

# Riproducibilità e utilità diagnostica dell'esame obiettivo ortopedico

<b>INTRODUZIONE</b>	<b>2</b>
<b>RIPRODUCIBILITÀ</b>	<b>3</b>
<b>UTILITÀ DIAGNOSTICA</b>	<b>3</b>
Tabella di contingenza 2×2	4
Accuratezza generale	4
Valore predittivo positivo e negativo	5
Sensibilità	5
Specificità	7
Rapporto di verosimiglianza (Likelihood Ratio, LR)	8
<b>INTERVALLO DI CONFIDENZA</b>	<b>10</b>
<b>PROBABILITÀ PRE-TEST E POST-TEST</b>	<b>10</b>
<b>CALCOLO DELLA PROBABILITÀ POST-TEST</b>	<b>11</b>
<b>VALUTAZIONE DELLA QUALITÀ DEGLI STUDI</b>	<b>12</b>
<b>RIEPILOGO</b>	<b>13</b>
<b>BIBLIOGRAFIA</b>	<b>14</b>

**INTRODUZIONE**

Le professioni mediche e sanitarie sostengono un cambiamento di paradigma verso la *pratica basata sulle prove di efficacia* (Evidence-Based Practice), definita come un'integrazione della ricerca del miglior esame disponibile e dell'esperienza clinica con il valore del paziente.<sup>1,2</sup> Le prove di efficacia potrebbero essere incorporate in tutti gli aspetti della fisioterapia nella gestione del paziente e del cliente, inclusi l'esame obiettivo, la valutazione, la diagnosi, la prognosi e la terapia. Probabilmente la componente più importante è un attento e breve esame clinico che può portare a una diagnosi accurata, alla scelta di una terapia adeguata e alla determinazione di una prognosi. Pertanto, è della massima importanza integrare le prove in grado di dimostrare la validità dei test e delle misurazioni cliniche nel distinguere i pazienti che presentano o meno specifiche patologie muscolo-scheletriche.<sup>1,2</sup>

Il processo diagnostico richiede un'anamnesi, lo sviluppo di un'ipotesi di lavoro e la scelta di specifici test e misurazioni per confermare o confutare l'ipotesi formulata. Il clinico deve determinare la probabilità pre-test (prima della valutazione) che il paziente abbia una particolare patologia. Basandosi su questa informazione, il clinico sceglie test e misurazioni adeguati, che lo potranno aiutare a determinare la probabilità post-test (dopo la valutazione) che un paziente abbia una particolare patologia, fino al raggiungimento di un grado di certezza tale che possa portare all'inizio della gestione del paziente (*soglia di trattamento*). La proposta di test clinici non serve per raggiungere la certezza diagnostica, ma piuttosto per ridurre il livello di incertezza fino al raggiungimento del trattamento limite.<sup>2</sup> I concetti di probabilità pre-test e post-test e di trattamento limite saranno approfonditi più avanti nel presente capitolo.

Con l'aumento del numero dei test clinici e delle misurazioni, è essenziale valutare in modo esauriente le proprietà diagnostiche di un test prima di introdurlo nella pratica clinica.<sup>3</sup> L'integrazione della migliore prova disponibile per l'utilità diagnostica di ciascun test clinico è essenziale per porre una diagnosi accurata e per scegliere un trattamento efficace. Sembra quindi ragionevole che i clinici e gli allievi siano informati delle proprietà diagnostiche dei test e delle misurazioni e sappiano quali hanno utilità clinica. Questo testo ha lo scopo di assistere i clinici nella selezione di test e misurazioni che assicurino un'appropriata classificazione dei pazienti e consentano la scelta del trattamento più efficace.

La valutazione dei test diagnostici implica l'esame di un certo numero di proprietà, tra cui la *riproducibilità* (o affidabilità) e l'*utilità* (o accuratezza) diagnostica. Un test viene considerato affidabile se fornisce informazioni precise, esatte e riproducibili. Si considera che un test sia accurato se ha la capacità di discriminare fra pazienti affetti e pazienti non affetti da una specifica patologia.<sup>4</sup> La valutazione scientifica dell'utilità clinica dei test e delle misurazioni in fisioterapia comporta il confronto tra i risultati dell'esame e i gold standard quali gli studi radiografici (che costituiscono la rappresentazione più vicina alla verità). Con l'uso dei metodi statistici generalmente applicati all'epidemiologia, si calcola l'accuratezza diagnostica – cioè la capacità di determinare quali pazienti hanno la patologia e quali no – del test. Questo capitolo mette a fuoco le caratteristiche che definiscono la riproducibilità e l'accuratezza diagnostica di specifici test e misurazioni. Il capitolo si conclude con una discussione relativa alla valutazione di qualità degli studi che indagano l'utilità diagnostica.

Un test clinico che fornisce informazioni che possono portare a prendere una decisione clinica deve essere riproducibile. La *riproducibilità* è il grado di coerenza con cui uno strumento o un indice misurano una particolare qualità.<sup>5</sup> Quando si ricerca la riproducibilità di una misurazione, si deve determinare in che proporzione quella misurazione sia una rappresentazione veritiera e in che proporzione sia frutto di un errore di misurazione.<sup>6</sup>

## RIPRODUCIBILITÀ

Quando si valuta un test clinico, è importante considerare due forme di riproducibilità: riproducibilità *intraesaminatore* e *interesaminatore*. La *riproducibilità intraesaminatore* è una misura della capacità di un singolo osservatore di ottenere una misurazione identica durante esecuzioni ripetute dello stesso test. La *riproducibilità interesaminatore* è una misura della capacità di due o più osservatori di ottenere risultati identici durante l'esecuzione dello stesso test.

