

Abréviations

Abréviations	Définition
Fl	Flexion
Ext	Extension
ABD	Abduction
ADD	Adduction
RM	Rotation Médiale
RL	Rotation Latérale
MS	Membre Supérieur
DD	Décubitus Dorsal
Se	Sensibilité
Spé	Spécificité
RV+	Rapport de vraisemblance positif
RV-	Rapport de vraisemblance négatif

Table des matières

1	Introduction	3
1.1	Description de la pathologie et de la population concernée	3
1.2	Anatomo-physio-pathologie	4
1.2.1	L'épaule	4
1.2.2	L'articulation gléno-humérale	4
1.2.3	Eléments en présence	4
1.2.4	Moyens d'union	5
1.2.5	Les ligaments	6
1.2.6	Eléments stabilisateurs	7
1.2.7	Biomécanique	8
1.3	Description des outils de mesure évalués	9
1.3.1	Test d'appréhension antérieure	9
1.3.2	Test de recentrage de Jobe (relocation test)	10
1.3.3	Release test ou surprise test	10
1.3.4	Load and shift test	11
1.4	Quel concept mesure l'outil de mesure ?	11
1.5	Intérêt et objectif de la revue de littérature	12
1.5.1	Intérêt de la revue de littérature	12
1.5.2	Objectif de la revue de littérature (modèle PICO)	12
2	Méthode	12
2.1	Critères d'éligibilité des études pour cette revue	12
2.2	Méthodologie de recherche des études	13
2.2.1	Sources documentaires investiguées	13
2.2.2	Equation de recherche utilisée (mots clefs)	13
2.3	Méthode d'extraction et d'analyse des données	15
2.3.1	Méthode de sélection des études	15
2.3.2	Evaluation de la qualité méthodologique des études sélectionnées	15
2.3.3	Grille de lecture utilisée, types de biais recherchés	16
2.3.4	Extraction des données	17
2.3.5	Méthode de synthèse des résultats	17
3	Résultats	18
3.1	Description des études	18
3.1.1	Diagramme de flux	18

3.1.2	Etudes exclues	18
3.1.3	Etudes incluses.....	19
3.2	Risques de biais des études incluses	20
3.3	Effets de l'intervention	24
4	Discussion	27
4.1	Analyse des principaux résultats	27
4.2	Applicabilité des résultats en pratique clinique	31
4.3	Qualité des preuves.....	32
4.4	Biais potentiels de la revue	33
5	Conclusion	34
5.1	Implication pour la pratique clinique	34
5.2	Implication pour la recherche	35
	Sources.....	36

1 Introduction

1.1 Description de la pathologie et de la population concernée

L'épaule instable est une pathologie fréquente du sujet jeune et sportif. Elle est le plus souvent d'origine traumatique comme l'épaule est une des articulations les plus luxées du corps (environ 45% des luxations en 2006)[1], de plus la luxation antérieure de l'articulation gléno-humérale est la plus fréquente (plus de 95% des luxations)[2]. Elle se fait avec l'épaule en abduction et rotation externe importantes. L'instabilité est une perte de contact de deux surfaces articulaires, elle peut venir soit d'une luxation, soit d'une subluxation et entraîner d'autres lésions sur les structures en présence (muscle, os, bourrelet glénoïdien etc.). Celles-ci favorisent l'apparition d'une instabilité de l'articulation ainsi que d'autres lésions sur les structures sanguines ou nerveuses qui peuvent avoir des conséquences plus graves.

L'instabilité de l'épaule est une pathologie complexe. L'instabilité multidirectionnelle est définie par une instabilité antérieure, inférieure et postérieure (donc par une translation gléno-humérale) avec atteinte de la capsule. Elle peut aussi se traduire par :

- Une luxation vraie (déboitement) très souvent d'origine traumatique et qui nécessite presque toujours une réduction par un médecin en milieu hospitalier et sous contrôle d'imagerie,
- Par des épisodes multiples de subluxations, de déboitement réductible qui correspondent à une mobilité anormale de la tête de l'humérus par rapport à la glène (se remet en place spontanément, le patient peut se faire une auto-réduction),
- Par une épaule douloureuse sans luxation ou subluxation avérée mais avec une sensation d'appréhension, et en particulier une appréhension pour mettre le bras en arrière et en rotation externe (hand-ball, baseball, service du tennis, position d'armé du bras).

Dans l'instabilité d'épaule, le **bourrelet glénoïdien** se détache de la glène dans l'immense majorité des cas en avant en cas de **luxation antéro-interne** (95% des cas) ou en arrière en cas de luxation postérieure (5% des cas).

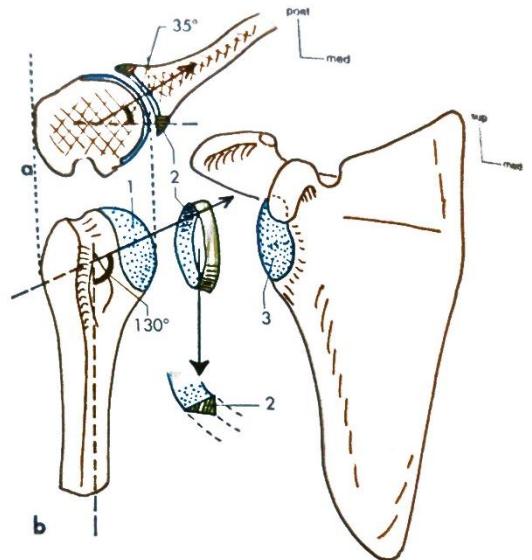
L'instabilité est un ensemble de symptômes ressenti par le patient (ressauts, blocages, engourdissements ou sensation de bras mort) différent de la laxité qui est un signe clinique objectivable par l'examineur [2] et n'est pas pathologique. Certains auteurs pensent que les instabilités antérieures, sans luxation ou subluxations, peuvent être secondaires à des étirements progressifs de la capsule par hyper sollicitation et/ou à des déséquilibres musculaires chez les sportifs de lancer. Par exemple, les mouvements répétés dans le sport tels que la natation (qui est un sport non traumatique), entraînent une modification du complexe articulaire de l'épaule avec hyperlaxité antéro-postérieure constatée. Ceci s'explique aussi par une faiblesse du grand dentelé et des rhomboïdes chez les nageurs de haut niveau, entraînant une instabilité de la scapula. De plus, l'instabilité antérieure chez le nageur pourrait s'expliquer par des assouplissements répétés avec impact entraînant un étirement excessif des ligaments compromettant la stabilité statique de l'articulation gléno-humérale. Le volume capsulaire serait plus important dans les épaules douloureuses instables

avec insuffisance de la capsule antérosupérieure. Avec ce volume plus important, on peut penser que la pression négative intra-articulaire de coaptation de l'articulation se voit moins efficace, perpétuant l'instabilité de l'épaule.

1.2 Anatomo-physio-pathologie

1.2.1 L'épaule

L'épaule est un complexe articulaire dont le mouvement à trois degrés de liberté dépend de cinq articulations : la gléno-humérale entre la glène scapulaire et la tête humérale, l'articulation acromio-claviculaire, la sterno-costo-claviculaire, la bourse sous deltoïdienne et la scapulo-thoracique qui sont des plans de glissement, la première sur une bourse séreuse et la deuxième sur un muscle : le dentelé antérieur.



► 2.2

Éléments en présence : vues transversale (a) et antérieure (b).

1. Tête humérale
2. Labrum
3. Gène

1.2.2 L'articulation gléno-humérale

Il s'agit d'une des deux grosses articulations du complexe articulaire de l'épaule. C'est une articulation sphéroïde non congruente et non concordante entre la tête de l'humérus et la glène de la scapula.

1.2.3 Eléments en présence

Dans le complexe articulaire de l'épaule, au niveau de l'articulation gléno-humérale, on retrouve :

- La glène humérale, c'est une surface articulaire de forme ovale à grand axe oblique en haut et en avant située à l'angle supéro-latéral de la scapula. Sa partie antérieure est légèrement incurvée. Elle est concave mais peu profonde. Le bord périphérique de la surface s'appelle le limbus et est couvert par un élément fibro-cartilagineux, le labrum. Elle regarde globalement en avant à 45° du plan sagittal du sujet et légèrement en haut. Elle répond à la tête humérale.
- Le labrum est triangulaire à la coupe avec une face axiale encroutée de cartilage, une face périphérique qui est la zone d'insertion de la capsule articulaire et une face profonde ancrée sur le limbus. Le rôle du labrum est d'augmenter les surfaces de contact entre la glène scapulaire et la tête humérale.

- La tête humérale, située à la partie supéromédiale de l'épiphyse supérieure de l'humérus, répond à la glène et à son labrum. Elle est aussi sphéroïde non congruente et non concordante. Elle a la forme d'un tiers de sphère d'environ cinq à six centimètres de diamètre. Elle regarde en dedans, en haut et en arrière.

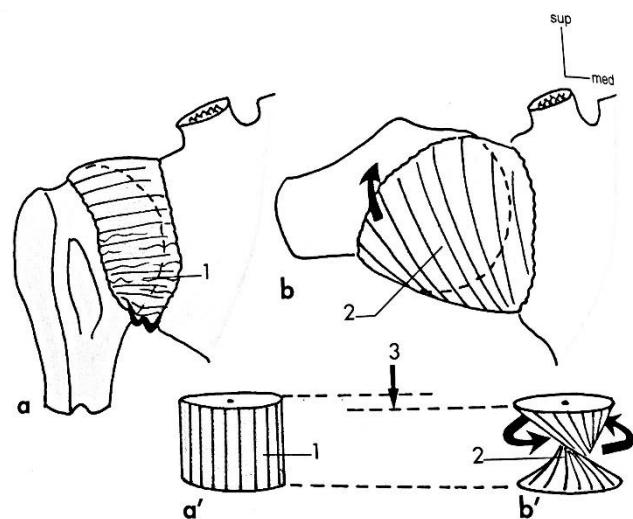
1.2.4 Moyens d'union

La capsule articulaire : elle s'insère au pourtour des surfaces cartilagineuses.

Sur la scapula, l'insertion suit le limbus mais englobe le tubercule supra-glénoïdien. Sur l'humérus, elle s'insère sur le col anatomique, c'est-à-dire au contact de la tête sauf en bas et en dedans où elle s'écarte un peu du cartilage pour descendre jusqu'au col chirurgical de l'os. Sur le labrum, elle s'insère sur l'ensemble de la face périphérique.

1.2.4.1 Caractéristiques

La capsule est lâche, elle permet donc les décoaptations articulaires jusqu'à 2 cm et autorise donc les mouvements complexes où ses fibres sont sollicitées simultanément dans plusieurs plans. Elle est constituée de fibres parallèles, scapulo-humérales, ainsi dans les mouvements tridimensionnels, ses fibres se tendent progressivement pouvant arriver à une position de tension maximale qui stabilise alors l'articulation. C'est ce qui se produit en extension, abduction, et rotation latérale, position où l'instabilité osseuse est compensée par un serrage capsulo-ligamentaire. La position d'abduction est d'autant plus freinée qu'elle met en tension les fibres inférieures de la capsule qui sont épaissies et forment des replis en position bras pendant, ce sont les freins de la capsule. Par ailleurs la capsule présente deux points faibles antérieurs, un supérieur qui permet parfois le passage de la synoviale, un inférieur, qui est parfois distendu par le passage de la tête humérale dans les luxations scapulo-humérales antéro-médiales.



► 2-4

Capsule en position de référence (a, a') et en rotation latérale – abduction (b, b').

1. Fibres détendues (et freins inférieurs)
2. Torsion et tension des fibres
3. Effet de serrage articulaire (stabilité)

La synoviale : elle tapisse la face profonde de la capsule. Elle présente deux caractéristiques : Elle communique parfois avec la bourse synoviale du muscle subscapulaire

Au niveau du tubercule supra-glénoïdien, elle se réfléchit en une gaine synoviale pour le tendon du long biceps. Cette gaine s'étend jusqu'à la sortie du sillon bicipital.

Le labrum, décrit plus haut est aussi un moyen d'union.

1.2.5 Les ligaments

La tête humérale n'étant pas couverte en avant, c'est là que seront les ligaments. Ils forment un double Z qui zèbre verticalement l'interligne antérieure de l'articulation. Les faisceaux ligamentaires sont regroupés en deux groupes principaux et deux autres ligaments :

- Ligament coraco-huméral : il prend son origine sur le processus coracoïde de la scapula, il a un trajet en deux faisceaux :

Faisceau supérieur oblique en dehors, légèrement en bas et en arrière

Faisceau inférieur oblique en dehors et en bas

Les deux faisceaux divergent et se terminent à l'extrémité supérieure de l'humérus,

Le faisceau supérieur sur la face supérieure du tubercle majeur et sur son bord médial, associé à la face profonde du tendon du muscle supra-épineux qui le renforce.

