

La fiabilité et l'utilité diagnostique de l'examen clinique de l'appareil locomoteur

1

FIABILITÉ, 2

PRÉCISION DU DIAGNOSTIC, 3

Table d'éventualité 2×2 , 3

Précision totale, 4

Valeurs prédictives positives ou négatives (VPP ou VPN), 4

Sensibilité, 4

Spécificité, 5

Ratios de vraisemblance, 6

INTERVALLES DE CONFIANCE, 8

PROBABILITÉ DU SIGNE PRÉLIMINAIRE ET PROBABILITÉ FINALE, 8

CALCULS DE LA PROBABILITÉ FINALE, 9

ÉVALUATION DE LA QUALITÉ DES ÉTUDES, 10

RÉSUMÉ, 11

RÉFÉRENCES, 12

À l'heure actuelle, les sciences et professions médicales mènent une révolution raisonnée vers la pratique fondée sur les niveaux de preuve, définis comme la combinaison de la meilleure recherche de preuves, disponible avec l'expérience clinique au service des patients [1, 2].

La preuve doit être incorporée dans tous les aspects de la kinésithérapie, du patient hospitalisé à la clientèle de cabinet, incluant l'examen, le bilan, le diagnostic, le pronostic et le traitement. Il se peut que la partie la plus importante soit un examen clinique à la fois rapide et prudent pouvant mener à un diagnostic précis, à un pronostic fiable et à un plan de traitement efficace. En conséquence, on ne doit pas sous-estimer l'importance d'incorporer la preuve de la pertinence des tests cliniques et des mesures, de façon à mettre en évidence les patients porteurs de tels ou tels désordres musculosquelettiques [1, 2].

Le processus du bilan kinésithérapique impose de recueillir l'histoire du patient, de développer des hypothèses de travail, de choisir des tests et des mesures spécifiques pour confirmer ou infirmer les hypothèses formulées.

Le clinicien doit déterminer la probabilité initiale (avant toute évaluation) que le patient ait tel ou tel problème. Suite à cette information, le clinicien choisit les mesures et tests appropriés qui l'aideront à déterminer la probabilité finale (après évaluation) que le patient ait ce problème-là. Le degré de certitude doit être suffisant pour que le traitement puisse commencer (idée de seuil de certitude à partir duquel le traitement peut être entrepris). Le but des tests cliniques n'est pas d'arriver à une certitude de diagnostic mais plutôt de réduire le degré d'incertitude jusqu'à ce que le seuil de certitude du traitement soit atteint [2].

Les concepts de probabilité initiale et finale et de seuil de certitude du traitement sont expliqués plus loin dans ce chapitre.

Comme la quantité de tests cliniques répertoriés augmente sans cesse, il est absolument essentiel d'évaluer soigneusement les propriétés de ces tests, avant de les introduire dans la pratique clinique [3]. L'intégration du meilleur niveau de preuve de chacun des tests disponibles en vue d'une utilité diagnostique est fondamentale pour pratiquer un diagnostic précis et complet, amenant un traitement pertinent et efficace.

Il est évident que les praticiens comme les étudiants doivent être attentifs aux propriétés diagnostiques et aux mesures obtenues par les tests employés lesquels sont d'une vraie utilité clinique.

Les lignes ci-dessous aident le praticien et/ou l'étudiant à sélectionner tests et mesures pour évaluer correctement les patients et permettre la mise en œuvre rapide d'une stratégie de soins efficace.

L'évaluation des tests utilisés pour le bilan implique l'examen de plusieurs propriétés, incluant la fiabilité et la précision du bilan. Un test est considéré comme fiable s'il produit une information reproductible, précise et spécifique. Un test est considéré comme précis s'il présente la possibilité de séparer les patients les uns des autres [4]. L'évaluation scientifique de l'utilité clinique des tests et mesures en bilan masso-kinésithérapique nécessite la comparaison des résultats avec des références indiscutables¹ comme des radiographies (lesquelles représentent la mesure la plus proche de la réalité).

Au moyen des outils statistiques issus du champ de l'épidémiologie, la précision d'un test – c'est-à-dire sa possibilité de déterminer quel patient présente la dysfonction et quel patient ne la présente pas – est finalement calculée.

Ce premier chapitre met l'accent sur les caractéristiques qui définissent la fiabilité et la précision des tests et mesures spécifiques.

Ce chapitre se termine par une discussion à propos de la qualité des études d'évaluation cherchant à évaluer l'utilité diagnostique.

Fiabilité

Pour qu'un test clinique donne une information utilisable pour guider une décision thérapeutique, il doit avoir une fiabilité acceptable. La fiabilité est le degré de confiance avec laquelle une méthode ou une échelle mesure un signe particulier [5]. Quand on quantifie la fiabilité d'une mesure, on détermine la proportion de ce qui est une représentation de la réalité par rapport à un résultat dû à une fausse mesure [6].

Quand le processus du bilan clinique est discuté, deux sortes de fiabilités doivent être envisagées : intra-examineur et inter-examineur. La fiabilité intra-examineur est la mesure de la capacité d'un unique évaluateur d'obtenir un résultat identique à la suite d'utilisations successives d'un même test. La fiabilité inter-examineur est la mesure de la capacité de deux ou plusieurs évaluateurs à obtenir des résultats identiques pour un même test.

Le coefficient kappa (κ) est une mesure de la proportion entre un accord ou un rejet des résultats, une fois le facteur hasard éliminé [1, 5, 7]. C'est le facteur de fiabilité le plus souvent utilisé pour les données en échelles

¹ Ce qu'on appelle communément un *gold standard* (NdT).