Il coefficiente kappa ( $\kappa$ ) è una misura della proporzione del potenziale accordo ottenibile dopo aver rimosso eventuali fattori casuali;<sup>1,5,7</sup> si tratta del coefficiente di riproducibilità utilizzato con maggiore frequenza per i dati categoriali (positivi o negativi).<sup>5</sup> Il coefficiente di correlazione utilizzato comunemente per determinare la riproducibilità dei dati continui in natura (per esempio, i dati relativi al range di movimento) è il coefficiente di correlazione intraclasse (CCD).<sup>7</sup> Sebbene esistano diverse interpretazioni relative alla riproducibilità, questi coefficienti vengono spesso valutati ricorrendo ai criteri descritti da Shrout,<sup>8</sup> in base ai quali valori inferiori a .10 indicano l'assenza di riproducibilità, valori compresi tra .11 e .40 indicano una bassa riproducibilità, valori compresi tra .41 e .60 indicano una riproducibilità sufficiente, valori compresi tra .61 e .80 indicano una riproducibilità discreta e valori superiori a .81 indicano una riproducibilità ottima. Per ogni test o misurazione specifica, è necessario che il clinico definisca il valore di "riproducibilità accettabile",<sup>9</sup> che deve essere determinato in base alla variabile testata, al motivo che giustifica l'importanza di un particolare test e al paziente su cui sarà eseguito il test.<sup>6</sup> Per esempio, un errore di misurazione pari al 5% può essere del tutto accettabile nel caso della determinazione del range di movimento articolare, mentre non è assolutamente ammissibile nel caso della misurazione della temperatura corporea interna in ambito pediatrico.

I test e le misurazioni cliniche non sono in grado di confermare o escludere in maniera assoluta la presenza di una specifica patologia.<sup>10</sup> Tuttavia, i test clinici possono essere utilizzati per modificare la stima fatta dal clinico in merito alla probabilità che un paziente sia affetto da uno specifico disturbo muscolo-scheletrico. L'accuratezza di un test è determinata dal grado di concordanza tra il test clinico e il gold standard.<sup>11,12</sup> Il gold standard è il criterio considerato come la rappresentazione più veritiera della presenza di una patologia.<sup>1</sup> Per determinare la percentuale di soggetti per i quali è stata emessa una diagnosi corretta, ovvero l'accuratezza diagnostica, si effettua un confronto tra i risultati ottenuti con il gold standard e quelli ottenuti utilizzando il test in esame.<sup>13</sup> Dal momento che i calcoli statistici relativi all'accuratezza diagnostica dipendono completamente sia dal gold standard utilizzato sia dalla popolazione in studio, in questo testo si è provveduto a indicare specificamente tali fattori al fine di fornire le informazioni di cui tenere conto nella selezione dei test e delle misurazioni riportati. L'accuratezza diagnostica viene spesso espressa in termini di valore predittivo positivo o negativo (PPV e NPV), sensibilità e specificità e rapporto di verosimiglianza (LR).<sup>1,14</sup>

## UTILITÀ DIAGNOSTICA

## UTILITÀ DIAGNOSTICA

### Tabella di contingenza 2x2

Per determinare l'utilità clinica di un test o di una misurazione, si devono comparare direttamente i risultati del gold standard con quelli del test che si sta analizzando tramite l'utilizzo di una tabella di contingenza 2x2, che fornisce un confronto diretto tra il gold standard e il test in esame.<sup>15</sup> Tale tabella permette di calcolare il valore associato all'accuratezza diagnostica che può aiutare il clinico nel determinare l'utilità del test clinico analizzato (*Tabella 1-1*).

La tabella di contingenza 2x2 è divisa in quattro celle (a, b, c, d) che consentono di determinare l'abilità di un test di identificare correttamente i veri positivi (cella a) e di eliminare i veri negativi (cella d). La cella b rappresenta i falsi positivi, in cui i test diagnostici sono positivi, mentre il gold standard è negativo. La cella c rappresenta i falsi negativi, in cui i test diagnostici sono negativi, mentre il gold standard è positivo.

Una volta che lo studio per analizzare l'utilità diagnostica di un test clinico è terminato ed è stato confrontato con il campione di riferimento nella tabella di contingenza, si può calcolare l'utilità diagnostica in termini di sensibilità, specificità, accuratezza generale, PPV, NPV e LR. Questi valori statistici sono utili al clinico per determinare se un test diagnostico è efficace per valutare la presenza o meno di una patologia.