Le faisceau inférieur se termine sur le bord supérieur du tubercle mineur.

Entre les deux faisceaux, un espace correspond au passage intra scapulaire du tendon du long biceps, venant du sillon intertuberculaire.

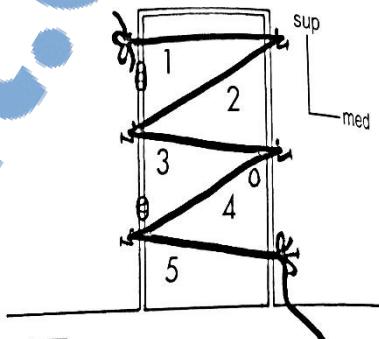
- Ligament gléno-huméral : il prend son origine sur le bord antérieur de la glène scapulaire et sur le labrum sur toute la hauteur.

Les fibres se divisent en trois faisceaux disposés en Z :

Le faisceau supérieur est dirigé presque horizontalement en dehors.

Le faisceau moyen est élargi en éventail, fortement oblique en dehors et en bas et renforcé par le muscle subscapulaire.

Le faisceau inférieur est dirigé horizontalement en dehors, il est large et renforcé par les freins scapulaires.

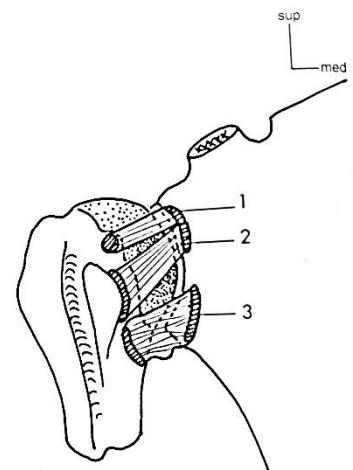


► 2-8

Zigzag des ligaments ant. à la manière d'un cordage empêchant l'ouverture d'une porte.

1 et 2 = les 2 faisceaux du ligament coraco-huméral

3, 4 et 5 = les 3 faisceaux du ligament gléno-huméral



► 2-9

Les 3 faisceaux du ligament gléno-huméral.

1. Faisceau supérieur

2. Faisceau moyen

3. Faisceau inférieur

Le ligament gléno-huméral se termine à l'extrémité supérieure de l'humérus. Le faisceau supérieur à la partie antérieure du col anatomique, juste au-dessus du tubercule mineur dans la fosse supra-tuberculaire.

Le faisceau moyen se termine sur le bord médian du tubercule mineur, juste en dedans du subscapulaire qui le renforce.

Le faisceau inférieur se termine à la partie antéroinférieure du col chirurgical.

- Ligament coraco-glénoïdien provient du processus coracoïde de la scapula, se dirige en éventail vers le bas et le dehors pour se terminer sur la partie postéro-supérieure de la capsule scapulo-humérale. C'est un ligament suspenseur de la capsule.

- Ligament coraco-acromial qui prend son origine sur le processus coracoïde de la scapula, se rétrécit dans son trajet vers le dehors et l'arrière mais reste épais et solide. Il se termine sur le processus acromial de la scapula, bord médial, en avant de la facette claviculaire. Il intervient dans la mécanique de la scapulo-humérale en tant que partie de la voûte coraco-acromiale surplombant la tête humérale. Il en forme la partie moyenne.

1.2.6 Eléments stabilisateurs

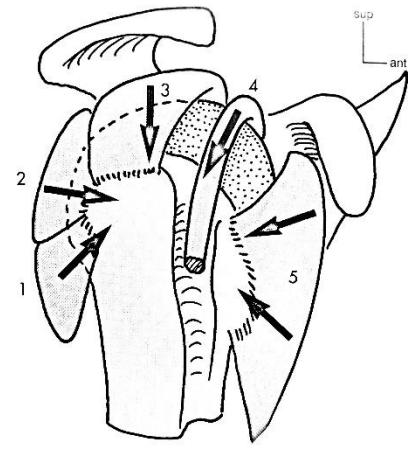
Ce sont les tendons qui entourent intimement la tête humérale ou le labrum.

Le tendon du long biceps a un trajet intra scapulaire même si extra-articulaire. Il s'insère sur la scapula et le labrum, accentuant donc la stabilité de ce dernier ;

Le tendon du subscapulaire renforce le ligament gléno-huméral moyen ;

Le tendon du supra-épineux renforce le ligament coraco-huméral, surtout le faisceau supérieur.

Toute la coiffe des rotateurs et le muscle deltoïde.



1.2.7 Biomécanique

Sur le plan mécanique et pathologique, cette articulation se distingue par la discordance entre la tête humérale et la glène avec des surfaces articulaires ni concordantes ni congruentes. Elle comprend des systèmes de contention passive, avec les ligaments, et active, avec les muscles de la coiffe, qui forment un néo-acétabulum semi-rigide avec la voûte coraco-acromiale.

Le muscle subscapulaire a une action de stabilisation antérieure de la tête humérale. Sur le plan mécanique et pathologique, il forme le verrou antérieur de l'épaule, face aux risques de luxations antérieures. A cause de cela, ce muscle est parfois sectionné puis suturé en position raccourcie, afin de brider la portion antérieure de l'articulation (intervention de Bankart), et limiter ainsi la récidive de certaines luxations.



Le muscle deltoïde est présent contre les faces antérieure, latérale et postérieure de la tête humérale, par sa position et sa forme englobante, il plaque la tête humérale contre la glène. Son tonus et ses contractions favorisent la coaptation de l'articulation lors d'un mouvement.

Le muscle supra épineux, sur le plan mécanique et pathologique, il assure une suspension de l'humérus, sa paralysie favorise la subluxation inférieure.

Le muscle infra épineux s'oppose à l'ascension possible de la tête humérale. Il assure un rôle statique d'abaissement de la tête humérale.

L'ensemble de la coiffe des rotateurs a un rôle de stabilisation de la tête humérale dans la glène par suspension, centrage et abaissement de la tête humérale.

La coiffe anatomique des rotateurs associée à la longue portion du muscle biceps brachial forme la coiffe fonctionnelle des rotateurs.

La longue portion du biceps joue un rôle de maintien et de coaptation de la tête humérale. Elle a un début de trajet horizontal puis, ce tendon change de direction et va dans la gouttière bicipitale. Dans son trajet, il crochète le trochin, ce qui permet de contrôler l'abaissement de la tête humérale. La longue portion du biceps passe entre l'acromion et le processus coracoïdien ce qui veut dire qu'elle est à l'aplomb du ligament acromio-coracoïdien qui est une nappe (plat).

Les pathologies traumatiques et la dégénérescence rhumatismale sont des phénomènes fréquents compte tenu du lourd cahier des charges de cette région anatomique.

1.3 Description des outils de mesure évalués

Selon les articles [2] et [3], il existe plusieurs tests pour le diagnostic d'une instabilité d'épaule :

- le test d'apprehension
- le test de recentrage de Jobe ou relocation test
- le release test ou surprise test
- le test du tiroir antérieur de Gerber et Ganz
- le load shift test
- le test d'hyper abduction
- le Fulcrum test ou augmentation test (parfois considéré comme étant le même que le crank test ou que le test d'apprehension [4])

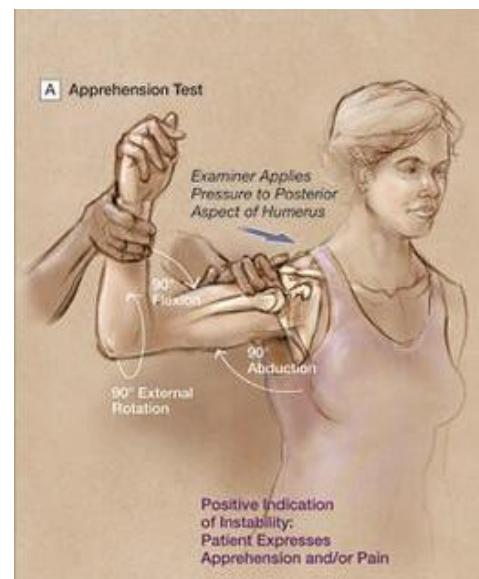
Le choix arbitraire a été fait, pour ce travail, de ne sélectionner que les tests au protocole semblable pour les comparer, ce seront donc les tests d'apprehension antérieure, le test de recentrage de Jobe (ou relocation test), le release test et le load and shift test. Ces 4 tests peuvent être utilisés pour mettre en évidence une instabilité gléno-humérale antérieure.

1.3.1 Test d'apprehension antérieure

Manœuvre : Le sujet est en décubitus dorsal. Le thérapeute place une main au niveau du coude (fléchi à 90°) du sujet et amène le bras à 90° d'abduction. Il place son autre main au niveau du poignet puis amène le bras en rotation externe maximale ou place sa deuxième main en arrière de la tête humérale pour lui imprimer une poussée vers l'avant.

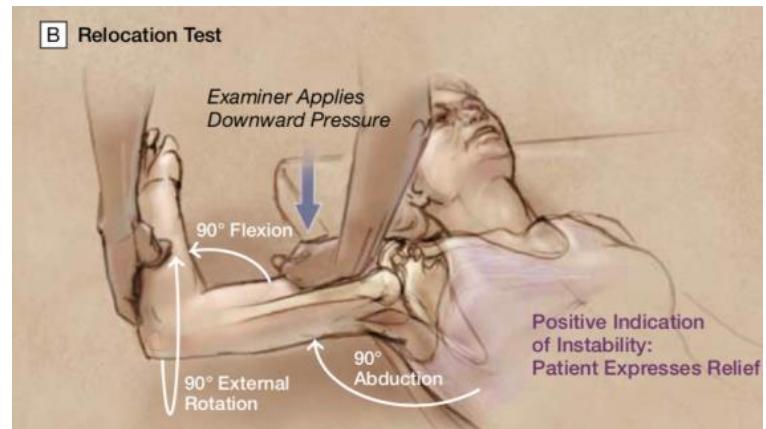
Positivité : Le test est considéré positif si le patient a l'impression que sa tête humérale va sortir de la glène ou qu'il présente une appréhension à continuer le mouvement de rotation externe. Cette appréhension peut être ou non accompagnée d'une douleur.

Remarque : Cette manœuvre peut aussi se réaliser en position assise.



1.3.2 Test de recentrage de Jobe (relocation test)

Manœuvre : Ce test s'effectue dans la continuité du test d'appréhension. Le sujet est en décubitus dorsal. Le thérapeute place une main au niveau du coude (fléchi à 90°) du sujet et amène le bras à 90° d'abduction. Il place son autre main au niveau de la face antérieure du moignon de l'épaule, en regard de la tête humérale puis exerce une force dirigée vers l'arrière afin d'effectuer une translation postérieure de la tête humérale. Tout en maintenant cette force, il amène le bras du sujet en rotation externe.

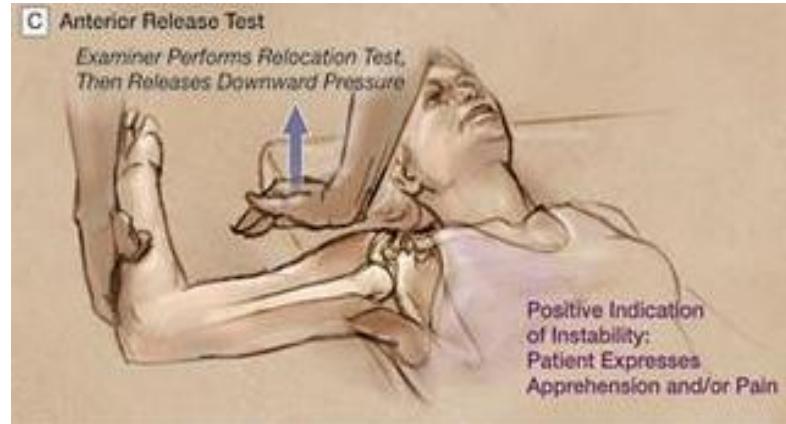


Positivité : Ce test est considéré positif si on remarque une augmentation de l'amplitude de la rotation externe (par rapport à celle du test d'appréhension antérieure) avant l'apparition de l'appréhension. Cette appréhension peut être accompagnée d'une douleur.

