pouvant donc être apparenté à une extension de QUADAS-2.

Pour faciliter la compréhension de l'outil, un modèle est présenté en **Annexe III**. Les grilles permettent une évaluation de la qualité méthodologique de l'article dans quatre domaines (sélection des patients, test(s) évalué(s), test de référence, déroulement et temporalité) en proposant une recherche spécifique de biais dans chacun des domaines, et une préoccupation concernant l'applicabilité sur la sélection des patients, le(s) test(s) évalué(s) et la référence. Les résultats sont présentés sous forme graphique avec des codes couleur, permettant une visualisation simplifiée et rapide de l'analyse.

2.2. Méta-analyse statistique

La méta-analyse a suivi les mêmes principes méthodologiques que la revue systématique qualitative, soutenue par les deux références bibliographiques précédemment citées (29–31).

Le logiciel utilisé pour les calculs statistiques a été le logiciel R, une application de codage *open source* gratuite, nécessaire pour mener à bien les calculs et les graphiques récapitulatifs souhaités. Le *pack* RStudio a été installé afin de faciliter l'utilisation de l'interface, et le choix des outils statistiques a été effectué grâce au *Handbook* de la *Cochrane Library*, en fonction de notre niveau de compétences et des résultats souhaités. La ressource *Comprehensive R Archive Network* (CRAN), par l'intermédiaire d'un « *Task View* » dédié aux méta-analyses (disponible à l'adresse : <https://cran.r-project.org/web/views/MetaAnalysis.html>), a fourni l'ensemble des *packages* nécessaires à la réalisation de la méta-analyse statistique de type DTA sur R : les *packages* mada (disponible à l'adresse : <https://cran.r-project.org/web/packages/mada/index.html>) pour les SROC, DTA *plots* (disponible à l'adresse : <https://cran.r-project.org/web/packages/DTAplots/index.html>) pour les *forest plots*

et metafor (disponible à l'adresse : <https://cran.r-project.org/web/packages/metafor/index.html>) pour les biais de publication globaux.

2.2.1. Extraction des données

Les articles de la revue systématique qualitative ont été inclus dans la méta-analyse statistique s'ils répondaient aux critères d'inclusion désirés, à savoir la présence de données chiffrées exploitables permettant à minima d'étudier la sensibilité, la spécificité et les effectifs correspondants. Cette deuxième extraction des données statistiques a été réalisée en format 2x2 sous forme de tableaux informatiques, dans les mêmes conditions et à l'aveugle de l'extraction des données de la revue puis confrontée, afin de tendre vers une analyse statistique optimale et non erronée. Les différents tests ont été ordonnés en sous-groupes.

2.2.2. Analyse des données

Si au moins deux études portaient sur un même test

La méta-analyse a été réalisée avec les données chiffrées extraites, test par test et en deux temps. Pour la synthèse statistique, une pondération par effet aléatoire a tout d'abord été choisie pour pouvoir harmoniser les populations étudiées. Comme recommandé par les statisticiens, un modèle bivarié a été utilisé afin de pouvoir analyser les sensibilités et les spécificités ensemble, et réaliser une synthèse chiffrée en prenant en compte la corrélation des deux paramètres statistiques (35). Les résultats de cette partie de la méta-analyse ont donc été présentés selon une visualisation en *forest plots*, chaque ligne rapportant l'étude dans laquelle le test est utilisé, en séparant la sensibilité et la spécificité. La partie centrale de chaque *forest plot* représente la statistique moyenne, et les traits horizontaux de part et d'autre représentent l'indice de confiance à 95%. La statistique récapitulative pondérée est présente à la partie basse du graphique, et sa largeur représente également l'indice de confiance calculé.

Pour les courbes récapitulatives, le modèle bivarié a également été choisi devant la probable hétérogénéité intra et inter-études, mais aussi pour prendre en compte la corrélation entre la sensibilité et la spécificité de chaque test. Des courbes SROC ont ainsi été générées pour chaque test méta-analysé sur logiciel informatique avancé, la courbe résumant le compromis entre sensibilité et 1-spécificité à travers les études concernées (36). Les deux axes du graphique correspondent donc à la 1-spécificité et à la sensibilité, et chaque triangle présent sur le graphique correspond à une étude qui a évalué le test diagnostique. Plus la courbe récapitulative est proche du coin en haut à gauche, plus le test est performant (haute sensibilité et haute 1-spécificité). Le point présent sur la courbe est une estimation pondérée

de la performance du test, et l'ellipse autour représente l'indice de confiance à 95% pondéré entre les études.

Il est possible aussi d'en tirer une aire sous la courbe (AUC) : plus elle est grande, plus le test est discriminant pour classer correctement un patient malade dans sa pathologie (37).

Un calcul de l'hétérogénéité statistique inter-études du modèle bivarié été également réalisé, en utilisant la méthode I^2 ajustée selon Holling qui tient compte de la corrélation entre sensibilité et spécificité (38). Le chiffre est en relation avec la variance graphique de la surface du SROC représentée par l'indice de confiance à 95%.

Si un test n'était présenté que sur une seule étude

Une synthèse descriptive a été réalisée devant l'impossibilité de réaliser la méta-analyse. Les données ont été ajoutées dans le tableau comparatif global de la revue, et des courbes visuelles « pseudo-ROC » ont été générées pour donner une vision d'ensemble.

2.2.3. Évaluation de l'hétérogénéité

L'hétérogénéité inter-études d'effet global étant un paramètre à prendre en compte dans une méta-analyse statistique, elle a été évaluée grâce à l'outil statistique I^2 : un I^2 inférieur à 25% est représentatif d'une hétérogénéité faible, un I^2 compris entre 25 et 50% d'une hétérogénéité modérée, et un I^2 supérieur à 50% d'une hétérogénéité importante.

L'effet seuil n'étant pas considéré par le test I^2 , une évaluation visuelle des *forest plots* a aussi été réalisée à la recherche de cette indication.

Pour compléter l'étude de l'hétérogénéité, une analyse en sous-groupes a été réalisée afin d'isoler chacun des tests collecté dans cette revue.

2.2.4. Estimation du biais de publication

Conjointement à l'exploration de la qualité méthodologique et du risque de biais de chaque article, le risque de biais de publication se doit d'être exploré dans le cadre d'une méta-analyse. Ce dernier apparaît lorsque les études aux résultats significatifs ont davantage de chances d'être publiées que celles rapportant des résultats négatifs ou non significatifs. Cela peut conduire à une surestimation de l'effet global, les études non publiées ou moins accessibles (souvent de petite taille ou présentant des résultats modestes) étant sous-représentées.

Pour évaluer ce biais, un *funnel plot*, appelé diagramme en entonnoir en Français, a été généré *via* R. Ce graphique a permis d'illustrer la répartition des études selon leur importance en abscisse (avec la taille effective des échantillons dans notre cas), et leur résultat

en ordonnée (l'*odd ratio* moyen par étude ici) : à chaque point correspond une étude, et en l'absence de biais, les points se répartissent symétriquement autour de l'effet global estimé. Une asymétrie dans le *funnel plot* peut suggérer un biais de publication, traduisant un manque d'études négatives ou non significatives.

Cette asymétrie a été vérifiée statistiquement grâce au test de Deeks (39). Une ligne graphique statistiquement significative associée à une répartition symétrique des études indique un *funnel plot* normal et, par conséquent, l'absence de biais de publication.

3. RESULTATS

3.1. Revue systématique de la littérature

3.1.1. Diagramme de flux

Notre recherche bibliographique a permis d'identifier 436 articles au total (**Fig. 9**).

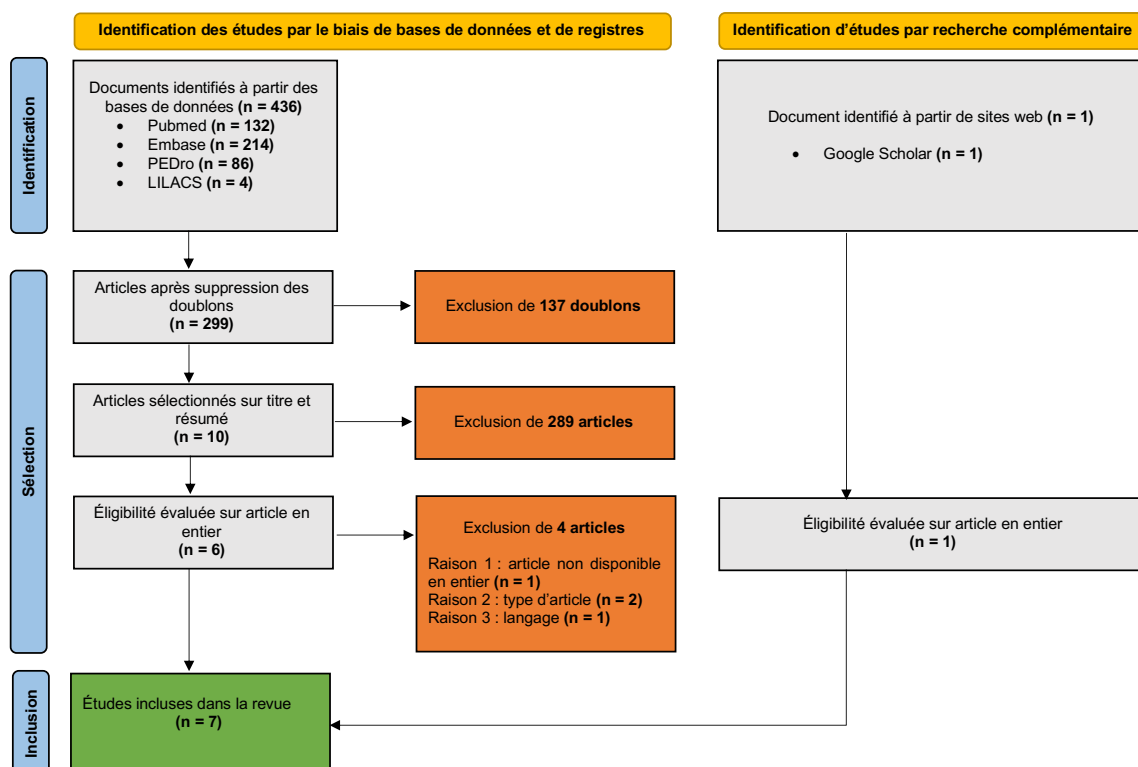


Figure 9 : Diagramme de flux du travail de recherche synthétisé selon la méthode PRISMA

La première étape de sélection, réalisée conjointement avec le logiciel automatique Rayyan et une vérification manuelle, a permis de supprimer 137 doublons et a donc laissé 299 articles disponibles pour la suite. L'étape suivante de lecture sur titres et résumés, réalisée en fonction des critères d'éligibilité, a été scindée en deux parties : une première partie avec un tri sur titres pour retirer les articles non concernés, et une deuxième partie sur lecture des résumés. Cette étape a permis d'écarter 289 articles supplémentaires de notre revue, laissant dix articles pour étude approfondie. Ces articles, répondant aux critères de sélection établis, ont été lus intégralement pour vérifier leur cohérence dans le cadre de notre recherche. A cette étape, quatre articles ont été exclus pour les raisons suivantes : un article n'était pas disponible en entier (malgré l'aide de la bibliothèque universitaire), deux articles étaient des revues systématiques sans statistiques exploitables et un article était rédigé en langue espagnole. Six articles ont donc été sélectionnés dans le cadre de notre démarche rigoureuse de

recherche, à laquelle nous avons ajouté manuellement un article issu de recherches complémentaires.

3.1.2. Présentation des articles inclus

Les sept articles constituant la revue systématique suivent tous une méthodologie de type étude clinique, et ont été rédigés entre 2010 et 2022 (**Tab. IV**). Les auteurs correspondants identifiés dans les articles sont par ordre alphabétique : Denise MJ Arnold (chirurgienne chercheuse), Robert M Choa (chirurgien de la main), Richard H Gelberman (chirurgien de la main), Melissa M Merritt (kinésithérapeute chercheuse), Zina Model (chirurgienne orthopédique), Samuel Ribak (chirurgien orthopédique) et Yaron Sela (chirurgien orthopédique) (40–46). Une grande majorité des études sont rédigées sous drapeau Américain, tandis qu'une source est Européenne et une autre est Sud-Américaine (**Tab. IV**). Il est à noter une tendance générale dans cette revue de niche : les auteurs les plus récents citent en général les auteurs les plus anciens pour alimenter la discussion et comparer les résultats des tests cliniques présentés.

3.1.2.1. Méthodes des études

Les études sélectionnées sont toutes prospectives, c'est-à-dire que les patients sont sélectionnés avant de rentrer dans le protocole, sauf celle de Sela,2019 qui est rétrospective et pour laquelle les patients ont été reconvoqués pour étude secondaire en étant choisis à partir d'une base de données (**Tab. IV**). Parmi les six études prospectives listées, celle de Choa,2014 est spécifiquement annoncée en cas-témoin, c'est-à-dire que les patients du groupe contrôle sont des individus non rhizarthrosiques avérés (**Tab. IV**). Le point commun de toutes les études est un recrutement de convenance basé sur l'accessibilité immédiate des patients (**Tab. IV**). Seul l'article de Merritt,2010 a bénéficié d'un soutien financier, et tous les auteurs ont réalisé les tests dans le cadre de la routine clinique excluant des pratiques discutables sur le plan de l'éthique médicale (**Tab. IV**).

Tableau IV : informations de base sur la méthode utilisée pour les articles

Article	Pays de l'étude	Type d'étude	Mode de recrutement	Conflits d'intérêt	Financement	Commission d'éthique
Arnold,2022	U.S.A.	prospectif	convenance	non	non	approuvé
Choa,2014	Angleterre	prospectif en cas-témoin	convenance	non	non	pas besoin (routine clinique)
Gelberman,2015	U.S.A.	prospectif	convenance	non	non	approuvé
Merritt,2010	U.S.A.	prospectif	convenance	non précisé	oui (American Society of Hand Therapy)	approuvé
Model,2016	U.S.A.	prospectif	convenance	non	non	approuvé
Ribak,2020	Brésil	prospectif	inconnu (convenance ?)	non	Non précisé	approuvé
Sela,2019	U.S.A.	rétrospectif	convenance	non	non	approuvé

3.1.2.2. Populations étudiées

Les articles de cette revue présentent une hétérogénéité relative des effectifs, puisqu'ils varient de 60 à 129 pouces inclus et étudiés (pour Choa,2014 et Gelberman,2015 respectivement)(**Tab. V**). Les lieux de consultation sont tous institutionnels, ils garantissent donc une approche rigoureuse et non-foraine des tests réalisés (**Tab. V**). Le but de la consultation pour toutes les études est une souffrance de la main ou du poignet nécessitant au minimum de réaliser des radiographies, incluant notamment un éventail allant de la suspicion à la pathologie avérée de rhizarthrose (**Tab. V**). C'est le paramètre radiologique qui sert de référence pour l'attribution aux groupes des participants, scindant les pouces inclus en deux à trois groupes selon les études (Gelberman,2015 a séparé en deux sous-groupes les pouces non rhizarthrosiques dans un objectif secondaire de discrimination aux diagnostics différentiels). Arnold,2022 inclue les pouces classés Eaton I dans le groupe « contrôle », tandis que dans les autres études détaillant la classification, ce même groupe est inclus dans le groupe « cas » (**Tab. V**). En détaillant la composition des groupes « cas » en fonction de la classification de Eaton, nous pouvons signaler, lorsque la statistique est rapportée, une faible proportion de Eaton I et Eaton IV (sauf pour l'étude de Merritt,2010) (**Tab. V**). L'âge moyen des participants s'échelonne de 42 ans dans le groupe « douleurs autres » de Gelberman,2015 à 73,1 ans dans le groupe « cas » d'Arnold,2022 (**Tab. V**). Certaines études proposent une répartition extrême des âges importante (peu en rapport avec l'histoire naturelle de la rhizarthrose, notamment chez Merritt,2010 et Sela,2019 avec 23 et 22 ans pour les plus jeunes participants respectivement), tout comme l'âge moyen et l'écart-type de certaines études à relativiser dans le contexte (groupe « douleur autre » de Gelberman,2015 à 42 +/- 17 ans) (**Tab. V**). La répartition par sexe des patients inclus, lorsqu'elle est annoncée, est équivalente chez Arnold,2022, avec une petite majorité de femmes chez Choa,2014, et une grande majorité de femmes chez Merritt,2010, Model,2016 et Sela,2019. Une fois la filtration par groupes réalisée et les données statistiques analysées, la répartition démographique des groupes « cas » reproduit le profil épidémiologique classique de la maladie débutante, caractérisé par une prédominance féminine et un âge supérieur à 60 ans (**Tab. V**).

Tableau V : description des populations étudiées

Article	Nombre de pouces inclus +/- répartition selon le sexe (H/F)	Répartition des groupes +/- répartition selon le sexe (H/F)	Age moyen (pouces inclus ou selon les groupes)	But de la consultation	Lieu de consultation
Arnold, 2022	Recrutés : n=80 Inclus : n=68 34H et 34F	Groupe cas (Eaton II-IV) : n=24 (8H et 16F) Groupe contrôle (0-I) : n=44 (26H et 18F)	Groupe cas : 73,1 (SD=8,9) Groupe contrôle : 66,1 (10,8)	Suspicion de rhizarthrose ou de pathologie autre	Dans un hôpital
Choa, 2014	Recrutés : n=60 25H et 35F	Groupe cas : n=30 (5H et 25F) Groupe contrôle : n=30 (20H et 10F)	Groupe cas : 66 (46-82) Groupe contrôle : 50 (36-56)	Groupe cas : rhizarthrose identifiée et traitée Groupe contrôle : asymptomatiques mais nécessitant des radiographies	Dans une clinique de la main
Gelberman, 2015	Éligibles : n=131 Recrutés : n=129	Groupe rhizarthrose : n=48 (12H et 36F) (Eaton I : n=0 ; II : n=7 ; III : n=26 ; IV : n=5) Groupe douleur radiale : n=44 (19H et 25F) Groupe douleur autre : n=47 (22H et 25F)	Groupe rhizarthrose : 62 +/-6 Groupe douleur radiale : 52 +/-16 Groupe douleur autre : 42 +/-17	Plainte traumatique de la main et du poignet	Dans deux établissements tertiaires de santé
Merritt, 2010	Recrutés : n=54 (16 bilatéral)	Pouces cas : n=55 Eaton I : n=6 ; II : n=15 ; III : n=10 ; IV : n=24	59,5 +/-13,4 (23-91)	Souffrance du pouce ou du poignet	Dans trois cliniques orthopédiques, un

	70 pouces étudiés 8H et 46F	Pouces témoins : n=15			centre de médecine du travail et la population générale
Model, 2016	Recrutés : n=62 (39 bilatéral et 23 unilatéral) Mains éligibles : n=124 mains incluses : n=121 21H et 41F	Mains symptomatiques : n=100 Mains asymptomatiques : n=21	62 +/-8 (45-86)	Douleur à la base du pouce	A priori dans un hôpital
Ribak, 2020	Recrutés : n=68	Groupe cas : n=34 (4H et 30F) Eaton I : n=1 ; II : n=13 ; III : n=19 ; IV : n=1 Groupe contrôle : n=34 (répartition non fournie)	Groupe cas : 60,6 (36-80) Groupe contrôle : âge non fourni	Groupe cas : patients avec symptomatologie typique et rhizarthrose confirmée Groupe contrôle : patients avec symptomatologie non typique de rhizarthrose	Dans un hôpital
Sela, 2019	Collectés : n=127 Éligibles : n=104 (23 bilatéral et 81 unilatéral) Pouces étudiés : n=127 33H et 71F	Groupe contrôle : n=22 Groupe cas : n=105 Sous-groupe Eaton I : n=0 Sous-groupe Eaton II : n=27 Sous-groupe Eaton III : n=56 Sous-groupe Eaton IV : n=22	59 (22-91)	Douleurs au bord radial de la main ou du poignet	Dans un hôpital

3.1.2.3. Tests comparés

Un ensemble de 22 tests sont répertoriés et bien décrits à travers les sept études, le grind test (ou test de broyage) se retrouvant notamment cité dans chaque étude car il fait office de test historique de la rhizarthrose (**Tab.VI**)(voir 3.2.1. **Synthèse décisionnelle**). L'étude de Merritt,2010 présente uniquement le grind test, tandis que d'autres études comme celle de Ribak,2020 présentent jusqu'à cinq tests différents. Seuls Ribak,2020 et Merritt,2010 donnent un temps de repos articulaire entre chaque test évalué, les autres auteurs enchaînant les évaluations (**Tab.VI**). En tant que test diagnostique, le but de la mobilisation est de reproduire le symptôme douloureux habituel, ce qui offre l'avantage d'une classification claire avec un seuil unique pour l'ensemble des études. Il est à noter qu'à notre échelle, il n'est pas possible de savoir si une rémanence de la douleur pourrait perturber la série de tests. Le nombre d'évaluateurs intervenant dans les études, lorsqu'il est explicitement connu, oscille entre un et quatre, ce qui peut décemment soulever un problème de reproductibilité et d'expertise inter-examineurs (**Tab.VI**). Pour préciser l'inter-opérabilité lorsqu'elle est pratiquée, Arnold,2022, Gelberman,2015 et Merritt,2010 décrivent un aveuglement réciproque, tandis que Sela,2019 et Model,2016 proposent un évaluateur unique pour chaque patient (**Tab.VI**). L'ordre des tests pour les études à tests multiples, lorsqu'il est explicitement annoncé méthodologiquement, est randomisé chez Arnold,2022 et Model,2016, et suit un ordre déterminé avec Choa,2014 et Sela,2019. L'opérateur n'a pas connaissance de la référence radiologique chez Arnold,2022, Gelberman,2015 et Merritt,2010, les autres auteurs n'ont en revanche pas explicitement précisé ce point de détail méthodologique (**Tab.VI**).