### Accuratezza generale

L'accuratezza generale di un test diagnostico è determinata dalla divisione delle risposte corrette (veri positivi e veri negativi) per il numero totale di pazienti.<sup>16</sup> Utilizzando la tabella di contingenza 2x2, l'accuratezza generale può essere determinata ricorrendo alla seguente equazione:

$$\text{Accuratezza generale} = 100\% \times (a+d)/(a+b+c+d) \quad (1-1)$$

Un test perfetto dovrebbe mostrare un'accuratezza generale del 100%. Questo valore è molto probabilmente irraggiungibile, in quanto nessun test clinico è perfetto e ciascun test mostra sempre almeno un piccolo grado di incertezza. L'accuratezza di un test diagnostico non dovrebbe essere usata per accertare l'utilità clinica di quel test, dal momento che l'accuratezza generale può essere un po' fuorviante. L'accuratezza di un test può essere significativamente influenzata dalla prevalenza, o presenza assoluta, di una patologia nella popolazione in un dato momento.<sup>5,6</sup>

**Tabella 1-1:** Tabella di contingenza 2x2 usata per comparare i risultati del gold standard con quelli del test analizzato

	Gold standard positivo	Gold standard negativo
Test clinico positivo	Vero positivo a	Falso positivo b
Test clinico negativo	Falso negativo c	Vero negativo d

**Valore predittivo positivo e negativo**

Il valore predittivo positivo (PPV) stima la possibilità che un paziente positivo a un test sia realmente affetto da una patologia.<sup>5,6,17</sup> Il PPV, che viene calcolato orizzontalmente nella tabella di contingenza 2x2 (*Tabella 1-2*), corrisponde alla percentuale di pazienti accuratamente identificati come aventi la patologia (veri positivi) divisa per tutti i risultati positivi del test in esame. Un PPV elevato indica che un risultato positivo è altamente predittivo per la presenza della patologia.<sup>5,6</sup> La formula per calcolare il PPV è la seguente:

$$PPV = 100\% \times a/(a+b) \tag{1-2}$$

Il NPV stima la verosimiglianza che un paziente negativo a un test non abbia realmente la patologia.<sup>5,6</sup> Anche il NPV viene calcolato orizzontalmente nella tabella di contingenza 2x2 (*Tabella 1-2*) e indica la percentuale di pazienti accuratamente identificati come non aventi la patologia (veri negativi) diviso tutti i risultati negativi del test in esame.<sup>11</sup>

La formula per calcolare il NPV è:

$$NPV = 100\% \times d/(c+d) \tag{1-3}$$

I valori predittivi sono influenzati in maniera significativa dalla prevalenza della condizione.<sup>11</sup> Perciò tali valori non sono stati riportati specificamente in questo testo.

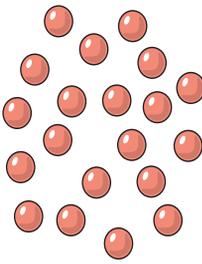
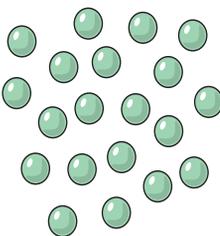
**Sensibilità**

La *sensibilità* di un test diagnostico è la capacità di un test, quando positivo, di identificare quali pazienti abbiano realmente la patologia, sulla base del gold standard. La sensibilità è chiamata anche *percentuale dei veri positivi*.<sup>1</sup> Test con alta sensibilità sono ottimi per escludere una particolare patologia. L'acronimo *SnNout* può essere usato per ricordare che un test con alta Sensibilità e un risultato Negativo è ottimo per escludere una patologia.<sup>1</sup>

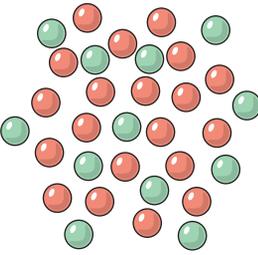
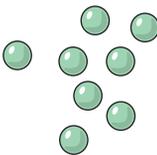
**Tabella 1-2:** Tabella di contingenza 2x2 che mostra il calcolo del valore predittivo positivo e negativo orizzontalmente e la sensibilità e la specificità verticalmente

	Gold standard positivo	Gold standard negativo	
Test clinico positivo	Veri positivi a	Falsi positivi b	PPV = a/(a+b)
Test clinico negativo	Falsi negativi c	Veri negativi d	NPV = d/(c+d)
	Sensibilità = a/(a+c)	Specificità = d/(b+d)	

UTILITÀ  
DIAGNOSTICA

20 pazienti con la patologia	20 pazienti senza la patologia
	

**Figura 1-1:** Esempio relativo alla sensibilità e alla specificità. 20 pazienti con e 20 pazienti senza la patologia.

100% Sensibilità	
20 pazienti con la patologia	20 pazienti senza la patologia
	

**Figura 1-2:** Una sensibilità del 100% implica che, se il test è positivo, tutti i soggetti con la patologia saranno inclusi. Tuttavia, si noti che anche se questo test ha identificato tutti i soggetti con la patologia, ha incluso molti soggetti esenti da patologia. Inoltre, se il risultato del test è negativo, siamo sicuri che il soggetto non sia affetto dalla patologia.

Si consideri, per esempio, un test clinico che, confrontato con il gold standard, mostri un'alta sensibilità per la diagnosi della stenosi lombare. Secondo la suddetta regola, se il test è negativo, si può attendibilmente eliminare la possibilità di essere di fronte a una stenosi lombare. Se il test è positivo, è possibile identificare accuratamente un'alta percentuale di pazienti che presentino una stenosi. Tuttavia, è possibile anche identificare molti pazienti senza patologia (falsi positivi). Quindi, sebbene un risultato negativo sia particolarmente indicativo, un risultato positivo del test non ci permette di trarre alcuna conclusione (*Figure 1-1 e 1-2*).