(positives ou négatives) [5]. Le coefficient de corrélation habituellement utilisé pour déterminer la fiabilité des données qui sont en continu dans la nature (par ex. les amplitudes articulaires) est le coefficient de corrélation intraclasse (CCI) [7]. Bien que l'interprétation de la fiabilité puisse varier, les coefficients sont souvent évalués par le critère décrit par Shrout [8] avec des valeurs inférieures à 0,10 indiquant pas de fiabilité, des valeurs entre 0,11 et 0,40 indiquant une fiabilité médiocre, des valeurs comprises entre 0,41 et 0,60 indiquant une fiabilité passable, des valeurs entre 0,61 et 0,80 montrant une fiabilité modérée, tandis que les valeurs supérieures à 0,81 indiquent une forte fiabilité.

Le niveau de «fiabilité acceptable» doit être décidé par le praticien qui utilise tel ou tel test spécifique ou telle ou telle mesure représentative [9]. Ce niveau doit être choisi en fonction de la variable testée, selon l'importance du test en question et qui utilise ce test [6]. Par exemple, 5 % d'erreur sur une mesure peut être parfaitement acceptable quand on mesure une amplitude articulaire, mais n'est pas acceptable quand on mesure la température centrale en pédiatrie.

Précision du diagnostic

En pratique clinique, les tests et les mesures ne peuvent jamais confirmer ou infirmer totalement la présence d'un trouble spécifique [10]. Cependant, les tests cliniques peuvent être utilisés pour modifier l'idée du clinicien sur la pathologie musculosquelettique du patient. La précision d'un test est évaluée en déterminant le degré d'accord entre le test clinique et une référence standard [11, 12]. Une référence standard est un critère considéré comme représentant la plus grande chance possible de pouvoir dire avec certitude que la pathologie est bien présente [1]. Les résultats obtenus avec cette référence standard sont comparés avec ceux obtenus par le test en question. De cette manière, le pourcentage de sujets correctement diagnostiqués, appelé la précision du diagnostic, peut être déterminé [13]. Puisque les statistiques relatives à l'utilité diagnostique sont totalement dépendantes de la référence standard et de la population étudiée, nous les avons listées dans le texte pour fournir des informations permettant de choisir à bon escient tel ou tel test ou mesure.

La précision du diagnostic est souvent exprimée en termes de valeurs prédictives positives ou négatives (VPP ou VPN), de sensibilité et de spécificité ou bien encore d'un ratio de vraisemblance (RV) [1, 14].

Table d'éventualité 2 × 2

Pour déterminer l'utilité d'un test ou d'une mesure, les résultats issus de la référence standard sont comparés avec ceux issus du test évalué, en utilisant une table d'éventualité 2 × 2. Celle-ci fournit une comparaison directe entre la référence standard et le test évalué [15]. Cela permet d'établir les valeurs associées à la précision du diagnostic de façon à aider le praticien à choisir le test approprié (tableau 1-1).

Une table d'éventualité 2 × 2 est divisée en quatre cellules (a, b, c et d). Cette division permet de déterminer la possibilité du test de diagnostic d'identifier correctement les résultats vrais positifs (cellule a) et les résultats vrais négatifs (cellule d). La cellule b montre les résultats faux positifs, c'est-à-dire les résultats positifs pour le test de diagnostic et négatifs pour la référence standard. La cellule c montre les résultats faux négatifs, dans laquelle le test de diagnostic se révèle faux alors que la référence standard donne un résultat positif.

Dès qu'une étude, menée pour connaître l'utilité diagnostique d'un test clinique, est achevée et qu'une comparaison avec une référence standard a été menée dans une table d'éventualité 2 × 2, on peut évaluer l'utilité clinique sous la forme de : précision totale, VPP, VPN de la sensibilité et la spécificité; ensuite les ratios de vraisemblance (RV) positif ou négatif peuvent être évalués. Ces statistiques sont utiles au clinicien pour déterminer si un test diagnostique est nécessaire pour retenir ou écarter un symptôme quelconque.

Tableau 1-1 Table d'éventualité 2 × 2 utilisée pour comparer les résultats d'une référence standard avec ceux d'un test étudié

	Référence standard – résultats positifs	Référence standard – résultats négatifs
Test de diagnostic, résultats positifs	Résultats vrais positifs a	Résultats faux positifs b
Test de diagnostic, résultats négatifs	Résultats faux négatifs c	Résultats vrais négatifs d

Précision totale

La précision totale d'un test est obtenue en divisant les réponses correctes (vrais positifs et vrais négatifs) par le nombre total de patients [16]. L'utilisation d'une table d'éventualité 2 × 2 nécessite l'utilisation de l'équation suivante :

$$\text{Précision totale} = 100 \% \times (a + d) / (a + b + c + d)$$

Un test parfait donnerait une précision totale de 100 %. Cela est impossible puisque aucun test clinique n'est parfait et tous possèdent au moins un petit degré d'incertitude. La précision d'un test de diagnostic ne doit pas être utilisée pour s'assurer de l'utilité clinique de ce test parce que la précision totale peut être trompeuse, même faiblement. La précision d'un test peut être significativement influencée par la prévalence totale d'une pathologie pour une population à un instant donné [5, 6].