Tableau VI : description des tests comparés

Article	Nom des techniques	Instructions détaillées et imagées	Classification claire	Évaluateur(s)	Synchronisation entre les tests	Mise en aveugle par rapport à la référence	Randomisation des tests	Mode d'évaluation
Arnold, 2022	Grind test Sesaw test	oui	oui (reproduction de la douleur)	2 (1 chirurgien + 1 parmi 5 fellows)	tests enchaînés	oui	oui	Évaluateurs aveugles l'un de l'autre
Choa, 2014	Grind test Traction-shift test	oui	oui (reproduction de la douleur)	2 (2 chirurgiens)	tests enchaînés selon un ordre déterminé	non précisé	non	Nombre d'évaluateurs indéterminé pour chaque patient
Gelberman, 2015	Extension stress test Adduction stress test Joint tenderness Grind test	partiellement (grind test et palpation non détaillés)	oui (reproduction de la douleur)	2 (1 chirurgien + 1 examinateur indépendant)	tests enchaînés	oui	non précisé	Évaluateurs aveugles l'un de l'autre
Merritt, 2010	Grind test	oui	oui (reproduction de la douleur)	2 +/- 3 MK (en cas de discordance)	5 minutes de repos entre chaque test	oui	NA	Évaluateurs aveugles l'un de l'autre
Model, 2016	Grind test Lever test MP extension test	partiellement (palpation non détaillée)	oui (reproduction de la douleur)	4 (chirurgiens de la main)	Tests enchaînés	non précisé	oui	Évaluateur unique pour chaque patient

Palpation test								
Ribak, 2020	Grind test Digit pressure test Shear test MC1 flexion test MC1 extension test	oui	oui (reproduction de la douleur)	Nombre inconnu	2 minutes de repos entre chaque test	non précisé	non précisé	Nombre d'évaluateurs indéterminé pour chaque patient
Sela, 2019	Grind test MC pressure-shear test MC flexion test MC extension test	oui	oui (reproduction de la douleur)	1 (chirurgien senior)	Tests enchaînés selon un ordre déterminé	non clair	non	Évaluateur unique pour chaque patient

3.1.2.4. Tests de référence utilisés

Comme annoncé dans la partie introductive, la référence standard varie pour objectiver et suivre la pathologie de la rhizarthrose (voir **1.6. Place de l'examen clinique de la rhizarthrose**). En partant du postulat que le triptyque histoire de la maladie (HDLM) + radiographies + examen clinique fait office de « *Gold standard* », Gelberman,2015 et Ribak,2020 utilisent l'outil de classement le plus complet, tandis que Arnold,2022, Merritt,2010 et Sela,2019 ne prennent en compte que la radiographie dans leur description (**Tab.VII**). Dans tous les cas, l'utilisation commune de la radiographie comme référence anatomique rend la classification claire. Une mise en aveugle de la réalisation du test de référence par rapport aux tests cliniques est décrite par Arnold,2022, Gelberman,2015, Merritt,2010 et Sela,2019, tandis que les autres auteurs ne précisent pas ce détail méthodologique (**Tab.VII**). Quelques spécificités sont à signaler : Arnold,2022 et Choa,2014 regroupent les stades Eaton II à IV en tant que rhizarthrosiques, mais le premier réalise une relecture à un mois pour vérifier tandis que le second réalise un geste invasif de contrôle avec une infiltration-test. Chez Merritt,2010, l'ordre entre la pratique des tests cliniques et celle du test référence n'est pas explicitement connu (**Tab.VII**).

Tableau VII : description des références utilisées

Article	Composition	Gold Standard optimal	Classification claire	Mise en aveugle par rapport aux tests	Commentaires
Arnold, 2022	Radiographie	-	oui	oui	Eaton II à IV pour positivité ; un mois d'attente avant relecture
Choa, 2014	Radiographie +/- injection test IA de confirmation	+/-	oui	non précisé	Eaton II à IV pour positivité ; geste invasif de vérification
Gelberman, 2015	HDLM + radiographies + examen clinique	+	oui	oui	Diagnostic final fait par l'évaluateur indépendant
Merritt, 2010	Radiographie	-	oui	oui	Eaton I à IV pour positivité
Model, 2016	HDLM + signes radiographiques	+/-	oui	Non précisé	Ordre non connu entre tests et référence
Ribak, 2020	HDLM + radiographies + QuickDASH	+	oui	non précisé	Regroupement de Eaton I avec II, et de III avec IV
Sela, 2019	Radiographie	-	oui	oui	2 lecteurs indépendants pour les radiographies ; Eaton I-IV pour positivité

3.1.2.5. Flux et temporalité des études

Les études statistiques transversales collectées dans cette revue offrent l'avantage pratique de ne pas avoir de patients considérés comme perdus de vue (**Tab.VIII**). La seule variation préjudiciable à noter concernant la synchronisation entre les tests réalisés et le test référence est l'absence de délai annoncé chez Choa,2014, alors que la technique du

test de référence est thérapeutique et son effet antalgique est normalement prolongé dans le temps (revoir **Tab.VII**). Une ambiguïté méthodologique demeure chez Gelberman,2015 et Ribak,2020 puisque le nombre de pouces testés varie par rapport au nombre de patients inclus, proposant donc un flux de données imprécis (**Tab.VIII**). Au sein des différentes études, les patients inclus ont tous reçu les tests cliniques et le test de référence (**Tab.VIII**).

Tableau VIII : description du flux et timing des études

Article	Données manquantes	Patients perdus de vue	Synchronisation entre les tests et la référence	Les patients inclus ont tous reçu les tests et la référence
Arnold, 2022	oui, non pris en compte	non	Dans le même temps à la visite initiale	oui
Choa, 2014	non	non	Délai inconnu	oui
Gelberman, 2015	oui (méthode imprécise)	non	Dans le même temps	oui
Merritt, 2010	non	non	Dans le même temps	oui
Model, 2016	non	non	Dans le même temps	oui
Ribak, 2020	oui (méthode imprécise)	non	Non précisé	oui
Sela, 2019	oui, non pris en compte	non	Non précisé	oui

3.1.2.6. *Tableau détaillé des tests cliniques*

Les tests cliniques détaillés, analysés et éventuellement regroupés sont visibles en **Annexe IV**. Ces tableaux permettent la compréhension pratique des tests, par l'intermédiaire d'illustrations et des descriptions issues des articles. Ils garantissent également la reproductibilité éventuelle, et permettent de visualiser les tests apparaissant plusieurs fois à travers les différentes études, et qui seront donc inclus dans la méta-analyse.

3.1.2.7. *Statistiques en tableaux 2x2*

Ces tableaux sont un travail préparatoire essentiel pour effectuer la méta-analyse statistique. Ils permettent de regrouper les informations de départ comme les tableaux en 2x2 individuels, mais aussi de garder un œil critique en vérifiant les chiffres proposés par les méthodes de calcul dans les différents logiciels (**Tab. IX**). Les statistiques ont été menées en double correction : les calculs ont été faits à la main grâce aux formules listées dans le **tableau III**, puis confrontés aux applications epiR sur R (package disponible à l'adresse : <https://cran.r-project.org/web/packages/epiR/index.html>) pour la sensibilité et la spécificité, accompagnées de leur indice de confiance à 95% (IC95). Les calculs ont été réalisés selon la méthode de Wilson, pertinente et reproductible pour la suite de l'analyse statistique. Le site Medcalc a été utilisé pour les calculs de RV+ et RV- (disponible à l'adresse : https://www.medcalc.org/en/calc/diagnostic_test.php) et pour les odd ratio (disponible à l'adresse : https://www.medcalc.org/en/calc/odds_ratio.php), ces derniers n'étant pas faisables avec la méthode de Wilson. Il est à noter que la description détaillée des statistiques

n'est pas réalisée à cette étape mais le sera à l'étape de la méta-analyse (partie 3.2. des résultats) pour éviter la redondance.

Tableau IX : statistiques complètes des tests individuels collectés dans les études

Article	Test Index	Sensibilité (%)	Spécificité (%)	RV+	RV-	Odd Ratio	Résultat Test Index	Résultat Référence standard	Total
Arnold, 2022	Grind test (chirurgien)	12	91	1,37	0,96	1,43 IC95 [0,29 ; 6,99] Z = 0,440 p = 0,6597	Test positif	Rhizarthrose	4
		IC95 [4 ; 30]	IC95 [79 ; 99]	IC95 [0,34 ; 5,64]	IC95 [0,81 ; 1,19]		Test négatif	Pas rhizarthrose	7
									61
									68
	Grind test (hand fellow)	17	98	7,33	0,85	IC95 [0,96 ; 81,97] Z = 3,871 p = 0,0004	Test positif	Rhizarthrose	5
		IC95 [7 ; 30]	IC95 [89 ; 100]	IC95 [0,87 ; 61,97]	IC95 [0,71 ; 1,03]		Test négatif	Pas rhizarthrose	63
									68
	Senseaw test (chirurgien)	71	82	3,90	0,36	IC95 [1,40 ; 35,10] Z = 4,017 p = 0,0001	Test positif	Rhizarthrose	25
		IC95 [61 ; 89]	IC95 [68 ; 91]	IC95 [1,98 ; 7,67]	IC95 [0,19 ; 0,69]		Test négatif	Pas rhizarthrose	43
									68
	Senseaw test (hand fellow)	42	96	3,06	0,88	IC95 [1,38 ; 14,77] Z = 2,5 p = 0,0124	Test positif	Rhizarthrose	16
		IC95 [29 ; 63]	IC95 [73 ; 99]	IC95 [1,27 ; 7,30]	IC95 [0,47 ; 0,97]		Test négatif	Pas rhizarthrose	52
									68
Chen, 2014	Grind test	30	97	9	0,72	IC95 [1,38 ; 105,74] Z = 2,307 p = 0,0211	Test positif	Rhizarthrose	10
		IC95 [17 ; 48]	IC95 [84 ; 100]	IC95 [1,21 ; 66,71]	IC95 [0,57 ; 0,93]		Test négatif	Pas rhizarthrose	50
									60
	Traction-shift test	67	100	•	0,33	• (119,09) IC95 [0,61 ; 2146,74] Z = 3,240 p = 0,0012	Test positif	Rhizarthrose	20
		IC95 [49 ; 81]	IC95 [99 ; 100]		IC95 [0,28 ; 0,55]		Test négatif	Pas rhizarthrose	40
									60
Guberman, 2015	Grind test	44	92	5,69	0,61	9,33 IC95 [0,59 ; 24,39] Z = 4,565 p < 0,0001	Test positif	Rhizarthrose	28
		IC95 [31 ; 58]	IC95 [85 ; 97]	IC95 [2,61 ; 12,43]	IC95 [0,47 ; 0,78]		Test négatif	Pas rhizarthrose	111
									139
	Joint tenderness test	94	81	5,02	0,08	IC95 [18,12 ; 236,33] Z = 6,388 p < 0,0001	Test positif	Rhizarthrose	62
		IC95 [83 ; 99]	IC95 [72 ; 93]	IC95 [3,25 ; 7,73]	IC95 [0,03 ; 0,23]		Test négatif	Pas rhizarthrose	77
									139
	Extension stress test	94	95	17,06	0,07	250 IC95 [88,96 ; 1128,98] Z = 7,373 p < 0,0001	Test positif	Rhizarthrose	50
		IC95 [83 ; 99]	IC95 [88 ; 99]	IC95 [7,25 ; 40,13]	IC95 [0,02 ; 0,20]		Test négatif	Pas rhizarthrose	89
									139
	Adduction stress test	94	93	14,32	0,07	212,5 IC95 [82,74 ; 589,98] Z = 7,334 p < 0,0001	Test positif	Rhizarthrose	51
		IC95 [83 ; 99]	IC95 [86 ; 97]	IC95 [8,54 ; 30,93]	IC95 [0,02 ; 0,20]		Test négatif	Pas rhizarthrose	88
									139
Marritt, 2010	Grind test (évaluateur 1)	42	92	6,27	0,62	10,06 IC95 [1,23 ; 82,04] Z = 2,197 p = 0,0310	Test positif	Rhizarthrose	24
		IC95 [30 ; 58]	IC95 [79 ; 99]	IC95 [0,92 ; 42,74]	IC95 [0,48 ; 0,87]		Test négatif	Pas rhizarthrose	46
									70
	Grind test (évaluateur 2)	53	80	2,64	0,59	4,46 IC95 [1,13 ; 17,58] Z = 2,137 p = 0,0325	Test positif	Rhizarthrose	32
		IC95 [40 ; 64]	IC95 [65 ; 93]	IC95 [0,93 ; 7,40]	IC95 [0,41 ; 0,80]		Test négatif	Pas rhizarthrose	38
									70
Model, 2016	Grind test	41	100	•	0,59	• (29,99) IC95 [0,74 ; 999,13] Z = 2,354 p = 0,0196	Test positif	Rhizarthrose	41
		IC95 [32 ; 51]	IC95 [95 ; 100]		IC95 [0,50 ; 0,66]		Test négatif	Pas rhizarthrose	29
									121
	Lever test	82	81	4,30	0,22	19,36 IC95 [0,81 ; 64,49] Z = 4,820 p = 0,0001	Test positif	Rhizarthrose	86
		IC95 [73 ; 89]	IC95 [69 ; 93]	IC95 [1,77 ; 10,49]	IC95 [0,14 ; 0,39]		Test négatif	Pas rhizarthrose	35
									121
	MP extension test	65	95	13,65	0,37	37,34 IC95 [4,78 ; 286,52] Z = 3,456 p = 0,0005	Test positif	Rhizarthrose	65
		IC95 [55 ; 74]	IC95 [87 ; 99]	IC95 [5,92 ; 92,94]	IC95 [0,28 ; 0,49]		Test négatif	Pas rhizarthrose	55
									121
	Palpation test	91	76	3,82	0,12	5,89 IC95 [0,99 ; 109,10] Z = 2,606 p < 0,0001	Test positif	Rhizarthrose	96
		IC95 [84 ; 99]	IC95 [65 ; 89]	IC95 [1,77 ; 8,23]	IC95 [0,08 ; 0,23]		Test négatif	Pas rhizarthrose	25
									121
Ritak, 2020	Grind test	60	76	2,80	0,53	4,71 IC95 [1,48 ; 15,06] Z = 2,618 p = 0,0088	Test positif	Rhizarthrose	33
		IC95 [46 ; 73]	IC95 [65 ; 89]	IC95 [1,12 ; 5,57]	IC95 [0,35 ; 0,87]		Test négatif	Pas rhizarthrose	25
									68
	Digit pressure test	85	95	17,87	0,16	114,28 IC95 [13,14 ; 993,93] Z = 4,294 p < 0,0001	Test positif	Rhizarthrose	41
		IC95 [72 ; 93]	IC95 [87 ; 99]	IC95 [3,63 ; 121,48]	IC95 [0,08 ; 0,31]		Test négatif	Pas rhizarthrose	27
									68
	Shear test	70	71	2,46	0,42	5,89 IC95 [1,89 ; 18,33] Z = 3,064 p = 0,0022	Test positif	Rhizarthrose	39
		IC95 [56 ; 81]	IC95 [59 ; 89]	IC95 [1,22 ; 4,90]	IC95 [0,25 ; 0,70]		Test négatif	Pas rhizarthrose	29
									68
	MC1 flexion test	43	81	2,23	0,71	3,15 IC95 [0,93 ; 10,80] Z = 1,823 p = 0,0683	Test positif	Rhizarthrose	24
		IC95 [30 ; 57]	IC95 [69 ; 93]	IC95 [0,87 ; 5,73]	IC95 [0,51 ; 0,98]		Test négatif	Pas rhizarthrose	44
									68
	MC1 extension test	64	86	4,47	0,42	10,59 IC95 [2,72 ; 41,23] Z = 3,402 p = 0,0007	Test positif	Rhizarthrose	33
		IC95 [50 ; 78]	IC95 [76 ; 95]	IC95 [1,53 ; 13,03]	IC95 [0,28 ; 0,64]		Test négatif	Pas rhizarthrose	35
									68
Seba, 2019	Grind test	64	100	•	0,36	• (73,96) IC95 [4,65 ; 1337,34] Z = 3,025 p = 0,0025	Test positif	Rhizarthrose	67
		IC95 [54 ; 73]	IC95 [95 ; 100]		IC95 [0,28 ; 0,47]		Test négatif	Pas rhizarthrose	22
									127
	MC pressure-shear test	99	95	21,79	0,01	2184 IC95 [1,51 ; 495,06] Z = 2,247 p = 0,0246	Test positif	Rhizarthrose	99
		IC95 [85 ; 100]	IC95 [87 ; 99]	IC95 [3,21 ; 147,89]	IC95 [0,01 ; 0,07]		Test négatif	Pas rhizarthrose	28
									127
	MC flexion test	36	100	•	0,64	• (25,67) IC95 [1,51 ; 495,06] Z = 2,247 p = 0,0246	Test positif	Rhizarthrose	38
		IC95 [27 ; 46]	IC95 [95 ; 100]		IC95 [0,55 ; 0,74]		Test négatif	Pas rhizarthrose	22
									127
	MC extension test	46	100	•	0,54	• (37,96) IC95 [2,34 ; 642,17] Z = 2,520 p = 0,0117	Test positif	Rhizarthrose	48
		IC95 [37 ; 58]	IC95 [95 ; 100]		IC95 [0,46 ; 0,60]		Test négatif	Pas rhizarthrose	79
									127

* en accord avec la méthodologie de la *Cochrane Library* (<https://training.cochrane.org/handbook/>), une cellule du tableau 2x2 éventuellement nulle n'autorisant pas le calcul de l'odd ratio a été modifiée à 0,5 pour réaliser son approximation. Pour calculs de Z et p : (47)

3.1.2.8. Diagrammes détaillés des études

Les diagrammes de flux des études, représentant la démarche complète en synthèse graphique, sont disponibles en **Annexe V**. Ces diagrammes essentiels représentent le

recrutement des patients dans les études incluses dans notre revue systématique portant sur l'efficacité des tests diagnostiques étudiés (31).

3.1.3. Risque de biais et applicabilité

L'outil QUADAS-C fait office de fondement en recherche scientifique pour évaluer les risques de biais et l'applicabilité clinique lors d'une comparaison d'efficacité de plusieurs tests diagnostiques par rapport à une norme de référence. Il accompagne et complète l'outil QUADAS-2, qui porte quant à lui son attention sur la présentation d'un seul test diagnostique par rapport à cette même norme de référence (le cas de figure se produit une seule fois dans notre travail puisque l'étude de Merritt, 2010 compare uniquement l'efficacité du *Grind test* par rapport à la norme de référence)(43). L'évaluation par QUADAS-C comporte donc une sous-partie QUADAS-2. En raison de sa complexité, cet outil nécessite une bonne maîtrise de son contenu et de sa structure. Il est ainsi recommandé de s'appuyer sur le chapitre dédié du guide méthodologique de la *Cochrane Library* pour saisir pleinement la démarche d'évaluation (48).

Pour l'étude des risques de biais, faite sur test unique et dans une démarche de comparaison de tests, quatre domaines sont évalués (**Fig. 10**) :

- 1) **Dans la sélection des patients** : il s'agit du domaine qui présente le plus grand risque puisque trois études ont un mode de sélection générateur de biais statistique (étude cas-témoins maquillée pour Model, 2016 ou vraie pour Choa, 2014, ou étude rétrospective pour Sela, 2019). De plus, la prévalence est supérieure à la norme de la population dans ces études, et biaise probablement l'ensemble des résultats.
- 2) **Pour le test index (ou test évalué)** : ce domaine est relativement respecté, bien qu'hétérogène dans les méthodes appliquées. Arnold, 2022, Gelberman, 2015 et Merritt, 2010 proposent une méthode claire sur les évaluateurs. Les quatre autres auteurs ont une méthode non claire quant aux biais potentiels lorsque plusieurs tests sont réalisés, avec un grand nombre d'évaluateurs différents ou un évaluateur unique sans confrontation. Des éclaircissements devraient être apportés sur l'aveuglement des évaluateurs par rapport à la norme de référence lorsque les tests comparés sont réalisés. Pour les tests en eux-mêmes, l'effet d'ordre des tests, ainsi que le délai entre chaque test lorsque l'étude porte sur plusieurs tests enchaînés, ne sont pas harmonisés entre les différentes études, et doivent donc pousser à la réflexion générale.
- 3) **Pour la norme de référence** : ce domaine est dans l'ensemble bien respecté. Le risque de biais est mineur dans quatre études, avec une mise en aveugle des résultats des tests cliniques au moment du test de référence, évitant ainsi un biais

d'incorporation. Les trois autres études n'ont pas mentionné explicitement leur méthodologie.

- 4) **Dans la gestion du flux et de la temporalité** : le niveau de risques est hétérogène. Chez Choa,2014, le manque d'informations entre les tests effectués et le traitement médical réalisé semble problématique, et le contrôle sur ces derniers dans le temps n'est pas mentionné par l'auteur. Pour Gelberman,2015 et Ribak,2020, un manque d'informations est à signaler puisque des patients apparaissent dans la comptabilité statistique alors qu'ils ne sont pas inscrits dans l'étude.

Pour l'étude de l'applicabilité clinique, faite sur tests uniques, trois domaines sont quant à eux évalués (**Fig. 10**) :

- 1) **Dans la sélection des patients** : l'applicabilité concernant la sélection des patients est hétérogène. Des points peuvent être discutés, notamment le groupe « contrôle » de Choa,2014 qui est peu représentatif de la réalité de la rhizarthrose, tandis que chez Model,2016 et Sela,2019, les mains opposées à celles symptomatiques sont utilisées comme groupe « contrôle ». Pour les autres études, l'effet de spectre (correspondant à une prévalence élevée par rapport à la normale), bien qu'il ne soit pas problématique, ne peut être écarté puisque les populations sont recrutées sur passage en institution spécialisée.
- 2) **Pour le test index** : la mise en place et l'application pratique des tests cliniques dans toutes les études correspond à notre question de recherche, il s'agit d'une preuve forte de la bonne sélection des articles.
- 3) **Pour la norme de référence** : les études de Gelberman,2015, Model,2016 et Ribak,2020 sont celles qui se rapprochent le plus du modèle communément admis, et où le rôle du kinésithérapeute peut être mis en valeur avec un regard davantage porté sur l'HDLM.

