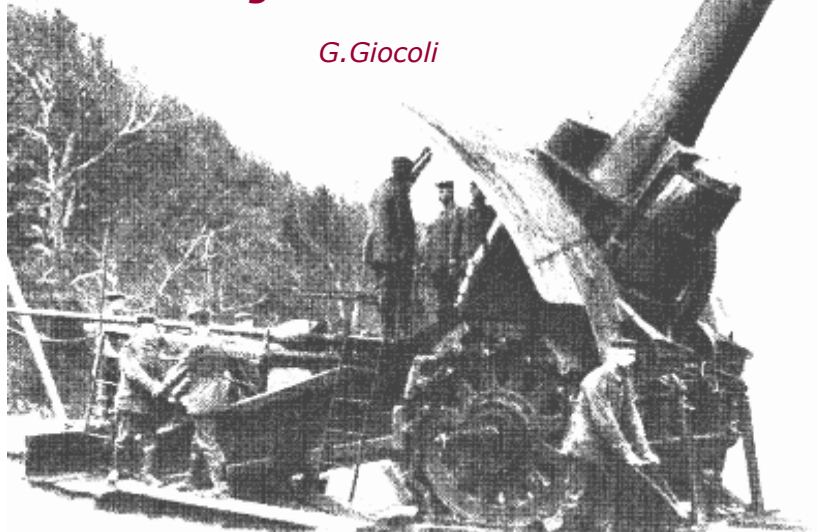


Cannoni d'agosto

Blog di mezza estate

G. Giocoli



"*Likelihood ratio*: non usateli per calcolare numeri senza significato, ma come misura della validità dei test. Considerate i loro limiti di confidenza come misura dell'accuratezza di questa stima ...".

Leggo sul *British Medical Journal* del 7 agosto questo incisivo commento di un oncologo canadese (1) ad un articolo sul corretto impiego dei *Likelihood ratio* (LR) (2) e l'associo alle voci di iniziative - attuate o in via di attuazione in alcuni laboratori italiani - per la refertazione di questi indici assieme ai risultati dei test.

A mio parere, i rilievi del ricercatore hanno colpito nel segno e, per quanto io possa aver capito dalla quasi quotidiana consultazione della letteratura sull'argomento, l'idea di segnalare i *likelihood ratio* nei referti di laboratorio mi sembra prematura, inopportuna e rischiosa.

Se - nell'attuale realtà italiana - questi indici (e relative istruzioni per l'uso) saranno a disposizione di tutti, quali saranno le conseguenze per i pazienti? Quale sarà il futuro dei *likelihood ratio* nella diagnostica? Li attende il riconoscimento di indicatori di qualità o una squallida fine nel dimenticatoio delle cose inutili? Oppure una sinistra fama di apportatori di danno iatrogeno?

Likelihood ratio

I *likelihood ratio* sono indici che esprimono la potenza (accuratezza) di un test diagnostico, ossia la capacità di distinguere chi ha una malattia da chi non ce l'ha. Qualsiasi test dicotomico (con risultato positivo/negativo, presente/assente, o simili) può dare risultati veri o falsi e, se ammettiamo che il numero UNO esprima l'equivalenza di queste possibilità, i test clinicamente più utili sono quelli caratterizzati da LR che si allontanano dall'unità, in direzione dell'infinito per i risultati positivi, in direzione dello zero per i risultati negativi. Il perchè è facile da comprendere, esprimendo entrambe le tendenze lo sbilanciamento in favore dei risultati veri nei confronti di quelli falsi.

Per le loro qualità, i *likelihood ratio* rendono realizzabile l'enunciato di Bayes "Le probabilità che il risultato di un test corrisponda o meno alla presenza della malattia dipendono dalle probabilità pre-test e dalla potenza del test".

Bayes: quando e come

Sackett ha auspicato che il medico, alle prese con il dilemma diagnostico, adoperi i *likelihood ratio* di un test per modificare le sue ipotesi sulla malattia. Egli può utilizzare un calcolo aritmetico o il nomogramma di Fagan oppure, in casi particolari, gli *SPin* o gli *SNout*.

SPin e *SNout* sono proprietà conferite ad alcuni test da una specificità o una sensibilità talmente prossime al 100% (o coincidenti con esso) da renderne assiomatici i risultati positivi o negativi, per cui una certa diagnosi sarebbe da essi confermabile o escludibile senza necessità di ulteriori indagini. Tuttavia, la validità di molti *SPin* e *SNout* riportati nel libro di Sackett e nel sito web del Centro EBM di Oxford risulta chiaramente contestabile, come provato dall'epidemiologo svizzero Daniel Pewsner (2).

Ma non basta: Andrew Robinson, il citato oncologo di Vancouver, ha esteso le sue critiche all'uso attuale e generalizzato dei *likelihood ratio* – grandi o piccoli che siano – per calcolare il valore predittivo dei test al letto del malato: i loro limiti di confidenza sono così ampi e le incertezze sulle probabilità pre-test così sensibili da annullare il significato pratico di tale procedura (1).

Statu quo

La logica bayesiana e i relativi strumenti di attuazione – LR compresi – sono oggetto di intense ricerche, soprattutto nei Paesi dell'area anglosassone – Australia, Canada, Regno Unito, USA – e dell'Europa settentrionale (Paesi Bassi). Non so quanto i *likelihood ratio* siano realmente entrati nella locale pratica clinica, ma è probabile che minoranze elette ne facciano uso.

Il problema è che la stima dell'accuratezza di un test (validazione) deve poggiare su solide fondamenta scientifiche edificate con studi di disegno appropriato, concernenti sia la validazione interna del test (corretto confronto con un idoneo gold standard, popolazione appropriata, numerosità adeguata), sia la sua validazione esterna (generalizzabilità e applicabilità dei risultati). Sono attualmente ben pochi i test che soddisfano a pieno tali requisiti.

I recenti protocolli STARD e QUADAS forniscono indicazioni sulle corrette procedure da seguire ed è probabile, dicevo, che in alcuni centri si ottengano stime affidabili dell'accuratezza (con relativi limiti di confidenza) dei test effettuati nei locali laboratori. Se a ciò si uniscono la disponibilità di rilievi epidemiologici e di regole predittive per definire il secondo termine dell'enunciato bayesiano (le probabilità pre-test), non può stupire la notizia che in ospedali australiani o canadesi i clinici usano il palmare o il diagramma di Fagan per "aggiornare" le loro opinioni sulle condizioni del malato all'arrivo (in tempo reale) del risultato di un test.