Valeurs prédictives positives ou négatives (VPP ou VPN)

Les VPP permettent d'estimer la vraisemblance qu'un patient ayant un résultat positif ait la pathologie [5, 6, 17]. Les VPP ou VPN sont calculées horizontalement dans une table d'éventualité 2 × 2 (tableau 1-2). Elles indiquent le nombre de patients correctement identifiés comme ayant la pathologie (vrais positifs) divisé par la somme des résultats positifs donnés par le test en évaluation. Une valeur élevée de cette VPP indique qu'un résultat positif donne l'assurance d'une forte prédiction que ce patient soit porteur de la pathologie [5, 6]. La formule pour calculer la VPP est la suivante :

$$\text{VPP} = 100 \% \times a / (a + b)$$

La valeur prédictive négative (VPN) estime la vraisemblance qu'un patient avec un résultat négatif n'ait pas la pathologie [5, 6]. La VPN est aussi calculée horizontalement dans la table d'éventualité 2 × 2 (voir tableau 1-2). Elle se calcule comme le nombre de patients correctement identifiés comme n'ayant pas la pathologie (vrais négatifs) divisés par tous les résultats négatifs du test en évaluation [11]. La formule pour calculer la VPN est la suivante :

$$\text{VPN} = 100 \% \times d / (c + d)$$

Les valeurs prédictives sont largement influencées par la fréquence de la maladie [11]. En conséquence, nous n'avons pas spécifiquement répertorié ce fait dans les lignes qui suivent.

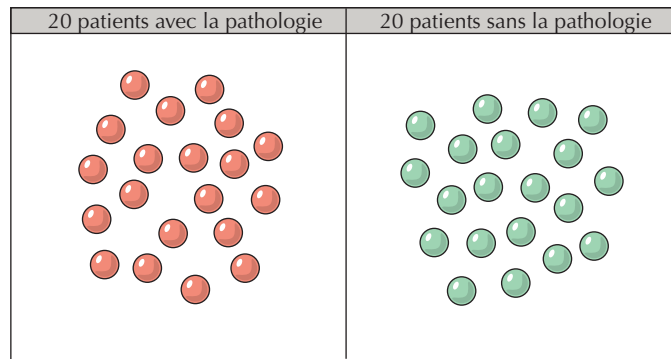
Sensibilité

La sensibilité d'un test de diagnostic montre la possibilité de ce test de détecter les patients porteurs de la pathologie, révélés par la référence standard. Cela fait aussi appel à un coefficient de vrai positif [1]. Les tests ayant une haute sensibilité sont bons pour exclure une pathologie particulière. L'acronyme SeNex peut être utilisé pour se souvenir qu'un test ayant une sensibilité élevée (Se) et un résultat négatif (N) est un bon test pour exclure (ex) la pathologie mesurée.

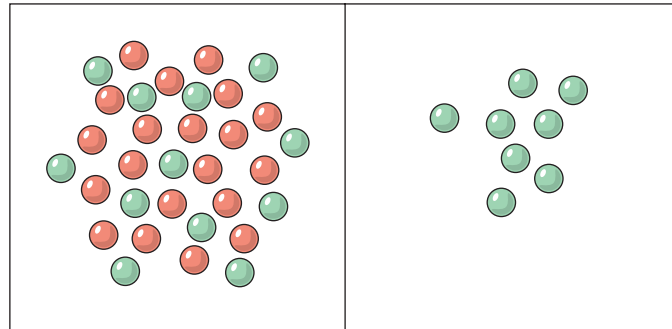
Considérons, par exemple, un test clinique qui, comparé avec la référence standard, montre une sensibilité élevée pour la détection d'une sténose spinolombale. Selon la règle énoncée ci-dessus, si le test est négatif, il élimine d'une manière fiable une sténose spinolombale. Si le test est positif, il est probable qu'il identifie efficacement un grand pourcentage de patients avec une sténose spinolombale. Cependant, le test peut aussi identifier comme positif beaucoup de patients ne présentant pas la pathologie (faux positifs). Par conséquent, bien qu'un résultat négatif soit performant, un résultat positif ne nous permet pas de conclure et de formuler des conclusions (figures 1-1 et 1-2).

Tableau 1-2 Table d'éventualité 2 × 2 montrant les calculs des valeurs prédictives positives et négatives (horizontalement) et les sensibilité et spécificité (verticalement)

	Références standard – résultats positifs	Références standard – résultats négatifs	
Test de diagnostic, résultats positifs	Résultats vrais positifs a	Résultats faux positifs b	VPP = a/(a + b)
Test de diagnostic, résultats négatifs	Résultats faux négatifs c	Résultats vrais négatifs d	VPN = d/(c + d)
	Sensibilité = a/(a + c)	Sensibilité = d/(b + d)	

**Figure 1-1**

Exemple de sensibilité et de spécificité : 20 patients avec et 20 patients sans la pathologie.

**Figure 1-2**

Sensibilité de 100 %, impliquant que si le test est positif, tous les porteurs de la pathologie seront identifiés. Cependant, bien que tous les porteurs de la pathologie soient identifiés, on remarque aussi que beaucoup de non-porteurs sont malgré tout détectés. Par contre, si le test est négatif, nous avons la certitude que la pathologie puisse être exclue (SeNex).