Remarque : Une variante de ce test consiste à réaliser le test d'appréhension antérieure jusqu'à ce qu'il devienne positif puis d'appliquer la force dirigée vers l'arrière sur la tête humérale. Cette variante est positive si l'appréhension du sujet diminue.

1.3.3 Release test ou surprise test

Manœuvre : Le sujet est en décubitus dorsal. Le thérapeute place une main au niveau du coude (fléchi à 90°) du sujet et amène le bras à 90° d'abduction. Il place son autre main au niveau de la face antérieure du moignon de l'épaule puis exerce une force dirigée vers l'arrière afin d'effectuer une translation postérieure de la tête humérale. Tout en maintenant cette force, il amène le bras en rotation externe jusqu'à l'appréhension du sujet. Dans cette position, il relâche subitement la force dirigée vers l'arrière sur la tête humérale.



Positivité : Le test est considéré positif si on remarque une augmentation de l'appréhension du sujet lors du relâchement de la translation postérieure de la tête humérale. Cette appréhension peut être ou non accompagnée d'une douleur.

Remarque : Le début de ce test est identique au test de recentrage de Jobe, il peut donc être effectué juste après.

1.3.4 Load and shift test

But : Mise en évidence d'une laxité gléno-humérale.

Manœuvre : Le sujet est assis avec les bras le long du corps. Le thérapeute, placé derrière le sujet, stabilise d'une main la scapula et la clavicule du côté à tester (en la plaçant au-dessus de l'épaule) et englobe de son autre main la tête humérale d'arrière en avant (de sorte que son pouce soit en regard de la partie postérieure de la tête et que ses doigts soient en regard de la partie antérieure de la tête humérale). Le thérapeute va compresser la tête humérale dans la glène afin de s'assurer du bon centrage de celle-ci. Tout en évitant des compensations scapulo-claviculaires, il va ensuite effectuer des translations antérieures de la tête humérale.

Positivité : Une translation plus importante de la tête humérale comparativement au côté opposé atteste d'une laxité plus importante. Le sens de la laxité correspond au sens de la translation anormale (dans cette étude on ne s'intéresse qu'à l'antérieure). On note ce test selon la cotation modifiée d'Hawkins [2] avec :

- grade 0, peu ou pas de mouvement
- grade 1, la tête humérale monte sur le bord du bourrelet glénoïdien
- grade 2, la tête humérale peut être luxée mais se remet en position dans la glène spontanément lorsqu'on enlève la pression
- grade 3, la tête humérale ne revient pas dans la glène lorsqu'on enlève la pression



Remarque : Cette manœuvre peut aussi se réaliser en décubitus dorsal avec le bras le long du tronc ou avec le bras à 90° d'abduction. Une compression préalable de la tête humérale dans la glène est aussi effectuée dans ces deux cas.

1.4 Quel concept mesure l'outil de mesure ?

Les différents tests servent pour le diagnostic de la même pathologie, l'instabilité antérieure d'épaule. Ils utilisent la symptomatologie de la pathologie comme critère pour déterminer si le sujet est sain ou non. Ce sont donc des tests discriminant l'état du sujet entre sain ou pathologique.

Ces tests évaluent la validité de critères donnés pour la diagnostic de la pathologie par rapport à un gold standard. Ici, ces critères seront surtout l'appréhension, la douleur, la sensation d'étirement ou la résistance au mouvement imprimé à l'articulation.

1.5 Intérêt et objectif de la revue de littérature

1.5.1 Intérêt de la revue de littérature

L'intérêt de cette revue de littérature réside dans le fait que l'instabilité d'épaule est une pathologie que l'on peut retrouver fréquemment chez le sujet jeune, d'autant plus avec le développement de sport de frappe, de contact ou d'hiver. L'apparition de la pratique du snowboard a fait augmenter l'incidence des blessures sur le membre supérieur, y compris l'instabilité de l'articulation gléno-humérale. De plus, les instabilités antérieures représentent une grande part de l'ensemble des instabilités gléno-humérales ; c'est pourquoi on s'y intéresse en priorité. Pour diagnostiquer cette pathologie, on peut retrouver plusieurs tests dont une partie est citée plus haut mais certains sont assez similaires dans leur protocole. Or, si dans l'exercice de ses fonctions, un masseur-kinésithérapeute se retrouve confronté à une situation où le patient présente une épaule pathologique et à risque d'instabilité où il doit effectuer un test diagnostique d'instabilité antérieure d'épaule pour savoir s'il doit autoriser son patient à poursuivre son activité sportive ou l'encourager à consulter son médecin, il ne parait pas judicieux ni temporairement optimal de faire passer au patient tous les tests décrits dans la littérature. Pour faire un choix parmi des tests semblables, le praticien peut choisir en fonction de son expérience personnelle ou se décider en comparant les propriétés métrologiques des outils diagnostiques à sa disposition.

1.5.2 Objectif de la revue de littérature (modèle PICO)

L'objectif défini de cette revue est d'établir de comparer et d'analyser la sensibilité et la spécificité de tests diagnostiques de l'instabilité antérieure d'épaule.

2 Méthode

2.1 Critères d'éligibilité des études pour cette revue

Pour une question diagnostique d'analyse des propriétés métrologiques d'outils de mesure, on recherchera des études transversales comparatives avec un étalon-or ou des études contrôlées randomisées. On cherche ici à déterminer la sensibilité et spécificité de tests cliniques d'instabilité antérieure d'épaule. Ces tests sont les tests d'appréhension antérieure, de recentrage de Jobe, le load and shift test et le release test.

Pour un test donné, on peut avoir deux résultats, positif ou négatif. Cependant, parmi les patients au résultat positif, il peut y avoir des vrais positifs et des faux positifs qui sont des patients diagnostiqués positifs par erreur du test. La même chose est possible chez les patients au résultat négatif dont certains peuvent être en réalité pathologiques.

La **sensibilité** d'un outil métrologique est sa capacité à être positif si le patient est pathologique.

La **spécificité** d'un test qualifie sa capacité à déterminer précisément quel patient n'est pas pathologique.

Plus ces deux indicateurs seront élevés, plus le diagnostic dépendant de ces tests sera **valide** et la quantité de faux positifs et de faux négatifs faible.

2.2 Méthodologie de recherche des études

2.2.1 Sources documentaires investiguées

Comme base de données électroniques de recherche on utilisera :

PubMed : Moteur de recherche médical de renommée internationale représentant sous forme d'archives une base de données bibliographiques qui concerne les domaines de la médecine et de la biologie. Il permet d'accéder à des millions d'articles et d'effectuer des recherches avec un degré de précision assez élevé.

Cochrane Library : Base de données produite par Cochrane, ce moteur regroupe trois sous-bases de données qui sont Cochrane Database of Systematic Reviews (CDSR), Cochrane Clinical Answers (CCAs) et enfin Cochrane central Register of Controlled Trials-Clinical Trials (CENTRAL). Les articles à disposition abordent les thèmes de la médecine et de la santé, elle contient plusieurs milliers de revues.

Pedro : Base de données dans le domaine de la physiothérapie fondée sur des preuves, celle-ci contient plusieurs milliers de revues systématiques, de recommandations, d'essais randomisés et autres. La qualité de certaines études est analysée par l'échelle Pedro qui est une grille de lecture critique d'article.

2.2.2 Equation de recherche utilisée (mots clefs)

Pour définir une équation de recherche la plus précise possible, on utilise le modèle de questions clinique **PICO** (pour Population, Intervention, Control et Outcome), sachant que la majorité de la littérature scientifique est rédigée en anglais pour une lisibilité mondiale, on devra donc effectuer la recherche avec des termes anglais qu'on pourra agrémenter ou changer selon les synonymes ou autres termes découverts au cours des recherches préliminaires.

PICO	A DEFINIR	MOTS CLEFS FRANÇAIS	ANGLAIS
POPULATION	Quelle est la population de l'étude.	Patient présentant une instabilité antérieure d'épaule	
INTERVENTION	Quel est l'outil de mesure utilisé.	Le test d'apprehension antérieure, de recentrage de Jobe, le release ou le load and shift test	Anterior apprehension test, relocation test, release test, load and shift test
CONTROL (COMPARATEUR)	Quel sera le comparateur utilisé		
OUTCOME (RESULTAT)	Quelles propriétés métrologiques seront analysées	Sensibilité et spécificité	Sensitivity, specificity

Comme équation de recherche, on utilisera des mots anglais à partir du tableau ci-dessus.

Sur Pubmed l'équation de recherche utilisée est (((((release test[Title/Abstract]) OR (anterior apprehension test[Title/Abstract])) OR (relocation test[Title/Abstract])) OR (load and shift test[Title/Abstract])) AND (sensitivity[Title/Abstract])) AND (specificity[Title/Abstract])) AND (shoulder[Title/Abstract]).

A partir de cette équation de recherche, on pourra ensuite affiner le tri des résultats en ne sélectionnant que les études transversales ou études contrôlées randomisées. Pour affiner encore la recherche on utilisera les opérateurs booléens pour lier entre eux les différents termes de la recherche et les synonymes. Ces opérateurs sont les mots « AND », « OR » et « NOT »

- «AND » permet d'obtenir des résultats contenant les mots qu'il lie entre eux
- « OR » permet d'obtenir des résultats pouvant ne contenir qu'un seul des mots qu'il lie
- « NOT » permet d'exclure les résultats de recherche qui contiennent les termes concernés

	Mot clef	Synonyme	Synonyme
Patient		OR	OR
Intervention	AND		
Comparateur	AND		
Objectif	AND		

2.3 Méthode d'extraction et d'analyse des données

2.3.1 Méthode de sélection des études

Après avoir lancé la recherche, il va falloir procéder à l'« écrémage » des résultats pour ne garder que les études les plus pertinentes. Ce tri va s'effectuer en plusieurs étapes qui sont :

- Utiliser les filtres de recherche pour que les résultats n'incluent que les études du type qui nous intéresse.
- Utiliser les filtres de recherche pour ne présenter que les résultats datant d'il y a moins de 20 ans pour ne garder que les données les plus actuelles de la médecine
- Exclusion des articles par lecture de leur titre. Ce premier tri de lecture de titre nous permet de voir quelles études ne semblent pas pertinentes du tout pour répondre à notre question.
- Inclusion des articles par lecture de leur titre, après exclusion des articles, cette fois on s'intéresse à sélectionner les études qui semblent pouvoir répondre à la question
- Suppression des doublons, certains articles peuvent apparaître comme résultats de la recherche sur plusieurs bases de données, il convient donc de ne les sélectionner qu'une seule fois pour ne pas voir les résultats statistiques biaisés par pondération des résultats en doublons.
- Sélection des études par lecture de leur abstract : présent au tout début des articles, il en est le résumé et peut nous informer sur le type d'étude menée, la méthodologie de celle-ci et nous donner un aperçu de l'intérêt de cet article par rapport à notre question clinique.
- Sélection des études par lecture du texte intégral, une fois les autres étapes accomplies, celle-ci sera très chronophage, il faudra en extraire les données qui nous intéressent et évaluer leur qualité méthodologique grâce à une échelle de lecture critique.

2.3.2 Evaluation de la qualité méthodologique des études sélectionnées

Une fois les études sélectionnées, puis triées selon des critères d'exclusion et inclusion, on s'intéresse au texte, et avant tout, à sa qualité méthodologique pour savoir si les résultats de l'étude analysée sont acceptables d'un point de vue qualitatif ou si la méthodologie de l'étude révèle des biais trop importants. Cette démarche est obligatoire, pour fournir une synthèse de résultats de bonne qualité il faut que les études dont sont issues ces résultats soient de bonne qualité. Si un ou plusieurs biais sont identifiés au cours de la lecture critique de l'article, il faut les prendre en compte lors de l'interprétation des résultats et les répertorier dans un tableau qui synthétise les biais présents dans les études analysées.

2.3.3 Grille de lecture utilisée, types de biais recherchés

L'évaluation de la qualité méthodologique d'une étude passe par l'analyse de sa validité interne, c'est sa capacité à mesurer et à communiquer une conclusion en accord avec la réalité des faits. Pour analyser le niveau de qualité générale d'une étude, il va alors falloir prêter attention aux erreurs possiblement commises. Deux grands types d'erreurs existent :

- Les erreurs aléatoires : Représentées par p, elles correspondent aux erreurs dues au hasard, une certaine quantité de hasard dans les résultats est inévitable mais il faut qu'il soit quantifié, qu'un intervalle de confiance soit donné.
- Les biais : ils peuvent venir de la sélection de l'échantillon des sujets, du test évalué de celui de référence ou de la temporalité de l'étude. Il faut essayer de n'en garder que le minimum et de les identifier.