Ed è certo che in diversi centri (specie universitari) del Nordamerica sono disponibili mini-protocolli (CATs) per la diagnosi rapida di malattie, anche da infezione (es. le meningiti). Nei CATs (*Critically Appraised Topics*) sono riportate in sintesi le prestazioni di alcuni test (sotto forma di LR) per aiutare i medici, specie nei reparti di emergenza, a ottenere risposte critiche mediante un semplice calcolo probabilistico.

Tali procedure restano tuttavia fuori portata dei più, sconosciute o addirittura avversate. Non tanto (in quest'ultima evenienza) per i motivi esposti, che presuppongono una non superficiale conoscenza dell'argomento, quanto per la riluttanza dei medici ad abbandonare tradizionali percorsi diagnostici ispirati alle opinioni personali piuttosto che all'evidenza obiettiva.

How good is a test?

Le argomentazioni di Pewsner e di Robinson non appartengono ad una delle tante campagne di demolizione dell'EBM, ma invocano un uso più adeguato alle conoscenze attuali di uno dei suoi strumenti più preziosi.

"... piuttosto serviamoci dei *likelihood ratio* per valutare la bontà dei test" (*as a barometer of how good a test is*)," esorta Robinson.

I migliori frutti della ricerca per una corretta definizione della potenza di un test - sono le Revisioni sistematiche (RS) e delle metanalisi in campo diagnostico. La revisione sistematica è una "summa" dei dati scientifici su un dato argomento, ottenuta dalla selezione della letteratura secondo criteri standardizzati e riproducibili. In molte RS sono ormai riportati sia la sensibilità e la specificità dei test che i relativi LR positivi e negativi, con i relativi limiti di confidenza.

Cito esempi di patologie per le quali sono state compiute profonde analisi critiche dei test pertinenti mediante le RS, spesso rivedute e chiosate da eminenti istituti quali il CRD dell'università di York: le infezioni genitali, ematiche, urinarie, da HIV1, il cancro della prostata e del colon, il diabete, l'insufficienza cardiaca, le fratture del malleolo e del piede ("regole di Ottawa"). Dopo anni dedicati alla revisione dei dati in campo terapeutico, anche la *Cochrane Collaboration* ha deciso di aprire una sezione per le RS in diagnostica.

Nelle revisioni sistematiche le stime dei *likelihood ratio* e dei relativi limiti di confidenza forniscono informazioni rilevanti per comprendere potenzialità e limiti dei test, che si traducono in elementi di giudizio e di confronto per istituzioni e singoli professionisti intenzionati a migliorare il proprio armamentario per la diagnosi e lo screening delle malattie.

Ma si è appena all'inizio di un duro cammino. Le "spietate" procedure di revisione hanno messo in luce enormi carenze nel campo diagnostico, da tempo bollate al vetriolo dalla rivista britannica *Bandolier*. L'aspetto positivo della vicenda è un vivace impulso alla standardizzazione delle procedure di validazione dei test, ad esempio i citati protocolli STARD e QUADAS, o altre iniziative dell'FDA statunitense. L'intento è di elevare il livello degli studi di validazione sì da rendere più agevole ed efficace il lavoro dei revisori.

Chi legge il barometro?

A chi è dunque rivolto l'invito di Robinson? A chi spetta la lettura del "barometro dei test"? Il ricercatore non lo dice, ma suppongo che la risposta sia "Ai clinici". Credo sia giusto. Ma, a mio parere, ad essi vanno affiancati coloro che i test li propongono e li eseguono: gli specialisti dei servizi diagnostici.

Proprio nello scorso luglio mi son permesso di proporre al *British Medical Journal* un *Future Theme Issue* sull'applicazione dell'EBM nei laboratori ospedalieri (*The EBM in the basement*). Appare opportuno che, nei nostri, non ci si affidi soltanto alla luce di lontane stelle polari. Le revisioni sistematiche sono ancora poche e solo alcune includono risultati ottenuti nel nostro Paese.

Ogni centro di riferimento dovrebbe avviare una supervisione critica dei test, ad iniziare dai più importanti, per stabilire la validità locale di dati ottenuti altrove, spesso in altri continenti. A questo scopo esprimere la potenza dei test sotto forma di *likelihood ratio* può essere assai utile, ma lasciandone l'interpretazione agli addetti ai lavori. Questi, a loro volta, dovrebbero avvalersi di tecniche per il trasferimento della conoscenza alla compagine sanitaria e ai pazienti (*knowledge translation, knowledge brokering, shared-decision making*).

"I fear me both are false" (Shakespeare, Richard III, I-ii)

Il problema, a mio modesto parere, è che non è possibile eseguire siffatti studi senza un organico e comune impegno di laboratoristi, clinici, epidemiologi e altri specialisti: basti solo pensare all'uso di gold standard clinici (o misti) per alcune malattie da infezione per le quali la coltura era considerata fino a ieri il non plus ultra dei riferimenti.

Certo, ormai le regole sono note ma (secondo me) gli ostacoli per la loro applicazione qui da noi risiedono non tanto nella complessità degli adempimenti quanto nella nostra atavica carenza di requisiti per assolverli. Perché sono ineludibili (cito a caso) un'oculata gestione delle risorse umane e del loro tempo, l'ordinato sviluppo di coerenti e stabili reti operative, un'affidabile organizzazione (anche spicciola). Pratiche non certo comuni dietro la facciata dei nostri palazzi della salute. Ma anche un serio impegno, coordinato, tenace e spesso oscuro e gregario, attributi agli antipodi dell'improvvisazione, del protagonismo sensazionalistico e del campanilismo che stimolano e modellano molte iniziative nel Bel Paese.

E' questo il motivo della mia forte contrarietà ad una prematura immissione dei *likelihood ratio* nella pratica clinica da parte dei laboratori italiani. Un conto è - in parallelo con il lavoro di revisione critica - ragionare insieme ai medici (e ai pazienti) su rischi e vantaggi di una procedura diagnostica, un conto è ritenersi "all'avanguardia" e ostentare un "prodotto" con la "griffe" di Fagan. E chi controllerebbe le inevitabili emulazioni, nel quadro della legittima concorrenza?