La sensibilité d'un test est aussi calculée à partir de la table d'éventualité 2 × 2. Cependant, elle est calculée verticalement (voir [tableau 1-2](#)). La formule pour calculer la sensibilité d'un test est la suivante :

$$\text{Sensibilité} = 100\% \times a / (a + c)$$

Spécificité

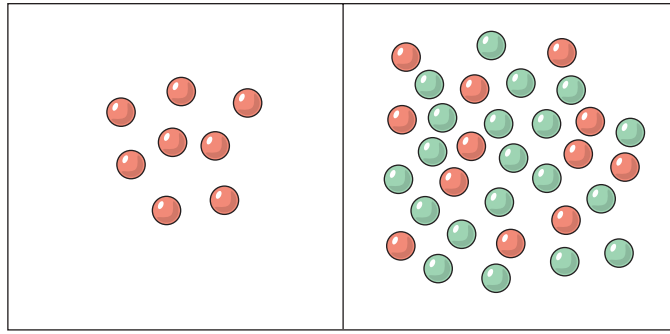
La spécificité d'un test de diagnostic signale simplement la possibilité pour le test de détecter les patients qui, en fait, n'ont pas la pathologie, indiquée par la référence standard. Cela fait aussi appel à un coefficient de vrai négatif [1]. Les tests ayant une haute spécificité sont bons pour inclure une pathologie particulière. L'acronyme SpePin peut être utilisé pour se souvenir qu'un test ayant une spécificité élevée (Spe) et un résultat positif (P) est un bon test pour inclure (in) la pathologie mesurée [16, 18, 19].

Considérons, par exemple, un test avec une spécificité élevée. Ce test permettra à coup sûr d'identifier précisément tous les patients qui ne présentent pas la pathologie. Si un test clinique, ayant une spécificité élevée est négatif, alors il est probable qu'un fort pourcentage de patients non porteurs de la pathologie sera détecté. Cependant, il est aussi possible qu'un test ayant une grande spécificité et un résultat négatif détecte un certain nombre de patients porteurs de la pathologie (faux négatifs). Par conséquent, on peut avoir confiance en un test couplant une spécificité élevée avec un résultat positif, il indique que la pathologie est bien présente ([figures 1-3](#)).

La formule pour calculer la spécificité d'un test est la suivante :

$$\text{Spécificité} = 100\% \times d / (b + d)$$

La sensibilité et la spécificité sont utilisées depuis fort longtemps pour déterminer la pertinence des tests de diagnostic. Cependant, elles sont associées à quelques limitations cliniques [11]. Bien que la sensibilité et la spécificité soient utiles comme aide aux cliniciens dans la sélection des bons tests pour inclure ou exclure

**Figure 1-3**

Spécificité de 100 % impliquant que si le test est négatif, tous les non-porteurs de la pathologie seront identifiés. Cependant, bien que tous les non-porteurs de la pathologie soient identifiés, beaucoup de porteurs de la pathologie sont malgré tout détectés. Par contre, si le test est positif, nous avons la certitude que les patients sont porteurs de la pathologie (SpePin).

une pathologie, peu de tests cliniques montrent à la fois une grande sensibilité et une grande spécificité [11]. De plus, la sensibilité et la spécificité ne fournissent pas d'indication sur un éventuel changement de la probabilité qu'un patient présente la dysfonction selon que les résultats du test sont positifs ou négatifs [18, 20]. En conséquence, les ratios de vraisemblance (RV) ont été mis en avant comme étant la meilleure technique statistique pour déterminer une modification de la probabilité du signe préliminaire qu'un patient soit porteur d'une pathologie précise.

Ratios de vraisemblance

Le résultat d'un test n'est valable que s'il change la probabilité d'un test initial montrant qu'un patient présente une pathologie [21]. Les ratios de vraisemblance (RV) combinent la sensibilité et la spécificité d'un test pour prédire la modification de la probabilité en exprimant le résultat spécifique de ce test. Les RV sont efficaces pour aider à la décision clinique [20]. Les RV sont une mesure précieuse qui peut augmenter ou réduire, de manière significative, la probabilité qu'un patient présente une maladie [22].

Les ratios de vraisemblance peuvent être positifs ou négatifs. Un RV positif indique un glissement de probabilité en faveur de l'existence d'une pathologie. Un RV négatif indique un glissement de probabilité en défaveur de l'existence d'une pathologie. Cependant les RV ne sont que rarement mentionnés dans les études menées pour évaluer l'utilité diagnostique d'un examen clinique. Ils peuvent être calculés facilement si la sensibilité et la spécificité d'un test sont indiquées. Tout au long de ce texte, pour les études qui ne mentionnent pas les RV tout en donnant la spécificité et la sensibilité de ces études, les RV ont été calculés par les auteurs.

La formule utilisée pour calculer un RV positif est :

$$\text{RV positif} = \text{sensibilité} / (1 - \text{spécificité})$$

La formule utilisée pour calculer un RV négatif est :

$$\text{RV négatif} = (1 - \text{sensibilité}) / \text{spécificité}$$

Le [tableau 1-3](#) montre un guide d'interprétation des résultats. Des RV positifs > 1 augmentent les chances de présenter une pathologie, donnant un test positif. Des RV négatifs < 1 diminuent les chances de présenter une pathologie, donnant un test négatif [22]. Cependant, c'est l'amplitude du glissement de la probabilité qui détermine l'utilité pratique d'un test. Des RV positifs et > 10 avec des RV négatifs proches de zéro représentent souvent un glissement important et significatif d'un test clinique. Un RV de 1 (positif ou négatif) ne changerait pas beaucoup la probabilité qu'un patient ait ou non une pathologie. Cela présente une faible valeur clinique [22]. Une fois que les RV ont été calculés, ils peuvent être appliqués dans un nomogramme [23] ([figures 1-4](#)). On peut aussi appliquer une équation mathématique [24] pour déterminer plus précisément les glissements de probabilité, dus aux RV et montrant le résultat d'un test particulier. Ces deux méthodes sont décrites en détail dans la suite du chapitre.