Une évaluation de la validité externe est également à garder à l'esprit. Plus subjective, par contre, la question qui se pose est : est-ce que les résultats sont extrapolables à une population plus large à la pratique clinique ?

Pour l'analyse des articles, de leurs biais et présentation des conclusions au sujet de la validité interne, on utilisera la grille Quadas-2, présentée en annexe (Annexe 1), qui est parmi les grilles recommandées lors de l'analyse critique d'articles pour l'élaboration d'une revue de littérature à partir d'études diagnostiques. On cherchera des biais de :

- Sélection des patients en répondant aux questions :
 - un échantillon de patients a-t-il été recruté de manière consécutive ou aléatoire ?
 - un schéma d'étude de type cas/témoin a-t-il été évité ?
 - l'étude a-t-elle évité des exclusions inappropriées ?
 - y a-t-il un risque que les patients sélectionnés ne correspondent pas à la question de recherche ?
- Test évalué : si plusieurs tests sont évalués, on répondra aux questions suivantes pour chaque test :
 - les résultats du test évalué ont-ils été interprétés sans avoir connaissance des résultats de référence ?
 - si un seuil de décision a été utilisé, a-t-il été spécifié ?
 - la réalisation ou l'interprétation du test évalué a-t-elle pu introduire un biais ?
 - y a-t-il une préoccupation quant au fait que le test évalué, dans sa réalisation, interprétation, diffère de la question de recherche ?
- Test de référence :
 - le test de référence est-il susceptible de correctement classer la condition recherchée ?
 - les résultats du test de référence ont-ils été interprétés sans avoir connaissance des résultats du test évalué ?
 - y a-t-il une préoccupation quant au fait que la pathologie telle qu'elle a été diagnostiquée par le test de référence diffère de celle de la question de recherche ?
- Déroulement et temporalité :
 - l'intervalle de temps entre la réalisation du test évalué et le test de référence est-il approprié ?

- tous les patients ont-ils reçu le test de référence ?
- tous les patients ont-ils reçu le même test de référence ?
- tous les patients sont-ils inclus dans l'analyse ?
- le déroulement auprès des patients a-t-il pu introduire un biais ?

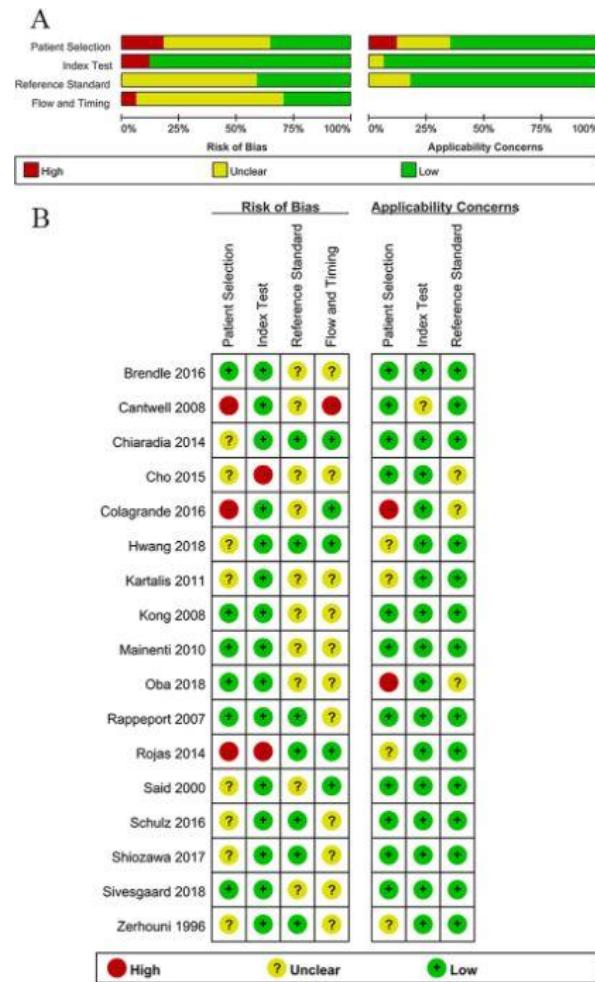
Puis le résultat de ces analyses sera présenté sous forme de diagramme comme dans l'exemple ci-contre.

2.3.4 Extraction des données

Au sein des études sélectionnées, toutes les informations ne seront pas pertinentes pour répondre à notre question clinique et leur traitement pourrait s'avérer trop long et fastidieux pour le lecteur de ce travail, il faut donc synthétiser les informations. Pour ce faire, on utilisera un tableau indiquant précisément quelle données extraire des articles et importer dans le tableau pour pouvoir ensuite les analyser de manière à formuler une conclusion à l'hypothèse de départ. Les données extraites proviennent de la partie résultats des articles lus, et non de la partie conclusion ou discussions qui présentent une analyse des résultats du point de vue de l'auteur de l'article et peuvent contenir un spin, qui est une exagération des résultats par l'auteur. Pour rester le plus objectif possible, on ne prendra que les résultats bruts présentés dans un tableau le plus souvent.

2.3.5 Méthode de synthèse des résultats

Pour synthétiser les résultats de tous ces articles, on utilisera un tableau par test clinique récapitulant les valeurs des différents indicateurs que l'on souhaite analyser selon les articles dont ils proviennent. Puis, on effectuera une moyenne des valeurs obtenues pour chaque indicateur qu'on présentera dans un autre tableau, permettant une comparaison plus simple des valeurs les unes par rapport aux autres pour chaque test. A partir de ce dernier tableau, on pourra formuler une conclusion par rapport à la question clinique initiale.

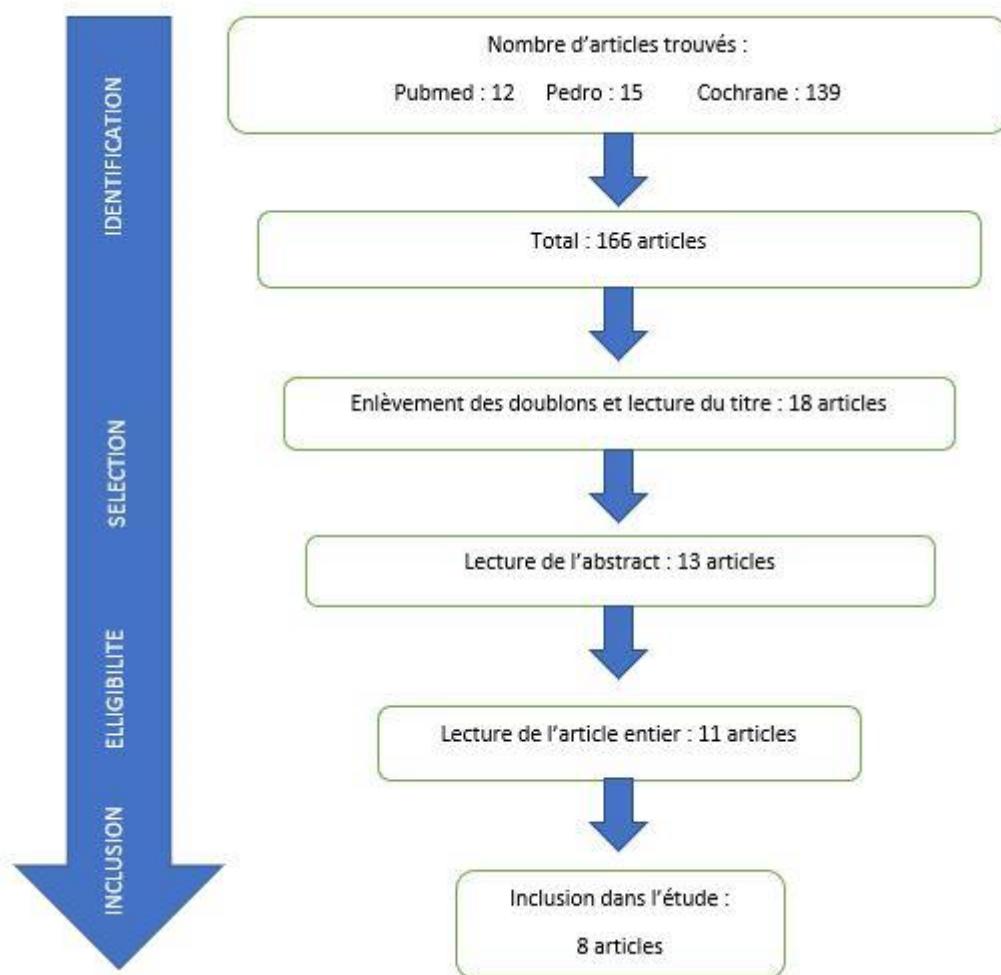


3 Résultats

3.1 Description des études

3.1.1 Diagramme de flux

Pour représenter le processus de sélection des études par exclusion puis inclusion, on utilisera un diagramme de flux, qui illustre chaque étape du processus et y indique le nombre d'études restant dans la sélection.



3.1.2 Etudes exclues

Une fois les recherches effectuées, nous avons procédé comme expliqué plus haut en commençant par trier les articles trouvés par la correspondance de leur titre avec la question de recherche initiale et la suppression des doublons, ce qui nous fait passer de 166 articles à seulement 13. Ensuite, en lisant l'abstract, 2 articles de plus sont exclus, 2 autres sont encore exclus à la lecture du texte. Les causes de ces 4 exclusions sont explicitées dans le tableau ci-dessous par souci de transparence.

Article	Cause de l'exclusion
Tzannes, 2002 [2]	Revue de littérature
Hegedus, 2012 [5]	Revue de littérature
Morey, 2016 [6]	Ne mesure pas la bonne condition
Lizzio, 2017 [7]	Revue de littérature
Pandya, 2008 [8]	Incohérence avec le sujet

3.1.3 Etudes incluses

A la suite de ces exclusions, on se retrouve avec 9 études qui correspondent à la question de recherche et au PICO ciblés. On va en extraire les résultats après avoir vérifié leur qualité méthodologique. Les caractéristiques principales de ces études sont résumées dans le tableau ci-dessous ; les résultats seront présentés dans un autre tableau.

Auteur et année	Type d'étude	Taille de l'échantillon	Population	Test de référence	Test évalué
Farber, 2006 [1]	Cas-témoin	363	46 cas et 317 témoins	Arthoscopie ou radiographie	-Appréhension (variante de Rowe et Zarins) -relocation
Lo, 2004 [4]	Cohorte prospective historique	46		Diagnostic préexistant	-Appréhension -Relocation -release
Safran, 2010 [9]	Cohorte prospective	52	Hommes entre 17 et 27 ans (20.3) dont 41 soldats	Survenue d'une relaxation dans le suivi	-Anterior apprehension test
Milgrom, 2014 [10]	Cohorte prospective	53	Hommes entre 17 et 27 ans (moyenne d'âge 20.2 ans) Dont 41 soldats	Survenue d'une relaxation dans le suivi	-Anterior Apprehension test
Guanche, 2003 [11]	Cohorte prospective	60 épaules	59 patients	Arthoscopie	-Anterior, apprehension test, -Relocation test

Loh, 2016 [12]	Cohorte rétrospective	227	Patient d'un hôpital avec une lésion de Bankart	Arthroscopie ou imagerie (IRM, ARM)	-Anterior apprehension test ET Load and shift
Kumar, 2015 [13]	Cohorte rétrospective	168	Entre 17 et 29 ans (23) ; 150H	Arthroscopie	-Anterior apprehension test
Van Kampen, 2013 [3]	Cohorte prospective	169		Arthrographie	-Appréhension -Relocation -Release -Load and shift

3.2 Risques de biais des études incluses

Afin d'évaluer la qualité méthodologique de chacune des études incluses, on utilisera la grille de lecture Quadas-2 qui est la plus adaptée pour les lectures critiques d'études diagnostiques. A partir des recommandations de cette grille, on évaluera les risques de biais principaux pour chacune des études avant de les synthétiser dans un tableau accompagné d'un graphique pour une meilleure lisibilité.