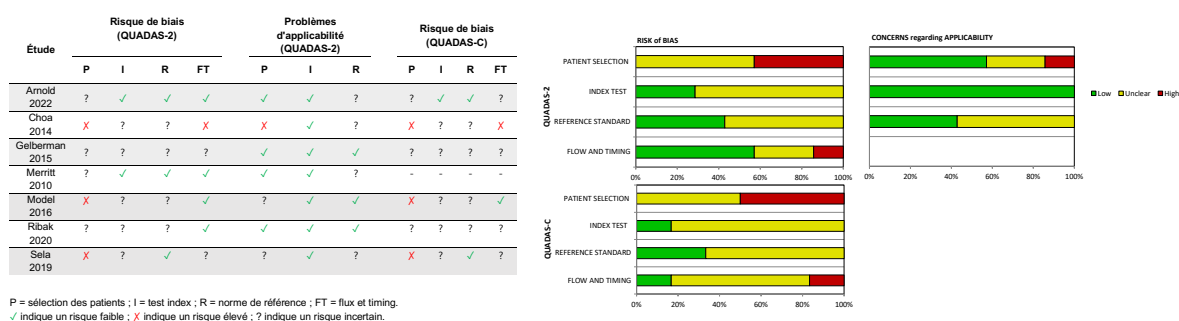


Figure 10 : récapitulatifs du risque de biais

3.2. Méta-analyse statistique

3.2.1. Synthèse décisionnelle

Afin de faciliter la compréhension pour le lecteur, un tableau récapitulatif permettant de visualiser les tests cliniques méta-analysables a été créé (**Tab.X**). Il s'agit donc des *Grind test*, *Palpation test*, *Lever test*, *Extension stress test*, *Adduction stress test* et *Seesaw test*, qui comportent respectivement entre 9 et 2 tableaux 2x2 en inter ou intra-études. Comme annoncé dans la partie **Matériel et méthodes**, les tests à échantillons de données uniques, à savoir les *MP extension test*, *MC extension test*, *MC flexion test* et *Traction-shift test*, sont traités à part pour une visualisation d'ensemble.

Tableau X : récapitulatif décisionnel de la méta-analyse

Test clinique (+ autres noms se rapportant au test)	Citation dans les études	Nombre d'évaluateurs	Échantillons de données correspondants
Grind test	Arnold, 2022	2 (chirurgien et hand fellow)	9
	Choa, 2014	1	
	Gelberman, 2015	1	
	Model, 2016	1	
	Merritt, 2010	2 (évaluateur 1 et évaluateur 2)	
	Ribak, 2020	1	
Palpation test (Joint tenderness test) (Digit pressure test)	Sela, 2019	1	3
	Gelberman, 2015	1	
	Model, 2016	1	
Lever test (MC pressure-shear test) (Shear test)	Ribak, 2020	1	3
	Model, 2016	1	
	Sela, 2019	1	
Extension stress test (MC1 extension stress test)	Gelberman, 2015	1	2
	Ribak, 2020	1	
Adduction stress test (MC1 flexion stress test)	Gelberman, 2015	1	2
	Ribak, 2020	1	
Seesaw test	Arnold, 2022	2 (hand fellow et chirurgien)	2
MP extension test	Model, 2016	1	1
MC extension test	Sela, 2019	1	1
MC flexion test	Sela, 2019	1	1
Traction-shift test	Choa, 2014	1	1

3.2.2. Forest plots

Les *forest plots*, réalisés avec le *package* DTApplots sur le logiciel R, ont permis de générer un récapitulatif selon un modèle bivarié pour chacun des tests méta-analysables (**Annexe VI**).

Le *Grind test* pondéré des 9 tableaux affiche une sensibilité de 42%, avec un indice de confiance à 95% (IC95) compris entre 31% et 53%. Sa spécificité est de 91%, avec un IC95 compris entre 84% et 95% (**Fig.11**). Par ailleurs, les indices de confiance relativement fins fournis ici matérialisent des valeurs statistiques fiables.

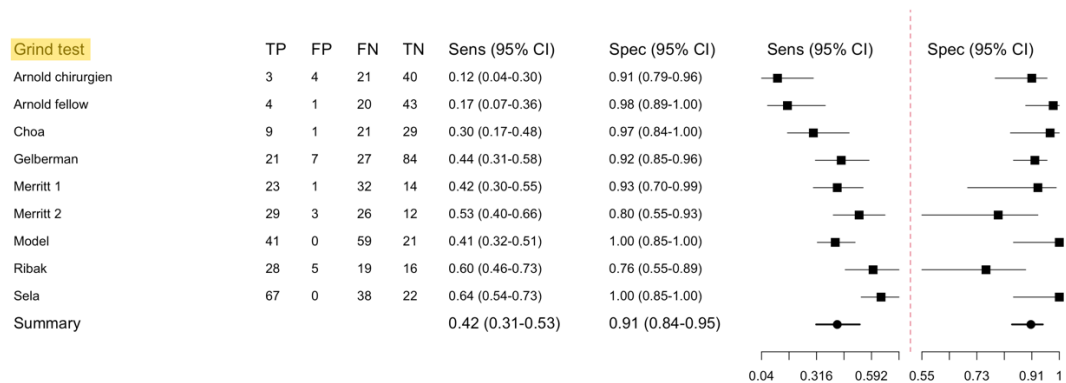


Figure 11 : méta-analyse statistique du Grind test

Le *Palpation test* pondéré des 3 tableaux présente une sensibilité de 90% et une spécificité de 83% (**Fig.12**). Les IC95 pour la sensibilité sont entre 83% et 94%, et entre 72% et 90% pour la spécificité. Ces valeurs resserrées montrent une fiabilité statistique intéressante sur ce test.

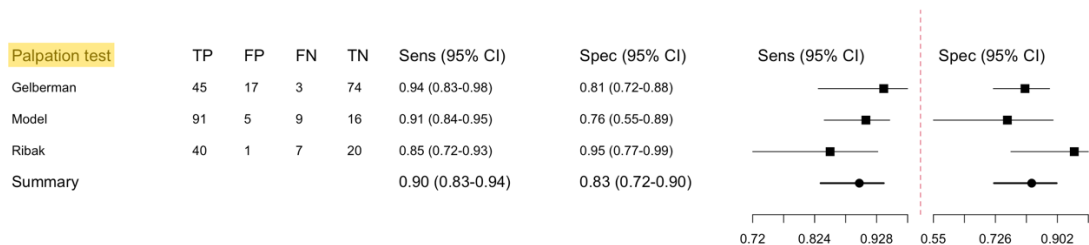


Figure 12 : méta-analyse statistique du Palpation test

Le *Lever test* pondéré des 3 tableaux propose une sensibilité de 90% et une spécificité de 85% (**Fig.13**). Les IC95 pour la sensibilité sont compris entre 52% et 99%, et ceux pour la spécificité sont entre 59% et 96%. On peut donc noter ici un élargissement franc des valeurs, dégradant la fiabilité du bornage statistique.

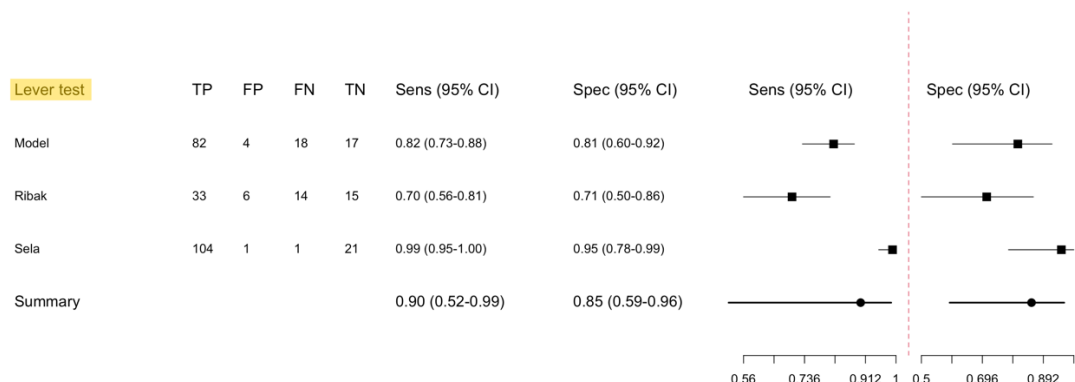


Figure 13 : méta-analyse statistique du Lever test

Le *Seesaw test* pondéré des 2 tableaux produit une sensibilité de 57% et une spécificité de 84% (**Fig.14**). Les IC95 pour la sensibilité sont entre 28% et 81%, et ceux de la spécificité entre 74% et 91%. Sur ce test, l'élargissement important des valeurs encadrant notamment la sensibilité pondérée sur ce test dégrade la fiabilité statistique.

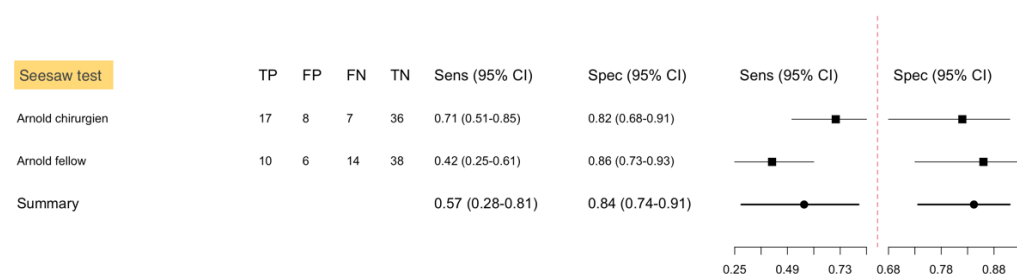


Figure 14 : méta-analyse statistique du Seesaw test

L'Adduction stress test pondéré des 2 tableaux annonce une sensibilité de 76% et une spécificité de 89% (**Fig.15**). Les IC95 pour la sensibilité sont entre 14% et 98%, et entre 67% et 97% pour la spécificité. Comme précédemment, l'élargissement important des valeurs encadrant notamment la sensibilité pondérée dégrade la fiabilité statistique.

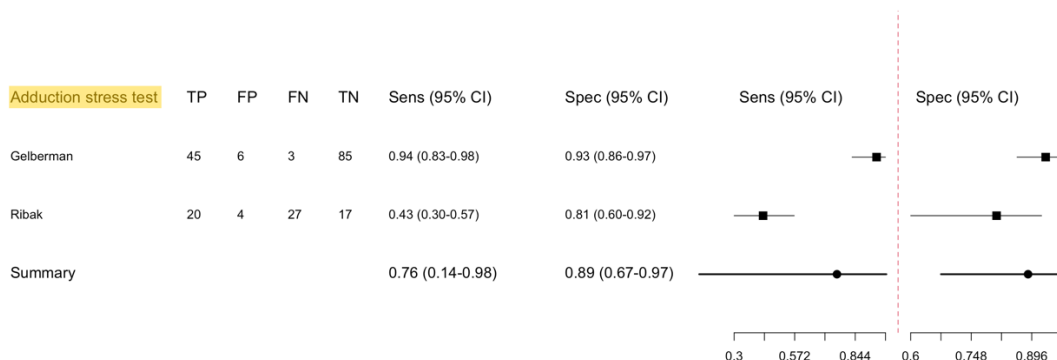


Figure 15 : méta-analyse statistique du Adduction stress test

L'Extension stress test pondéré des 2 tableaux montre une sensibilité de 83% et une spécificité de 91% (**Fig.16**). Les IC95 pour la sensibilité sont compris entre 37% et 98%, ceux pour la spécificité sont entre 75% et 97%. La fiabilité statistique est dégradée devant un élargissement des limites hautes et basses, surtout pour les valeurs encadrant la sensibilité pondérée.

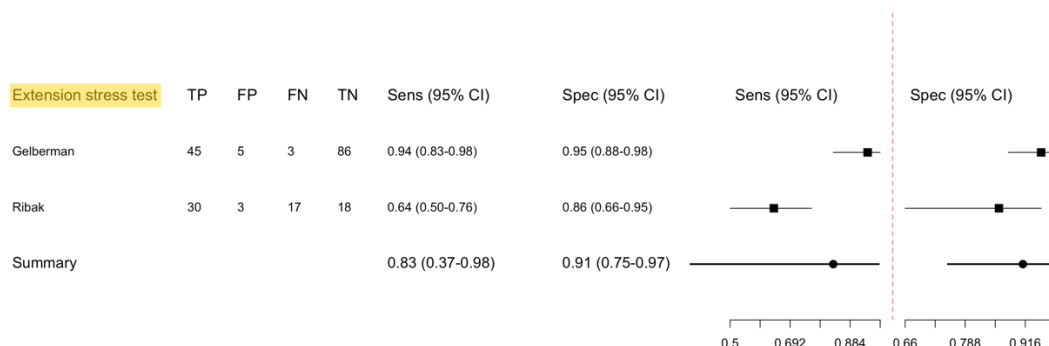


Figure 16 : méta-analyse statistique du Extension stress test

3.2.3. SROC

La seconde étape de la méta-analyse statistique, réalisée grâce au module spécifique mada sur R, donne lieu à un récapitulatif graphique des tests pondérés (**Annexe VI**).

Sur chaque courbe SROC, l'axe des abscisses représente le taux de faux positifs (par le calcul $1 - \text{spécificité}$) et l'axe des ordonnées représente le taux de vrais positifs (directement par la sensibilité). Les sensibilités et $1 - \text{spécificités}$ des tests pondérés sont donc visibles de manière graphique par un point récapitulatif sur la courbe, en complément des *forest plots* vus précédemment (revoir partie 3.2.2. pour résultats numériques). Le point récapitulatif représente ainsi, pour chaque test méta-analysé, les valeurs de sensibilité en ordonnée et de $1 - \text{spécificité}$ en abscisse (**Fig.17**).

Les IC95 modélisés sur le graphique sont par ailleurs représentés par les ellipses entourant le point récapitulatif, et la différence de taille de chaque ellipse est identique aux IC95 présentés par les *forest plots*. Les IC95 les plus importants sont donc ceux du *Lever test*, *Adduction stress test* et *Extension stress test*, expliqué par le fait qu'elles comportent peu d'études et des niveaux de performance hétérogène à travers les études.

L'hétérogénéité statistique I^2 ajustée selon Holling, qui a l'avantage de corriger les effectifs et donc d'atténuer l'effet de taille des échantillons, est davantage en relation avec l'hétérogénéité statistique pure inter-études que le petit nombre d'études ou les petits échantillons (38). Il y a donc ici une relation directe entre l'hétérogénéité statistique importante et les IC95 larges pour les trois tests évoqués plus haut. Les *Grind test* et *Palpation test* présentent des ellipses IC95 minimales, en rapport avec une hétérogénéité statistique faible voire absente, et donc des résultats concordants dans l'ensemble.

Le rendu graphique du *Seesaw test* est original puisque la symétrie de sa présentation graphique est quasiment parfaite. L'explication tient au fait que les tableaux 2x2 sont issus de la même étude et produits par deux opérateurs indépendants. L'IC95 du *Seesaw test* relativement large tient alors probablement son explication d'une méta-analyse pratiquée sur une population identique avec 2 tableaux statistiques significativement différents.

D'autre part, une application pratique des courbes SROC est le calcul des aires sous la courbe AUC représentent la performance globale des tests cliniques étudiés (49). Elle correspond au calcul de la surface virtuelle couverte par la partie sous la courbe, comprise entre 0 et 1. Plus le point récapitulatif se situe en haut à gauche du graphique, plus la performance du test est importante, et plus la valeur de l'AUC se rapproche de 1. Les tests méta-analysés présentent des AUC comprises entre 0,803 et 0,940 qui les rendent performants sur le plan statistique. La notion de probabilité de détecter la pathologie entre un individu sain et un individu malade est également apportée par la *Cochrane Library* avec les

AUC, les tests présentés sont donc probablement assez bons pour détecter la rhizarthrose chez un individu malade d'après les calculs réalisés (50).

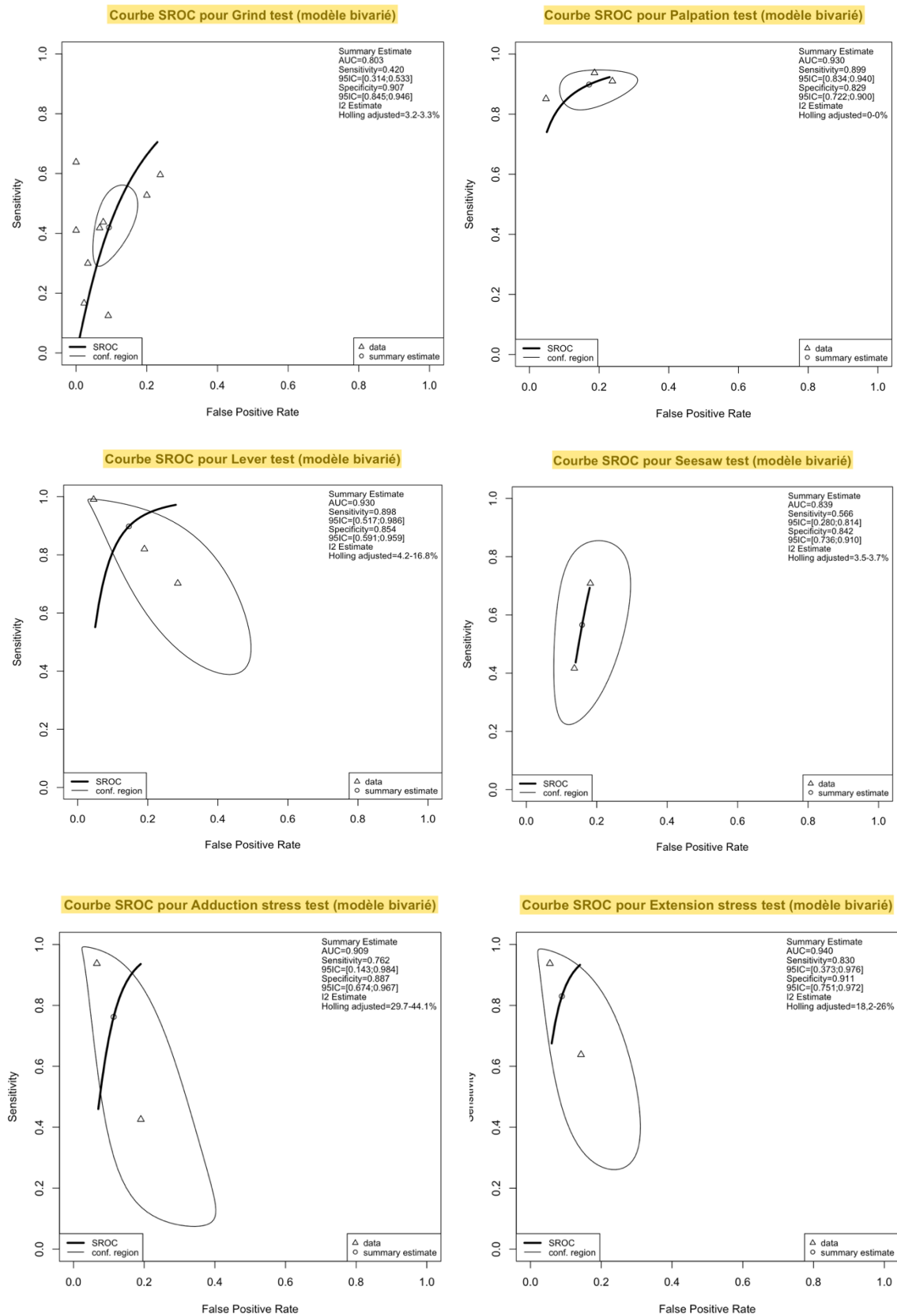


Figure 17 : courbes récapitulatives SROC de la méta-analyse statistique

3.2.4. « Pseudo-ROC »

Ces courbes ont été réalisées sur R en suivant une procédure de création graphique simple (**Annexe VI**). Elles permettent une présentation à titre de comparaison visuelle, sans rapport direct avec les SROC précédemment développées.

Les tests uniques présentent une AUC calculée comprise entre 0,68 et 0,83 reflétant le niveau de performance évalué dans leur étude individuelle (**Fig.18**). Pour information, la « ligne du hasard » en pointillés vaut 0,5 et rend un test non informatif, et le coin en haut à gauche vaut 1 et rend ce même test parfait sur le plan de la performance (50). Sur le même principe que les courbes SROC, le point d'intersection de chaque courbe colorée représente, pour un test isolé, les valeurs de sensibilité en ordonnée et de 1-spécificité en abscisse (valeurs numériques consultables dans le **Tableau IX**).

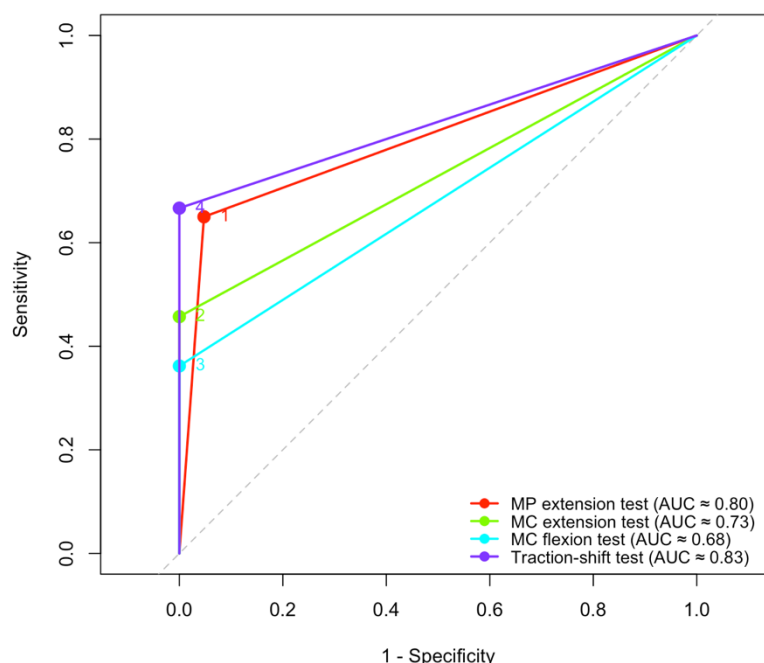


Figure 18 : courbes "pseudo-ROC" des tests uniques

3.2.5. Recherche de biais de publication

Le *package* metafor, utilisé sur R, a permis la réalisation technique de cette recherche. Elle permet d'avoir un regard critique concernant la sélection des articles et des tests qui ont servi à réaliser la méta-analyse statistique secondaire à ce travail de revue (disponible à l'adresse : <https://cran.r-project.org/web/packages/metafor/index.html>). La procédure des packages et appels utilisés dans le cadre de cette recherche est disponible en **Annexe VI**.