A chi mi conosce come fautore dei *likelihood ratio* potranno sembrare strane e incoerenti queste critiche. È che vorrei per questi numeretti un adeguato posto nello strumentario diagnostico, non il futuro ostracismo per malasania.

La Spezia, ferragosto dell'anno 2004

1. Andrew G Robinson. Out with Snout. Rapid response to 'Daniel Pewsner, et al. Ruling a diagnosis in or out with "SpPIn" and "SnNOut": a note of caution. BMJ 2004; 329: 209-213 (7 August 2004)
2. Daniel Pewsner, et al. Ruling a diagnosis in or out with "SpPIn" and "SnNOut": a note of caution. BMJ 2004; 329: 209-213

Livelli di Prova e Forza delle Raccomandazioni

*Un commento a proposito di **Evidence-based Guidelines in Laboratory Medicine: Principles and Methods***

di Wytze P. Oosterhuis et al. pubblicato a maggio su Clinical Chemistry 50:5; 806-818, 2004

Nel Manuale come produrre, diffondere e aggiornare raccomandazioni per la pratica clinica del **Manuale Metodologico del progetto Progetto Nazionale Linee Guida**, pubblicato nel 2002 ed aggiornato nel maggio del 2004 (http://www.pnlg.it/doc/Manuale_PNLG.pdf), le raccomandazioni sono qualificate con un grado di Livello di Prova (LDP) e Forza delle raccomandazioni (FDR).

Livello di Prova e Forza delle Raccomandazioni sono espressi nel primo caso come numeri romani e nel secondo da lettere, come illustrato nel capitolo 8 dal titolo "sintesi del metodo adottato dal PNLG per la produzione di Linee Guida".

Il livello di prova si riferisce al tipo di studio dal quale è derivata l'informazione utilizzata e conseguentemente alla validità dell'informazione possibilmente priva di errori sistematici e metodologici.

La forza della raccomandazione si riferisce alla probabilità che l'applicazione alla pratica clinica determini un miglioramento dello stato di salute

I livelli di prova e la forza delle raccomandazioni possono essere così sintetizzati:

Livelli di prova

- I** Prove ottenute da più studi clinici controllati e/o revisioni sistematiche di studi randomizzati
- II** Prove ottenute da un solo studio randomizzato di disegno adeguato
- III** Prove ottenute da studi di coorte non randomizzati con controlli concorrenti o storici o loro metanalisi
- IV** Prove ottenute da studi retrospettivi tipo caso controllo o loro metanalisi
- V** Prove ottenute da studi di casistica "serie di casi" senza gruppo di controllo
- VI** Prove basate sull'opinione di esperti, o comitati di esperti come indicato in linee guida o consensus conference

Forza delle Raccomandazioni:

- A** Indica una particolare raccomandazione sostenuta da prove scientifiche di buona qualità, anche se non necessariamente di tipo I o II
- B** Si nutrono dubbi sul fatto che quella particolare procedura o intervento debba essere sempre raccomandato ma si ritiene che la sua esecuzione debba essere attentamente considerata

- C** Esiste una sostanziale incertezza a favore o contro la raccomandazione di eseguire la procedura od intervento
- D** L'esecuzione della procedura non è raccomandabile
- E** Si sconsiglia fortemente l'esecuzione della procedura

Tuttavia altre metodologie sono proposte e correntemente utilizzate nella formulazione di linee guida orientate al momento diagnostico in ambito internazionale.

E' questo il caso della prestigiosa National Academy of Clinical Biochemistry, la recente linea guida dell'American College of Cardiology/ American Heart Association/American Association of Clinical Chemistry con oggetto "**Biomarkers of acute coronary syndromes and heart failure**" presentata all'AACC di Los Angeles a Luglio del 2004, <http://www.nacb.org/lmpg/main.stm>, propone questa classificazione:

Indicazioni e forza delle raccomandazioni

- I** Condizioni per le quali ci sono evidenze e/o un consenso generale che una determinata procedura o trattamento è utile ed efficace
- II** Condizioni per le quali vi sono evidenze in contrasto e/o divergenze d'opinione relativamente l'utilità e/o efficacia di una procedura od un trattamento
 - II a** Il peso delle evidenze/opinioni appare in favore dell'efficacia/efficienza
 - II b** L'utilità/efficacia è di minore consistenza nelle evidenze/opinione
- III** Condizioni per le quali c'è evidenza e/o un generale consenso che le procedure/trattamento non è utile/efficace ed in alcuni casi può essere pericoloso

Peso delle evidenze

- A** Dati che derivano da studi clinici randomizzati che coinvolgono un largo numero di pazienti
- B** Dati che derivano da un numero limitato di studi che coinvolgono un piccolo numero di pazienti o da corrette analisi di studi non randomizzati o registri osservazionali
- C** Il consenso tra esperti è a base della raccomandazione

Come esempio dei diversi approcci al problema si possono citare le **National Clinical Guidelines for Type 2 Diabetes** formulate dal Royal College of General Practitioners, Effective Clinical Practice Unit, ScHAAR University of Sheffield, consultabili all'indirizzo:

<http://www.shef.ac.uk/guidelines/>

In questa linea guida, nella parte relativamente al metodo utilizzato nello sviluppo delle linee guida, la classificazione delle evidenze ed il grading delle raccomandazioni è il seguente:

Classificazione delle evidenze

- Ia** Evidenze da meta-analisi o trias controllati e randomizzati
- Ib** Evidenze da almeno un trial controllato e randomizzato
- IIa** Evidenze da almeno uno studio controllo non randomizzato
- IIb** Evidenze da almeno uno studio di altro disegno quasi sperimentale
- III** evidenze da uno studio descrittivo, come studio comparativo, studi di correlazione e studi caso controllo
- IV** evidenze prodotte da report di comitati di esperti od opinioni od esperienze di cliniche autorevoli

Grading delle raccomandazioni

- A** basate su evidenze di categoria I
- B** basate su evidenze di categoria II, o estrapolate da evidenze di categoria I
- C** basate su evidenze di categoria III o estrapolate da evidenze di categoria I o II
- D** basate su evidenze di categoria I, II o III

Se si valutano le raccomandazioni riportate nei vari capitoli di queste linee guida che hanno l'avvallo del prestigioso NICE, tutto quanto riguarda le raccomandazioni relative alle misure di laboratorio, come ad esempio l'HbA_{1c} la forza delle raccomandazioni è D, B mai A.