Tableau 1-3 Interprétation des ratios de vraisemblance

Ratio de vraisemblance positif	Ratio de vraisemblance négatif	Interprétation
> 10	< 0,1	Montre un glissement de probabilité étendu et le plus souvent significatif
5–10	0,1–0,2	Montre un glissement de probabilité modéré
2–5	0,2–0,5	Montre un glissement de probabilité pouvant être quelquefois important
1–2	0,5–1	La probabilité change peu et rarement

Adapté de Jaeschke R, Guyatt GH, Sackett DL III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA* 1994;271 :703-07.

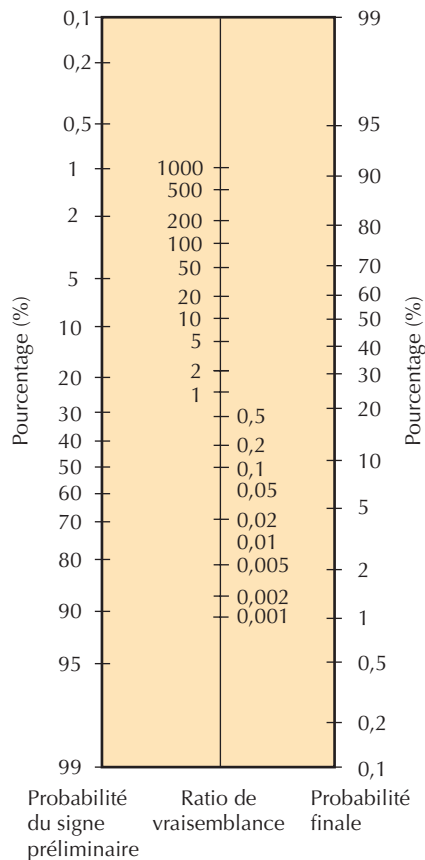


Figure 1-4

Nomogramme de Fagan. (Adapté avec la permission de Fagan TJ. Nomogram for Baye's theorem. *N Engl J Med* 1975 ;293:257. Copyright 2005, Massachusetts Medical Society. All rights reserved.)

Si un test diagnostique possède une spécificité de 1, le RV positif ne peut pas être calculé puisque le dénominateur de l'équation sera égal à zéro. Dans ce cas, il a été montré que la table d'éventualité 2 × 2 doit être modifiée en ajoutant 0,5 à chaque cellule de la table. Cela permet le calcul des RV [25].

À titre d'exemple, considérons l'utilité diagnostique du test de Crank [5, 26], permettant de déterminer une déchirure du labrum gléno-huméral, en le comparant à un examen arthroscopique, la référence standard. Le [tableau 1-4](#) montre les résultats sous la forme d'une table d'éventualité 2 × 2. L'impossibilité de calculer le RV positif est évidente dans la formule suivante :

$$RV \text{ positif} = \text{sensibilité} / (1 - \text{spécificité}) = 1 / (1 - 1) = 1 / 0$$

Intervalles de confiance

Tableau 1-4 Résultats du test de Crank permettant de détecter les déchirures du labrum gléno-huméral en comparaison avec l'examen arthroscopique servant de référence standard

	Examen arthroscopique Le test est positif ($n = 12$)	Examen arthroscopique Le test est négatif ($n = 3$)	
Test de Crank positif	10 a	0 b	Valeur prédictive positive = $100 \% \times 10/10$ = 100 %
Test de Crank négatif	2 c	3 d	Valeur prédictive négative = $100 \% \times 3/5$ = 60 %
	Sensibilité = $100 \% \times 10/12$ = 83 %	Spécificité = $100 \% \times 3/3$ = 00 %	

Il est impossible d'avoir une fraction avec un dénominateur égal à zéro, aussi, on modifie la table d'éventualité 2×2 en ajoutant 0,5 à chacune des cellules.

Bien qu'une addition de 0,5 à chaque cellule soit la seule méthode décrite pour modifier la table d'éventualité en évitant un zéro au dénominateur d'un calcul de RV, le changement que cela entraîne pour les propriétés de diagnostic telles que la sensibilité, la spécificité et les valeurs prédictives, est trop important, les erreurs sont telles que cette technique n'a pas été appliquée dans ce livre. Quand la spécificité est de zéro et que le RV positif ne peut pas être calculé, il sera indiqué comme non disponible (ND). Bien que le RV positif ne puisse pas être calculé, le lecteur devra comprendre que le test donnera, malgré tout, un important glissement de la probabilité.

Intervalles de confiance

Les calculs de sensibilité, de spécificité et de RV sont habituellement exprimés sous la forme d'estimation, c'est-à-dire de simples valeurs représentatives d'un échantillon d'une population [5]. Toutefois, puisque ces estimations sont évaluées sur de petits échantillons de population, il est peu probable qu'elles soient la parfaite représentation d'une population plus large. Il est par conséquent plus pertinent de mettre en avant un écart de valeurs (intervalle de confiance) à l'intérieur duquel on est quasi certain que la population se trouve.

Un intervalle de confiance (IC) est un écart de valeurs autour de la valeur centrale calculée et les valeurs qui représentent la population sont probablement à l'intérieur de cet écart [27]. Par convention, un IC de 95 % est évalué dans les études portant sur l'utilité diagnostique et les examens cliniques.

Un IC de 95 % indique que la dissémination des résultats à 95 % de chance de se trouver à l'intérieur de l'écart donné [5]. Dans tout le texte qui suit, un IC de 95 % est évalué pour toutes les études dont il est fait état.