La sensibilità di un test può essere calcolata anche con tabelle di contingenza 2x2. Tuttavia, è calcolata verticalmente (*Tabella 1-2*). La formula per calcolare la sensibilità di un test è la seguente:

$$\text{Sensibilità} = 100\% \times a/(a+c) \quad (1-4)$$

**Specificità**

La *specificità* di un test diagnostico indica la capacità di un test, quando negativo, di individuare quei pazienti che effettivamente non presentano la patologia, sulla base del gold standard.

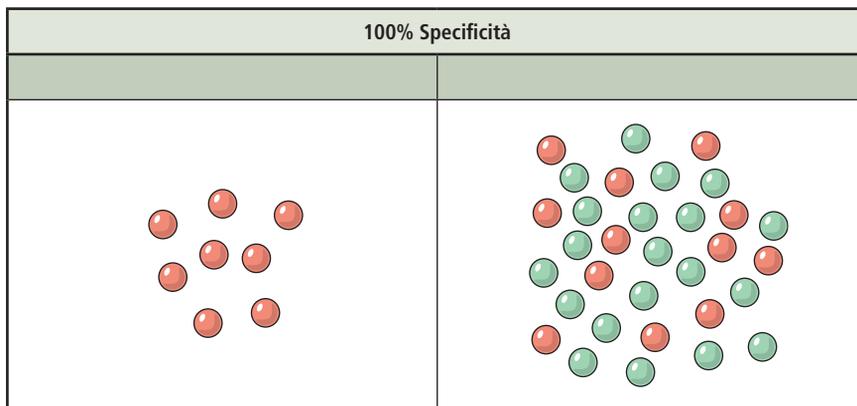
È chiamata anche *percentuale di veri negativi*.<sup>1</sup> Test con alta specificità sono ottimi per accertare la presenza di una patologia. L'acronimo *SpPin* può essere utilizzato per ricordare che un test con alta Specificità e un risultato *Positivo* è ottimo per accertare la patologia.<sup>16,18,19</sup>

Si consideri un test con un'alta specificità. Dovrebbe identificare accuratamente tutti i pazienti che non presentano la patologia. Se un test altamente specifico è negativo, è possibile identificare un'alta percentuale di pazienti che non presentano la patologia. Tuttavia, è possibile che un test altamente specifico possa identificare come negativi un numero di pazienti realmente affetti dalla patologia (falsi negativi). Quindi, è possibile essere ragionevolmente sicuri che un test altamente specifico che risulti positivo indica la presenza di patologia (*Figura 1-3*).

La formula per calcolare la specificità di un test è la seguente:

$$\text{Specificità} = 100\% \times d/(b+d) \tag{1-5}$$

La sensibilità e la specificità sono state usate per decenni per determinare l'utilità diagnostica di un test; tuttavia, sono associate a poche limitazioni cliniche.<sup>11</sup> Anche se la sensibilità e la specificità possono essere utili al clinico per scegliere il test migliore per accertare o escludere una particolare patologia, pochi test clinici dimostrano sia un'elevata specificità sia un'elevata sensibilità.<sup>11</sup> Inoltre, la sensibilità e la specificità non forniscono informazioni riguardanti la variazione di probabilità che un paziente abbia una patologia se il test risulta positivo o negativo.<sup>18,20</sup> Invece, il LR è considerato il migliore strumento statistico per determinare le variazioni della probabilità pre-test che un paziente abbia una specifica patologia.



**Figura 1-3:** Una specificità del 100% implica che, se il risultato del test è negativo, tutti i soggetti senza la patologia saranno inclusi. Tuttavia, si noti che anche se questo test ha identificato tutti i soggetti senza la patologia, ha incluso molti soggetti affetti dalla patologia. Inoltre, se il test è positivo, siamo sicuri che il paziente abbia la patologia (*SpPin*).

## UTILITÀ DIAGNOSTICA

### Rapporto di verosimiglianza (Likelihood Ratio, LR)

Il risultato di un test è utile solo se modifica la probabilità pre-test che un paziente abbia una patologia.<sup>21</sup> Il LR combina la sensibilità e la specificità di un test in modo da fornire un'indicazione della variazione di probabilità associata a uno specifico risultato del test ed è utile nel guidare la formulazione della decisione clinica.<sup>20</sup> Il LR è un metodo di misurazione efficace che può aumentare o ridurre significativamente la probabilità che un paziente presenti una patologia.<sup>22</sup>

Il LR può essere positivo o negativo. Un LR positivo indica uno spostamento della probabilità a favore dell'esistenza di una patologia, mentre un LR negativo indica uno spostamento della probabilità a favore dell'assenza di una patologia. Anche se spesso il LR non viene riportato negli studi che esaminano l'utilità diagnostica dell'esame clinico, può essere calcolato facilmente se sono disponibili la specificità e la sensibilità del test. In questo testo, per gli esami che non riportano il LR ma che dispongono della sensibilità e della specificità, gli autori hanno calcolato il LR.