Quando viceversa si valutano le raccomandazioni relative all'uso dei farmaci utilizzati nel controllo glicemico nel paziente, molte delle raccomandazioni proposte sono classificate come A. Tutti i lavori su cui si fondano queste raccomandazioni A si basano sostanzialmente sulla misura dell'HbA_{1c} come indicatore di conduzione ottimale della terapia; l'emoglobina glicata è l'indicatore di esito della terapia!

Se ne può dedurre come il modello proposto è sicuramente studiato più per i trial farmacologici che per il test diagnostico ed è conseguentemente di difficile applicazione nel campo della Medicina di Laboratorio.

In **Evidence-based Guidelines in Laboratory Medicine: Principles and Methods** P. Oosterhuis et al. propongono un modello formalmente diverso forse metodologicamente più utile a definire i livelli di evidenza e le raccomandazioni per valutare una procedura diagnostica.

Qualità degli studi primari e revisioni sistematiche: rating del livello delle evidenze

- Ia** Metaanalisi o review sistematiche basate su più studi di livelli **Ib**
- Ib** Trial diagnostici or studi di esito di buona qualità
- II** Trial diagnostici or studi di esito di media qualità, numero insufficiente di pazienti, o altri trials (case-control, altri designi)
- III** Studi descrittivi, case report ed altri studi
- IV** Indicazioni di comitati, opinioni di esperti, e così di seguito (reviews non sistematiche etc.)

Rating della forza delle evidenze a supporto delle raccomandazioni nelle linee guida

- A** Supportati da almeno due studi di livello Ib o da una review di livelli Ia (“E’ stato dimostrato”)
- B** Supportate da almeno due studi indipendenti di livello II o estrapolazioni da studi di livelli I (“E’ plausibile”)
- C** Non supportati da adeguati studi di livello I o II (“indicazioni”)
- D** Indicazioni di esperti (“non ci sono prove”)

E’ doveroso considerare le specificità del test diagnostico quando lo si voglia valutare sia come livello di evidenza della pubblicazione scientifica esaminata sia come conseguente raccomandazione da inserire in una linea guida.

La proposta appare di grande interesse ed utilità per una valutazione EBLM del test diagnostico e presumibilmente sarà proposta dal Comitato IFCC sulla Evidence Based Laboratory Medicine per essere adottata in Medicina di Laboratorio.

(Tommaso Trenti, Dipartimento di Patologia Clinica, Ausl di Modena, t.trenti@ausl.mo.it)

8

Sintesi del metodo adottato dal PNLG per la produzione di linee guida

Il Programma nazionale per le linee guida (PNLG) ha adottato la metodologia che è qui riassunta. Gli esempi sono tratti dalla *Linea guida sul trattamento della sindrome influenzale*.

Scelta degli argomenti. Il metodo epidemiologico (rilevanza in termini di incidenza, mortalità eccetera), quello della disponibilità delle prove (analizzando su banche dati la disponibilità di ultimi dati di efficacia di interventi sui problemi sanitari individuati) e pareri di operatori vengono combinati in modo semi-strutturato. Seguendo questo metodo misto si è individuata la sindrome influenzale come tema sanitario ricorrente, di grande impatto per il paese e per il quale vi è una base di prova strutturata crescente.

Scelta dei gruppi di lavoro. La scelta si basa sulle competenze specifiche degli esperti, sul loro curriculum e sulla rappresentatività in relazione all'argomento. Per il trattamento della sindrome influenzale sono stati scelti epidemiologi, pneumologi, infettivologi, farmacoepidemiologi e clinici pediatrici, rappresentanti di società scientifiche di medici e pediatri di famiglia, farmacisti e igienisti.

Definizione dei quesiti scientifici e piano di lavoro. I quesiti scientifici a cui dare risposta vengono definiti in sede di gruppo di lavoro durante la prima riunione del *panel* di esperti, nella quale vengono anche definiti i principali quesiti per la gestione del problema clinico. Il primo nodo decisionale nella gestione della sindrome influenzale ha riguardato l'esame della possibilità di diagnosticare con certezza ragionevole l'agente infettivo in questione. Fa parte di questa fase l'esame della banca dati CeVEAS alla ricerca di analoghe LG e dei più importanti quesiti clinici. La definizione dei quesiti scientifici viene discussa con i rappresentanti delle società scientifiche. Il primo quesito clinico individuato per la gestione della sindrome influenzale ha riguardato la possibilità per il medico di famiglia di fare diagnosi differenziale dell'agente infettivo in tempo reale tramite *kit* diagnostici rapidi.

Raccolta delle prove. Il PNLG ha adottato una struttura gerarchica di ricerca e valutazione delle evidenze. Le prove di efficacia e sicurezza vengono ricercate innanzitutto nella Cochrane Library (Clib)¹, e più precisamente nella banca dati di revisioni sistematiche (CDSR) prodotte dai Gruppi Cochrane o da altri (DARE). Qualora questo tentativo si riveli infruttuoso viene di nuovo consultata la Clib e altre banche dati bibliografiche per identificare singoli *trial* e/o studi comparativi non randomizzati.

Redazione ed evoluzione della bozza di LG. Una volta raccolte e valutate le prove (che vengono sintetizzate su apposite schede), il comitato di redazione procede a una prima stesura della LG e delle raccomandazioni. Questa bozza viene fatta circolare tra tutti coloro che contribuiscono al progetto (*panel*, rappresentanti società scientifiche e comitato di redazione del PNLG) per eventuali commenti. La bozza viene rivista e cambiata sulla base dei commenti ricevuti; pro-

blemi e opinioni ulteriori vengono esaminati durante la seconda e ultima riunione plenaria del gruppo di lavoro e con i rappresentanti delle società scientifiche.