Probabilité du signe préliminaire et probabilité finale

La probabilité du signe préliminaire² mesure la vraisemblance qu'un patient montre un signe pathognomonique d'une pathologie quelconque avant que le bilan clinique ne soit fait. Souvent, les taux de prévalence sont utilisés comme une indication de la probabilité du signe préliminaire. Toutefois, dans certaines circonstances quand le taux de prévalence est inconnu, la probabilité du signe préliminaire est fondée sur une combinaison de l'interrogatoire du patient, de la catamnèse (résultats des bilans précédents) et de l'expérience clinique du praticien [16]. La détermination de la probabilité du signe préliminaire est la première étape dans le processus de prise de décision pour les cliniciens. La probabilité du signe préliminaire est une estimation par le praticien pouvant être exprimée par un pourcentage (par exemple : 75 %, 80 %) ou bien par une mesure qualitative (« probable » ou « tout à fait probable ») [11, 16]. Dès que la probabilité préliminaire qu'un patient présente le signe précis d'une pathologie est calculée, les tests et les mesures ayant la faculté de changer cette probabilité doivent être sélectionnés en vue du bilan.

La probabilité finale mesure la vraisemblance qu'un patient montre un signe pathognomonique d'une pathologie quelconque comme résultat final du bilan clinique.

² La notion de signe préliminaire exprime l'idée de l'existence d'un signe clinique précurseur de tout désordre orthopédique; la probabilité finale représente la chance de mettre en évidence un signe clinique, spécifique d'un désordre orthopédique à la fin du bilan clinique (NdT).

Calculs de la probabilité finale

Comme déjà mentionné, les RV aident le clinicien à déterminer le glissement probable de la probabilité apparaissant après que les résultats d'un test ont été enregistrés, selon les ratios de RV de ce test-là. La méthode la plus rapide pour déterminer ce glissement de probabilité, une fois que le RV d'un test quelconque est connu, est l'utilisation d'un nomogramme [23] (figures 1-5). Un nomogramme est un diagramme qui illustre la probabilité du signe préliminaire à gauche et la probabilité finale à droite. Les RV se trouvent au milieu. Pour déterminer le changement de probabilité, on procède comme suit. Un repère représentant la probabilité du signe préliminaire est placé sur la ligne verticale à gauche du nomogramme. Puis un repère est placé sur la ligne au centre du nomogramme au niveau du RV calculé (qu'il soit positif ou négatif). Les deux repères sont reliés par une ligne droite prolongée vers la droite du nomogramme. Cette ligne droite croise la ligne représentant la probabilité finale en un point. Ce point indique le changement de probabilité recherché.

Un calcul plus précis du glissement de probabilité peut être fait algébriquement au moyen des formules suivantes [16] :

Étape 1 Estimation du signe préliminaire = probabilité du signe préliminaire / (1 - probabilité du signe préliminaire)

Étape 2 Estimation du signe préliminaire \times RV = estimation finale

Étape 3 Estimation finale / (estimation finale + 1) = probabilité finale

Quand la probabilité finale est basse, le clinicien doit décider d'éliminer la réalité de cette pathologie pour envisager d'autres investigations. Ou bien, quand la probabilité finale est élevée, le clinicien peut envisager avec confiance la réalité de cette pathologie. Le niveau à partir duquel on cesse l'évaluation pour commencer un traitement est appelé le seuil de traitement [16] (figures 1-6).

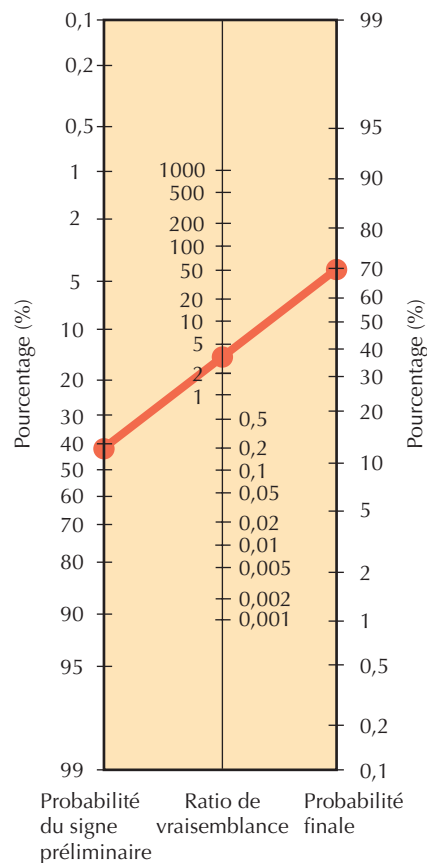


Figure 1-5

Nomogramme représentant le changement de la probabilité du signe préliminaire (42 %) si le test est positif (ratio de vraisemblance positif = 4,2) vers une probabilité finale de 71 %. (Adapted with permission from Fagan TJ. Nomogram for Baye's theorem. *N Engl J Med* 1975;293:257. Copyright 2005, Massachusetts Medical Society. All rights reserved.)

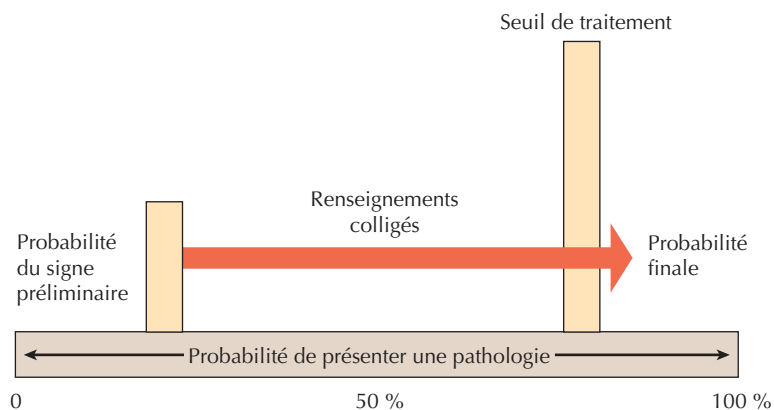


Figure 1-6

Les cliniciens doivent utiliser la probabilité du signe préliminaire et les ratios de vraisemblance pour déterminer le seuil de traitement, comme montré sur cette illustration.