La formula usata per determinare un LR positivo è la seguente:

$$\text{LR} = \text{Sensibilità}/(1-\text{Specificità}) \quad (1-6)$$

La formula usata per determinare un LR negativo è la seguente:

$$\text{LR} = (1-\text{Sensibilità})/\text{Specificità} \quad (1-7)$$

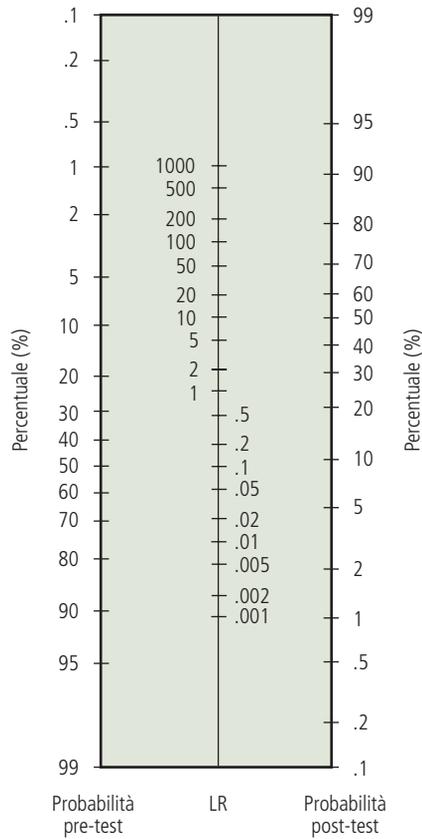
La *Tabella 1-3* fornisce la guida per interpretare i risultati del test. LR positivi maggiori di 1 aumentano la probabilità di esistenza di una patologia dato un test positivo; LR negativi inferiori a 1 diminuiscono la probabilità di una patologia dato un test negativo.<sup>22</sup> Tuttavia, è l'ampiezza della variazione della probabilità a determinare l'utilità di un test clinico. LR positivi maggiori di 10 e LR negativi vicini allo zero spesso rappresentano una grande e significativa variazione della probabilità. Un LR pari a 1 (sia negativo sia positivo), che non altera la probabilità che il paziente presenti o meno una particolare patologia, ha poco valore clinico.<sup>22</sup> Una volta calcolato il LR, si può applicare quest'ultimo al nomogramma (*Figura 1-4*)<sup>23</sup> oppure si può utilizzare un'equazione matematica<sup>24</sup> per determinare precisamente le variazioni della probabilità

**Tabella 1-3:** Interpretazione dei rapporti di verosimiglianza

Rapporto di verosimiglianza positivo	Rapporto di verosimiglianza negativo	Interpretazione del rapporto
> 10	< 0.1	Grande e decisiva variazione della probabilità
Da 5 a 10	Da 0.1 a 0.2	Moderata variazione della probabilità
Da 2 a 5	Da 0.2 a 0.5	Piccola ma a volte importante variazione della probabilità
Da 1 a 2	Da 0.5 a 1.0	Piccola e raramente importante variazione della probabilità

(Modificata da: 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;271:703-707.)

UTILITÀ  
DIAGNOSTICA



**Figura 1-4:** Nomogramma di Fagan.

(Modificato da: Fagan TJ. Letter: nomogram for Bayes theorem. *N Engl J Med.* 1975;293:257. Copyright 2005, Massachusetts Medical Society. Tutti i diritti riservati.)

dato il risultato di uno specifico test. Entrambi i metodi sono descritti nel dettaglio più avanti nel capitolo.

Se un esame diagnostico mostra una specificità pari a 1, il LR positivo non può essere calcolato perché l'equazione risulterà zero per il denominatore; in questo caso, si suggerisce di modificare la tabella di contingenza 2x2, aggiungendo .5 a ciascuna cella per permettere il calcolo del LR.<sup>25</sup>

Si consideri, per esempio, l'utilità diagnostica del Crank test<sup>5,26</sup> nel rilevare la presenza di lesioni del cerniere glenoideo rispetto alla valutazione artroscopica, attuale standard di riferimento. Questo confronto viene effettuato utilizzando una tabella di contingenza 2x2 (Tabella 1-4). Applicando la formula sotto riportata, risulta evidente l'impossibilità di calcolare un LR positivo:

$$\text{LR positivo} = \text{Sensibilità}/(1-\text{Specificità}) = 1/(1-1) = 1/0 \quad (1-8)$$

Dal momento che è impossibile calcolare il risultato di una frazione avente per denominatore lo zero, la tabella di contingenza 2x2 viene modificata aggiungendo il valore .5 a ciascuna cella.

Anche se l'aggiunta di .5 a ciascuna cella è l'unico metodo di modifica della tabella di contingenza riportato per prevenire lo zero al denominatore nel calcolo di un LR, considerate le modifiche a carico delle proprietà diagnostiche della sensibilità,

**UTILITÀ  
DIAGNOSTICA**

**Tabella 1-4:** Risultato del Crank test nell'accertamento delle lesioni del cingolo glenoide quando comparate con la valutazione artroscopica (gold standard)<sup>32</sup>

	Esame artroscopico positivo (n = 12)	Esame artroscopico negativo (n = 3)	
<b>Crank test positivo</b>	10 a	0 b	Valore predittivo positivo = $100\% \times 10/10 = 100\%$
<b>Crank test negativo</b>	2 c	3 d	Valore predittivo negativo = $100\% \times 3/5 = 60\%$
	Sensibilità = $100\% \times 10/12 = 83\%$	Specificità = $100\% \times 3/3 = 100\%$	

della specificità e del valore predittivo, questa tecnica non è stata usata in questo testo. Nelle circostanze in cui la specificità è zero e non si può calcolare il LR positivo, il risultato sarà indicato come “indefinito”. In questi casi, anche se non calcoliamo il LR positivo, il test potrà essere indicativo di una grande variazione della probabilità.