Assegnazione *grading*. Nelle linee guida, le raccomandazioni vengono qualificate con un certo grado di Livello di prova (LDP) e di Forza della raccomandazione (FDR), espressi rispettivamente in numeri romani (da I a VI) e in lettere (da A a E). Il livello di prova si riferisce alla probabilità che un certo numero di conoscenze sia derivato da studi pianificati e condotti in modo tale da produrre informazioni valide e prive di errori sistematici. La FDR si riferisce invece alla probabilità che l'applicazione nella pratica di una raccomandazione determini un miglioramento dello stato di salute della popolazione obiettivo cui la raccomandazione è rivolta. Esistono diversi sistemi di gradazione per le prove di efficacia e per la forza delle raccomandazioni riportati in letteratura. Il sistema adottato dal PNLG si basa sulla rielaborazione messa a punto dal Centro per la valutazione della efficacia dell'assistenza sanitaria (CeVEAS). Questo sistema ha come principale caratteristica il fatto che la forza delle raccomandazioni non si basa soltanto sul tipo di disegno di studio ma tiene conto anche di altri fattori quali la fattibilità, l'accettabilità e l'economicità dell'intervento.

Presentazione editoriale della LG. L'autorevolezza e la chiarezza delle raccomandazioni, l'uso di un linguaggio non ambiguo e di termini ben definiti (anche attraverso un glossario) sono condizioni necessarie ad assicurare il successo di uno strumento che si propone di assistere e indirizzare le decisioni cliniche.

Occorre anche che le raccomandazioni siano presentate in una forma che ne renda facile e rapida la consultazione, come pure l'apprendimento e la memorizzazione. E' dunque opportuno definire la struttura generale dei documenti e le caratteristiche di forma e stile (dal punto di vista dei testi e della grafica) che dovranno sempre essere rispettate, pur con le variazioni che potranno essere suggerite dalla specificità dei temi trattati. A questo proposito si propone la seguente struttura.

Il corpo del documento deve essere preceduto da alcune informazioni essenziali:

1. L'indicazione degli autori delle raccomandazioni, con la specificazione della qualifica e del ruolo svolto nell'elaborazione.
2. L'indicazione di eventuali finanziamenti.
3. Un'avvertenza («Note per gli utilizzatori», da ripetere sempre uguale) dove si riassume il significato delle linee guida nazionali. Nell'avvertenza si ricorda che le decisioni cliniche sul singolo paziente richiedono l'applicazione delle raccomandazioni, fondate sulle migliori prove scientifiche, alla luce dell'esperienza clinica e di tutte le circostanze di contesto.
4. La data in cui il documento è stato redatto e, quando possibile, l'indicazione della durata di prevedibile validità, prima degli aggiornamenti.
5. Introduzione al documento nella quale si definiscono, con dettaglio maggiore rispetto a quello possibile nel titolo, l'argomento e l'ambito delle linee guida. Inoltre si definiscono gli scopi e i destinatari.

6. Corpo del documento. L'esposizione deve iniziare con una esplicita identificazione delle decisioni chiave che si pongono nella pratica clinica a proposito della categoria di pazienti e degli interventi in esame. A ogni snodo decisionale rilevante si associano le raccomandazioni che gli autori delle linee guida hanno formulato sulla base di una valutazione critica delle migliori prove empiriche disponibili. Ogni raccomandazione è accompagnata da *grading* e, per quanto possibile, da una esplicita formulazione del vantaggio atteso, in termini di miglioramento della prognosi o, in forma subordinata, di altri obiettivi intermedi e dagli indicatori di monitoraggio.
7. In appendice verranno riportate le sintesi delle principali prove, corredate dal metodo seguito.
8. La bibliografia chiude la pubblicazione.

LIVELLI DI PROVA

Prove di tipo

I	Prove ottenute da più studi clinici controllati randomizzati e/o da revisioni sistematiche di studi randomizzati.
II	Prove ottenute da un solo studio randomizzato di disegno adeguato.
III	Prove ottenute da studi di coorte non randomizzati con controlli concorrenti o storici o loro metanalisi.
IV	Prove ottenute da studi retrospettivi tipo caso-controllo o loro metanalisi.
V	Prove ottenute da studi di casistica («serie di casi») senza gruppo di controllo.
VI	Prove basate sull'opinione di esperti autorevoli o di comitati di esperti come indicato in linee guida o consensus conference, o basata su opinioni dei membri del gruppo di lavoro responsabile di queste linee guida.

FORZA DELLE RACCOMANDAZIONI

A	L'esecuzione di quella particolare procedura o test diagnostico è fortemente raccomandata. Indica una particolare raccomandazione sostenuta da prove scientifiche di buona qualità, anche se non necessariamente di tipo I o II.
B	Si nutrono dei dubbi sul fatto che quella particolare procedura o intervento debba sempre essere raccomandata, ma si ritiene che la sua esecuzione debba essere attentamente considerata.
C	Esiste una sostanziale incertezza a favore o contro la raccomandazione di eseguire la procedura o l'intervento.
D	L'esecuzione della procedura non è raccomandata.
E	Si sconsiglia fortemente l'esecuzione della procedura.

Promulgazione e disseminazione della LG. La LG viene promulgata in data prestabilita tramite il sito www.pnlg.it, riviste delle società scientifiche e altri media cartacei e corsi di formazione per operatori sanitari. La valutazione dell’impatto e rilevanza della LG viene fatto tramite feed-back di utenti sul sito www.pnlg.it e in futuro con studi *ad hoc*.

Aggiornamento della LG. Le linee guida del Programma nazionale linee guida vengono aggiornate ogni due anni, eccetto nei casi in cui vi siano stati errori, omissioni importanti ovvero quando la base di evidenza sia mutata in maniera tale da rendere le raccomandazioni o il loro *grading* obsoleti.

8.1 Note

1. Consultabile al sito internet www.cochranelibrary.com

APPLICAZIONE PRATICA DELL'APPROCCIO EBM IN LABORATORIO: UN CASO DI SPONDILITE ANCHILOSANTE.