Présentation globale

L'option retenue en premier lieu est de réaliser une agrégation conventionnelle globale, avec les sept études mises en revue (**Fig. 11**).

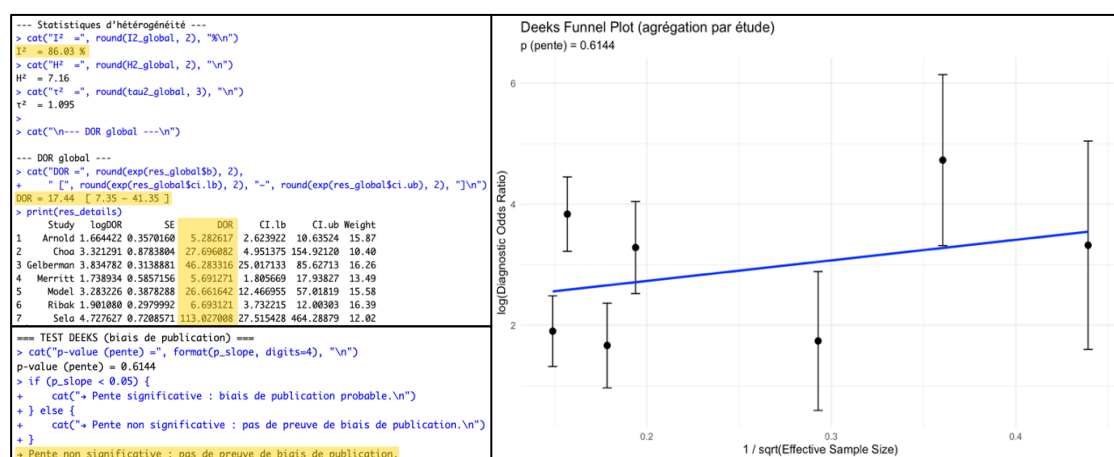


Figure 19 : recherche de biais de publication globale

Tout d'abord, il convient de signaler que le DOR équivaut à l'*odd ratio*. C'est ce paramètre, corrélé à la taille effective des échantillons des différents articles, qui est évalué pour analyser le biais de publication lors de comparaisons portant sur l'efficacité de tests diagnostiques. Le DOR moyen est pondéré à 17,44 : trois études se situent en-dessous de la valeur pondérée, alors que les quatre autres sont au-dessus. Nous pouvons donc dire en premier lieu qu'il existe un équilibre dans les publications. Ce paramètre brut doit être discuté par rapport à l'étude de Sela, 2019, qui est significativement plus haute que toutes les autres, et qui biaise probablement le chiffre final donné, ainsi que les indices de confiance larges qui sont parfois attribués aux DOR des études (notamment chez Choa, 2014, Sela, 2019 et Gelberman, 2015). Conjointement aux données chiffrées, l'analyse graphique permet de dire qu'une légère pente est visualisée, avec une relative symétrie des études de part et d'autre. Pour information, le *funnel plot* selon Deeks est calculé en log(DOR), les valeurs par étude sont données dans le tableau récapitulatif et peuvent donc être retrouvées individuellement sur le graphique. La seconde analyse visuelle permet donc de corroborer les informations données par les DOR quant aux biais. La valeur p du test de Deeks égale à 0,6144 rend la pente non significative, suggérant l'absence de biais de publication. Le dernier élément permettant d'évaluer la pertinence des chiffres annoncés est le calcul de l'hétérogénéité I². Ce chiffre de 86,03%, signe une hétérogénéité statistique importante. Corrélé à la faible puissance statistique rendu par le petit nombre d'études mis en revue, il convient donc de réaliser une analyse complémentaire en sous-groupes afin d'affiner la recherche de biais de publication.

Présentation en sous-groupes

Cette recherche en sous-groupes permet à la fois d'optimiser les chances d'homogénéisation des résultats des tests présentés dans les études, et d'analyser la performance des tests individuellement. Une approche visuelle par *forest plots* pour les tests regroupés a été conçue dans R, afin d'accompagner le tableau récapitulatif (**Fig. 12**).

Test	logDOR	se[,1]	ci.lb	ci.ub	DOR[,1]	DOR_lb	DOR_ub	tau2	I2
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 Adduction ...	3.24	2.11	-0.891	7.36	25.4	0.410	1578.	8.41e+0	9.48e+1
2 Extension ...	3.94	1.60	0.814	7.07	51.6	2.26	1179.	4.57e+0	8.97e+1
3 Grind	1.91	0.267	1.39	2.44	6.77	4.01	11.4	2.58e-6	3.72e-4
4 Lever	3.92	1.70	0.601	7.25	50.6	1.82	1403.	7.77e+0	9.31e+1
5 MC extensi...	3.61	1.44	0.783	6.44	37.1	2.19	627.	0	0
6 MC flexion	3.22	1.44	0.386	6.05	25.0	1.47	423.	0	0
7 MP extensi...	3.61	1.05	1.56	5.66	37.1	4.78	289.	0	0
8 Palpation	3.94	0.417	3.12	4.76	51.5	22.8	117.	0	0
9 Seesaw	1.96	0.441	1.09	2.82	7.07	2.98	16.8	2.96e-2	7.60e+0
10 Traction s...	4.79	1.48	1.89	7.68	120	6.63	2172.	0	0

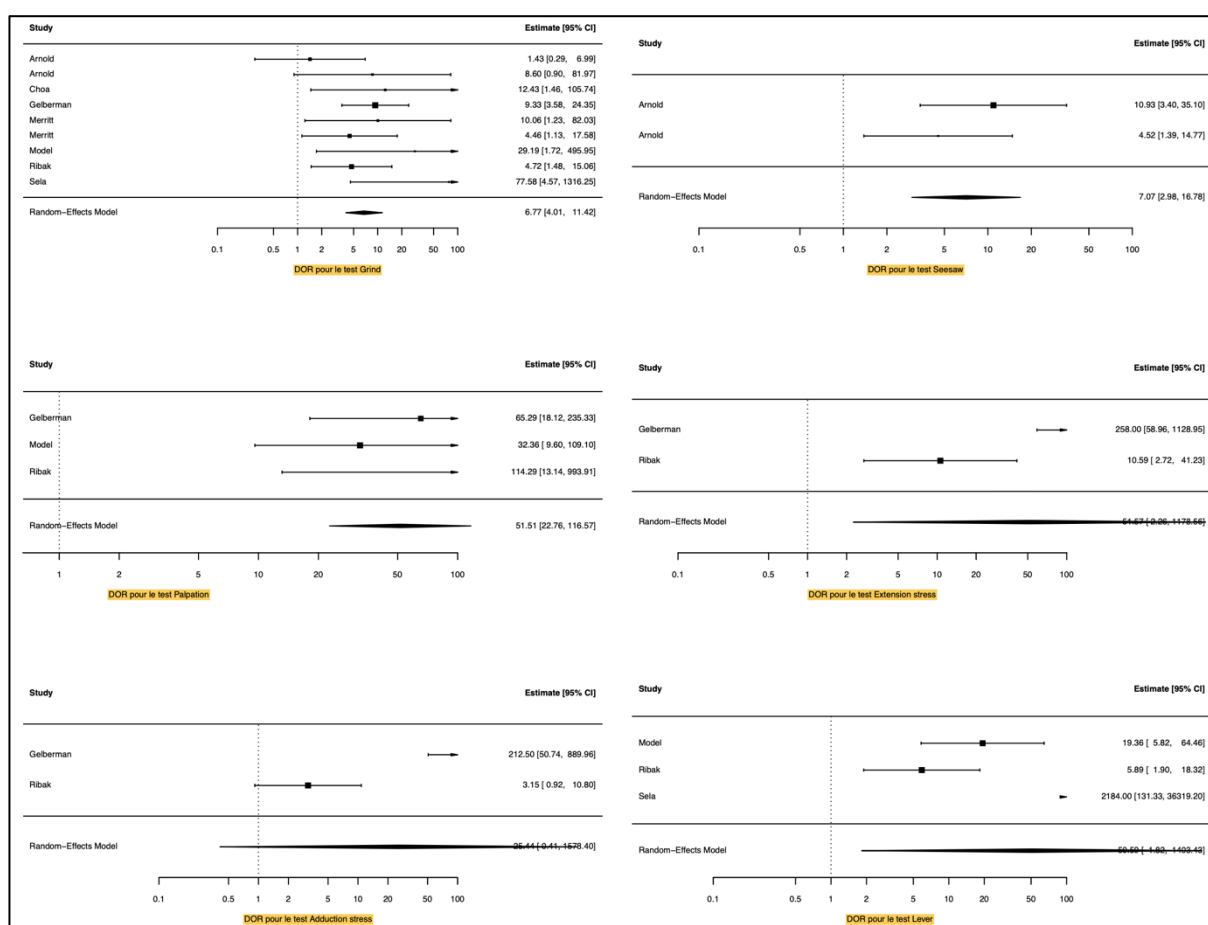


Figure 20 : récapitulatifs de l'analyse en sous-groupe

Les conclusions que nous pouvons faire, qui doivent être traitées avec la plus grande prudence, sont que le *Grind test* doit être considéré comme un test introductif. Il possède un DOR pondéré relativement faible à 6,77 mais avec un IC95 contenu et une hétérogénéité très faible : il peut donc d'ores et déjà être considéré solide sur le plan statistique, mais intermédiaire sur le plan diagnostique. Le *Palpation test* possède un DOR pondéré à 51,5 avec un IC95 assez large et une limite inférieure de l'IC95 intéressante à 22,8. De plus, l'hétérogénéité y est absente, ce qui en fait un test robuste à travers les différentes études. L'*Adduction stress test* et l'*Extension stress test* possèdent des performances diagnostiques significatives (25,4 et 51,6 respectifs en DOR) mais avec des IC95 extrêmes. Ils ont également une hétérogénéité élevée (89,7 et 94,8% respectivement), et peut-être correspondent-ils à une sur-efficacité diagnostique évaluée de la part d'un de leurs auteurs (Gelberman, 2015 à chaque fois). Le *Lever test* possède un DOR à 50,6, un IC95 extrême et une hétérogénéité à 93,1%. Model, 2016 et Ribak, 2020 en avaient fait un outil diagnostique correct (19,36 et 5,89 de DOR respectivement et IC95 ajustés), mais Sela, 2019 a peut-être surévalué l'apport de ce test dans son étude (2184 de DOR avec un IC95 très extrême). Le *Seesaw test* possède un DOR moyen à 7,07 avec un IC95 contenu, et une hétérogénéité faible à 7,6%. L'originalité de ce test étudié dans notre méta-analyse est qu'il fait partie de la même étude, qui comportait deux évaluateurs aveugles l'un de l'autre. Il est donc recevable tant sur le plan statistique que diagnostique, mais demanderait tout de même à être réévalué dans une prochaine étude clinique. Les quatre autres tests isolés, qui ne présentent pas d'hétérogénéité car seuls, présentent des très bonnes performances diagnostiques : de 25 à 120 en DOR mais possèdent des IC95 systématiquement très importants. Lorsque nous remontons à la source en regardant les tableaux 2x2 correspondants, l'explication tient au fait que pas ou très peu de faux positifs ont été déclarés pour ces tests (entre zéro et un). L'idéal serait naturellement de retrouver ces tests isolés dans d'autres études, avec un recrutement de malades et de non-malades significatif, pour envisager une confirmation ou un ajustement de l'apport diagnostique élevé annoncé.

4. DISCUSSION

4.1. Forces et faiblesses des articles inclus

Comme expliqué précédemment, les articles de ce travail fonctionnent en général en vase clos sur ce sujet de niche, et les auteurs les plus récents citent et discutent leur travail en fonction des études les plus anciennes. Il s'agit donc en quelque sorte pour les auteurs de présenter le nouveau test clinique le plus puissant de la littérature. Cette pratique fut intéressante en premier lieu puisqu'elle nous a permis d'avoir rapidement une vue d'ensemble de la majorité des tests cliniques existants. Par ailleurs, des revues systématiques retrouvées dans le cadre de la méthode appuyaient ces recherches, mais rendaient aussi le travail initial de notre revue déjà au moins en partie réalisé (51,52). Les résultats préliminaires du travail de revue nous ont aussi permis de trouver des articles prometteurs présentant d'autres tests, mais non exploitables car incomplets sur le plan statistique (53,54). C'est l'article de Ribak, 2020 issu de recherches annexes sur le sujet qui nous a permis de pouvoir passer un cap intéressant, et de s'offrir la possibilité d'utiliser la méthode de la méta-analyse statistique dans ce travail. Cet article compare en effet ses résultats avec un grand nombre d'auteurs, et donc de tests cliniques présents dans la revue. Nous avons par ailleurs été particulièrement attentifs à la qualité de cet article avant de l'ajouter à la revue systématique, puisque son journal n'était pas indexé dans les bases de données principales.

4.2. Forces et faiblesses des tests analysés

Dans le cadre de ce travail, un axe supplémentaire qui a guidé notre réflexion aura été la mise en pratique réelle des tests récoltés, et donc de réfléchir notamment à la structure anatomique mise en jeu sur chaque test proposé, la facilité de mise en œuvre du test et la reproductibilité inter et intra-opérateur. Tout d'abord, nos lectures préalables ont soulevé des interrogations sur la structure anatomique la plus importante à évaluer pour rechercher une rhizarthrose. Naturellement, la physiologie de l'arthrose laisse à penser que l'évaluation de l'architecture articulaire et l'environnement capsulo-ligamentaire sont primordiales, alors que certains auteurs discutent aussi d'autres structures mises en cause comme les ligaments à distance (55,56). En ce sens, les tests en compression articulaire comme le *Grind test*, et ceux de mise en tension capsulo-ligamentaire comme les *Lever test*, *Adduction stress test* et *Extension stress test* paraissent des options fortes. *A contrario*, le *Traction-shift test*, avec son étape préalable de traction articulaire, paraît aller à l'encontre de la physiopathologie purement articulaire de l'arthrose, et semble techniquement difficile à réaliser. La nature des

tests pratiqués a aussi été un élément de réflexion de notre part. Les tests faits en actif pour une recherche articulaire de provocation douloureuse nous paraissent moins pertinents que ceux réalisés en passif par le thérapeute. Le seul « test » actif adéquat serait probablement une mesure de type empan, pour objectiver une perte d'amplitude en abduction du pouce dans les stades avancés de la maladie. Le *MP extension test*, pratiqué en actif et en interposant l'articulation métacarpo-phalangienne, ne semble ainsi pas le meilleur choix. Le *Seesaw test* paraît quant à lui difficile à reproduire dans le cadre de l'inter-opérabilité avec son mouvement en balançoire, et donc pas le choix le plus adéquat dans le cadre de notre travail.

De plus, les comparaisons de tests cliniques ne nous ont pas toujours semblé pertinentes. Par exemple, Ribak,2020 dans la discussion de son étude compare le *Traction-shift test* de Choa,2014 avec son *Lever test* sans pratiquer la traction longitudinale préalable. Selon notre interprétation, il ne s'agit pas de la même contrainte biomécanique qui est engagée, puisque Choa,2014 propose un mouvement dorso-palmaire du pouce alors que Ribak,2020 propose un mouvement médio-latéral. Nous avons choisi, de notre côté, de séparer ces deux tests pour la méta-analyse statistique devant une différence méthodologique nette. D'un point de vue plus global, des variantes se dégagent dans le positionnement des mains du thérapeute ou par interposition articulaire de l'articulation métacarpo-phalangienne lors de certains tests, il s'agit d'une faiblesse à prendre en compte et une harmonisation des pratiques serait à prévoir (voir **Annexe IV**).

4.3. Intérêt et limites de la démarche méthodologique présentée

La démarche méthodologique utilisée dans ce travail se doit d'être commentée. Tout d'abord, les grilles QUADAS-2 et QUADAS-C sont des outils puissants, mais avec une mise en œuvre difficile. Une application stricte de la méthode requiert en théorie la participation d'au moins deux correcteurs aveugles l'un de l'autre, avec une confrontation et une mise en commun des résultats. Il est également à noter que QUADAS-2 s'utilise en fonction de la question de recherche concernant l'applicabilité, et est donc personnalisée selon le contexte de la recherche (48). Ce paramètre a été à la fois un avantage et un inconvénient, puisqu'il nous a mis face à notre responsabilité d'évaluateur inexpérimenté dans le cadre d'un sujet technique, avec notre propre interprétation à donner. Les grilles ont par ailleurs mis en lumière un certain nombre de limites sur le plan des risques de biais, ne permettant pas de garantir des résultats irréprochables au niveau de la démarche méthodologique dans ce travail de comparaison d'efficacité de tests cliniques.

De plus, le choix de l'outil de calcul statistique sélectionné est à expliquer. Dans une méta-analyse statistique, deux grandes approches par effets aléatoires existent aujourd'hui, la méthode par effet fixe étant obsolète. Le modèle bivarié estime la sensibilité et la spécificité

en tenant compte de leur corrélation ; il s'agit du modèle classique et le plus utilisé en recherche pour la méta-analyse de tests diagnostiques. Ce modèle donne des intervalles de confiance et la possibilité de réaliser des courbes ROC de synthèse, il est donc adapté mais parfois limité lorsque peu d'études sont méta-analysées. L'approche bayésienne est l'autre filière exploitable pour mener à bien une méta-analyse statistique. Elle utilise les mêmes équations que le modèle bivarié, mais avec une logique probabiliste ; elle nécessite de définir des *priors* (ou probabilités *a priori*) avant l'étape de calcul. Cette méthode fournit des intervalles de confiance souvent plus stables quand il y a peu d'études, elle est aussi plus flexible mais plus complexe à mettre en place (57). Pour notre travail, c'est le modèle bivarié qui a été choisi car il est standard, simple, accepté par les revues et très documenté. L'approche bayésienne a été écartée car elle demandait plus de justifications informatiques et une haute technicité statistique. Parmi les applications bivariées non bayésiennes listées par la *Cochrane Library*, le *Handbook* privilégie l'utilisation de la fonction *glmer* sur R ; ainsi des essais ont été effectués mais ce logiciel n'était pas accessible à notre niveau de compétences. L'application mada choisie n'est pas la plus recommandée par la *Cochrane Library*, mais elle est la plus utilisée en recherche clinique d'efficacité des tests diagnostiques (58). Notre choix final s'est donc porté sur cette fonction, puisqu'elle représentait le bon compromis technicité/temps disponible/résultats escomptés pour notre travail.

Par souci de transparence face au lecteur, une information importante à transmettre était la mise en garde par le logiciel R des résultats donnés lorsque deux tableaux 2x2 seulement étaient méta-analysés. Cela met probablement en lumière une limite de ce travail sur le plan de la pertinence statistique.

4.4. Intérêt et limites des résultats pour la pratique et la profession

L'arthrose n'est pas un *Red Flag* urgent à proprement parler, le kinésithérapeute peut donc exercer son art de la science clinique et, grâce à l'ensemble de ses connaissances et son expérience, devenir un maillon essentiel dans l'accompagnement de cette maladie sans nuire au patient (59). L'intervention dans le cas précis de recherche de rhizarthrose peut donc être bénéfique pour potentiellement accélérer la prise en charge médicale, au contact d'une maladie présentant un tableau douloureux et de perte fonctionnelle pouvant être importants. Ce *cluster* de tests a donc potentiellement un intérêt pour la pratique, tout en se limitant à notre champ de compétences propre. De plus, l'étape palpatoire du bilan, évoquée dans l'Introduction du Mémoire, devient un réel test clinique mis en évidence par les statistiques et supporté par une marche à suivre (voir **Annexe IV**). Par ailleurs, les aspects biomécaniques, à travers les amplitudes articulaires ; la fonction, à travers les questionnaires adaptés ; la proprioception, à travers les différentes sensibilités ; ou l'environnement neuro-musculaire, à

travers les tests de force musculaire, n'ont pas été abordés dans ce sujet mais sont des éléments essentiels d'un bilan kinésithérapique complet. D'autre part, l'aspect du diagnostic différentiel devrait occuper une place non négligeable de la prise en charge dans le cas précis du raisonnement clinique. Cette étape importante aurait sa place dans un autre Mémoire à lui seul, et il est essentiel de se rapporter aux écrits de référence pour exhaustivité (18,21). Nous pouvons citer à titre d'exemple les maladies rhumatismales inflammatoires, les différentes fractures possibles dont celle du scaphoïde, la ténosynovite de De Quervain, les différentes entorses possibles, un SDRC, une maladie de Dupuytren localisée au pouce, le syndrome du canal carpien, un kyste synovial ou encore une névrite. La limite principale de nos résultats pour la pratique professionnelle est ainsi l'impact du diagnostic différentiel sur les tests proposés, tout en sachant qu'une anamnèse et une inspection bien menées, ainsi que des tests cliniques dédiés, puissent déjà théoriquement en éliminer la plupart.