**INTERVALLO  
DI CONFIDENZA**

I calcoli relativi alla sensibilità, alla specificità e al LR vengono definiti *stime puntuali*. In altre parole, si tratta delle singole stime maggiormente rappresentative dei valori di una popolazione.<sup>5</sup> Tuttavia, poiché le stime puntuali si basano su piccoli sottogruppi di individui (campioni), è improbabile che costituiscano una rappresentazione perfetta della popolazione generale. È pertanto più accurato includere un intervallo di valori (*stima intervallare*) all'interno del quale è probabile che ricada il valore relativo alla popolazione. L'*intervallo di confidenza* (IC) corrisponde a un range di valori costruito intorno alla stima puntuale, che contiene con una certa probabilità il valore relativo alla popolazione.<sup>27</sup> Generalmente, negli studi che indagano l'utilità diagnostica di un esame clinico, si calcola l'IC al 95%. L'IC al 95% rappresenta l'intervallo di dispersione dei valori entro il quale ricade, con una certezza del 95%, il valore della popolazione.<sup>5</sup> Nel presente testo, viene riportato l'IC al 95% per tutti gli studi che forniscono questa informazione.

**PROBABILITÀ  
PRE-TEST  
E POST-TEST**

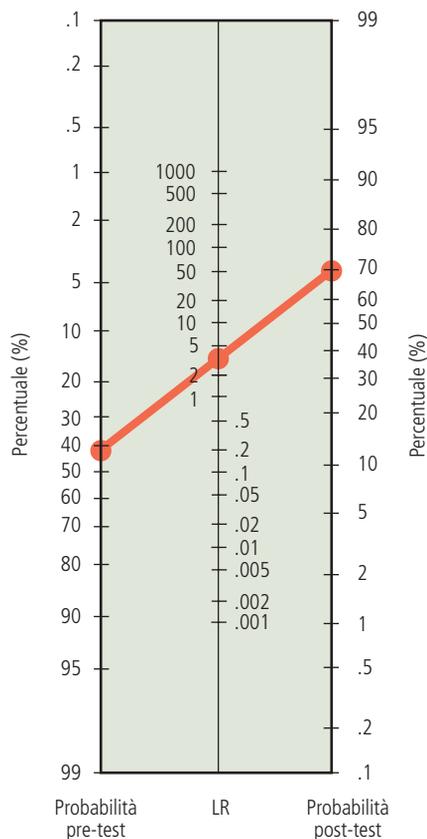
La probabilità pre-test indica la possibilità che un paziente presenti una specifica patologia prima dell'esecuzione dell'esame clinico. Spesso, come misura indicativa della probabilità pre-test, vengono impiegati i tassi di prevalenza; tuttavia, quando i tassi di prevalenza sono sconosciuti, la determinazione della probabilità pre-test si basa su una combinazione di dati derivati dall'anamnesi, dai risultati di test precedenti e dall'esperienza del medico.<sup>16</sup> La determinazione della probabilità pre-test è il primo passo nel processo decisionale del medico. La probabilità pre-test è una stima fatta dal medico e può essere espressa in termini percentuali (per esempio, 75% o 80%) o come misura qualitativa (per esempio, abbastanza probabile, molto probabile).<sup>11,16</sup>

In seguito alla determinazione della probabilità pre-test che un paziente abbia una specifica patologia, è opportuno scegliere per l'esame obiettivo i test e le misurazioni in grado di alterare tale probabilità. La probabilità post-test è la possibilità che un paziente abbia una specifica patologia dopo l'esecuzione dell'esame obiettivo.

**PROBABILITÀ  
PRE-TEST  
E POST-TEST**

Come precedentemente menzionato, il LR può aiutare il clinico a determinare la variazione della probabilità che dovrebbe presentarsi dopo aver ottenuto il risultato di un test e che dipende dal valore del LR del test. Il metodo più veloce per determinare la variazione della probabilità, una volta ottenuto il LR, è l'utilizzo del nomogramma (Figura 1-5).<sup>23</sup> Il nomogramma è un diagramma che mostra la probabilità pre-test sulla sinistra e la probabilità post-test sulla destra. Al centro sono indicati gli LR. Per determinare lo spostamento della probabilità, si pone un segno sul nomogramma per rappresentare la probabilità pre-test. Quindi, si colloca un segno sul nomogramma a livello del LR (sia negativo sia positivo). I due segni vengono uniti con una linea che viene prolungata verso la destra del diagramma. Il punto in cui la linea incrocia la scala della probabilità post-test indica la variazione della probabilità.