Évaluation de la qualité des études

Dès qu'un article intéressant a été trouvé, l'étape suivante est de faire une analyse critique du contenu au moyen d'une méthode rigoureuse. Il a été démontré que la qualité méthodologique des études relatives à l'utilité diagnostique de l'examen clinique est souvent inférieure à celle des études portant sur l'efficacité thérapeutique [28, 29]. Malheureusement, les études ayant une bonne méthodologie peuvent cacher une faille : certains des tests spécifiques habituellement utilisés peuvent entraîner une utilisation de procédures inefficaces. Il en résulte des diagnostics non fiables et une mauvaise prise en charge des patients. À l'opposé, l'identification et l'utilisation de tests cliniques rigoureusement évalués peuvent entraîner une prise en charge de qualité et des résultats finaux meilleurs [29].

La qualité de la fiabilité des études en utilisant le protocole QAREL (*Quality Appraisal of Reliability Studies*) a été développée pour évaluer la qualité et la fiabilité diagnostique des études [30]. Le protocole QAREL est une liste de contrôle comportant 11 questions mises au point par un groupe d'experts de référence en recherche diagnostique et appréciation de la qualité. Il est utilisé pour valider la qualité méthodologique des études. Chaque question est évaluée sous forme de « oui », « non », « NR = non répertorié ou manque de clarté », « N/A = non applicable ». Le protocole QAREL a prouvé qu'il était un outil d'évaluation fiable à condition que les relecteurs donnent la clé de discussion des critères avec lesquels ils interprètent chacune des questions [31]. La fiabilité de 9 questions sur 11 a été évaluée comme bonne, alors que la fiabilité de 2 questions sur 11 a été évaluée comme faible [31].

Nous avons utilisé le protocole QUAREL pour évaluer chacune des études liées à la fiabilité et recensées dans l'ouvrage. Dans ce but, nous avons pris en compte une bonne qualité comme étant ≥ 75 et une faible qualité comme étant comprise entre 74 % et 50 %. Ces pourcentages ont été calculés en divisant le nombre de réponses « oui » par 11, moins le nombre de réponses N/A. Les symboles de couleur verte indiquent un haut niveau de qualité méthodologique impliquant que le lecteur peut avoir confiance dans les résultats de ces études. Les symboles de couleur orange indiquent un niveau de qualité méthodologique passable impliquant que le lecteur doit interpréter les résultats de ces études avec précaution. Les études de mauvaise qualité n'ont pas été incluses dans les tables d'utilité diagnostique tout au long des chapitres suivants.

L'évaluation qualitative des études de fiabilité diagnostique (*Quality Assessment of Diagnostic Accuracy Studies* ou QUADAS) a été développée pour évaluer la qualité de la fiabilité diagnostique des études [32]. Quatre sessions Delphi ont identifié un panel de 14 critères utilisés pour évaluer la qualité méthodologique d'une étude. Chaque élément est répertorié comme « positif », « négatif » ou « indéterminé ». Le QUADAS ne permet pas de quantifier une échelle pour chaque étude mais fournit plutôt une évaluation qualitative de l'étude avec une identification de ses faiblesses [32].

Le QUADAS a montré un accord satisfaisant pour chacun des éléments de la liste de contrôle [33].

Nous avons utilisé le QUADAS pour évaluer chacune des études référencées dans le livre présenté. Dans ce but, nous avons pris en compte une bonne qualité comme étant ≥ 75 et une faible qualité comme étant comprise entre 74 % et 50 %. Ces pourcentages ont été calculés en divisant le nombre de réponses « oui » par 14 (le nombre total de critères évalués). Les symboles de couleur verte indiquent un haut niveau de qualité méthodologique impliquant que le lecteur peut avoir confiance dans les résultats de ces études. Les symboles

Tableau 1-5 Table d'éventualité 2×2 et propriétés statistiques utilisées pour déterminer l'utilité diagnostique des tests ou des mesures

	Références standard – résultats positifs	Références standard – résultats négatifs
Test de diagnostic, résultats positifs	Résultats vrais positifs a	Résultats faux positifs b
Test de diagnostic, résultats négatifs	Résultats faux négatifs c	Résultats vrais négatifs d



Nom de la propriété statistique	Formules	Description
Précision totale	$(a + d)/(a + b + c + d)$	Pourcentage de patients correctement diagnostiqués
Sensibilité	$a/(a + c)$	Proportion de patients ayant une pathologie et un résultat positif
Spécificité	$d/(b + d)$	Proportion de patients ayant une pathologie et un résultat négatif
Valeur prédictive positive	$a/(a + b)$	Proportion de patients ayant un résultat positif et ayant la pathologie
Valeur prédictive négative	$d/(c + d)$	Proportion de patients ayant un résultat négatif et n'ayant pas la pathologie
Ratio de vraisemblance positif	Sensibilité/(1 - spécificité)	Si le test est positif, il y a une augmentation de l'estimation en faveur de la présence de la pathologie
Ratio de vraisemblance négatif	(1 - sensibilité)/spécificité	Si le test est positif, il y a une augmentation de l'estimation en défaveur de la présence de la pathologie

de couleur jaune indiquent un niveau de qualité méthodologique passable impliquant que le lecteur doit interpréter les résultats de ces études avec précaution. Les études de mauvaise qualité n'ont pas été incluses dans les tables d'utilité diagnostique tout au long des chapitres suivants.