Farber, 2006

	Risk of bias				Applicability		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Farber 2006	High	Low	Unclear	Low	High	Low	Low

Comme on le voit sur le tableau ci-dessus, la sélection des patients présente un haut risque de biais alors que le test de référence présente un risque de biais incertain. La sélection des patients peut biaiser le résultat par le fait que l'étude suit un schéma cas-témoin avec un échantillon de patients qui diffère en plus de manière significative du groupe témoin. Par ailleurs, les études concernant les patients avec un diagnostic de la pathologie testée préétabli et un groupe contrôle sans cette pathologie, peuvent exagérer la précision du diagnostic du test étudié. [14]

Le test de référence présente un risque de biais incertain car il n'est pas précisé dans l'étude s'il a été effectué ou interprété sans avoir eu connaissance du test évalué.

La sélection des patients peut aussi présenter un biais en termes d'applicabilité car l'étude exclut aussi les patients avec une instabilité non franche (patient avec douleur mais pas d'histoire de luxation ou subluxation de l'épaule) alors que ces patients peuvent aussi présenter une instabilité antérieure d'épaule.

Lo, 2004

	Risk of bias				Applicability		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Lo 2004	High	Low	Low	Unclear	High	Low	Low

Cette étude présente des risques de biais :

- de sélection élevés, elle exclut les patients n'ayant pas subi une luxation ainsi que tout patient avec une instabilité d'origine non traumatique
- de temporalité incertains ; il y est écrit que les patients sont recrutés selon leur dossier après qu'un diagnostic soit établi, ce diagnostic sert de référence, mais on ne sait pas quel est l'intervalle de temps qui sépare ce diagnostic et la participation du patient au test évalué.

Safran, 2010

	Risk of bias				Applicability		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Safran 2010	High	Low	Low	Unclear	High	High	Unclear

Cette étude présente des risques de biais de sélection élevés, tous les patients sont des militaires et les luxations dues aux accidents en véhicule motorisé sont exclues. La temporalité présente des biais incertains. Le test de référence est le suivi pour voir si les patients se relaxent l'épaule après avoir passé le test, la période de suivi ne s'étend que jusqu'à 4 ans après la première luxation, ça peut ne pas être suffisant.

Au niveau de l'applicabilité de l'étude on a des biais élevés concernant les patients qui sont tous hommes et militaires avec une activité physique élevée, le test de référence qui n'est utilisé ici que pour évaluer un risque de relaxation d'épaule au lieu d'une instabilité pouvant être traduite par d'autres symptômes. Le test de référence présente aussi des biais incertains pour les raisons de temporalité précédemment évoquées.

Milgrom, 2014

	Risk of bias				Applicability		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Milgrom 2014	High	Low	Low	Unclear	High	High	Unclear

Cette étude présente les mêmes risques de biais que l'étude précédente car il s'agit d'un suivi de celle-ci avec une extension de suivi des patients allant jusqu'à 75 mois.

Guanche, 2003

	Risk of bias				Applicability		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Guanche	Low	Low	Low	Low	Low	Low	Low

Cette étude présente des risques de biais faibles en général.

Loh, 2016

	Risk of bias				Applicability		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Loh 2016	High	Low	Low	Low	Low	Low	Low

Cette étude présente un biais de sélection des patients élevé du fait que les patients ont été sélectionnés rétrospectivement sur dossier parmi les patients ayant subi une opération de réparation de Bankart, lésion très souvent due à une luxation d'épaule. Le risque est donc que cette étude montre des résultats biaisés car elle n'inclut que des patients qui ont été opérés, donc plus probablement des patients présentant une lésion et susceptible d'avoir eu un test positif, on peut avoir une surestimation des propriétés des tests. Ce biais peut affecter les résultats de spécificité des tests évalués. De plus, les tests d'instabilité antérieure sont ici utilisés pour diagnostiquer une conséquence possible mais pas obligatoire de l'instabilité d'épaule, différant légèrement de la pathologie initiale pour lesquels on veut évaluer les tests.

Kumar, 2015

	Risk of bias				Applicability		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Kumar 2015	High	Low	Low	Low	High	Low	Low

Cette étude présente des biais de sélection élevés ainsi que des biais d'applicabilité élevés du fait qu'elle n'inclue que des patients avec une histoire de luxation antérieure d'épaule récurrente et exclue les patients qui se la luxent pour la première fois. Or l'instabilité antérieure d'épaule peut faire suite directement à la première luxation et s'exprimer par des symptômes autres que la reluxation. De plus, elle exclut les patients qui ressentent de la douleur au test d'appréhension quand certaines études utilisent la douleur comme critère.

Van Kampen, 2013

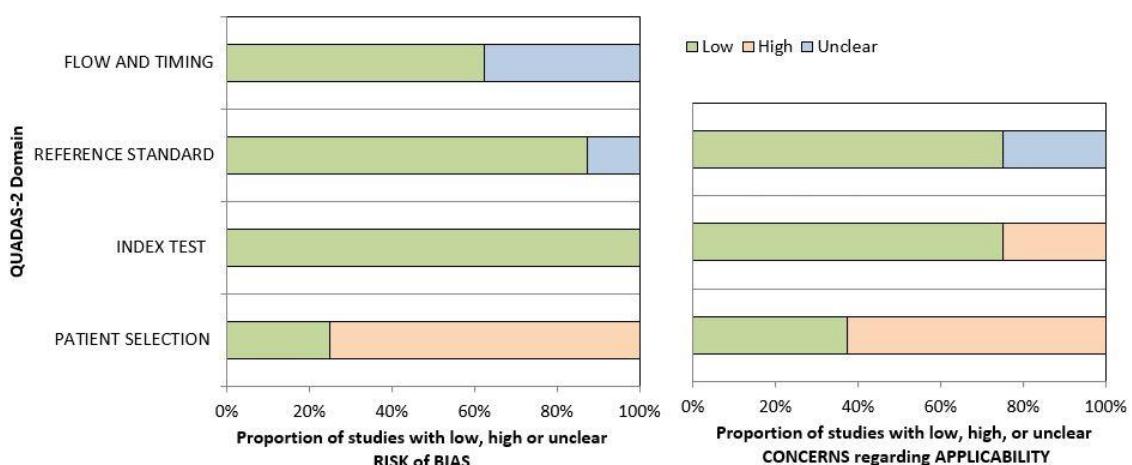
	Risk of bias				Applicability		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Van Kampen 2013	Low	Low	Low	Low	Low	Low	Low

Cette étude ne présente que des risques de biais faibles.

En résumé :

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Farber 2006	😊	😊	?	😊	😊	😊	😊
Lo 2004	😊	😊	😊	?	😊	😊	😊
Safran 2010	😊	😊	😊	?	😊	😊	?
Milgrom 2014	😊	😊	😊	?	😊	😊	?
Guanche 2003	😊	😊	😊	😊	😊	😊	😊
Loh 2016	😊	😊	😊	😊	😊	😊	😊
Kumar 2015	😊	😊	😊	😊	😊	😊	😊
Van Kampen 2013	😊	😊	😊	😊	😊	😊	😊

😊 Low Risk 😊 High Risk ? Unclear Risk



Sur ce graphique, on peut voir la synthèse des résultats de chacune des études. Nous pouvons notamment voir que les risques de biais de temporalité sont faibles en majorité mais 40% d'études présentent des risques incertains.

Le risque de biais concernant le test de référence est faible malgré 20% d'études avec des risques incertains.

Le risque de biais concernant les tests évalués est bas dans toutes les études.

Le risque de biais de sélection des patients est majoritairement haut avec 80% des études concernées.

Pour l'applicabilité des résultats, on retrouve aussi 20% de risques incertains mais une majorité de risque faible pour le test de référence, 20% d'études avec un risque élevé et 80% avec risque faible concernant les tests évalués. Pour le risque de biais de sélection des

patients, seules 40% des études ont un risque de biais faible alors que le reste présente un risque de biais élevé.

Pour conclure, on peut voir que nos études présentent majoritairement un faible risque de biais, malgré certains paramètres qui présentent un risque de biais incertain et d'autres paramètres qui seront à prendre en compte à cause de leur risque de biais majoritairement élevé (sélection des patients). Ce biais rendra l'extrapolation des résultats des études sur la population générale plus imprécise.

3.3 Effets de l'intervention

Après avoir estimé les risques de biais de chaque étude, on a procédé à l'extraction des résultats, en nous intéressant aux données chiffrées qui étaient présentées. Comme notre étude ne porte que sur la sensibilité et la spécificité de certains tests d'instabilité antérieure d'épaule, on ne va extraire que ces données. Lorsque celles-ci ne seront pas disponibles nous pourront les calculer à condition d'avoir assez de données.

Pour calculer la sensibilité il nous faut savoir le nombre total de patients soumis au test, le nombre de patients avec un résultat positif et parmi eux ceux avec un vrai positif (diagnostic confirmé par le test de référence) et le nombre de faux négatifs (patients dont le résultat du test évalué est négatif mais sont diagnostiqués positifs par le test de référence).

La formule mathématique est donc :

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

Pour calculer la spécificité, il nous faut le nombre de vrais négatifs (patients diagnostiqués négatifs donc le test de référence confirme le diagnostic négatif) et le nombre de faux positifs (patients que le test évalué classe positifs mais dont le diagnostic est invalidé par le test de référence).

La formule mathématique est donc :

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

		STATUT RÉEL		
		Cas	Non-cas	Total
RÉSULTAT	Positif	VP (vrais positifs)	FP (faux positifs)	VP+FP
	Négatif	FN (faux négatifs)	VN (vrais négatifs)	FN+VN
Total		VP+FN	FP+VN	

On utilisera aussi ces formules pour vérifier les chiffres fournis quand on aura les données nécessaires à la vérification.

Comme cette étude est une analyse de validité de critères à variable binaire (test positif ou négatif), il peut aussi être intéressant de calculer le rapport de vraisemblance (RV). Le rapport de vraisemblance positif (RV+) traduit la probabilité d'avoir un test positif chez un sujet malade par rapport à un sujet sain, il donne un résultat compris entre un et l'infini. Le rapport de vraisemblance négatif traduit la probabilité d'avoir un test négatif chez un sujet sain par rapport à un sujet malade, il donne une valeur comprise entre 0 et 1.

On les calcule à partir des valeurs de la sensibilité et de la spécificité selon les formules :

$$RV+ = Se / (1 - Sp)$$

$$RV- = (1 - Se) / Sp$$

On interprète ces valeurs selon des seuils qui sont :

Après avoir recueilli les données, les avoir vérifiées et avoir comblé les différents manques grâce au calcul lorsque les informations nécessaires étaient présentes, on peut résumer les résultats des études dans un tableau.

Dans ce tableau, nous pouvons voir la sensibilité et la spécificité de chaque test dans chaque étude, pour certaines études, si plusieurs critères de jugement sont utilisés pour un seul test et que les sensibilité et spécificité sont disponibles pour chaque critère de jugement alors ces valeurs seront aussi présentées. On y fera aussi figurer les rapports de vraisemblance positif et négatif associé à chaque couple sensibilité-spécificité.

RV+	Interprétation	RV-
>10	Bonne	< 0,1
5,0–10,0	Modérée	0,1–0,2
2,0–5,0	Faible	0,2–0,5
1,0–2,0	Rarement important	0,5–1,0

Test	Étude	Critère de jugement	Se %	Spé %	RV+	RV-
Appréhension antérieure	Farber	Appréhension	72	96	18	0.30
		Douleur	50	56	1.14	0.9
	Lo	Douleur, appréhension ou étirement	52.8	98.9	48	0.48
	Safran	Appréhension, résistance	41.6	85.7	2.91	0.68
	Milgrom	Appréhension, résistance	45.8	47.4	0.87	1.14
	Guanche	Douleur	40	87	3.08	0.69
	Kumar	Appréhension	86.3	14.3	1	0.96
	Van Kampen	Appréhension	98.3	71.6	3.46	0.02

Test	Étude	Critère de jugement	Se %	Spé %	RV+	RV-
Relocation	Farber	Appréhension	81	92	10.13	0.21
		Douleur	30	90	3	0.78
	Lo	Appréhension	31.9	100	X	0.54
		Appréhension ou douleur	45.9	54.4	1	0.99
	Douleur	40	42.7	0.7	1.41	
	Guanche	Douleur	44	87	3.38	0.64
	Van Kampen	Appréhension	96.7	78	4.40	0.04

Test	Étude	Critère de jugement	Se %	Spé %	RV+	RV-
Release	Lo	Douleur, appréhension ou étirement	63.9	98.9	58.1	0.37
	Van Kampen	Appréhension	91.7	83.5	5.56	0.1

Test	Étude	Critère de jugement	Se %	Spé %	RV+	RV-
Load and shift	Loh	Appréhension	94	X	X	X
	Van Kampen	Appréhension ou grade 2/3	71.7	89.9	7.1	0.31

Dans ce tableau, nous pouvons voir les différentes valeurs calculées pour chaque test. Les valeurs des tests reposant sur le critère de la douleur sont grisées pour les mettre en évidence, on peut ainsi voir que certains auteurs différencient bien les résultats de tests utilisant la douleur comme critère de jugement de ceux utilisant l'appréhension, la plupart des auteurs n'utilisant pas la douleur pour le diagnostic, certains excluant même les patients éprouvant de la douleur au moment du test [13]. Nous allons nous y intéresser au moment de l'analyse des résultats pour savoir en quoi choisir la douleur comme critère de jugement affecte les propriétés d'un de ces tests.