Renzo Gianello, Danilo Gervasoni**, *Responsabile Laboratorio Analisi chimico-cliniche e microbiologiche, Fondazione "S. Maugeri"-Clinica del lavoro e della Riabilitazione-I.R.C.C.S.-sede di Lumezzane (Bs); ** Responsabile Reparto di ortopedia, Ospedale di Esine (Bs)*

►DESCRIZIONE DEL CASO CLINICO:

Il paziente G.A.P., di anni 53, giungeva alla nostra osservazione, per la prima volta, circa 10 anni fa, in seguito ad una lunga storia clinica di dolore alla colonna vertebrale che accusava dall'età di 5 anni, allorché gli veniva diagnosticata una cifoscoliosi con compromissione della postura e della deambulazione.

La terapia fisica e l'adozione di un corsetto portava ad una parziale risoluzione del quadro clinico. All'età di 24 anni presenta una importante lombosciatalgia che induce il medico curante al ricovero presso gli O.C. di Brescia.

Dopo varie cure mediche, senza apprezzabili risultati, ed accertamenti diagnostici viene sottoposto ad intervento chirurgico di laminectomia bilaterale (L5-S1) che risolve completamente la sciatalgia ma non il dolore lombo-sacrale, solo parzialmente ridotto da alcuni cicli di terapia fisica e dall'uso di FANS.

Successivamente il paziente presenta una graduale e progressiva rigidità della colonna vertebrale ed all'età di 38 anni manifesta la comparsa di lesioni cutanee ascrivibili, a detta dello specialista dermatologo, ad una dermatite psoriasica.

Dopo varie traversie, giunge al ricovero presso la nostra struttura ospedaliera all'età di 44 anni, per una recrudescenza della sintomatologia dolorosa a carico della colonna lombo-sacrale.

Dopo uno studio clinico-anamnestico approfondito, vengono eseguiti opportune ed appropriate indagini ematochimiche (che illustreremo in seguito) e, quindi, una consulenza dello Specialista Reumatologo.

Quest'ultimo, sulla base del quadro clinico, degli esami di laboratorio e del quadro Radiologico, pone diagnosi di Spondilite Anchilosante di probabile origine psoriasica, consigliando una terapia a base di Salazopirina e di FANS (Indometacina).

Nel corso degli anni la terapia con Salazopirina è stata abbandonata perché ritenuta inefficace, mentre il paziente continua tuttora la terapia con Indometacina, con benefici alterni.

D'altro canto, la psoriasi non ha trovato soluzione con terapia alcuna, tanto che, attualmente, essa interessa circa 1/3 della superficie cutanea.

Due anni fa circa, in seguito ad incidente automobilistico, il soggetto si procura la frattura di D12, di L1 e dell'omero destro.

In questa occasione viene eseguita una indagine RNM che conferma la fusione dei corpi vertebrali al tratto dorso-lombare ed una stenosi secondaria del canale vertebrale.

L'esame clinico attuale evidenzia una fusione della colonna in toto, la presenza di algie a numerose articolazioni ed una evidente compromissione della capacità deambulatoria che peggiora sensibilmente la qualità di vita.

E' inoltre rilevabile un incremento ponderale importante (110 kg circa), ipercifosi dorsale e iperlordosi cervicale.

L'autonomia deambulatoria è scarsa e possibile solo con l'uso di bastoni.

Il rachide è molto rigido, soprattutto nel tratto lombo-sacrale mentre la dermatopatia desquamativa è ormai diffusa a dorso, braccia, addome, cosce e gambe.

I parametri funzionali articolari di gomiti, spalle, polsi e ginocchia sono assai ridotti su tutti i piani.

►DATI DI LABORATORIO:

I risultati delle indagini di laboratorio rivelano un modestissimo incremento degli indici di attività infiammatoria, con VES: 20 mm, F.R. negativo, PCR negativa, acido urico: 8,3 mg%ml.

Sono stati eseguiti anche test di carattere immunologico che hanno evidenziato una lieve positività dell'ANA test, un anti-DNA negativo, anti-mitocondri negativi, ma soprattutto la positività dell'HLA B27 (Antigene di Istocompatibilità B27).

► **APPROCCIO EBM AL CASO CLINICO**

Le indagini di laboratorio appena descritte impongono una diagnosi differenziale tra varie patologie facenti parte del capitolo delle spondiloartriti (o spondiloartropatie) sieronegative.

Tali artropatie sono state recentemente riordinate dal punto di vista classificativo da Amor (1990) e dall' E.S.S.G. (European Spondyloarthropaty Study Group) nel 1991 con lo scopo di poter agevolmente collocare forme patologiche che, specie nelle fasi iniziali di malattia, risulterebbero indifferenziate tanto da non trovare una adeguata collocazione nosologica.

Nel nostro caso la diagnosi differenziale si pone tra la Spondilite Anchilosante (associata a Psoriasi), l'Artrite Reattiva, l'Artrite Psoriasica con variante Spondilitica e le Artriti associate a patologie infiammatorie autoimmuni dell'intestino.

Tra queste abbiamo escluso l'Artrite Reattiva per la mancanza, nell'Anamnesi, di una recente infezione, di una artrite acuta o di una poliartrite associata ad uretrite e congiuntivite (S. di Reiter); Per analoghe ragioni abbiamo escluso l'Artrite associata a malattie infiammatorie intestinali (come la Rettocolite Ulcerosa e il M. di Crohn) non presentando il paziente i segni di una compromissione intestinale, alla quale peraltro, in questi casi, la spondiloartrite si associa solitamente nel 5-10% dei casi (1).

Molto più arduo, nel nostro caso, è distinguere tra la Spondilite Anchilosante (S.A.) ed una forma di Artrite Psoriasica (A.P.) a prevalente manifestazione spondilitica.

Quest'ultima, in caso di marcata componente spondilitica associata a positività dell'antigene HLAB27, è indistinguibile dalla Spondilite Anchilosante vera e propria (come nel caso clinico qui presentato).

E' inoltre da sottolineare che nella percentuale variabile del 10-15% la Psoriasi possa insorgere dopo la patologia articolare (1).

Di una qualche utilità diagnostica è la presenza di una Sacroileite, che manifesta aspetti diversi nelle due patologie: infatti nella S.A. la localizzazione sacroileitica si manifesta di solito bilateralmente (simmetrica all'indagine radiologica), mentre nella A.P. è più spesso monolaterale (asimmetrica).

Per questi motivi, si è ricorso alla stesura dei cosiddetti "Criteri di New York "(1982) che considerano sia le manifestazioni sacroileitiche che precisi criteri diagnostici clinico-biochimici.