Résumé

La fiabilité et l'utilité diagnostique des tests et mesures doivent être étudiées avant de les inclure dans le processus de l'examen clinique. Les tests et les mesures doivent montrer la pertinence de leur fiabilité avant d'être utilisés pour toute conclusion diagnostique. Tout au long de ce livre, nous montrons toutes les données sur la fiabilité de nombreux tests et mesures. Il est essentiel que les cliniciens apprécient ces niveaux de fiabilité dans le contexte clinique de leur pratique quotidienne.

Avant que des tests et des mesures soient intégrés aux bilans en kinésithérapie, l'utilité diagnostique de chacun d'entre eux doit être mesurée. Le [tableau 1-5](#) résume les statistiques expliquées en ce qui concerne la précision du diagnostic, aussi bien que les équations mathématiques et les définitions opératoires. L'utilité ou la pertinence d'un test ou d'une mesure sont le plus souvent évaluées selon les propriétés diagnostiques propres à chacun des outils, en d'autres termes selon les propriétés de sensibilité, spécificité, les valeurs prédictives positives et négatives (VPP et VPN). Cependant, la propriété diagnostique la plus importante d'un test est sans doute le ratio de vraisemblance (RV) qui permet d'évaluer le changement de probabilité entre le signe préliminaire et la probabilité finale d'une pathologie donnée.

Aucun test clinique ou mesure ne fournit la certitude absolue de la présence ou l'absence d'un désordre quelconque. Toutefois, quand suffisamment de données de bilan ont été rassemblées pour évaluer la probabilité du seuil de traitement, les praticiens peuvent déterminer quand cesser l'évaluation pour démarrer le traitement. De plus, une évaluation méthodologique soignée fournit un meilleur aperçu de la rigueur scientifique de chaque étude, accompagné de ces performances, de son applicabilité, de sa fiabilité et de sa reproductibilité dans un contexte clinique donné.

1. Sackett DL, Straws SE, Richardson WS, et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd ed. London: Harcourt Publishers Limited; 2000.
2. Kassirer JP. Our stubborn quest for diagnostic certainty: a cause of excessive testing. *N Engl J Med* 1989;320:1489–91.
3. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6.
4. Schwartz JS. Evaluating diagnostic tests: what is done – what needs to be done. *J G Intern Med* 1986;1:266–7.
5. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 2nd ed. Upper Saddle River, NJ: Prentice Hall Health; 2000.
6. Rothstein JM, Ehternach JL. *Primer on Measurement: An Introductory Guide to Measurement Issues*. Alexandria, VA: American Physical Therapy Association; 1999.
7. Domholdt E. *Physical Therapy Research*. 2nd ed. Philadelphia: WB Saunders; 2000.
8. Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res* 1998;7:301–17.
9. Van Genderen F, De Bie R, Helders P, Van Meeteren N. Reliability research: towards a more clinically relevant approach. *Phys Ther Rev* 2003;8:169–76.
10. Bossuyt PMM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Chem* 2003;49:1–6.
11. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther* 2001;81:1546–64.
12. Jaeschke R, Guyatt GH, Sackett III DL. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA* 1994;271:389–91.
13. Bossuyt PMM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;49:7–18.
14. McGinn T, Guyatt G, Wyer P, et al. Users' guides to the medical literature XXII: how to use articles about clinical decision rules. *JAMA* 2000;284:79–84.
15. Greenhalgh T. Papers that report diagnostic or screening tests. *BMJ* 1997;315:540–3.
16. Bernstein J. Decision analysis (current concepts review). *J Bone Joint Surg* 1997;79:1404–14.
17. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. *Phys Ther* 1985;65:1671–5.
18. Boyko EJ. Ruling out or ruling in disease with the most sensitive or specific diagnostic test: short cut or wrong turn? *Med Decis Making* 1994;14:175–80.
19. Riddle DL, Stratford PW. Interpreting validity indexes for diagnostic tests: an illustration using the Berg balance test. *Phys Ther* 1999;79:939–48.
20. Hayden SR, Brown MD. Likelihood ratio: a powerful tool for incorporating the results of a diagnostic test into clinical decision making. *Ann Emerg Med* 1999;33:575–80.
21. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763–70.
22. Jaeschke R, Guyatt GH, Sackett DL. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA* 1994;271:703–7.
23. Fagan TJ. Letter: nomogram for Bayes theorem. *N Engl J Med* 1975;293:257.
24. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston: Little, Brown; 1991.
25. Wainner RS, Fritz JM, Irrgang JJ, et al. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine* 2003;28:52–62.
26. Mimori K, Muneta T, Nakagawa T, Shinomiya K. A new pain provocation test for superior labral tears of the shoulder. *Am J Sports Med* 1999;27:137–42.
27. Fidler F, Thomason N, Cumming G, et al. Editors can lead researchers to confidence intervals, but can't make them think. *Psychol Sci* 2004;15:119–26.
28. Moons KGM, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem* 2004;50:473–6.
29. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. *JAMA* 1995;274:645–51.
30. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010;63(8):854–61.
31. Lucas N, Macaskill P, Irwig L, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol* 2013;13:111.
32. Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol* 2005;5:19.
33. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;6:9.