Pour chaque test, nous avons choisi de mettre en évidence les valeurs extrêmes qui semblaient sortir de la moyenne générale qui se dégageait de l'ensemble les comprenant. Les valeurs affichées en vert sont les valeurs les plus hautes et en rouge sont les valeurs les plus basses pour la sensibilité et la spécificité.

Pour les RV+ et RV-, nous avons mis en rouge les valeurs les moins bonnes à l'interprétation et en vert celles avec une interprétation « bonne ».

Pour le test d'appréhension

On peut voir que la majorité des valeurs pour la sensibilité semble être autour de 50% alors que pour la spécificité elle se situe autour de 90%.

Pour la sensibilité, les études présentant des valeurs supérieures à la moyenne sont les études de Farber (basée sur l'appréhension), Kumar et Van Kampen.

Celles présentant des valeurs inférieures à la moyenne sont celles de Farber (basée sur la douleur) et Guanche.

Pour la spécificité, les études présentant des valeurs supérieures à la moyenne sont les études de Farber (basée sur l'appréhension) et Lo.

Celles présentant les valeurs les plus basses sont celles de Farber (basée sur la douleur), Milgrom et Kumar.

Pour le RV+, les études présentant les valeurs interprétables comme « bonne » sont celles de Farber (basée sur l'appréhension) et Lo, et celles présentant des valeurs interprétables comme « rarement importantes » sont celles de Farber (basée sur la douleur), Milgrom et Kumar.

Pour le RV-, l'étude avec des valeurs interprétables comme « bonne » est celle de Van Kampen et celles avec des valeurs interprétables comme « rarement important » sont celles de Farber (basée sur la douleur), Milgrom et Kumar.

Pour le relocation test

On peut voir que pour la sensibilité, les valeurs hautes sont celles des études de Farber (basée sur l'appréhension) et Van Kampen. Les valeurs basses sont celles des études de Farber (basée sur la douleur) et Lo (basée sur l'appréhension).

Pour la spécificité, les études présentant des valeurs hautes sont celles de Farber (basée sur la douleur) et Lo (basée sur l'appréhension) alors que les valeurs les plus basses sont celles de l'étude de Lo avec comme critère de jugement l'appréhension/douleur et celle avec la douleur.

Pour le RV+, la valeur « bonne » est celle de l'étude de Farber (basée sur l'appréhension), alors que les valeurs « rarement importantes » sont celles de l'étude de Lo basée sur appréhension/douleur et celles basées sur la douleur.

Pour le RV-, la valeur « bonne » est celle de l'étude de Van Kampen alors que les valeurs « rarement importantes » sont celles de l'étude de Lo basée sur appréhension/douleur et celles basées sur la douleur.

Pour le release test

Pour la sensibilité, on peut voir que la meilleure des deux valeurs est celle de l'étude de Van Kampen, pour la spécificité c'est celle de Lo, pour le RV+ c'est aussi celle de Lo qui est « bonne » alors que la valeur de l'étude de Van Kampen est « modérée », pour le RV- la valeur de l'étude de Van Kampen est « bonne » alors que celle de Lo est « faible ».

Pour le load and shift test

Pour la sensibilité, c'est l'étude de Loh qui présente la plus grande valeur, mais étant donné qu'elle ne fournit pas d'autres valeurs que celle de la sensibilité on ne peut pas faire de comparaison pour les autres valeurs.

4 Discussion

Le but de cette revue est d'établir un état des lieux le plus récent possible concernant la validité de 4 tests d'instabilité antérieure d'épaule. Après l'extraction des résultats des études présentée dans la partie précédente, nous allons, dans cette partie, présenter les analyses de ces résultats. Pour ce faire, nous regrouperons les données extraites pour chaque test afin de calculer pour chaque test la moyenne et l'intervalle de confiance à 95% de sa sensibilité, sa spécificité, son rapport de vraisemblance positif et négatif. Nous regrouperons ces résultats dans un tableau pour les rendre plus lisibles afin d'en tirer une conclusion.

4.1 Analyse des principaux résultats

Nous avons vu précédemment que les études présentant les résultats les plus hauts pour les tests sont les études de Farber (pour les tests basés sur l'appréhension), l'étude de Lo (pour le test d'appréhension, et le release test) et l'étude de Van Kampen alors que les études avec

les résultats les plus faibles sont celles de Milgrom, Lo (pour le relocation test) et Farber (tests basés sur la douleur).

Or on sait que pour ces études : Farber peut présenter une surestimation des résultats à cause de son schéma d'étude cas-témoin (voir partie 3.2). Mais cette étude présente aussi des résultats parmi les plus bas quand le critère de jugement utilisé pour le test était la douleur. On peut donc penser que ce critère de jugement peut ne pas être le plus approprié sur les tests évalués par Farber (appréhension et relocation test) par rapport à l'appréhension.

L'étude de Lo présente les résultats les plus extrêmes parmi les études, avec des résultats parmi les plus hauts ainsi que d'autres parmi les plus bas. On avait vu précédemment qu'elle présentait un biais de sélection des patients élevé ainsi qu'un risque de biais incertain au niveau de la temporalité de l'expérience.

L'étude de Van Kampen présentait un risque de biais assez faible en général, le seul biais que l'on pourrait lui reprocher éventuellement est d'être une étude réalisée uniquement sur des sujets de la même nationalité.

L'étude de Milgrom présente des valeurs parmi les plus basses mais comme on a vu précédemment que le test de référence utilisé était la présence de relaxation à long terme chez des militaires on peut avancer l'hypothèse qu'un échantillon de militaires n'a pas les mêmes contraintes physiques au quotidien que des civils, même des sportifs, et donc que le test peut ne pas correspondre à ce type de population.

Synthèse des résultats

Test	Critère de positivité	Effectif	Sensibilité %	Spécificité %	RV+	RV-	Test de référence
Apprehension	Appréhension et douleur	6	66,13 (47,65 ; 84,62)	68,98 (42,77 ; 95,2)	10,61 (- 2,45 ; 23,66)	0,6 (0,29 ; 0,91)	Arthroscopie, relaxation, diagnostic préexistant, imagerie
	Douleur	2	45 (35,2 ; 54,8)	61,5 (11,52 ; 111,48)	0,62 (0,47 ; 0,78)	1,3 (0,85 ; 1,75)	
Relocation	Appréhension	3	69,87 (31,61 ; 108,12)	89,67 (77,57 ; 101,77)	11,61 (- 2,62 ; 25,47)	0,31 (- 0,02 ; 0,64)	
	Appréhension OU Douleur	1	45,9	54,4	1	0,99	
	Douleur	3	38 (29,84 ; 46,16)	73,23 (43,26 ; 103,2)	1,52 (- 1,42 ; 4,46)	1,31 (0,12 ; 2,49)	

Release	Appréhension, douleur, étirement	2	77,8 (50,56 ; 105)	91,2 (76,11 ; 106,29)	31,84 (- 19,66 ; 83,31)	0,23 (-0,02 ; 0,49)	
Load and shift	Appréhension et grade 2 ou 3 de la cotation	1	71,7	89,9	7,1	0,31	

Nous avons fait le choix de séparer les valeurs des tests basés sur la douleur de ceux basés sur l'appréhension quand cela était possible, pour leur comparaison. Nous avons isolé dans ces calculs les valeurs données dans l'étude de Lo lorsque le test avait pour critère de jugement appréhension OU douleur car l'inclusion de celles-ci dans n'importe lequel des ensembles de valeurs (critère de jugement : appréhension/ douleur) aurait risqué de biaiser la comparaison. Cette comparaison met en évidence le fait qu'utiliser la douleur comme critère de jugement dans les tests d'appréhension et de recentrage est moins efficace qu'utiliser l'appréhension. On peut notamment voir qu'en utilisant la douleur on obtient des valeurs beaucoup moins élevées pour tous les paramètres évalués pour ces tests. Néanmoins, pour l'apprehension test on peut se permettre de nuancer ce jugement car nous n'avons eu que 2 études qui décrivaient le test selon ce critère de jugement contre 6 études basées sur l'appréhension. En revanche, pour le relocation test, le nombre d'études selon chaque critère de jugement est identique et leur qualité méthodologique étant à peu près équivalente on peut estimer cette comparaison comme étant proche de la réalité, pour pouvoir donner un avis plus tranché, il faudrait avoir plus d'études sur le sujet.

On peut cependant retenir que l'appréhension semble être un critère de jugement plus efficace que la douleur pour ces deux tests. Il serait par ailleurs intéressant de pouvoir comparer l'efficacité diagnostique des autres tests présents selon ces deux critères de jugement, malheureusement, le manque d'informations à ce sujet ne nous le permet pas. De plus, en regardant les résultats, on voit que le relocation test est aussi décrit dans l'étude de Lo avec comme critère appréhension OU douleur et que la sensibilité du test est meilleure selon cette combinaison de critères que pour chaque critère isolément. La spécificité est, elle, située entre les valeurs des critères isolés comparée avec les résultats pour ces critères dans la même étude. De cette observation, on peut déduire que la combinaison des critères permet d'identifier la pathologie chez plus de sujets que chaque critère pris indépendamment, en revanche on obtient un nombre de faux positifs plus élevé qu'en utilisant juste l'appréhension, mais toujours moins élevé qu'en utilisant la douleur seule. La combinaison des critères pourrait donc être un bon moyen d'augmenter la sensibilité du test au détriment de sa spécificité.

Pour le Load and shift test nous n'avons pas pu calculer toutes les valeurs voulues à cause du manque de données présentes dans les études (1 seule valeur pour la spécificité). De plus les données de l'étude de Loh n'ont pas été utilisées étant donné que cette étude ne donnait que la sensibilité de deux tests combinés, le load and shift test associé au test d'appréhension. Nous avons jugé que ces chiffres n'étaient pas pertinents du fait qu'on ne peut pas savoir quelle est la contribution de chaque test à ce résultat, le prendre en compte pourrait donc

fauisser grandement l'analyse faite dans cette étude, surtout que pour le load and shift test le nombre d'études est très faible étant donné qu'il n'est décrit que dans une seule autre étude. Or, compte tenu du très faible nombre d'études, on risque d'influencer d'autant plus les valeurs qu'on pourrait calculer si on incluait les résultats de l'étude de Loh.

Etant donné la faible différence dans la quantité de biais relevés entre chaque étude, nous avons choisi d'inclure toutes les autres études dans la synthèse des résultats mais il faut garder à l'esprit que seules 2 d'entre elles avaient un risque de biais général faible, donc que la synthèse des résultats ici présentée ne doit pas être estimée comme étant absolument exacte.

De plus, parmi les études, toutes n'ont pas le même test de référence, pas la même taille d'échantillon ou caractéristiques de l'échantillon, pas le même protocole, ni le même critère de jugement pour un même test. Mais étant donné le faible nombre d'études trouvé, nous avons choisi d'utiliser leurs résultats en prenant en compte ces risques de biais et d'hétérogénéité des études.

Ces différences entre les études peuvent entraîner une grande variabilité dans les résultats. Nous avons donc pour chaque test calculé les moyennes et intervalles de confiance à 95% le concernant pour nous permettre une estimation de ces résultats dans des populations plus importantes.

En analysant les résultats, nous pouvons voir que les tests présentant les meilleures propriétés sont dans l'ordre du meilleur au moins bon :

- le release test,
- le load and shift test
- le relocation test basé sur l'appréhension
- l'apprehension test basé sur l'appréhension
- Le relocation test basé sur la douleur
- l'apprehension test basé sur la douleur

Le release test est celui avec les meilleures sensibilité et spécificité, ainsi que les meilleurs rapports de vraisemblance. Néanmoins, il est nécessaire de prendre en compte le fait que les données pour ce test ne proviennent que de 2 études différentes, ce qui est très peu et peut représenter un biais important. Plus il y a d'études documentant un test, plus les estimations données pour ce test sur une population plus large seront précises. Par ailleurs les études qui documentent ce test sont assez différentes, le rapport de vraisemblance est 11 fois plus élevé dans une étude que dans l'autre (5 contre 58), le test de référence utilisé diffère aussi ainsi que la population des études ce qui rend une certaine prudence nécessaire à leur interprétation.