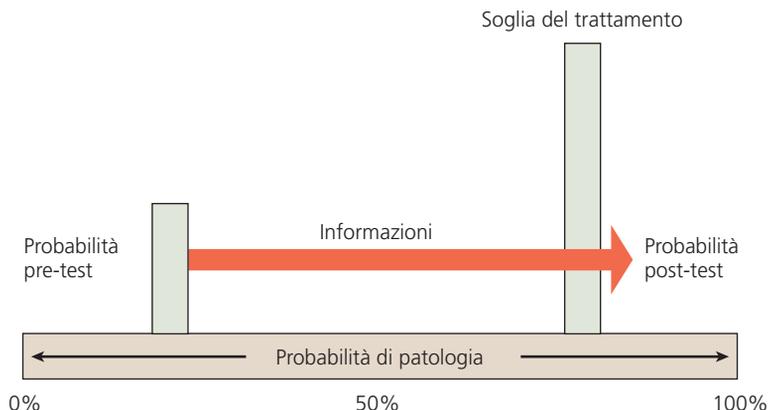
**CALCOLO  
DELLA  
PROBABILITÀ  
POST-TEST**



**Figura 1-5:** Nomogramma che rappresenta la variazione da una probabilità pre-test del 42% se il test è positivo (LR positivo = 4,2) a una probabilità post-test del 71%.

(Modificato da: Fagan TJ. Letter: nomogram for Bayes theorem. *N Engl J Med.* 1975;293:257. Copyright 2005, Massachusetts Medical Society. Tutti i diritti riservati.)

## CALCOLO DELLA PROBABILITÀ POST-TEST



**Figura 1-6:** Soglia di trattamento. Il clinico deve usare la probabilità pre-test e il LR per determinare la soglia del trattamento, come indicato in questa figura.

Una determinazione più precisa della variazione di probabilità può essere ottenuta algebricamente utilizzando la seguente formula:<sup>16</sup>

$$\text{Step 1. ODDS pre-test} = \text{Probabilità pre-test} / (1 - \text{Probabilità pre-test}) \quad (1-9)$$

$$\text{Step 2. ODDS pre-test} \times \text{LR} = \text{post-test ODDS} \quad (1-10)$$

$$\text{Step 3. ODDS post-test} / (\text{ODDS post-test} + 1) = \text{Probabilità post-test} \quad (1-11)$$

Il clinico deve determinare quando la probabilità post-test è abbastanza bassa per escludere la presenza di una certa patologia o quando la probabilità post-test è abbastanza alta da poter affermare con certezza la presenza della patologia. Il livello a cui cessa la valutazione e inizia il trattamento è noto come *soglia di trattamento* (Figura 1-6).<sup>16</sup>

## VALUTAZIONE DELLA QUALITÀ DEGLI STUDI

Dopo aver recuperato gli articoli scientifici pertinenti, il passo successivo consiste nell'analisi critica del loro contenuto per valutare l'adeguatezza del rigore metodologico. È stato riportato che la qualità metodologica degli studi che indagano l'utilità diagnostica dell'esame obiettivo è generalmente inferiore a quella degli studi diretti a valutare l'efficacia delle terapie.<sup>28,29</sup> Purtroppo, la presenza di studi con errori metodologici significativi che riportano l'utilità di specifici test e misurazioni può portare all'utilizzo di test inefficaci. Ciò può condurre a diagnosi errate e a una gestione inadeguata dei pazienti. Al contrario, l'identificazione e l'uso di test clinici valutati con rigore scientifico possono migliorare l'assistenza ricevuta dai pazienti e i relativi esiti.<sup>29</sup>

Per valutare la qualità degli studi di riproducibilità diagnostica, è stato sviluppato uno strumento chiamato QAREL (Quality Appraisal for Reliability Studies).<sup>30</sup> Il QAREL è costituito da una checklist di 11 item elaborata in consultazione con un gruppo di esperti nel campo della ricerca diagnostica e della valutazione della qualità, che viene utilizzato per valutare la qualità metodologica di uno studio di riproducibilità. A ogni item viene assegnato un valore di "sì", "no", "non chiaro" o "N/A". Il QAREL si è dimostrato uno strumento di valutazione affidabile quando agli esaminatori viene data l'opportunità di discutere i criteri in base ai quali interpretare ciascun item.<sup>31</sup> La

riproducibilità di 9 item su 11 viene considerata come una buona riproducibilità diagnostica, mentre 2 item su 11 indicano una debole riproducibilità.<sup>31</sup> Il QAREL è stato utilizzato per valutare gli studi di affidabilità citati nel volume. A questo scopo sono stati considerati di buona qualità i punteggi  $\geq 75\%$  e di qualità discreta quelli compresi tra il 74% e il 50%. Le percentuali sono state calcolate dividendo il numero di risposte "sì" per 11 e sottraendo il numero di risposte "N/A". I simboli verdi indicano un elevato livello di qualità metodologica e suggeriscono che i risultati dei relativi studi siano affidabili. I simboli gialli indicano una qualità metodologica discreta e riferiscono che i risultati devono essere interpretati con cautela. Gli studi giudicati di scarsa qualità metodologica non sono stati inclusi nelle tabelle di utilità diagnostica nei vari capitoli.