4.5. Intérêt et limites des résultats par rapport à l'état des lieux de la littérature

Les revues de la littérature récentes de McGee,2024 et He,2025 permettent de voir que la prise en charge de la rhizarthrose trouve un intérêt auprès de la communauté scientifique dans le cadre du bilan et du suivi global (51,52). Les aspects mis en lumière par l'étude McGee,2024 portent sur la clinimétrie, dans laquelle certains tests de notre revue sont retrouvés et associés à des mesures avec pinchmètre, dynamomètre de main, goniométrie ou centimétrie. L'étude de He,2025 porte sur les aspects majoritaires du bilan de la rhizarthrose à travers l'histoire de la maladie, les tests cliniques, dont certains sont retrouvés dans notre revue, et les examens complémentaires. Notre travail complète donc ces revues en se focalisant sur un aspect évoqué par les autres auteurs. D'autre part, la dernière mise à jour internationale pour harmoniser une classification de l'arthrose de la main modernise la vision de la pathologie développée dans le rapport *Task Force* de EULAR,2009, dans lequel les signes d'arthrose radiologique étaient le facteur majoritaire de diagnostic de l'arthrose de la main (60). La douleur, l'inspection clinique, l'histoire de la maladie, la raideur, la recherche de diagnostic différentiel sont vus comme des critères décisionnels, au même titre que les ostéophytes ou autres signes radiologiques. Ces aspects soutiennent donc l'importance de développer la clinique comme nous le faisons à travers notre travail.

4.6. Réflexion autour des questions de recherche

Rappel de la question principale : Quels tests cliniques existants sont statistiquement fiables pour détecter une rhizarthrose sur une population à risque ?

Rappel des sous-questions : La constitution d'un cluster de confirmation statistiquement significatif est-elle possible lorsqu'un doute clinique apparait ? Dans un avenir proche, cette démarche peut-elle permettre aux kinésithérapeutes, à travers l'utilisation de ce type d'outil, de prendre part à des actes de prévention, mais aussi d'affiner leurs compétences face au triage des patients qui prendra une part de plus en plus importante dans leur activité avec l'accès direct qui devient possible ?

L'article rédigé par Michener en 2009 offre la structure d'article la plus à même de répondre à la création pratique d'un *cluster* de tests (61). Ainsi, une étude prospective semble obligatoire pour la mise en application, mais la base du travail pour la rhizarthrose serait préparée avec une sélection préalable de tests. Un guidage méthodologique est nécessaire pour comprendre le cheminement ayant abouti à notre réflexion (**Fig.21**)(26). Ainsi, grâce à l'uniformisation des résultats permise par la méta-analyse statistique et notre raisonnement clinique, une tendance a ainsi pu d'ores et déjà se dégager.

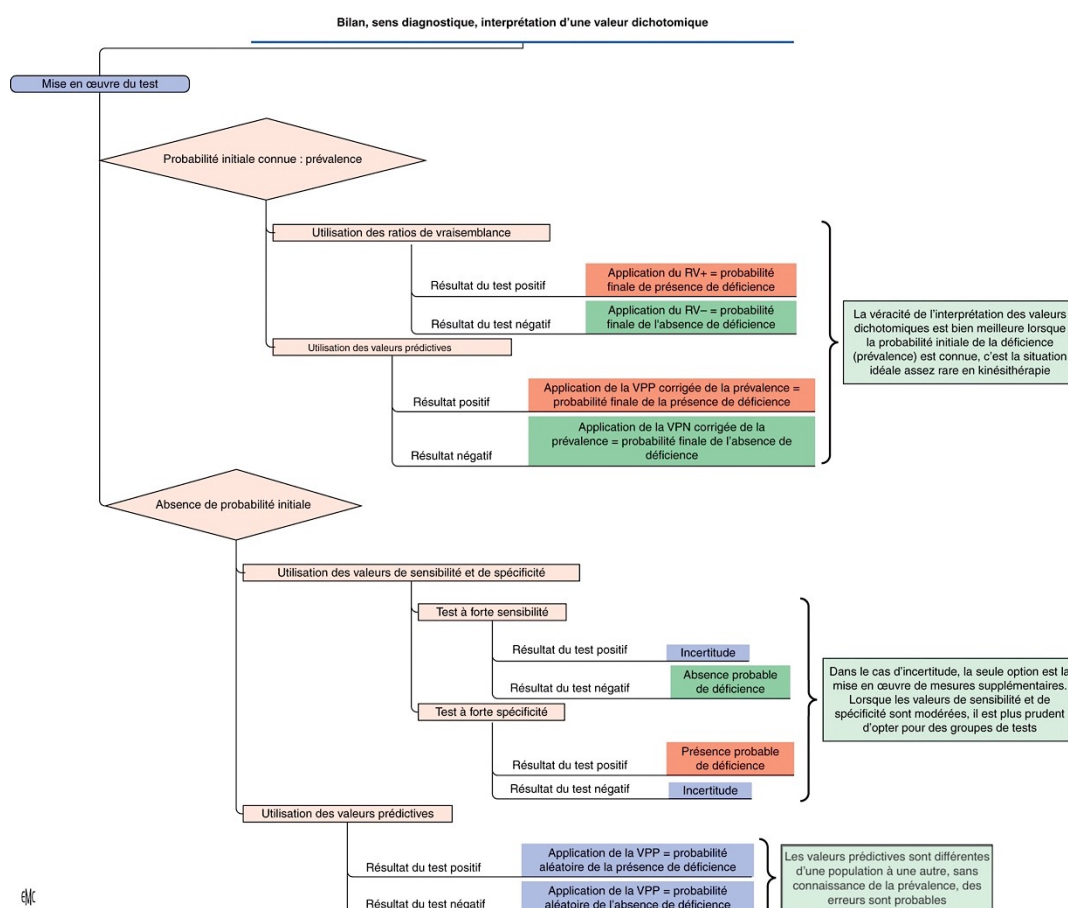


Figure 21 : démarche dans le cadre de la clinimétrie raisonnée en kinésithérapie (source : Piette, 2021 corrigé)

La première étape fut de valoriser de manière chiffrée les grilles QUADAS réalisées à travers les différentes études par un système de points. Le système mis en place représente un indice de confiance attribué à chaque article en fonction des paramètres QUADAS (**Tab.XI**).

Chaque item est noté sur 2 points selon le barème suivant, en fonction des coches des différents QUADAS : risque faible = 2 points ; risque incertain = 1 point ; risque élevé = 0 point (voir **Fig.10**). Ce système de notation a servi plus tard de base pour évaluer la qualité méthodologique des différents tests à travers leurs études initiales. La moyenne est comprise entre 0 et 1, et les études de Arnold,2022, Gelberman,2015, Merritt,2010 et Ribak,2020 sont celles qui sont le mieux valorisées avec une moyenne comprise entre 0,75 et 0,85.

Tableau XI : grilles QUADAS valorisées et moyennées

Un indice de confiance proche de 0 indique une qualité méthodologique médiocre alors qu'un indice de confiance proche de 1 révèle une qualité méthodologique très bonne.

Article	Grilles QUADAS		
	Risque de biais (chaque item noté sur 2)	Applicabilité clinique (chaque item noté sur 2)	Moyenne de l'article (uniformisation des QUADAS)
Arnold,2022	13/16	5/6	0,82
Choa,2014	4/16	3/6	0,38
Gelberman,2015	8/16	6/6	0,75
Merritt,2010	7/8	5/6	0,85
Model,2016	8/16	5/6	0,67
Ribak,2020	9/16	6/6	0,78
Sela,2019	8/16	4/6	0,58

L'étape suivante a été de calculer les rapports de vraisemblance positif RV+ et négatif RV-, qui sont des critères d'efficacité diagnostique d'un test pour discriminer les patients sains de patients malades, et s'utilisent en binôme. Un report du **Tab.IX** a été fait pour les tests uniques, et pour les autres, des calculs ont été réalisés manuellement (voir **Tab.III** pour les formules), puis confrontés au calculateur du site *MSD Manuals* (disponible à l'adresse : <https://www.msdmanuals.com/fr/professional/multimedia/clinical-calculator/multicalculatrice-de-rapport-de-vraisemblance>). Les résultats sont disponibles dans le **tableau XII** (page suivante), et en suivant la méthodologie fournie par Piette,2021 les tests avec RV+ entre 5 et 10 et ceux avec RV- entre 0,1 et 0,2 sont pertinents mais à effet modéré seul (26). Les tests remplissant les conditions sont les *Palpation test*, *Lever test* et *Extension stress test*.

D'autre part, la prévalence, qui s'appelle aussi probabilité pré-test en statistiques, représente le nombre de malades parmi une population à un instant déterminé. Une prévalence moyenne a été choisie à 15% pour mener à bien les calculs sur les tests et permettre d'homogénéiser le rendu statistique. Le chiffre a été volontairement minimisé en fonction de l'objectif précis du *cluster* de tests : nous justifions donc ce choix pour une amélioration du repérage de la pathologie, en fonction des données récoltées dans la littérature (voir **1.4.1**)(10). La prévalence aurait pu être reprise à partir des différentes études de la revue, mais les méthodes de recrutement différentes auraient probablement faussé les résultats avec une prévalence surévaluée (62).

A partir de la prévalence déterminée, il a été possible de calculer les RV+ et RV- post-test, qui sont aussi nommés valeurs prédictives positive (VPP) et négative (VPN), et sont exprimées en pourcentage. Ces chiffres représentent respectivement la probabilité qu'un individu soit malade si le test est positif pour la VPP, et la probabilité qu'un individu soit non-

malade si le test est négatif pour la VPN. Les calculs ont été faits manuellement, puis confrontés aux résultats acquis avec module de calcul de l'HAS (disponible à l'adresse : <https://has-sante.pages.has-sante.fr/public/infographie-vpp-vpn/>). Les formules ayant servi pour les calculs ont été : $VPP = (Sn \times Pr) / (Sn \times Pr + (1 - Pr) \times (1 - Sp))$ et $VPN = (Sp \times (1 - Pr)) / (Sp \times (1 - Pr) + Pr \times (1 - Sn))$. Les RV post-test, dont le Nomogramme de Fagan est une représentation populaire chez les cliniciens ayant besoin de résultats visuels et rapides, peuvent en effet servir de support pour justifier la proposition d'un *cluster* de tests (**Fig.22**).

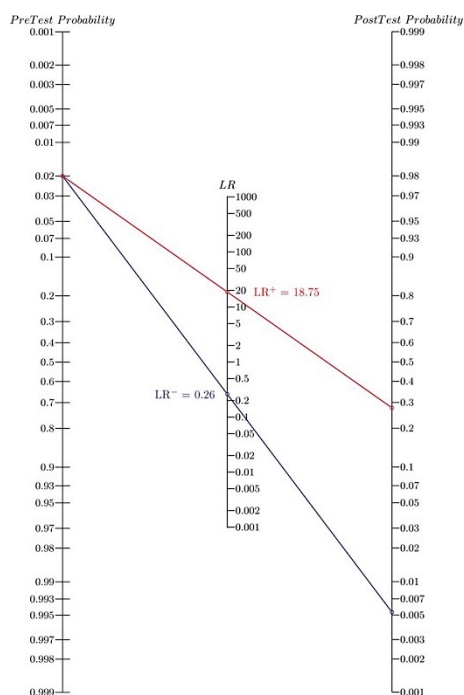


Figure 22 : Nomogramme de Fagan projetant un exemple à partir de la prévalence et des RV+ et RV- (Source : Caraquel CGB, 2013)

Les VPP des 6 tests méta-analysés sont relativement bas dans notre projection avec prévalence réduite, avec une VPP comprise entre 39% et 62%, et ne pourraient pas garantir l'assurance de détecter correctement la pathologie en étant utilisés isolément. Ce paramètre appuie l'utilité d'un *cluster* de tests puisque les valeurs de l'étude de Michener, 2009 étaient comprises entre 40% et 64% (61). Les tests uniques proposent de bien meilleures performances sur le plan de la VPP, mais ont l'inconvénient d'être issus de tableaux 2x2 uniques (**Tab.XII**).

Tableau XII : tableau des rapports de vraisemblance des tests

Nom du test	RV+	RV-	RV+ post-test (VPP)	RV- post-test (VPN)
Grind test	4,67	0,64	45%	90%
Palpation test	5,29	0,12	48%	98%
Lever test	6	0,12	51%	98%
Extension stress test	9,22	0,19	62%	97%
Adduction stress test	6,91	0,27	55%	96%
Seesaw test	3,56	0,51	39%	92%
MP extension test	13,65	0,37	70%	94%
MC extension test	∞	0,54	100%	91%
MC flexion test	∞	0,64	100%	90%
Traction-shift test	∞	0,33	100%	95%

Pour la dernière étape, un récapitulatif général par tests a été créé afin d'acquérir une prise d'informations globale liant la méthodologie uniformisée, les statistiques projetées sur une population homogène, et une réflexion personnelle sur la mise en pratique (**Tab.XIII**). Les items développés reprennent les RV, l'indice de Youden, le QUADAS moyen et une appréciation autour de la biomécanique et du niveau de difficulté de réalisation du test.

Le QUADAS moyen des dix tests cliniques de la revue a été calculé pour générer une vue d'ensemble sur la pertinence méthodologique des articles qui composent chaque test (voir **Tab.X**). La valeur des QUADAS moyens de chaque étude a été ajoutée, puis divisée par le nombre d'études mentionnant le test. Ce système a donc produit une note comprise entre 0 et 1. Un seuil arbitraire a été attribué à 0,70 en fonction des QUADAS étudiés dans les différents articles, défavorisant notamment les modèles méthodologiques de Choa,2014, Model,2016 et Sela,2019 (cf. partie **3.1.3.**). Les tests les mieux valorisés furent les tests inclus dans notre méta-analyse, à l'exception des *Grind test* et *Lever test*.

L'indice de Youden Y apporte une information sur la précision de la méthode diagnostique, mais sa valeur reste théorique puisqu'elle ne prend pas en compte la prévalence de la maladie. Néanmoins, son calcul est intéressant puisqu'il représente le point de coupure idéal entre les valeurs de sensibilité et de spécificité de chaque test. Il est calculé par la formule : $Y = Sn + Sp - 1$ et sa valeur est comprise entre 0 et 1 (37,63). Nous avons calculé cet indice de Youden pour chaque test, et les avons confrontés à l'aide d'une application de calcul automatique présente sur le site internet mdapp (disponible à l'adresse : <https://www.mdapp.co/youden-index-calculator-479/>). L'exploration de la littérature scientifique ne nous pas permis de déterminer une valeur minimale utile de l'indice de Youden en pratique clinique. Aussi, nous avons déterminé un chiffre de 0,5 correspondant à une sensibilité et une spécificité de 75%, qui représentent à titre d'exemple la valeur de coupure du *cluster* de Michener,2009 lorsque 3 tests cliniques des 5 tests sont positifs (61). Les tests valorisés par l'indice sont les *Palpation test*, *Lever test*, *Extension stress test*, *Adduction stress test* et *Traction-shift test*.

La justification de la valorisation pour la réalisation pratique et la pertinence biomécanique ont, quant à elles, été données plus haut dans la Discussion.

Tableau XIII : récapitulatif général par tests

Nom du test	RV+	RV-	Indice de Youden	QUADAS moyen du test à travers leurs études	Facilité de réalisation et pertinence biomécanique
Grind test	4,67	0,64	0,33	0,69	+
Palpation test	5,29	0,12	0,73	0,73	+
Lever test	6	0,12	0,75	0,68	+
Extension stress test	9,22	0,19	0,74	0,77	+
Adduction stress test	6,91	0,27	0,65	0,77	+
Seesaw test	3,56	0,51	0,41	0,82	-
MP extension test	13,65	0,37	0,6	0,67	-
MC extension test	∞	0,54	0,46	0,58	+
MC flexion test	∞	0,64	0,36	0,58	+
Traction-shift test	∞	0,33	0,67	0,38	-

Au final, les résultats très avantageux des tests réalisés dans l'étude de Sela,2019 et l'unicité des tableaux correspondants tendent à écarter les *MC flexion test* et *MC extension test* du choix final, tout comme le *Traction-shift test* présenté uniquement par Choa,2014 dans son étude en cas-témoin.

La facilité et la pertinence de réalisation donnent par ailleurs avantage au *Grind Test* sur les *Seesaw test* et *MP extension test*. La sélection se porte donc sur les *Grind test*, *Palpation test*, *Lever test*, *Extension stress test* et *Adduction stress test*.

5. LIMITES DE L'ETUDE

Cette méta-analyse comporte des limites qu'il est important de soulever pour tempérer les résultats fournis. Tout d'abord, pour le *Grind test* pondéré, une correction des cases nulles des tableaux 2x2 a été faite afin de pouvoir réaliser les calculs souhaités sur R, en dehors des recommandations de la *Cochrane Library* pour la méthode. Cette étape, nommée *correction control* dans le *package mada*, est justifiée par le livre utilisateur disponible à l'adresse : <https://cran.r-project.org/web/packages/mada/vignettes/mada.pdf>. Elle a été réalisée avec l'option « *single* » et non « *all* », et évite ainsi de générer un biais supplémentaire en remplaçant uniquement 0 par 0,5 aux cellules concernées plutôt que d'appliquer +0,5 à l'ensemble des cellules. En pratique, les échelles de sensibilité et spécificité ont donc été très légèrement modifiées pour 2 tableaux sur les 9 présents pour le *Grind test*, mais ne dégradent pas l'ensemble des calculs de manière franche.

D'autre part, des limites philosophiques peuvent être soulevées, puisque la littérature récente tend parfois à remettre en cause les *clusters* de tests de certaines régions du corps et discutent le fait de prendre une décision clinique sur des statistiques (64). De plus, d'un point de vue méthodologique, les progrès constants en statistiques tendent à éliminer la méthode bivariée sélectionnée au profit de la méthode bayésienne, cependant beaucoup plus exigeante au niveau des connaissances théoriques et pour l'application pratique (58,65). Sur le plan statistique, une limite doit être mentionnée sur la pertinence de réaliser une méta-analyse à partir de seulement deux tableaux 2x2. Il existe donc un manque net de puissance statistique pour certains tests.

Au niveau des études réalisées, la question du biais de recrutement est importante. D'une part, la prévalence de la rhizarthrose est extrêmement forte dans les différentes études. D'autre part, les populations issues des études sont en partie déjà fortement suspectées d'être atteintes de la pathologie recherchée et sont dans une démarche de consultation, ainsi la démarche philosophique de mener ces tests en cabinet de kinésithérapie sur une population élargie est différente et doit requérir une expertise accrue. A la lumière des statistiques de la

méta-analyse, un besoin d'approfondir également certains tests « nouveaux » comme les *Adduction stress test* et *Extension stress test* serait nécessaire pour augmenter l'homogénéité inter-études, avec un recrutement de population significatif et éventuellement davantage d'études comparant ces tests. Conjointement à cela, certains axes d'amélioration méthodologique seraient fondamentaux : une rigueur évaluative standardisée serait nécessaire, ainsi qu'une uniformisation de la référence standard utilisée.

De manière intrinsèque, la prise de décision en tant qu'examineur unique de cette méta-analyse est à risque de générer des erreurs à chaque étape ; elle nécessite donc une rigueur extrême, et dans l'idéal d'être en équipe de chercheurs et confronter chaque étape afin d'éliminer l'erreur humaine (66).

6. PERSPECTIVES DE RECHERCHE

Ce travail de recherche préliminaire, faisant l'état des lieux d'une filière spécifique de la pratique clinique, pourrait avoir permis de construire les bases de recherches futures. En effet, la finalité pratique de ce *cluster* de tests théorique est sa mise en place pratique, auprès des patients. Pour aller plus loin, de grandes étapes resteraient donc à franchir, comme la mise en application réelle du *cluster*, la récolte des données épidémiologiques complètes permettant de pouvoir étudier la légitimité de la proposition face à la population et aux différents professionnels de santé, la création de statistiques dont l'efficacité globale et l'étude du *cut-off* ou seuil clinique dans le groupe de tests, l'évaluation de la fiabilité inter-opérateurs, ou encore l'évaluation générale de la pertinence statistique et pratique de la proposition.

Ce modèle s'inscrit plus globalement dans un circuit de raisonnement et de triage, et ne peut fonctionner qu'avec une formation continue adaptée et solide, ainsi qu'une volonté des professionnels de pratiquer sur le terrain. Ces perspectives de recherche s'inscrivent donc dans une logique d'amélioration des compétences et d'assurance dans la pratique professionnelle. En effet, l'apprentissage des *Red Flags* et autres mécanismes de réflexion et réorientation font partie intégrante de la formation initiale en kinésithérapie depuis l'obtention du grade Master, mais représentent un socle de connaissances qu'il est important de perpétuellement conforter et approfondir.

7. CONCLUSION

La prise en charge communément admise dans le cadre du repérage de la rhizarthrose laisse place à des pans entiers accessibles au kinésithérapeute clinicien. C'est en ce sens que

notre travail s'est tourné, dans un but de prévention et dépistage au contact d'une population dont le vieillissement augmente, avec une pénurie médicale grandissante, face à une pathologie sous-diagnostiquée et potentiellement génératrice de limitations fonctionnelles importantes. Cette proposition de travail de recherche fait sens pour améliorer les conditions de prise en charge du patient en kinésithérapie, pour aider le kinésithérapeute dans son raisonnement clinique et pour une réorientation médicale éventuelle. Notre interprétation de l'état actuel de la recherche scientifique permet de dégager un *cluster* théorique de 5 tests cliniques pertinents, relativement significatifs statistiquement et facilement reproductibles que sont les *Grind test*, *Palpation test*, *Lever test*, *Extension stress test* et *Adduction stress test*.

8. FINANCEMENT

Cette étude n'a bénéficié d'aucun financement spécifique et n'a reçu aucun soutien financier de la part d'organismes publics, commerciaux ou à but non lucratif.

9. DECLARATION DE CONFLIT D'INTERET

L'auteur déclare l'absence de tout conflit d'intérêt financier en lien avec cette recherche.