Nous savons par ailleurs que le test d'appréhension antérieure avec comme critère de jugement l'appréhension est le plus documenté parmi ceux présentés ici, donc le plus à même d'avoir des valeurs proches de la réalité. Cependant, les intervalles de confiance pour sa sensibilité et sa spécificité sont très larges (entre 40 et 50% d'intervalle), ce qui démontre que pour avoir des résultats plus précis, il faudrait avoir plus d'études le décrivant pour pouvoir

les inclure aux calculs. Or, les autres tests sont encore moins documentés, on a donc plus de mal à émettre un avis qui soit vraiment tranché quant à l'efficacité de ces divers tests.

Parmi nos résultats, nous observons que la sensibilité des tests se situe pour tous les tests entre 65 et 80% et la spécificité entre 70 et 95%.

Pour le test d'apprehension antérieure basé sur l'apprehension, le rapport de vraisemblance positif est bon et le rapport de vraisemblance négatif est rarement important alors que si on prend la douleur comme critère le rapport de vraisemblance positif comme le négatif est rarement important.

Pour le test de recentrage basé sur l'apprehension seule, le rapport de vraisemblance positif est bon alors que le rapport de vraisemblance négatif est faible. Pour les autres critères pour ce test(apprehension OU douleur ou douleur seule) les rapports de vraisemblance sont rarement importants.

On observe à nouveau une différence statistique entre les tests basés sur l'apprehension et ceux basés sur la douleur, en faveur de ceux utilisant l'apprehension.

Pour le release test, le rapport de vraisemblance positif est bon alors que le négatif est faible.

Pour le load and shift test, le rapport de vraisemblance positif est modéré et le négatif est faible.

Selon ces paramètres, le release test paraît avoir les meilleures propriétés, suivi par le test de recentrage basé sur l'apprehension. Ces deux tests présentent les meilleures propriétés métrologiques, ce qui se traduit par une meilleure capacité discriminative entre les sujets sains et ceux présentant une instabilité antérieure d'épaule.

4.2 Applicabilité des résultats en pratique clinique

En pratique, ces tests ne demandent aucun investissement matériel couteux contrairement à l'imagerie qui est utilisée comme « gold standard », ils ne demandent qu'un peu de temps et de formation, qui peut aussi être une auto-formation pour savoir les réaliser. Ces tests peuvent être réalisés par un médecin aussi bien qu'un kinésithérapeute et les effectuer sur les patients à risques ou après une luxation pourrait permettre d'adresser un patient avec une instabilité antérieure d'épaule chez un chirurgien assez tôt pour éviter une autre luxation ou des lésions associées (lésions nerveuses, artérielles, de Bankart, SLAP lésion etc). En revanche, les critères de jugement de ces tests étant subjectifs, ils doivent au maximum être effectués en suivant un protocole précis pour éviter des biais d'interprétation de la part de l'examineur.

Pour le patient, ces tests peuvent provoquer un inconfort ou une douleur selon le critère de jugement choisi par l'examineur.

Pour pratiquer un de ces tests, il faut aussi choisir si on l'effectue sur un patient allongé sur le dos ou assis, ce qui implique que le patient puisse se mettre dans ces positions et que le

praticien puisse se positionner de manière à effectuer le test. En cabinet, il faut donc une table ou un tabouret/chaise mais on peut aussi effectuer les tests en extérieur dans le cas de patient sportif ou militaire sur le terrain d'entraînement ressentant une instabilité ou des subluxations dans l'épaule par exemple.

Cette étude semble nous permettre de désigner le release test comme étant le meilleur des 4 test par rapport à ses propriétés métrologiques. Mais étant donné la proximité des valeurs de ces tests, on pourrait aussi choisir de préférer un test un peu plus documenté et de valeur à peu près équivalente avec le test de recentrage basée sur l'appréhension, ou choisir le test le plus documenté qui est le test d'appréhension basé sur l'appréhension avec un nombre d'études le décrivant plus important.

Chaque test présente une sensibilité moins grande que l'ARM (arthrographie par rayonnement magnétique) (91%) mais à peu près équivalente à l'IRM (70%) lues par des radiologues musculosquelettiques selon l'article [13].

4.3 Qualité des preuves

Afin d'estimer la pertinence des études incluses dans notre revue, nous allons établir leur niveau de preuve et leur grade de recommandation par la HAS. Pour cela nous allons suivre les recommandations du système GRADE (Annexe 2). A partir du schéma d'étude, nous pouvons établir le niveau de preuve scientifique ainsi que le grade et le niveau d'évidence scientifique correspondant. Nous obtenons alors ce tableau :

Etude	Niveau de preuve	Evidence scientifique	Grade
Farber	3	Faible niveau de preuve scientifique	C
Lo	2	Présomption scientifique	B
Safran	2	Présomption scientifique	B
Milgrom	2	Présomption scientifique	B
Guanche	2	Présomption scientifique	B
Kumar	4	Faible niveau de preuve scientifique	C
Van Kampen	2	Présomption scientifique	B

Ainsi, nous pouvons voir que le niveau de preuve prépondérant est un niveau 2 (1 étant le meilleur et 4 le pire), mais notre étude comporte aussi deux études de moindre niveau de preuve dont une avec le pire niveau de preuve. Pour le grade, nous avons 5 études de Grade B et 2 de grade C ; ce qui représente une majorité de présomption scientifique avec un peu de faible niveau de preuve scientifique. Nous pouvons conclure de ce niveau de preuve modéré

et des différents biais présents dans les articles et la rédaction de notre revue que notre étude présente elle aussi un faible niveau de preuve. Pour obtenir un niveau de preuve plus élevé, il aurait fallu inclure plus d'études de cohorte et exclure les schéma d'étude cas-témoin (niveau de preuve 4) et cohorte rétrospective (niveau de preuve 3).

4.4 Biais potentiels de la revue

Afin de réaliser une analyse critique des biais méthodologiques de notre travail, nous avons choisi de nous appuyer sur la grille AMSTAR qui cote la validité interne des revues de littérature. Nous avons utilisé la version donnée par [14] et présentée en Annexe (Annexe 3). Chaque item peut avoir comme réponse « Oui », « Non » ou « Incertain »

Item	Cotation	Commentaire
1	Oui	Voir les parties équations de recherche et tableau PICO
2	Non	La revue est un mémoire de fin d'étude sans protocole pré-établi
3	Oui	
4	Oui	Recherche sur au moins 2 bases de données ainsi que dans les références des articles
5	Non	Travail réalisé par une seule personne
6	Non	Travail réalisé par une seule personne
7	Oui	Les études exclues sont présentées dans un tableau partie 3
8	Incertain	Il faudrait pouvoir comparer les populations des études
9	Oui	Quadas 2
10	Non	Sources de financement non investiguées
11	Oui	Les valeurs intéressantes pour la validité de critère sont la sensibilité, spécificité et RV+/RV-
12	Non	
13	Oui	Voir la partie 4
14	Oui	Voir la partie 4
15	Non	
16	Oui	Pas de financement reçu ni conflit d'intérêt
Total	9/16	Le résultat témoigne d'une qualité méthodologique moyenne

La grille d'AMSTAR servant à l'analyse des biais méthodologiques présents dans la revue de littérature, on peut rechercher d'autres biais non méthodologiques qui peuvent avoir affecté notre étude :

- Un biais de sélection vis-à-vis du langage des études trouvées, nous n'avons inclus que les études en français ou anglais
- L'hétérogénéité des études, qui n'utilisent pas toutes les tests de la même manière, avec des comparateurs différents entre les études et des critères de jugement qui peuvent être différents. De plus, la pathologie recherchée par les tests n'est pas toujours décrite de la même façon.

5 Conclusion

5.1 Implication pour la pratique clinique

En conclusion, on peut retenir que dans la pratique, les tests d'appréhension antérieure, de recentrage, de surprise et le load and shift présentent chacun une sensibilité pouvant rivaliser avec les IRM pour le diagnostic des instabilités antérieures d'épaule. Il peut donc être intéressant, en cas de doute par rapport à l'épaule d'un patient, de le soumettre à l'un de ces examens en première intention. Cela permettrait ainsi de ne pas exposer le patient aux rayonnements de l'imagerie et d'avoir un diagnostic à un coût moindre pour la sécurité sociale.

De plus, soumettre le patient à plusieurs de ces tests à la fois peut permettre d'affiner son bilan. Il peut aussi être intéressant de choisir le critère de jugement du test en fonction des besoins pour les tests d'appréhension et de recentrage. Si on désire un résultat plus sensible pour ces tests, on utilisera l'appréhension ou la douleur comme critère, si on veut un résultat plus spécifique on n'utilisera que l'appréhension par exemple.

Par ailleurs, les tests de recentrage et de surprise sont des tests qui ne s'effectuent qu'après un test d'appréhension car ils sont jugés sur la base de l'aggravation ou disparition des symptômes ressentis lors de ce premier test, ils sont donc dépendants de celui-ci.

Le test d'appréhension antérieure étant le seul test effectuable indépendamment des autres, il peut être réalisé seul si on veut économiser du temps mais il faut garder à l'esprit que le résultat d'un test seul est moins précis que celui d'un ensemble de tests concernant la même pathologie.

Pour le load and shift test, on ne peut pas émettre de conclusion tranchée à cause du trop faible nombre d'études le décrivant comme test diagnostique de cette pathologie.

5.2 Implication pour la recherche

Du côté de la recherche, on s'aperçoit en définitive qu'assez peu d'études ont été faites sur le sujet, surtout parmi des populations standard sans critères d'exclusion qui n'induisent un biais de sélection des patients important.

De plus, on voit que contrairement au test d'apprehension antérieure, les release test et load and shift sont très peu documentés. On retrouve également peu d'études sur la différence d'efficacité diagnostique en fonction du critère de jugement choisi pour chaque test.

Il pourrait être intéressant de faire davantage d'études sur le sujet, afin d'en avoir plusieurs avec un protocole similaire, ce qui permettrait de comparer les résultats selon chaque critère de jugement. Comme dit précédemment, plus d'études permettrait d'affiner les intervalles de confiance pour chaque valeur et ainsi d'obtenir des estimations plus précises.

Pour finir sur une question d'ouverture, il pourrait être intéressant de comparer l'efficacité diagnostique de chaque test selon chacun des critères de jugement pris isolément pour déterminer s'il est plus intéressant d'utiliser les critères de jugement isolément ou en les combinant.