Il Quality Assessment of Diagnostic Accuracy Studies (QUADAS) è stato sviluppato per valutare la qualità degli studi di accuratezza diagnostica.<sup>32</sup> Tramite un panel Delphi articolato in quattro round, il QUADAS valuta la qualità metodologica di uno studio secondo 14 item, a ciascuno dei quali viene assegnato un valore di "sì", "no", "non chiaro". Il QUADAS non ha lo scopo di produrre un punteggio numerico per ciascuna valutazione, ma piuttosto di fornire una valutazione qualitativa con l'identificazione delle debolezze metodologiche.<sup>32</sup> Il QUADAS ha dimostrato di essere adeguatamente conforme ai singoli item inclusi nella checklist.<sup>33</sup>

Il QUADAS è stato utilizzato per valutare tutti gli studi citati nel volume. A questo scopo sono stati considerati di buona qualità i punteggi  $\geq 75\%$  e di qualità discreta quelli compresi tra il 74% e il 50%. Le percentuali sono state calcolate dividendo il numero di risposte "sì" per 14 (il numero totale degli item). I simboli verdi indicano un elevato livello di qualità metodologica e suggeriscono che i risultati siano affidabili. I simboli gialli indicano una qualità metodologica discreta e riferiscono che i risultati devono essere interpretati con cautela. Gli studi giudicati di scarsa qualità metodologica non sono stati inclusi nelle tabelle di utilità diagnostica nei vari capitoli.

È importante considerare riproducibilità e utilità diagnostica di test e misurazioni prima che essi vengano inclusi nell'esame clinico. I test e le misurazioni dovrebbero mostrare un'adeguata riproducibilità, prima di essere utilizzati per guidare le scelte cliniche. In questo testo sono stati riportati i dati sulla riproducibilità dei più importanti test e misurazioni. È essenziale che il clinico consideri i livelli di riproducibilità nel contesto.

Prima di integrare i test e le misurazioni nell'esame obiettivo ortopedico, bisogna considerare la riproducibilità di ciascun test. La *Tabella 1-5* riassume i valori statistici correlati all'accuratezza diagnostica, così come le equazioni matematiche e le descrizioni tecniche per ciascun test.

L'utilità di un test e di una misurazione è più comunemente considerata in termini di proprietà diagnostiche del rispettivo test. Queste possono essere descritte in termini di sensibilità, specificità, PPV e NPV. Tuttavia, la proprietà diagnostica più utile è il LR, che può aiutare nello stabilire la variazione della probabilità che un paziente abbia una specifica patologia.

I test clinici e le misurazioni non forniscono una certezza assoluta di assenza o presenza della patologia. Tuttavia, i clinici possono determinare quando sufficienti dati sono stati raccolti per aumentare la probabilità di presenza della patologia oltre la "soglia di trattamento" e quindi terminare la valutazione e iniziare la terapia. Inoltre,

## RIEPILOGO

un'attenta valutazione metodologica garantisce una migliore comprensione del rigore scientifico di ciascuno studio, nonché dei parametri a esso associati, come la resa, l'applicabilità, l'affidabilità e la riproducibilità nell'ambito di una data pratica clinica.

**Tabella 1-5:** Tabella di contingenza 2x2 e valori statistici utilizzati per determinare l'utilità diagnostica del test e della misurazione

	Gold standard positivo		Gold standard negativo	
Test clinico positivo	Veri positivi	a	b	Falsi positivi
Test clinico negativo	Falsi negativi	c	d	Veri negativi



Valore statistico	Formula	Descrizione
Accuratezza generale	$(a+d)/(a+b+c+d)$	Percentuale di individui che sono stati sottoposti a diagnosi corretta
Sensibilità	$a/(a+c)$	Proporzione di pazienti con patologia che hanno avuto un risultato positivo ai test
Specificità	$d/(b+d)$	Proporzione di pazienti senza patologia che hanno avuto un risultato negativo ai test
Valore predittivo positivo	$a/(a+b)$	Proporzione di pazienti che hanno avuto un risultato positivo ai test che presentano la patologia
Valore predittivo negativo	$d/(c+d)$	Proporzione di pazienti che hanno avuto un risultato negativo ai test che non presentano la patologia
LR positivo	$\text{sensibilità}/(1 - \text{specificità})$	Se il test è positivo, aumenta la probabilità della presenza di una patologia
LR negativo	$(1 - \text{sensibilità})/\text{specificità}$	Se il test è positivo, diminuisce la probabilità della presenza di una patologia

## BIBLIOGRAFIA

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;320:1489-1491.
3. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282:1061-1066.
4. Schwartz JS. Evaluating diagnostic tests: what is done—what needs to be done. *J G Intern Med*. 1986;1:266-267.
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, Echternach 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;7:301-317.

9. Van Genderen F, De Bie R, Helders P, Van Meeteren N. Reliability research: towards a more clinically relevant approach. *Physical Therapy Reviews*. 2003;8:169-176.
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;81:1546-1564.
12. Jaeschke R, Guyatt GH, Sackett DL III. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA*. 1994;271:389-391.
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;315:540-543.
16. Bernstein J. Decision analysis (current concepts review). *J Bone Joint Surg*. 1997;79:1404-1414.
17. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. *Phys Ther*. 1985;65:1671-1675.
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;14:175-180.
19. Riddle DL, Stratford PW. Interpreting validity indexes for diagnostic tests: an illustration using the Berg balance test. *Phys Ther*. 1999;79:939-948.
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;33:575-580.
21. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol*. 1991;44:763-770.
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;271:703-707.
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;27:137-142.
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;15:119-126.
28. Moons KGM, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem*. 2004;50:473-476.
29. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. *JAMA*. 1995;274:645-651.
30. 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.
31. 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.
32. 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):854-861.
33. 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.