10. BIBLIOGRAPHIE

1. Roux C-H. Traitement médical de l'arthrose. EMC - Appareil locomoteur. janv 2017;12(1):1-12. [Article 14-003-C-40].
2. Runge M. Arthrose du coude, du poignet et de la main. EMC - Radiologie et imagerie médicale - Musculosquelettique - Neurologique - Maxillofaciale. janv 2010;5(4):1-12. [Article 31-312-B-10].
3. Dufour M. Anatomie de l'appareil locomoteur. Tome 2. Membre supérieur. 3ème édition. Issy-Les-Moulineaux: Elsevier Masson; 2016. 536 p.
4. Dufour M, Pillu M. Biomécanique fonctionnelle. Membres-Tête-Tronc. Issy-Les-Moulineaux: Elsevier Masson; 2011. 563 p.
5. Kapandji AI. Physiologie articulaire. Schémas commentés de mécanique humaine. 6ème édition. Paris: Maloine; 2005. 351 p.
6. Cooney WP, Chao EY. Biomechanical analysis of static forces in the thumb during hand function. J Bone Joint Surg Am. janv 1977;59(1):27-36.
7. Boutan M, Baladron R, Rezzouk J. Evolution du protocole de rééducation post-opératoire de rhizarthrose. Les feuillets du GEMMSOR. [Internet]. [cité 22 sept 2025]. Disponible sur: <https://www.sfrm-gemmsor.fr/file/medtool/webmedtool/gemmttool01/botm0053/pdf00012.pdf>
8. Chevalier X, Compaore C, Eymard F, Flipo RM. Rhizarthrose. Revue du Rhumatisme Monographies. avr 2012;79(2):110-5.
9. Cushnaghan J, Dieppe P. Study of 500 patients with limb joint osteoarthritis. I. Analysis by age, sex, and distribution of symptomatic joint sites. Ann Rheum Dis. janv 1991;50(1):8-13.
10. Zhang Y, Niu J, Kelly-Hayes M, Chaisson CE, Aliabadi P, Felson DT. Prevalence of Symptomatic Hand Osteoarthritis and Its Impact on Functional Status among the Elderly: The Framingham Study. American Journal of Epidemiology. déc 2002;156(11):1021-7.
11. INSERM. Arthrose. La maladie articulaire la plus répandue. En ligne [Internet]. 14 nov 2022 [cité 1 oct 2025]; Disponible sur: <https://www.inserm.fr/dossier/arthrose/>
12. Kloppenburg M, Van Beest S, Kroon FPB. Thumb base osteoarthritis: A hand osteoarthritis subset requiring a distinct approach. Best Pract Res Clin Rheumatol. oct 2017;31(5):649-60.
13. Zhang W, Doherty M, Leeb BF, Alekseeva L, Arden NK, Bijlsma JW, et al. EULAR evidence-based recommendations for the diagnosis of hand osteoarthritis: report of a task force of ESCISIT. Annals of the Rheumatic Diseases. janv 2009;68(1):8-17.

14. Levadoux M. Arthrose STT. In: Fontaine C, Liverneaux P, Masmajeau E, directeurs. Cours européen de pathologie chirurgicale du membre supérieur et de la main. Montpellier: Sauramps Médical; 2010. p. 511-9.
15. Athlani L, De Almeida YK, Martins A, Seaourt AC, Dap F. La rhizarthrose en 2023. *Revue de Chirurgie Orthopédique et Traumatologique*. oct 2023;109(6, Supplement):S8-17.
16. Kennedy CD, Manske MC, Huang JI. Classifications in Brief: The Eaton-Littler Classification of Thumb Carpometacarpal Joint Arthrosis. *Clin Orthop Relat Res*. déc 2016;474(12):2729-33.
17. Batra S, Kanvinde R. Osteoarthritis of the thumb trapeziometacarpal joint. *Current Orthopaedics*. avr 2007;21(2):135-44.
18. Dufour M, Tixa S, Del Valle Acedo S. Démarche clinique et diagnostic en kinésithérapie. Issy-Les-Moulineaux: Elsevier Masson; 2018. 386 p.
19. Forman TA, Forman SK, Rose NE. A Clinical Approach to Diagnosing Wrist Pain. *American Family Physician*. nov 2005;72(9):1753-8.
20. Sauvé PS, Rhee PC, Shin AY, Lindau T. Examination of the Wrist: Radial-Sided Wrist Pain. *The Journal of Hand Surgery*. oct 2014;39(10):2089-92.
21. Colonna S, Borghi C. Rhizarthrosis Part I: A Literature Review. *Cureus*. janv 2024;16(1):1-23.
22. Normand M, Tang TS, Brismée JM, Sobczak S. Clinical evaluation of thumb base osteoarthritis: A scoping review. *Hand Ther*. juin 2021;26(2):63-78.
23. Hemmati S, Ponich B, Lafreniere AS, Genereux O, Rankin B, Elzinga K. Approach to chronic wrist pain in adults: Review of common pathologies for primary care practitioners. *Can Fam Physician*. janv 2024;70(1):16-23.
24. Glickel SZ. Clinical assessment of the thumb trapeziometacarpal joint. *Hand Clin*. mai 2001;17(2):185-95.
25. Delacour H, Mouline S, Ceppa F. Mesure des performances et méthode de valorisation des tests biologiques. *EMC - Biologie médicale* 2013;8(1):1-8 [Article 90-60-0207-A].
26. Piette P. Clinimétrie appliquée à la kinésithérapie : mesures, tests et bilans, concepts fondamentaux. *EMC - Kinésithérapie-Médecine physique-Réadaptation* 2022;35(2):1-18 [Article 26-007-A-40].
27. Wainner RS, Fritz JM, Irrgang JJ, Boninger ML, Delitto A, Allison S. Reliability and Diagnostic Accuracy of the Clinical Examination and Patient Self-Report Measures for Cervical Radiculopathy: *Spine*. janv 2003;28(1):52-62.
28. Desjardins T, Kerambrun JM, Chauvet G. Triage, accès direct et raisonnement bayésien en kinésithérapie. *Kinésithérapie, la Revue*. janv 2024;24(265):16-22.

29. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. Statistical Methods of Meta-Analysis. Korean J Radiol. déc 2015;16(6):1188-96.
30. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part I. General Guidance and Tips. Korean Journal of Radiology. déc 2015;16(6):1175-87.
31. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated july 2023). [Internet]. 2023. Disponible sur: <https://training.cochrane.org/handbook-diagnostic-test-accuracy/current>
32. Beaumier M, Deshêtres A, Despatis MA, Bronchti G, Laurencelle L. Traduction en langue française de la grille STARD 2015 : Standards for Reporting of Diagnostic Accuracy Studies. Revue Francophone de Cicatrisation. oct 2019;3(4):36-40.
33. Gedda M. Traduction française des lignes directrices STARD pour l'écriture et la lecture des études sur la précision des tests diagnostiques. Kinésithérapie, la Revue. janv 2015;15(157):45-9.
34. Rostagno S, Pallot A. QUADAS-2 : traduction française de l'échelle de qualité méthodologique pour les études diagnostiques de validité. Kinésithérapie, la Revue. juill 2021;21(235):10-2.
35. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology. oct 2005;58(10):982-90.
36. Arends LR, Hamza TH, Van Houwelingen JC, Heijenbrok-Kal MH, Hunink MGM, Stijnen T. Bivariate Random Effects Meta-Analysis of ROC Curves. Med Decis Making. sept 2008;28(5):621-38.
37. Çorbacioğlu ŞK, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. Turkish Journal of Emergency Medicine. oct 2023;23(4):195-8.
38. Holling H, Böhning W, Masoudi E, Böhning D, Sangnawakij P. Evaluation of a new version of I^2 with emphasis on diagnostic problems. Communications in Statistics - Simulation and Computation. avr 2020;49(4):942-72.
39. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of Clinical Epidemiology. sept 2005;58(9):882-93.
40. Arnold DMJ, Gottlieb R, Wilkens SC, Garg R, Nazzal A, Chen NC, et al. Evaluation of the Seesaw Test as New Diagnostic Test for Thumb CMC Osteoarthritis. Hand (N Y). juill 2022;17(4):706-13.

41. Choa RM, Parvizi N, Giele HP. A prospective case-control study to compare the sensitivity and specificity of the grind and traction-shift (subluxation-relocation) clinical tests in osteoarthritis of the thumb carpometacarpal joint. *J Hand Surg Eur* Vol. mars 2014;39(3):282-5.
42. Gelberman RH, Boone S, Osei DA, Cherney S, Calfee RP. Trapeziometacarpal Arthritis: A Prospective Clinical Evaluation of the Thumb Adduction and Extension Provocative Tests. *J Hand Surg Am*. juill 2015;40(7):1285-91.
43. Merritt MM, Roddey TS, Costello C, Olson S. Diagnostic value of clinical grind test for carpometacarpal osteoarthritis of the thumb. *J Hand Ther*. 2010;23(3):261-7.
44. Model Z, Liu AY, Kang L, Wolfe SW, Burket JC, Lee SK. Evaluation of Physical Examination Tests for Thumb Basal Joint Osteoarthritis. *Hand (N Y)*. mars 2016;11(1):108-12.
45. Ribak S. Rhizarthrosis: diagnostic assessment of rhizarthrosis and clinical-radiological correlation. *International Physical Medicine & Rehabilitation Journal*. juill 2020;5(4):158-66.
46. Sela Y, Seftchick J, Wang WL, Baratz ME. The diagnostic clinical value of thumb metacarpal grind, pressure-shear, flexion, and extension tests for carpometacarpal osteoarthritis. *J Hand Ther*. 2019;32(1):35-40.
47. Sheskin DJ. *Handbook of Parametric and Nonparametric Statistical Procedures: Third Edition*. New York, NY: Chapman and Hall/CRC; 2003. 1193 p.
48. Reitsma JB, Rutjes A, Whiting P, Yang B, Leeflang MM, Bossuyt PM, Deeks JJ. Chapter 8: Assessing risk of bias and applicability. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 2.0 (updated July 2023). Cochrane, 2023. Disponible sur: <https://training.cochrane.org/handbook-diagnostic-test-accuracy/current>
49. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol*. mars 2018;63(7):07TR01.
50. Deeks JJ, Takwoingi Y, Macaskill P, Bossuyt PM. Chapter 4: Understanding test accuracy measures. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 2.0 (updated July 2023). Cochrane, 2023. Disponible sur: Available from <https://training.cochrane.org/handbook-diagnostic-test-accuracy/current>
51. He Y, Krastman P, Bierma-Zeinstra SMA, Kraan G, Mathijssen NM, Runhaar J. Diagnostic accuracy of history taking, physical examination, and auxiliary examination for thumb osteoarthritis: a systematic review. *Annals of Medicine*. 31 déc 2025;57(1):2524086.

52. McGee C, Valdes K, Bakker C, Ivy C. Psychometric properties of body structures and functions measures in non-surgical thumb carpometacarpal osteoarthritis: A systematic review. *J Hand Ther.* 2024;37(1):22-37.
53. Kafury-Goeta AA, Novillo ME, Angarita-Maldonado F, Castañeda JF, Rojas-Neira JA, Muñoz SP. Validation of a New Clinical Test for Diagnosis of Trapeziometacarpal Arthritis. *J Wrist Surg [Internet].* févr 2025; Disponible sur: <http://www.thieme-connect.de/products/ejournals/abstract/10.1055/a-2498-0718>
54. Kafury-Goeta AA, Angarita-Maldonado FA, Rojas-Neira JA, Gil-Henao GA, Minaya-Gomez ED. Description of thumb compression and flexion test for the diagnosis of rhizarthrosis. Cross sectional Cohort Study. *Rev Colomb Ortop Traumatol.* 2020;34(4):391-5.
55. Lin, JD, Karl JW, Strauch RJ. Trapeziometacarpal joint stability: the evolving importance of the dorsal ligaments. *Clin Orthop Relat Res.* avr 2014;472(4):1138-45.
56. North ER, Eaton RG. Degenerative joint disease of the trapezium: A comparative radiographic and anatomic study. *The Journal of Hand Surgery.* mars 1983;8(2):160-7.
57. Wade S. 2023 Bayesian cluster analysis. *Phil Trans R Soc A.* mai 2023;381(2247):20220149.
58. Takwoingi Y, Dendukuri N, Schiller I, Rücker G, Jones HE, Partlett C, Macaskill P. Chapter 10: Undertaking meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.* Version 2.0 (updated July 2023). Cochrane, 2023. Disponible sur: Available from <https://training.cochrane.org/handbook-diagnostic-test-accuracy/current>
59. Panchout E. Capacité de triage par raisonnement clinique en accès direct : apport de la didactique professionnelle. *Kinésithérapie, la Revue.* mars 2023;23(255):41-2.
60. Haugen IK, Felson DT, Abhishek A, Berenbaum F, Bierma-Zeinstra S, Dziedzic KS, et al. 2023 EULAR classification criteria for hand osteoarthritis. *Ann Rheum Dis.* oct 2024;83(11):1428-35.
61. Michener LA, Walsworth MK, Doukas WC, Murphy KP. Reliability and Diagnostic Accuracy of 5 Physical Examination Tests and Combination of Tests for Subacromial Impingement. *Archives of Physical Medicine and Rehabilitation.* nov 2009;90(11):1898-903.
62. Safari S, Baratloo A, Elfil M, Negida A. Evidence Based Emergency Medicine; Part 4: Pre-test and Post-test Probabilities and Fagan's nomogram. *Emerg (Tehran).* 2016;4(1):48-51.
63. O'Rourke C, Welaratne I, Cournane S, McLoughlin LC, Reynolds JV, Johnston C, et al. Diagnostic accuracy of SUVmax in predicting malignancy of supraclavicular

lymph nodes from primary oesophageal cancer. *European Journal of Radiology*. avr 2020;125:108860.

64. Saueressig T, Owen PJ, Diemer F, Zebisch J, Belavy DL. Diagnostic Accuracy of Clusters of Pain Provocation Tests for Detecting Sacroiliac Joint Pain: Systematic Review With Meta-analysis. *Journal of Orthopaedic & Sports Physical Therapy*. sept 2021;51(9):422-31.
65. Menten J, Lesaffre E. A general framework for comparative Bayesian meta-analysis of diagnostic studies. *BMC Med Res Methodol*. déc 2015;15(1):70.
66. Dinnes J, Deeks JJ, Leeflang MM, Li T. Chapter 7: Collecting data. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 2.0 (updated July 2023). Cochrane, 2023. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 2.0 (updated July 2023). Cochrane, 2023. Disponible sur: Available from <https://training.cochrane.org/handbook-diagnostic-test-accuracy/current>

ANNEXES

Annexe I : équation de recherche dans les bases de données bibliographiques

PubMed	((test[Title/Abstract]) OR (assessment[Title/Abstract]) OR (evaluation[Title/Abstract]) OR (examination[Title/Abstract])) AND ((sensitivity[Title/Abstract]) OR (specificity[Title/Abstract]) OR (diagnosis[Title/Abstract]) OR (accuracy[Title/Abstract])) AND ((thumb[Title/Abstract]) OR (basal joint[Title/Abstract]) OR (trapeziometacarpal joint[Title/Abstract]) OR (carpometacarpal joint[Title/Abstract])) AND ((osteoarthritis[Title/Abstract]) OR (arthritis[Title/Abstract]) OR (rhizarthrosis[Title/Abstract]))
Embase	#5. #1 AND #2 AND #3 AND #4 #4. 'osteoarthritis':ti,ab,kw OR 'arthritis':ti,ab,kw OR 'rhizarthrosis':ti,ab,kw #3. 'thumb':ti,ab,kw OR 'trapeziometacarpal joint':ti,ab,kw OR 'carpometacarpal joint':ti,ab,kw #2. sensitivity:ti,ab,kw OR specificity:ti,ab,kw OR 'accuracy':ti,ab,kw OR 'diagnosis':ti,ab,kw #1. 'test':ti,ab,kw OR 'assessment':ti,ab,kw OR 'evaluation study':ti,ab,kw OR 'examination':ti,ab,kw
LILACS	(test) OR (assessment) OR (evaluation) OR (examination) AND (sensitivity) OR (specificity) OR (diagnosis) OR (accuracy) AND (thumb) OR (basal joint) OR (trapeziometacarpal joint) OR (carpometacarpal joint) AND (osteoarthritis) OR (arthritis) OR (rhizarthrosis) AND instance:"regional"
PEDro	Abstract & title: ("arthritis" OR "osteoarthritis") AND ("assessment" OR "evaluation") Body part: « hand or wrist »

Annexe II : ligne directrice d'écriture STARD 2015 en Français

Guide méthodologique de lecture et d'écriture STARD 2015			
Item de la checklist	N°	Attendu théorique	Retrouvé à l'analyse
Titre ou résumé	1	Identification en tant qu'étude de l'exactitude d'un test diagnostique avec au moins une mesure d'exactitude (la sensibilité, la spécificité, les paramètres prédicteurs, l'aire ROC).	
Résumé	2	Résumé structuré du devis de l'étude, sa méthodologie, ses résultats et ses conclusions (voir « STARD for abstracts » pour informations supplémentaires).	
Introduction	3	Mise en contexte scientifique et clinique, incluant les utilisations prévues et le rôle clinique du test diagnostique.	
	4	Objectifs et hypothèses de l'étude.	
Méthodologie	Devis de l'étude	5	La collecte des données a-t-elle été planifiée avant la mise à l'épreuve du test diagnostique avec le test de référence (étude prospective) ou après la mise à l'épreuve (étude rétrospective) ?
		6	Critères d'inclusion.
	Participants	7	Sur quels critères de base ont été identifiés/recrutés les participants (p. ex. symptômes, résultats d'examens précédents, consultation d'un registre de patients).
		8	Où et quand les participants éligibles ont été identifiés (lieu, contexte et dates).
		9	Les participants ont-ils été recrutés suivant un échantillonnage probabiliste (aléatoire) ou accidentel (dit "de convenance") ?
		10a	Le test diagnostique est suffisamment documenté/détaillé pour assurer la reproductibilité.
	Méthodes et techniques de tests	10b	Le test de référence est suffisamment documenté/détaillé pour assurer la reproductibilité.
		11	La logique ou la raison sous-jacente au choix du test de référence est énoncé (notamment s'il existe d'autres tests possibles).
		12a	La définition et le rationnel sous-jacent pour la positivité du test diagnostique concernant ses valeurs de point de coupure (ou valeurs seuils) ou les résultats catégoriels (ou de hiérarchisation) sont distinguées qu'elles soient pré-spécifiées ou exploratoires.
		12b	La définition et le rationnel sous-jacent pour la positivité du test de référence pour ses valeurs de point de coupure (ou valeurs seuils) ou les résultats catégoriels (ou de hiérarchisation) sont distinguées qu'elles soient pré-spécifiées ou exploratoires.
		13a	Les données cliniques et les résultats du test de référence étaient disponibles aux examinateurs/lecteurs du test diagnostique.
		13b	Les données cliniques et les résultats du test diagnostique étaient disponibles aux examinateurs/lecteurs du test de référence.
	Analyse	14	Les méthodes utilisées pour estimer ou comparer l'exactitude diagnostique.
		15	Comment ont été traités les résultats ambigus du test diagnostique et du test de référence.
		16	Comment ont été traitées les données manquantes du test diagnostique et du test de référence.
		17	La variabilité des données diagnostiques est considérée, et analysée, en distinguant les résultats découlant des seuils pré-spécifiées des seuils exploratoires.
		18	Taille prévue de l'échantillon et la méthode par laquelle elle a été fixée.
Résultats	Participants	19	Répartition des participants illustrée dans un diagramme.
		20	Données démographiques et cliniques des participants
		21a	Les degrés de sévérité sont répartis (distribués) chez les participants ayant la condition ciblée.
		21b	Les diagnostics alternatifs sont répartis parmi les participants n'ayant pas la condition ciblée.
		22	Intervalle de temps et liste des interventions cliniques entre le temps du test diagnostique et du test de référence.
	Résultats des tests	23	Tableau comparatif (tableau croisé) des résultats du test diagnostique et du test de référence.
		24	Estimation de l'exactitude et de la précision diagnostique (p. ex. intervalle de confiance à 95 %).
Discussion		25	Effets secondaires indésirables imputables à la réalisation des test diagnostique ou du test de référence.
		26	Limites de l'étude, incluant biais potentiels, incertitude statistique et généralisabilité.
Autres informations		27	Implications pour la pratique, incluant l'utilisation prévue et le rôle clinique du test diagnostique.
		28	Numéro d'enregistrement et nom du registre.
		29	Référence pour récupérer le protocole complet de l'étude.
		30	Sources de financement et autres soutiens logistiques ; rôle des donateurs.

Annexe III : échelle méthodologique QUADAS-C traduite en Français

QUADAS-C : risque de biais dans les études comparatives de précision diagnostique

Question de l'étude comparative

Les patients :	
Test d'indexation A :	
Test d'indexation B :	
Norme de référence et condition cible :	

Ajouter des lignes pour des tests d'index supplémentaires si nécessaire

Conception d'une étude comparative

Auquel des modèles d'étude suivants l'étude primaire ressemble-t-elle le plus ?	<i>#N°1 Entièrement jumelé</i> <i>#N°2 Randomisé</i> <i>#N°3 Partiellement apparié avec un sous-ensemble aléatoire</i> <i>#N°4 Partiellement apparié avec un sous-ensemble non aléatoire</i> <i>#N°5 Non apparié, non randomisé</i> <i>Autre (veuillez décrire la conception de l'étude) :</i>
---	---

Le [document d'orientation QUADAS-C](#) contient des exemples de diagrammes de flux pour chaque conception.