In particolare, questi ultimi evidenziano l'associazione tra una sacroileite monolaterale di grado III-IV o bilaterale di grado II-IV con una lombalgia infiammatoria presente da almeno 3 mesi, oppure una limitazione funzionale del Rachide sia sul piano sagittale che su quello frontale, oppure una ridotta espansibilità toracica (2).

Ciò presuppone una stretta collaborazione con lo Specialista Radiologo, che deve essere particolarmente sensibilizzato ed informato sui criteri clinico-radiologici che si manifestano in tali patologie.

Dal punto di vista epidemiologico l'incidenza della A.P. varia dal 5-7% sui pazienti ambulatoriali al 30% sui pazienti ospedalizzati (che generalmente presentano manifestazioni cliniche più serie (2), mentre la incidenza della S.A. è nettamente inferiore, con percentuali variabili dallo 0,1 allo 0,5 % e con incidenza prevalente nelle fasce di età giovanile (15-35 anni), nel sesso maschile (9:1), a differenza della A.P (2).

Alla luce di questi dati ed alla lunga storia clinica del paziente in oggetto, possiamo ragionevolmente porre una probabilità pre-test, al fine di una adeguata applicazione del Nomogramma di Fagan e delle indicazioni della Medicina basata sulle evidenze (EBM), di almeno il 25-30 %.

Per quanto attiene ai test diagnostici che abbiamo considerato più significativi al fine di una diagnosi il più precoce possibile abbiamo scelto le indagini radiologiche (Radiografia del bacino e delle art. sacroiliache) e la ricerca dell'HLA B27.

Purtroppo la letteratura reperibile su questo argomento si è rivelata avara di dati di un certo rilievo e pertanto cercheremo di discutere in modo critico i dati a disposizione.

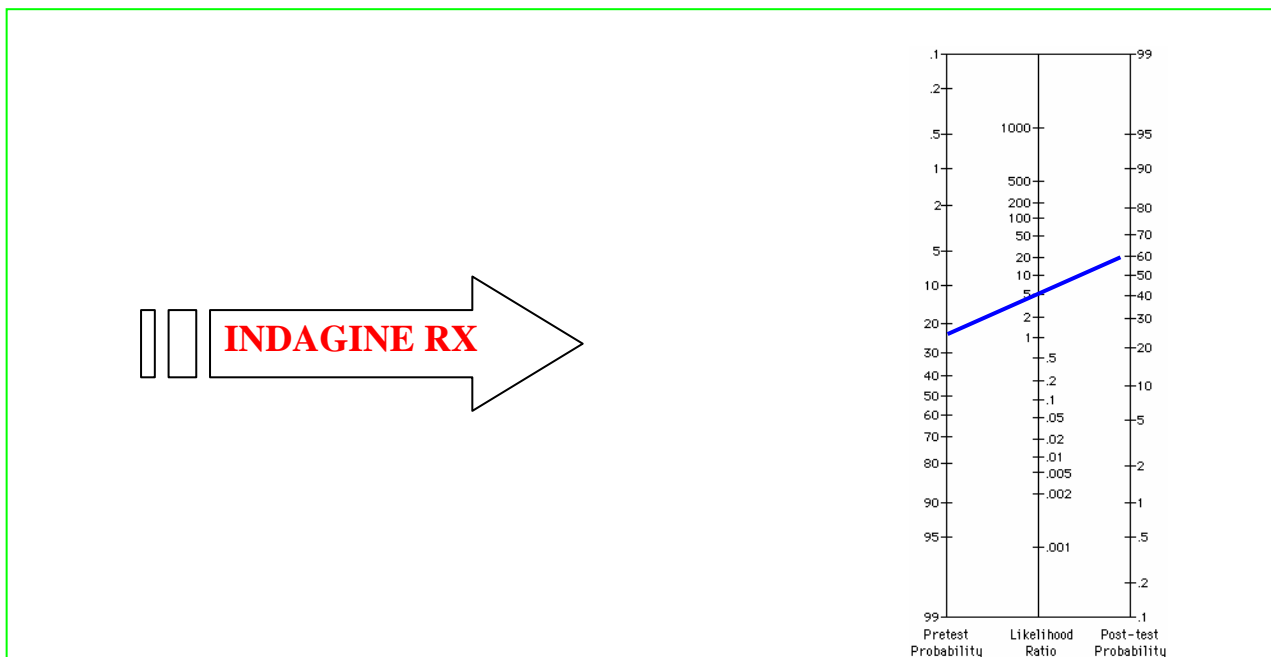
A) Il dato radiologico di Sacroileite risulta associato alla Psoriasi, secondo alcuni A.A., nel 20% dei casi (di cui l'8% è asintomatico) (3), secondo altri nel 14% dei casi (di cui il 7% affetti da S.A. giàdiagnosticata) (4), ed infine addirittura nel 78% dei casi in un recente studio randomizzato multicentrico (livello I° in base all'approccio EBM) (5).

Riguardo invece all'associazione tra Sacroileite e S.A., essa è stata riscontrata nel 79% dei casi da alcuni A.A.(6), nel 19% dei casi da altri (7), ed infine nel 72% dei casi in uno studio multicentrico del 1985 (8).

L'ACP ha pubblicato recentemente uno studio (9) sulla correlazione tra indagine radiologica classica e presenza di spondiloartropatia: la sensibilità è relativamente bassa (50%), la specificità molto buona (90%) mentre l' LR+ è discreto (5,0) e l'LR- è scarso (0,56).

Possiamo concludere quindi che tale tipo di indagine può essere considerato buono ai fini di uno screening, dato la bassa sensibilità ed il basso potere di esclusione della malattia.

L'introduzione della TAC ha migliorato sensibilità e capacità di esclusione ma ha ridotto la specificità dell'esame radiologico (sensibilità: 80%, specificità: 70%, LR+: 2,7, LR-: 0,29) (9). L'applicazione pratica del Nomogramma di Fagan, nel nostro caso specifico, presupponendo la presenza di segni radiologici precoci di spondiloartropatia, porta ad un risultato non particolarmente esaltante (60%) relativamente alla probabilità post-test, e rende necessaria l'introduzione di un test diagnostico (che vedremo successivamente) che abbia maggiori possibilità di discriminazione.