Sources

- [1] FARBER AJ, CASTILLO R, CLOUGH M, BAHK M, MCFARLAND EG. Clinical Assessment of Three Common Tests for Traumatic Anterior Shoulder Instability. *J Bone Jt Surgery-American Vol* 2006;88:1467–74. <https://doi.org/10.2106/00004623-200607000-00006>.
- [2] Tzannes A, Murrell GAC. Clinical examination of the unstable shoulder. *Sport Med* 2002;32:447–57. <https://doi.org/10.2165/00007256-200232070-00004>.
- [3] van Kampen DA, van den Berg T, van der Woude HJ, Castelein RM, Terwee CB, Willems WJ. Diagnostic value of patient characteristics, history, and six clinical tests for traumatic anterior shoulder instability. *J Shoulder Elb Surg* 2013;22:1310–9. <https://doi.org/10.1016/j.jse.2013.05.006>.
- [4] Lo IKY, Nonweiler B, Woolfrey M, Litchfield R, Kirkley A. An Evaluation of the Apprehension, Relocation, and Surprise Tests for Anterior Shoulder Instability. *Am J Sports Med* 2004;32:301–7. <https://doi.org/10.1177/0095399703258690>.
- [5] Hegedus EJ, Goode AP, Cook CE, Michener L, Myer CA, Myer DM, et al. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. *Br J Sports Med* 2012;46:964–78. <https://doi.org/10.1136/bjsports-2012-091066>.
- [6] Morey VM, Singh H, Paladini P, Merolla G, Phadke V, Porcellini G. The Porcellini test: a novel test for accurate diagnosis of posterior labral tears of the shoulder: comparative analysis with the established tests. *Musculoskeletal Surg* 2016;100:199–205. <https://doi.org/10.1007/s12306-016-0422-3>.
- [7] Lizzio VA, Meta F, Fidai M, Makhni EC. Clinical Evaluation and Physical Exam Findings in Patients with Anterior Shoulder Instability. *Curr Rev Musculoskelet Med* 2017;10:434–41. <https://doi.org/10.1007/s12178-017-9434-3>.
- [8] Pandya NK, Colton A, Webner D, Sennett B, Huffman GR. Physical Examination and Magnetic Resonance Imaging in the Diagnosis of Superior Labrum Anterior-Posterior Lesions of the Shoulder: A Sensitivity Analysis. *Arthrosc - J Arthrosc Relat Surg* 2008;24:311–7. <https://doi.org/10.1016/j.arthro.2007.09.004>.
- [9] Safran O, Milgrom C, Radeva-Petrova DR, Jaber S, Finestone A. Accuracy of the anterior apprehension test as a predictor of risk for redislocation after a first traumatic shoulder dislocation. *Am J Sports Med* 2010;38:972–5. <https://doi.org/10.1177/0363546509357610>.
- [10] Milgrom C, Milgrom Y, Radeva-Petrova D, Jaber S, Beyth S, Finestone AS. The supine apprehension test helps predict the risk of recurrent instability after a first-time anterior shoulder dislocation. *J Shoulder Elb Surg* 2014;23:1838–42. <https://doi.org/10.1016/j.jse.2014.07.013>.
- [11] Guanche CA, Jones DC. Clinical testing for tears of the glenoid labrum. *Arthrosc - J Arthrosc Relat Surg* 2003;19:517–23. <https://doi.org/10.1053/jars.2003.50104>.
- [12] Loh B, Lim JBT, Tan AHC. Is clinical evaluation alone sufficient for the diagnosis of a Bankart lesion without the use of magnetic resonance imaging? *Ann Transl Med* 2016;4. <https://doi.org/10.21037/atm.2016.11.22>.
- [13] Kumar K, Makandura M, Leong NJ, Gartner L, Hwee Lee C, Ng DZ, et al. Is the Apprehension Test Sufficient for the Diagnosis of Anterior Shoulder Instability in Young Patients without

Magnetic Resonance Imaging (MRI)? n.d.

- [14] Pallot A, Rostagno S. AMSTAR-2 : traduction française de l'échelle de qualité méthodologique pour les revues de littérature systématiques. *Kinesither Rev* (2019), <http://dx.doi.org/10.1016/j.kine.2019.12.050>
- [15] Dufour, anatomie de l'appareil locomoteur membre supérieur. 3^e édition, Elsevier Masson
- [16] Jonathan Leconte. Instabilité de l'épaule du sportif : " Pourquoi ça fait mal ? " Revue systématique de la littérature : Quel est le lien entre douleur et instabilité de l'épaule ? Médecine humaine et pathologie. 2018. ffdumas-01906012f
- [17] <http://physiotherapytest.com/tests/orthopedie/epaule/>
- [18] http://chirurgie-epaule-fontvert.fr/instabilite_luxation.html
- [19] The development of QUADAS-2: A revised tool for the quality assessment of diagnostic test accuracy studies (bristol.ac.uk)

Annexe

Annexe 1 :

Grille de lecture critique QUADAS-2

QUADAS-2

Phase 1: State the review question:

<i>Patients (setting, intended use of index test, presentation, prior testing):</i>
<i>Index test(s):</i>
<i>Reference standard and target condition:</i>

Phase 2: Draw a flow diagram for the primary study

--

Phase 3: Risk of bias and applicability judgments

QUADAS-2 is structured so that 4 key domains are each rated in terms of the risk of bias and the concern regarding applicability to the research question (as defined above). Each key domain has a set of signalling questions to help reach the judgments regarding bias and applicability.

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Describe methods of patient selection:

- ❖ Was a consecutive or random sample of patients enrolled? Yes/No/Unclear
- ❖ Was a case-control design avoided? Yes/No/Unclear
- ❖ Did the study avoid inappropriate exclusions? Yes/No/Unclear

Could the selection of patients have introduced bias? RISK: LOW/HIGH/UNCLEAR

B. Concerns regarding applicability

Describe included patients (prior testing, presentation, intended use of index test and setting):

Is there concern that the included patients do not match the review question? CONCERN: LOW/HIGH/UNCLEAR

DOMAIN 2: INDEX TEST(S)

If more than one index test was used, please complete for each test.

A. Risk of Bias

Describe the index test and how it was conducted and interpreted:

- ❖ Were the index test results interpreted without knowledge of the results of the reference standard? Yes/No/Unclear
- ❖ If a threshold was used, was it pre-specified? Yes/No/Unclear

Could the conduct or interpretation of the index test have introduced bias? RISK: LOW /HIGH/UNCLEAR

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW /HIGH/UNCLEAR

DOMAIN 3: REFERENCE STANDARD**A. Risk of Bias**

Describe the reference standard and how it was conducted and interpreted:

- ❖ Is the reference standard likely to correctly classify the target condition? Yes/No/Unclear
- ❖ Were the reference standard results interpreted without knowledge of the results of the index test? Yes/No/Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias?

RISK: LOW /HIGH/UNCLEAR

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?

CONCERN: LOW /HIGH/UNCLEAR

DOMAIN 4: FLOW AND TIMING**A. Risk of Bias**

Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):

Describe the time interval and any interventions between index test(s) and reference standard:

- ❖ Was there an appropriate interval between index test(s) and reference standard? Yes/No/Unclear
- ❖ Did all patients receive a reference standard? Yes/No/Unclear
- ❖ Did patients receive the same reference standard? Yes/No/Unclear
- ❖ Were all patients included in the analysis? Yes/No/Unclear

Could the patient flow have introduced bias?

RISK: LOW /HIGH/UNCLEAR

Annexe 2 :

Haute autorité de Santé : Niveau de preuve et Grade de recommandation des études

Grade des recommandations	Niveau de preuve scientifique fourni par la littérature
A Preuve scientifique établie	Niveau 1 - essais comparatifs randomisés de forte puissance ; - méta-analyse d'essais comparatifs randomisés ; - analyse de décision fondée sur des études bien menées.
B Présomption scientifique	Niveau 2 - essais comparatifs randomisés de faible puissance ; - études comparatives non randomisées bien menées ; - études de cohortes.
C Faible niveau de preuve scientifique	Niveau 3 - études cas-témoins. Niveau 4 - études comparatives comportant des biais importants ; - études rétrospectives ; - séries de cas ; - études épidémiologiques descriptives (transversale, longitudinale).

Annexe 3 :

Grille AMSTAR [14]

Tableau I. Traduction française de AMSTAR-2.

Items	Cotation
1	Est-ce que les questions de recherche et les critères d'inclusion de la revue ont inclus les critères PICO ?
2	Est-ce que le rapport de la revue contenait une déclaration explicite indiquant que la méthode de la revue a été établie avant de conduire la revue ? Est-ce que le rapport justifiait toute déviation significative par rapport au protocole ?
3	Les auteurs ont-ils expliqué leur choix de schémas d'étude à inclure dans la revue ?
4	Les auteurs ont-ils utilisé une stratégie de recherche documentaire exhaustive ?
5	Les auteurs ont-ils effectué en double la sélection des études ?
6	Les auteurs ont-ils effectué en double l'extraction des données ?
7	Les auteurs ont-ils fourni une liste des études exclues et justifié les exclusions ?
8	Les auteurs ont-ils décrit les études incluses de manière suffisamment détaillée ?
9	Les auteurs ont-ils utilisé une technique satisfaisante pour évaluer le risque de biais des études individuelles incluses dans la revue ?
10	Les auteurs ont-ils indiqué les sources de financement des études incluses dans la revue ?
11	Si une méta-analyse a été effectuée, les auteurs ont-ils utilisé des méthodes appropriées pour la combinaison statistique des résultats ?
12	Si une méta-analyse a été effectuée, les auteurs ont-ils évalué l'impact potentiel des risques de biais des études individuelles sur les résultats de la méta-analyse ou d'une autre synthèse des preuves ?
13	Les auteurs ont-ils pris en compte le risque de biais des études individuelles lors de l'interprétation / de la discussion des résultats de la revue ?
14	Les auteurs ont-ils fourni une explication satisfaisante pour toute hétérogénéité observée dans les résultats de la revue, et une discussion sur celle-ci ?
15	S'ils ont réalisé une synthèse quantitative, les auteurs ont-ils mené une évaluation adéquate des biais de publication (biais de petite étude) et ont discuté de son impact probable sur les résultats de la revue ?
16	Les auteurs ont-ils rapporté toute source potentielle de conflit d'intérêts, y compris tout financement reçu pour réaliser la revue ?

Résumé

Introduction

L'instabilité antérieure d'épaule est une pathologie très fréquente, surtout chez le sujet jeune, elle est le plus souvent d'origine traumatique et peut entraîner des répercussions à long terme sur la vie d'un patient. Pour la diagnostiquer, on dispose de plusieurs outils dont certains accessibles au kinésithérapeute, comme les tests physiques.

Objectif(s)

Nous avons choisi 4 tests parmi ceux utilisés dans le diagnostic de cette pathologie : le test d'apprehension antérieure, le test de recentrage de Jobe, le test de surprise et le load and shift test. A partir des études les décrivant, nous allons essayer de comparer la validité de critère de ces tests par rapport à un test de référence. Le but de cette comparaison est de déterminer s'il existe un test plus efficace que les autres pour le diagnostic de l'instabilité antérieure d'épaule.

Méthodologie de recherche

Nous avons effectué des recherches dans les 3 bases de données Pubmed, Pedro et Cochrane pour réunir des études datées entre 2001 et 2020. Nous avons ensuite extrait les données concernant la sensibilité et spécificité des 7 articles inclus après avoir fait une analyse de la qualité méthodologique de chacun d'eux. Les données extraites ont ensuite, à leur tour été analysées.

Résultats et analyses

Les études incluses ont mis en évidence des propriétés à peu près équivalentes entre les divers tests mais qui pouvait grandement différer en fonction du critère de jugement sur lequel s'appuyait le test. On retrouve une plus grande efficacité du test d'apprehension et du test de recentrage utilisant l'apprehension plutôt que la douleur comme critère de jugement. Même si le test de surprise paraît être le plus efficace, il ne peut s'effectuer sans un test d'apprehension au préalable et une combinaison des tests offre la meilleure efficacité diagnostique.

Discussion

La qualité méthodologique de notre revue étant seulement moyenne selon la grille AMSTAR ainsi que le faible nombre d'articles trouvés pour certains tests, nous amènent à tempérer notre conclusion. De plus, les études incluses n'utilisaient pas toutes le même test de référence ni les mêmes critères de jugement et les populations sont parfois très différentes. Le risque de biais en découlant est aussi une menace de plus pour la validité de la conclusion établie dans ce travail. Un nombre plus important d'études comparables serait nécessaire pour chaque test pour obtenir des résultats fiables.

Mots clefs : instabilité antérieure d'épaule/test/diagnostic/validité/sensibilité/spécificité

Abstract

Background

The anterior shoulder instability is a very common pathology, especially for the young people ; it's very often from a traumatic injury and it can have long-term consequences for the patient. To diagnose it, we could use many tools among which some that the physiotherapist can use, the physical examination tests, such as the anterior apprehension test, the relocation test, the release test and the load and shift test.

Objectives

We found several tests to diagnose this pathology, among which we chose 4 tests : the anterior apprehension test, the relocation test, the release test and the load and shift test. From the studies describing them, we will try to compare the validity of these tests against a reference test. The goal of this is to determine if one of these tests is more efficient than the others to diagnose the anterior shoulder instability.

Methods

We searched through the 3 databases Pubmed, Pedro and Cochrane to select studies published between 2000 and 2020. We analysed the methodological quality of each of the 7 studies included, then, extracted and analysed the data concerning the sensitivity and specificity for each test.

Results

The included studies showed that the different tests have almost the same usefulness, but it could differ a lot according to the criteria used to assess the patient. We found a greater usefulness for the apprehension test and the relocation test using apprehension instead of pain as a criteria. Even if the release test seems to be the most useful, it cannot be performed within an apprehension test. A combination of tests have a greater diagnostic usefulness.

Discussion

The methodological quality of our review, judged to be adequate according to the AMSTAR evaluation model and the few articles we found for some tests lead us to be tempered with our conclusion. Moreover, the included studies didn't all use the same reference test nor the same assessment criteria ; The population included may presents a lot of differences. The risk of biases coming from it is another threat for this work's conclusion validity. More studies are needed for each test to obtain more accurate results.

Keywords : shoulder anterior instability/test/diagnosis/validity/sensitivity/specificity