Organigramme

Dessinez un organigramme pour l'étude primaire

Domaine 1 : Sélection des patients			
Informations à l'appui du jugement	<i>Décrire les méthodes de sélection des patients. Décrire les patients inclus (tests antérieurs, présentation, utilisation prévue du test de référence et contexte). Décrivez comment les patients ont été répartis pour recevoir chacun des tests d'index. Si la randomisation a été utilisée pour affecter des patients individuels (ou des groupes de patients) aux tests de référence, décrivez le processus de randomisation.</i>		
Précision du test unique (QUADAS-2)		Réponses pour _____ (test A)*	Réponses pour _____ (test B)*
Questions relatives à la signalisation	1.1 Un échantillon consécutif ou aléatoire de patients a-t-il été recruté ?	Oui/Non/Non clair	Oui/Non/Non clair
	1.2 A-t-on évité une conception cas-témoins ?	Oui/Non/Non clair	Oui/Non/Non clair
	1.3 L'étude a-t-elle évité les exclusions inappropriées ?	Oui/Non/Non clair	Oui/Non/Non clair
Risque de biais	1.4 La sélection des patients a-t-elle pu introduire des biais ?	Faible/élevé/non clair	Faible/élevé/non clair
Préoccupations concernant l'applicabilité	1.5 Y a-t-il des inquiétudes quant au fait que les patients inclus ne correspondent pas à la question de l'étude ?	Faible/élevé/non clair	Faible/élevé/non clair
Précision comparative (QUADAS-C)		Réponses pour le comparaison des tests	
Questions relatives à la signalisation	C1.1 Le risque de biais pour chaque test index a-t-il été jugé "faible" pour ce domaine ?	Oui/Non	
	C1.2 A-t-on utilisé un modèle entièrement apparié ou randomisé ?	Oui/Non/Non clair	
	C1.3 La séquence d'attribution était-elle aléatoire?†	Oui/Non/Non clair/ Sans objet	
	C1.4 La séquence d'allocation a-t-elle été dissimulée jusqu'à ce que les patients aient été recrutés et assignés aux tests d'indexation?†	Oui/Non/Non clair/ Sans objet	
Risque de biais	C1.5 La sélection des patients a-t-elle pu introduire un biais dans la comparaison ?	Faible/élevé/non clair	

* Exemple lorsque la comparaison porte sur deux tests d'index. Des colonnes supplémentaires peuvent être ajoutées pour chaque essai supplémentaire dans la comparaison.

† Uniquement applicable aux plans de sondage aléatoires

Voir le [document d'orientation QUADAS-C](#) pour des explications plus détaillées.

C1.1 : Répondre "oui" si le risque de biais pour la précision d'un seul test (question 1.4 de QUADAS-2) est "faible" pour chaque test index.

C1.2 : Répondez "oui" si l'une des méthodes suivantes a été utilisée pour répartir les patients entre les tests d'index : (1) chaque patient reçoit tous les tests d'index (modèle entièrement apparié) ou (2) les patients sont répartis de manière aléatoire entre les tests d'index (modèle randomisé).

C1.3 : Répondez "oui" si l'étude a généré une séquence d'allocation réellement aléatoire, par exemple des nombres aléatoires générés par ordinateur et des tables de nombres aléatoires.

C1.4 : Répondez "oui" si l'étude a utilisé des méthodes appropriées pour dissimuler l'allocation, telles que des systèmes de randomisation centralisés et des enveloppes scellées opaques.

C1.5 : Le risque de biais peut être jugé "faible" si les réponses aux questions C1.1 à C1.4 sont "oui" (les questions C1.3 et C1.4 ne s'appliquent qu'aux modèles randomisés). S'il a été répondu "non" à au moins une question, les utilisateurs doivent envisager un jugement de "risque élevé de biais" si le biais associé à la caractéristique de conception est si préoccupant que l'ensemble du domaine est considéré comme problématique. S'il a été répondu "non" à la question C1.2, il convient d'envisager fortement un "risque élevé de biais".

Domaine 2 : Test d'indexation			
Informations à l'appui du jugement	<i>Décrire les tests d'indexation et la manière dont ils ont été menés et interprétés. Pour les études comparatives par paires, décrire l'ordre dans lequel les tests d'index ont été effectués.</i>		
Précision du test unique (QUADAS-2)		Réponses pour _____ (test A)	Réponses pour _____ (test B)
Questions relatives à la signalisation	2.1 Les résultats du test d'index ont-ils été interprétés sans connaître les résultats de l'étalon de référence ?	Oui/Non/Non clair	Oui/Non/Non clair
	2.2 Si un seuil a été utilisé, a-t-il été spécifié à l'avance ?	Oui/Non/Non clair	Oui/Non/Non clair
Risque de biais	2.3 La conduite ou l'interprétation du test d'indexation pourrait-elle avoir introduit un biais ?	Faible/élevé/non clair	Faible/élevé/non clair
Préoccupations concernant l'applicabilité	2.4 Existe-t-il des inquiétudes quant au fait que le test d'indexation, sa conduite ou son interprétation diffèrent de la question de l'examen ?	Faible/élevé/non clair	Faible/élevé/non clair
Précision comparative (QUADAS-C)		Réponses pour le comparaison des tests	
Questions relatives à la signalisation	C2.1 Le risque de biais pour chaque test index a-t-il été jugé "faible" pour ce domaine ?	Oui/Non	
	C2.2 Les résultats du test d'indexation ont-ils été interprétés sans connaître les résultats de l'autre ou des autres tests d'indexation?‡	Oui/Non/Non clair/ Sans objet	
	C2.3 La réalisation d'un test d'indice <u>ne</u> <u>risque</u> -t-elle <u>pas</u> d'affecter les performances de l'autre ou des autres tests d'indice?‡	Oui/Non/Non clair/ Sans objet	
	C2.4 Les tests d'indexation ont-ils été menés et interprétés sans avantager l'un des tests ?	Oui/Non/Non clair	
Risque de biais	C2.5 La conduite ou l'interprétation des tests d'indexation pourrait-elle avoir introduit un biais dans la comparaison ?	Faible/élevé/non clair	

‡ Uniquement applicable si les patients ont reçu des tests d'indexation multiples (modèles entièrement ou partiellement appariés)

C2.1 : Répondre "oui" si le risque de biais pour la précision d'un seul test (question 2.3 de QUADAS-2) est "faible" pour chaque test index.

C2.2 : Répondre "oui" si le test d'indexation A a été interprété en aveugle par rapport aux résultats du test d'indexation B et vice versa.

L'aveuglement n'est pas nécessaire si aucun des tests d'index n'implique une interprétation subjective.

C2.3 : Répondre "oui" si un test d'indexation ne peut pas influencer ou interférer avec les résultats d'un ou de plusieurs tests d'indexation effectués ultérieurement. Parmi les exemples d'influence ou d'interférence, on peut citer la distorsion de la zone d'échantillonnage (biopsies) et la fatigue du patient (questionnaires).

C2.4 : Répondez "oui" s'il n'y a pas de différences entre les tests d'indexation susceptibles d'avantager injustement l'un des tests. Un exemple d'une telle différence est lorsque le test d'index A a été réalisé par un expert et le test d'index B par un non-expert. Les différences entre les tests qui reflètent la pratique clinique sont acceptables, auquel cas la réponse "oui" est appropriée.

C2.5 : Le risque de biais peut être jugé "faible" si les réponses aux questions de signalisation C2.1 à C2.4 sont "oui" (C2.2 et C2.3 ne s'appliquent qu'aux modèles entièrement ou partiellement appariés). Si au moins une question a reçu une réponse négative, les utilisateurs doivent envisager un jugement de "risque élevé de biais" si le biais associé à la caractéristique de conception est si préoccupant que l'ensemble du domaine est considéré comme problématique.

Domaine 3 : Norme de référence			
Informations à l'appui du jugement	Décrivez la norme de référence, la manière dont elle a été réalisée et interprétée, et indiquez si l'un des tests d'index faisait partie de la norme de référence.		
Précision du test unique (QUADAS-2)		Réponses pour _____ (test A)	Réponses pour _____ (test B)
Questions relatives à la signalisation	3.1 La norme de référence est-elle susceptible de classer correctement l'affection cible ?	Oui/Non/Non clair	Oui/Non/Non clair
	3.2 Les résultats de l'étalon de référence ont-ils été interprétés sans connaître les résultats du test de référence ?	Oui/Non/Non clair	Oui/Non/Non clair
Risque de biais	3.3 La norme de référence, sa conduite ou son interprétation ont-elles pu introduire un biais ?	Faible/élevé/non clair	Faible/élevé/non clair
Préoccupations concernant l'applicabilité	3.4 Peut-on craindre que l'état cible défini par la norme de référence ne corresponde pas à la question de l'examen ?	Faible/élevé/non clair	Faible/élevé/non clair
Précision comparative (QUADAS-C)		Réponses pour le comparaison des tests	
Questions relatives à la signalisation	C3.1 Le risque de biais pour chaque test index a-t-il été jugé "faible" pour ce domaine ?	Oui/Non	
	C3.2 La norme de référence a-t-elle évité d'incorporer l'un ou l'autre des tests d'indexation ?	Oui/Non/Non clair	
Risque de biais	C3.3 La norme de référence, sa conduite ou son interprétation ont-elles pu introduire un biais dans la comparaison ?	Faible/élevé/non clair	

C3.1 : Répondre "oui" si le risque de biais pour la précision d'un seul test (question 3.3 de QUADAS-2) est "faible" pour chaque test index.

C3.2 : Répondre "oui" si aucun des tests d'index ne faisait partie de la norme de référence. Notez que cette question est différente de celle de l'insu (question de signalisation 3.2 dans QUADAS-2).

C3.3 : Le risque de biais peut être jugé "faible" si les réponses aux questions de signalisation C3.1 et C3.2 sont "oui". Si au moins une question a reçu une réponse négative, les utilisateurs doivent envisager un jugement de "risque élevé de biais" si le biais associé à la caractéristique de conception est si préoccupant que l'ensemble du domaine est considéré comme problématique.

Domaine 4 : Flux et timing			
Informations à l'appui du jugement	<i>Décrivez les patients qui n'ont pas reçu les tests de référence ou la norme de référence ou qui ont été exclus de l'analyse.</i> <i>Décrivez l'intervalle de temps et les interventions éventuelles entre les tests de référence et la norme de référence.</i> <i>Décrivez l'intervalle de temps et les interventions éventuelles entre les tests d'indexation comparés.</i>		
Précision du test unique (QUADAS-2)		Réponses pour _____ (test A)	Réponses pour _____ (test B)
Questions relatives à la signalisation	4.1 L'intervalle entre les tests d'index et la norme de référence était-il approprié ?	Oui/Non/Non clair	Oui/Non/Non clair
	4.2 Tous les patients ont-ils reçu un étalon de référence ?	Oui/Non/Non clair	Oui/Non/Non clair
	4.3 Tous les patients ont-ils reçu la même norme de référence ?	Oui/Non/Non clair	Oui/Non/Non clair
	4.4 Tous les patients ont-ils été inclus dans l'analyse ?	Oui/Non/Non clair	Oui/Non/Non clair
Risque de biais	4.5 Le flux de patients a-t-il pu introduire des biais ?	Faible/élevé/non clair	Faible/élevé/non clair
Précision comparative (QUADAS-C)		Réponses pour le comparaison des tests	
Questions relatives à la signalisation	C4.1 Le risque de biais pour chaque test index a-t-il été jugé "faible" pour ce domaine ?	Oui/Non	
	C4.2 L'intervalle entre les tests d'indexation était-il approprié ?	Oui/Non/Non clair	
	C4.3 La même norme de référence a-t-elle été utilisée pour tous les tests d'indexation ?	Oui/Non/Non clair	
	C4.4 Les proportions et les raisons des données manquantes sont-elles similaires pour tous les tests d'indexation ?	Oui/Non/Non clair	
Risque de biais	C4.5 Le flux de patients a-t-il pu introduire un biais dans la comparaison ?	Faible/élevé/non clair	

C4.1 : Répondre "oui" si le risque de biais pour la précision d'un seul test (question 4.5 de QUADAS-2) est "faible" pour chaque test index.

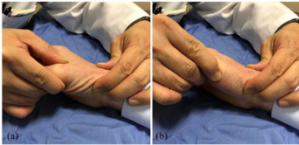
C4.2 : Pour de nombreux tests d'index, le terme "approprié" signifie que les tests doivent être effectués au même moment après l'inscription du patient. Cela exclut la possibilité d'une progression de la maladie ou d'un changement dans la prise en charge du patient. Certains tests d'index ont des "fenêtres diagnostiques" différentes et sont idéalement réalisés à des moments différents ; une expertise en la matière est nécessaire pour le déterminer.

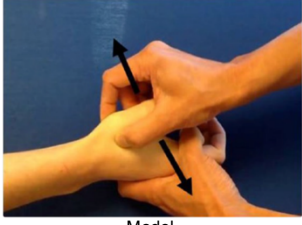
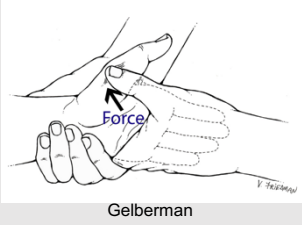
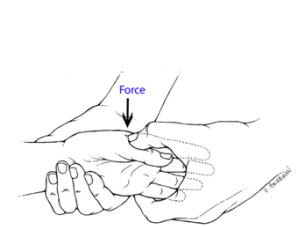
C4.3 : Répondez "oui" si (1) une seule norme de référence a été utilisée pour tous les patients ou (2) plusieurs normes de référence ont été utilisées (par exemple, chirurgie ou suivi) et que ces normes de référence étaient les mêmes pour les patients recevant le test index A et les patients recevant le test index B.

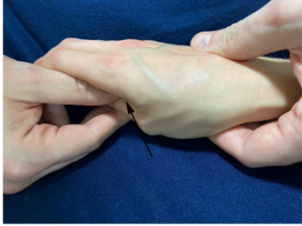
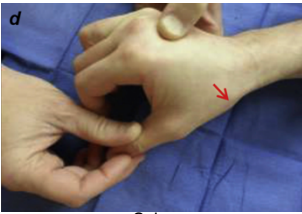
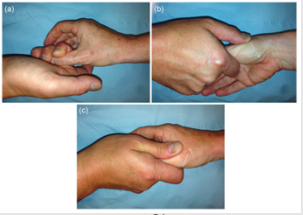
C4.4 : Il y a des données manquantes si les résultats des tests ne sont pas disponibles, invalides ou non concluants, ou si les patients sont exclus de l'analyse. Répondez "oui" s'il n'y a pas de données manquantes ou si la proportion et les raisons des données manquantes sont similaires pour le test index A et le test index B.

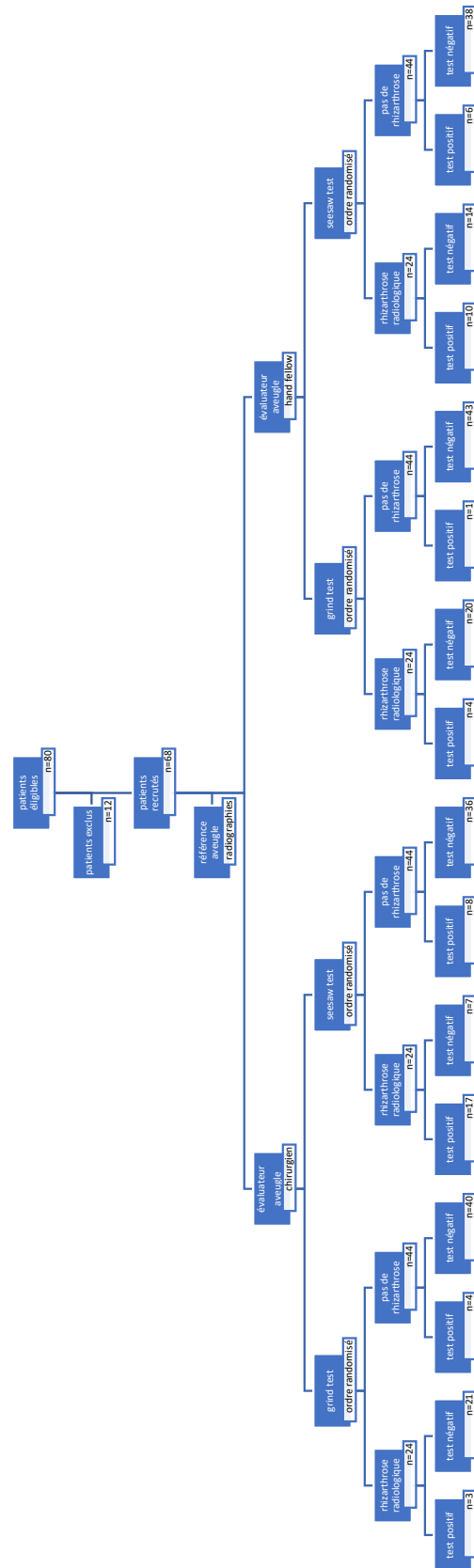
C4.5 : Le risque de partialité peut être jugé "faible" si les réponses aux questions de signalisation C4.1 à C4.4 sont "oui". S'il a été répondu "non" à au moins une question, l'utilisateur doit envisager un jugement de "risque élevé de biais" si le biais associé à la caractéristique de conception est si préoccupant que l'ensemble du domaine est considéré comme problématique.

Annexe IV : tableau détaillé des tests cliniques identifiés

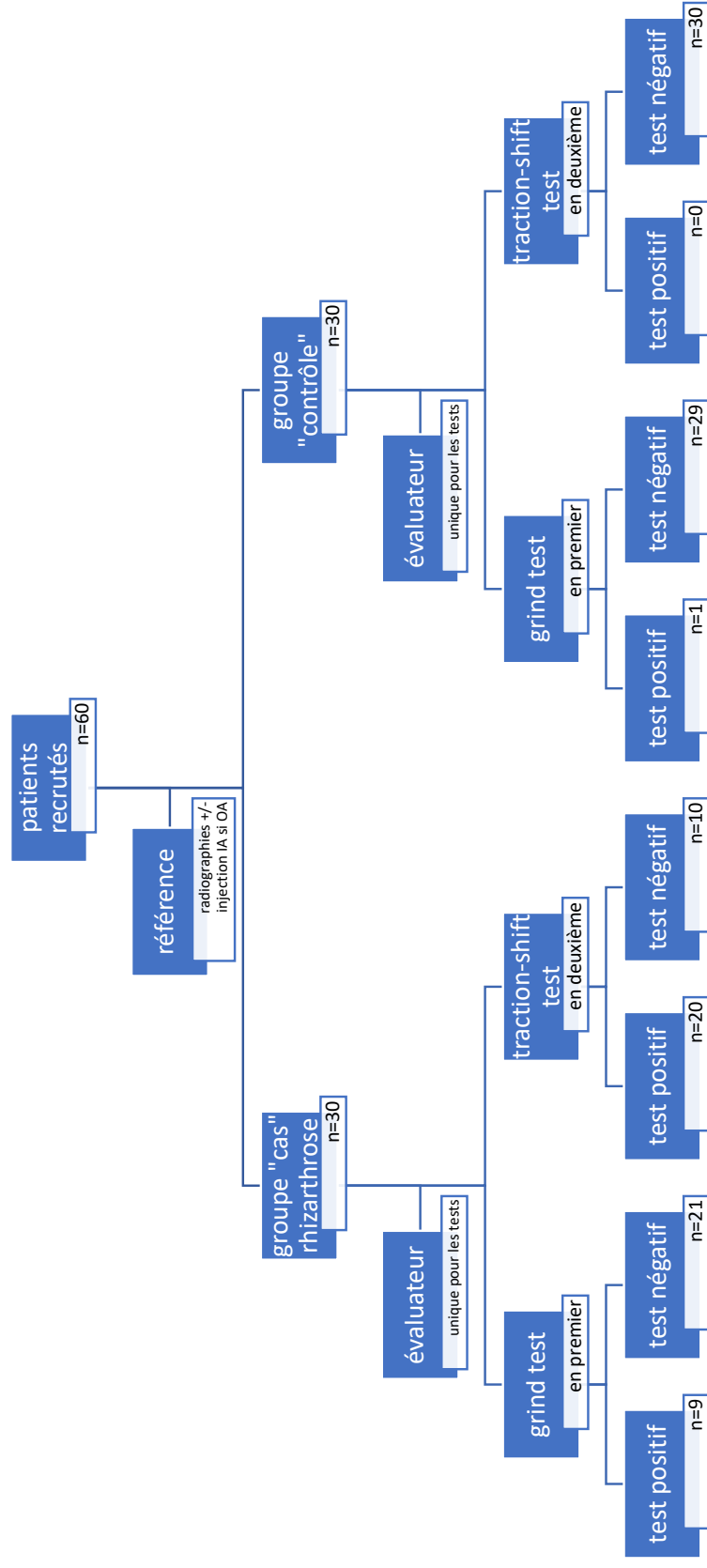
Test clinique	Citation	Commentaires	Illustration/photographie	Description globale
Grind test	Arnold, 2022 Choa, 2014 Gelberman, 2015 Model, 2016 Merritt, 2010 Ribak, 2020 Sela, 2019	Traduction : test de broyage articulaire	 Arnold	Le grind test consiste à saisir le métacarpien du pouce tout en stabilisant le trapèze, et à exercer une rotation combinée à une compression axiale de l'articulation. Le test est considéré positif si une provocation de douleur est retrouvée pendant la manœuvre. Le grincement ou le bruit ressenti sont également à décrire. NB : technique globalement homogène selon les auteurs (contre-preise analytique au trapèze pas toujours décrite ni réalisée).
			  Choa	
			 Merritt	
			 Model	
			 Ribak	
			 Sela	

Digit pressure test / Joint tenderness test / Palpation test	Gelberman, 2015 Model, 2016 Ribak, 2020	Palpation localisée de l'articulation	 Ribak	Une palpation ferme est effectuée au dos de l'articulation trapézo-métacarpienne. Le test est considéré comme positif lorsque le patient signale une douleur dans la région trapézo-métacarpienne. NB : technique homogène selon les auteurs.
MC pressure-shear test / Lever test / Shear test	Model, 2016 Ribak, 2020 Sela, 2019	Traduction : MC pressure-shear test : test de pression-cisaillement ; Lever test : test du levier Même test avec noms différents. NB : test +/- en compression initiale, à différencier du traction-shift.	 Model  Ribak  Sela	L'examineur place son pouce et son index des deux côtés de l'articulation basale du pouce et fait lever sur la première articulation métacarpienne radialement et ulnairement jusqu'à leurs fins de course au niveau de l'articulation basale. NB : technique assez homogène selon les auteurs (analytique trapézo-métacarpien pour Ribak, contre-prise au poignet ou à la main pour les autres).
MC1 extension stress test / Extension stress test	Gelberman, 2015 Ribak, 2020	Test en passif, mouvement défini dans l'axe du pouce	 Gelberman  Ribak	L'examineur est assis face au patient. Le patient place la main affectée sur la table d'examen avec le coude fléchi à 90° et l'avant-bras en rotation neutre. L'examineur place son pouce ipsilatéral le long du bord radial de M1, 5 à 10 mm proximale à l'articulation métacarpophalangienne du pouce, et étend le pouce jusqu'à ce que M1 se trouve dans un plan parallèle à la paume ou jusqu'à ce qu'une fin de course dure soit atteinte. La reproduction de la douleur au niveau de l'articulation TMC est considérée comme un test positif. NB : technique assez homogène selon les auteurs (analytique pour Gelberman, interposition de P1 pour Ribak).
MC1 flexion stress test / Adduction stress test	Gelberman, 2015 Ribak, 2020	Test en passif, mouvement défini dans l'axe du pouce ; Gelberman le nomme Adduction stress test	 Gelberman	L'examineur est assis face au patient. Le patient place la main affectée sur la table d'examen avec le coude fléchi à 90° et l'avant-bras en rotation neutre. L'examineur place sa main ipsilatérale de manière à ce que le pouce de l'examineur repose dorsalement sur la tête de M1. La main controlatérale de l'examineur soutient le côté ulnaire de la main du patient pour maintenir le poignet du patient en position neutre afin d'éviter la déviation ulnaire du poignet du patient. L'examineur dirige fermement une force d'adduction vers le bas sur la tête métacarpienne du patient jusqu'à ce que le M1 du patient soit parallèle à l'axe médian de M2 ou jusqu'à ce qu'une fin de course dure soit atteinte. La reproduction de la douleur au

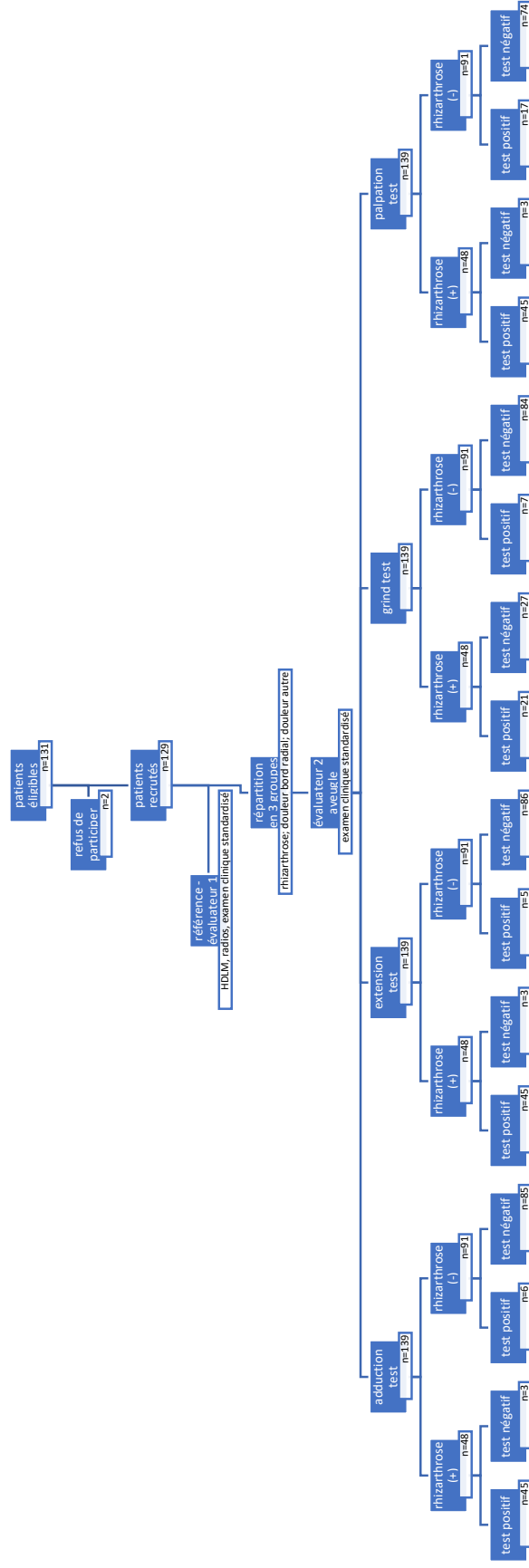
				niveau de l'articulation TMC est considérée comme un test positif. NB : technique assez homogène selon les auteurs (analytique pour Gelberman, interposition de P1 pour Ribak).
Seesaw test	Arnold, 2022	Traduction : test de la balançoire		Le seesaw test commence par la saisie et la stabilisation du métacarpien du pouce, en le réduisant (a) dans l'articulation trapézo-métacarpienne, tout en stabilisant (b) le trapèze avec l'autre main. Avec une pression axiale continue sur le métacarpien pour le maintenir réduit dans l'articulation trapézo-métacarpienne, une extension (c) et une flexion (d) du métacarpe sur le trapèze sont effectuées (tout en évaluant la douleur, le grincement ou le bruit).
MP extension test	Model, 2016	Test en actif		Le patient essaie de tendre le pouce tandis que l'examineur fournit une résistance contre l'extension en plaçant un doigt sur l'articulation IP du pouce.
MC extension test	Sela, 2019	Test en passif, mouvement défini dans l'axe de la main		Le MC extension test se réalise sous la forme d'un enchaînement avec le MC flexion test. Les tests de flexion et d'extension trapézo-métacarpiennes sont effectués avec l'examineur qui saisit la main du participant avec sa main pour obtenir un support sur la table d'examen. Les métacarpes et le poignet du patient sont placés dans une position neutre. L'autre main de l'examineur place M1 en flexion. Après avoir sollicité une réponse pour savoir si la position fléchie crée de la douleur, M1 est ensuite mis en extension maximale. Les tests sont considérés comme positifs si le sujet signale une douleur (même une légère sensation de douleur) dans la région de l'articulation basale du pouce.
MC flexion test	Sela, 2019	Test en passif, mouvement défini dans l'axe de la main		Le MC flexion test se réalise sous la forme d'un enchaînement avec le MC extension test. Les tests de flexion et d'extension trapézo-métacarpiennes sont effectués avec l'examineur qui saisit la main du participant avec sa main pour obtenir un support sur la table d'examen. Les métacarpes et le poignet du patient sont placés dans une position neutre. L'autre main de l'examineur place M1 en flexion. Après avoir sollicité une réponse pour savoir si la position fléchie crée de la douleur, M1 est ensuite mis en extension maximale. Les tests sont considérés comme positifs si le sujet signale une douleur (même une légère sensation de douleur) dans la région de l'articulation basale du pouce.
Traction-shift test	Choa, 2014	Traction puis aller-retour palmaire/dorsal		Le test de traction-shift (subluxation-relocalisation) se réalise avec une seule main, l'examineur applique une traction longitudinale au pouce (souvent en repositionnant la posture subluxée de l'articulation malade) (a). Une pression alternée dorsale (b) et palmaire (c) sur la base du métacarpien est ensuite appliquée pour provoquer une subluxation puis un déplacement de l'articulation. Le test est positif s'il provoque une douleur dans l'articulation.

Arnold, 2022

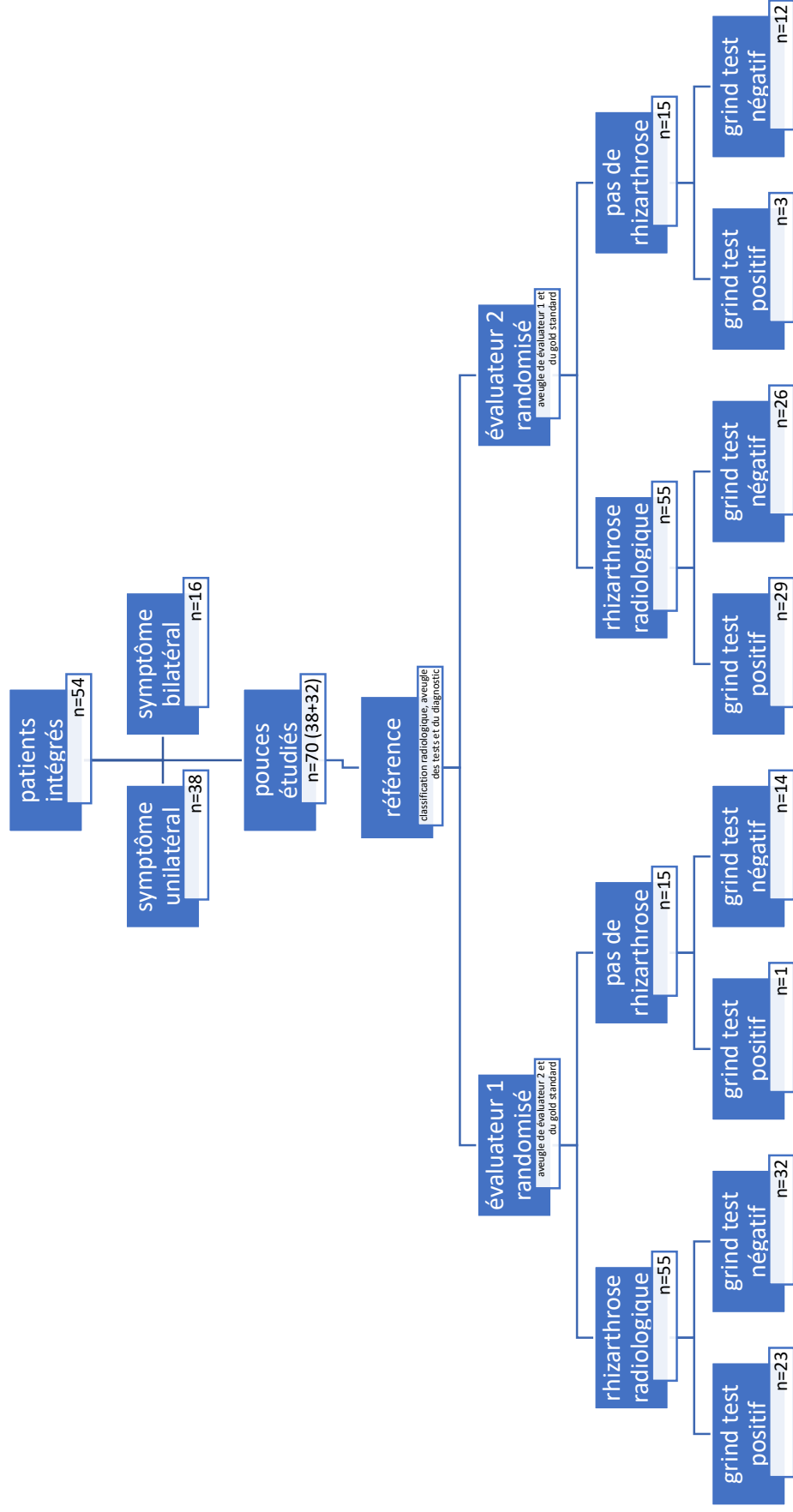
Choa, 2014



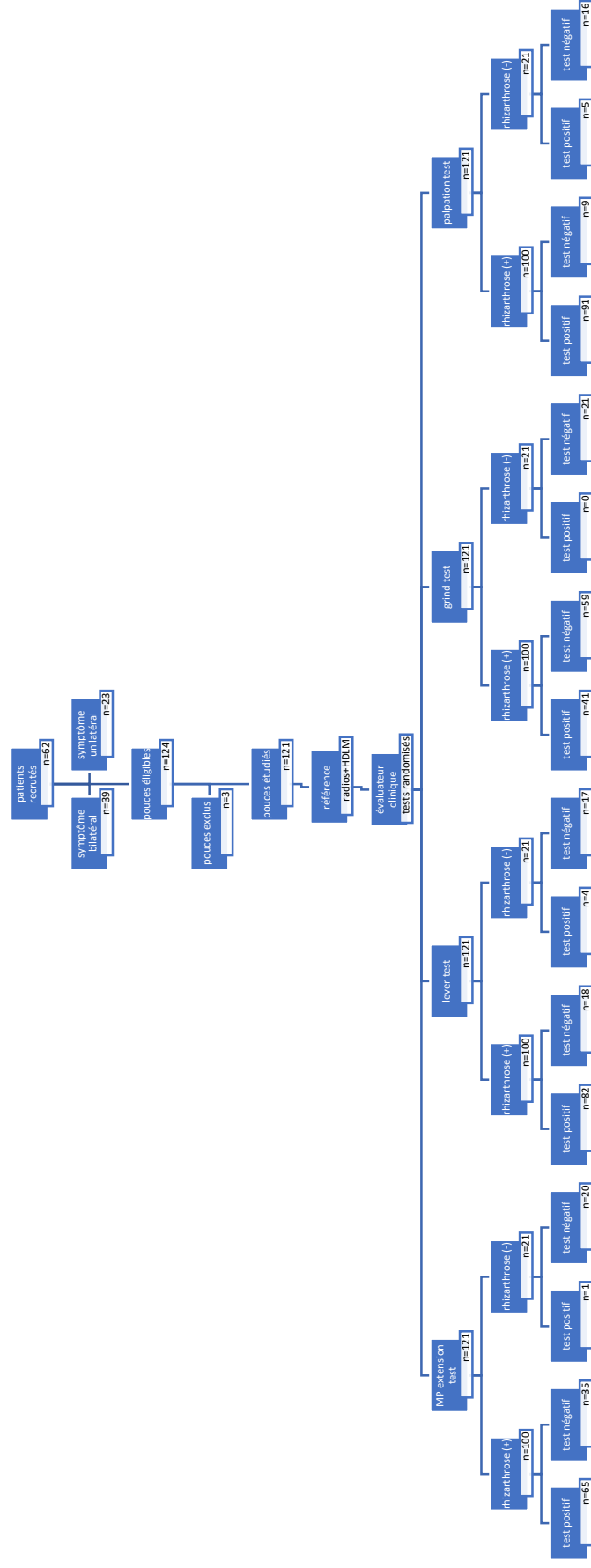
Gelberman, 2015



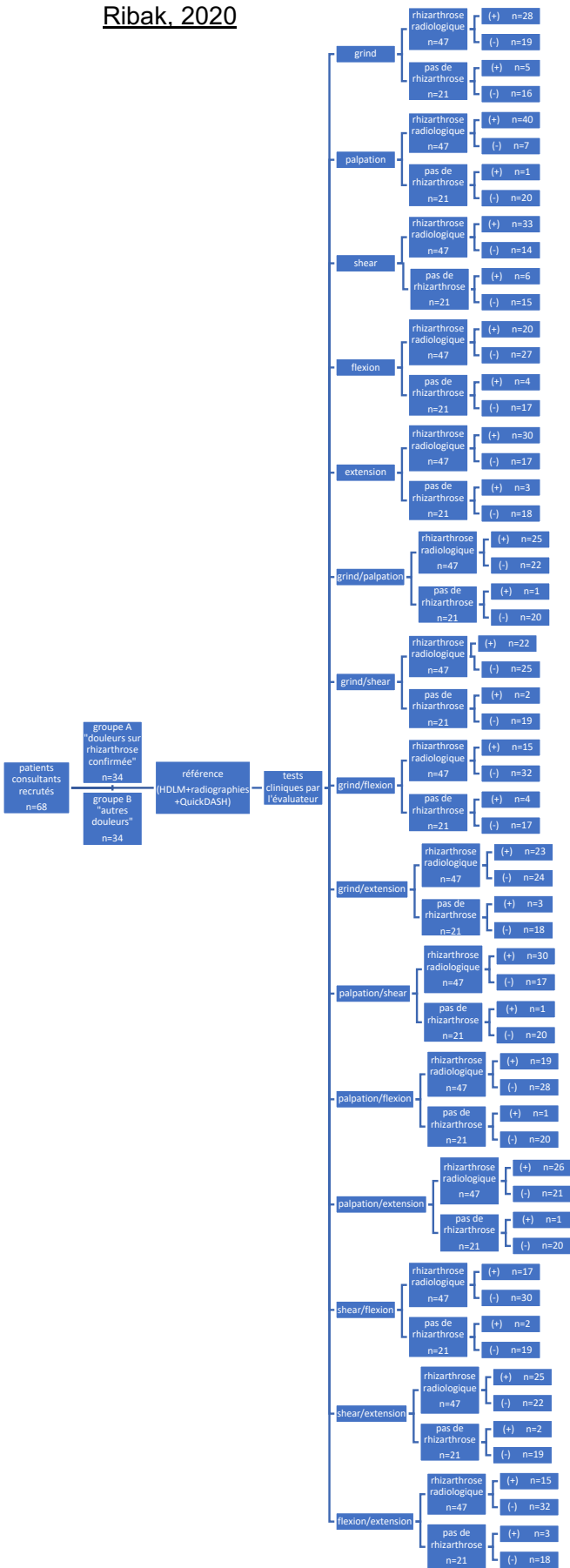
Merritt, 2010



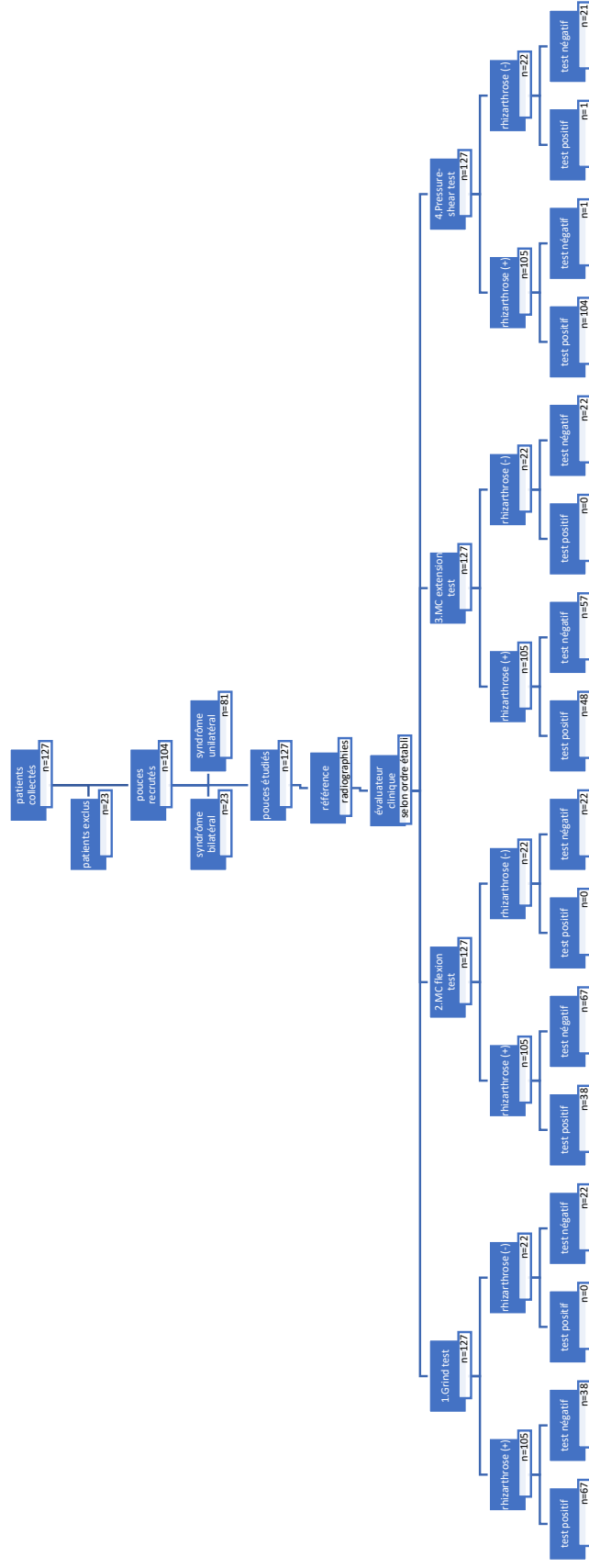
Model, 2016



Ribak, 2020



Sela, 2019



Annexe VI : packages et appels sur R

Courbes SROC

Packages : readxl (transcripteur automatique de tableaux Excel) et mada (package central, dédié aux SROC de DTA)

Appels : readxl (import des tableaux dédiés), reitsma (calculs avec modèle bivarié), plot (charte graphique de la courbe récapitulative pondérée), point (tableaux 2x2 en visualisation graphique), legend (légendes du graphique)

Forest plots

Packages : readxl et DTApplots (package central, option avec méta-analyse bivariée)

Appels : readxl (imports des tableaux dédiés), forest (calculs et graphisme du forest plot), posterior_samples_Bivariate dataset (option de calcul avec modèle bivarié), summary (ajout de la ligne de méta-analyse)

« Pseudo-ROC »

Package : seulement readxl (aucun outil spécifique)

Appels : readxl (imports des tableaux dédiés), df (tous les calculs diagnostiques), calc_auc (calcul des aires sous la courbe), plot (charte graphique des courbes), for (tracé des courbes), palette_colors : rainbow (choix automatique des couleurs), abline (diagonale du hasard), legend

Recherche de biais de publication (*funnel plot* de Deeks et *forest plots* globaux/ en sous-groupes)

Packages : readxl, metafor (package central : contient le modèle à effets aléatoires, les éléments de variance et d'hétérogénéité groupées et le calcul de Deeks), dplyr (transformation des données test par test), forcats (organisation des données par ordre et nom), ggplot2 (grammaire graphique avancée)

Appels : readxl (imports des tableaux dédiés), dat (calculs diagnostiques), study_summary (agrégation par étude), res_global : rma (résultat de la méta-analyse globale en mode aléatoire), test_summary agrégation par tests groupés), res_tests : rma (résultat par tests groupés en mode aléatoire), fit_deeks (test du biais de publication), ggplot (graphisme), forest (calculs et graphisme des *forest plots*), addpoly (ajout de la ligne de méta-analyse)