B) Per quanto concerne il test diagnostico da noi preso in considerazione, il dosaggio dell'HLA B27, esso si associa nel 30% dei casi all'Artrite Psoriasica (soprattutto nei pazienti affetti dalla variante spondilitica) (1).

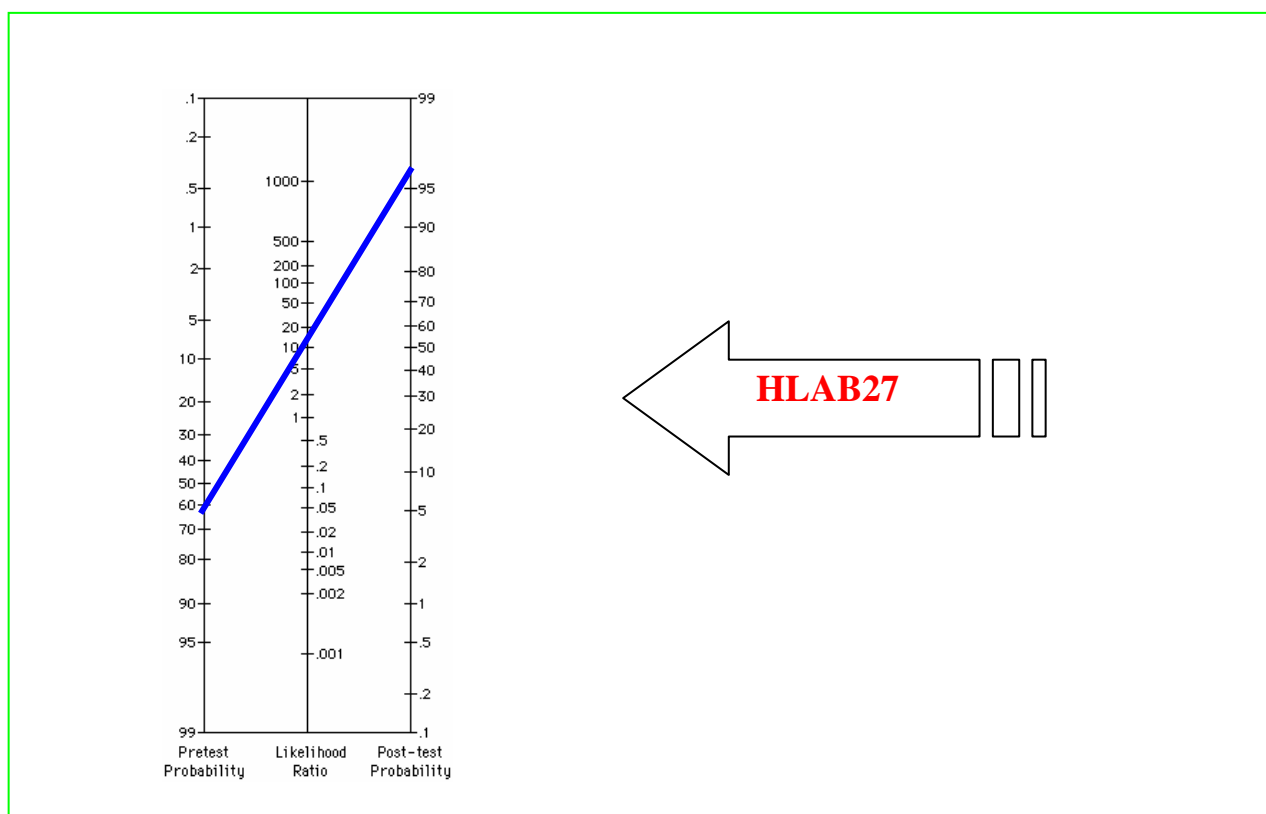
Ben più elevata è l'associazione tra S.A. e positività dell'HLA B27: nella popolazione caucasica, infatti, tale associazione può arrivare anche al 96% (2; 8;10;11;16) a seconda delle casistiche.

Dati gli elevati costi dell'esecuzione di questo test è improponibile il suo utilizzo 2 "a tappeto" ma deve essere riservato ai casi di difficile interpretazione, laddove risultano insufficienti i criteri di New York.

Dati cumulativi pubblicati qualche anno fa (12) hanno evidenziato una elevata sensibilità e specificità (92%), un valore di LR+ molto buono (11,5), ed un ottimo LR-, in grado di apportare delle modifiche conclusive alla probabilità post-test della malattia (LR-: 0,09).

Dati di questo tipo permettono al test di avere un elevato potere discriminante ed una elevata capacità di esclusione della malattia.

Se, infatti, applichiamo il Nomogramma di Fagan al nostro caso clinico, utilizzando come punto di partenza la probabilità pre-test già calcolata dalla indagine radiologica (60%), otteniamo una probabilità post-test molto promettente, ed altamente significativa (96% circa), cioè:



E' di fondamentale importanza comunque sottolineare quanto sia importante pervenire, in casi simili, ad una diagnosi quanto più precoce possibile.

A tal riguardo un recente studio epidemiologico (13) ha preso in considerazione un gruppo di 92 pazienti affetti da S.A. della durata media di 15 anni e nei quali la diagnosi era stata posta dopo circa 6 anni dall'esordio (e quindi piuttosto tardivamente): negli ultimi 10 anni la loro qualità di vita, come nel caso clinico qui illustrato, era notevolmente peggiorata non essendo stati sottoposti a terapia adeguata.

Nonostante la recente introduzione di farmaci innovativi (ad es. Etanercept), la terapia fisica rappresenta un sistema efficace per il controllo della progressione della S.A..

Essa si avvale del controllo posturale, della ginnastica respiratoria, della mobilizzazione della colonna vertebrale, della rieducazione funzionale e dell'idroterapia: se introdotte nelle fasi iniziali, la loro efficacia risulta migliore nel ridurre il livello di disabilità e nel raggiungere una qualità di vita più che accettabile.

Tutto questo viene ottenuto, oltretutto, con un minor dispendio di risorse economiche da parte del S.S.N..

BIBLIOGRAFIA

- 1) Betterle C. : Le malattie autoimmuni, 2001, Ed. Piccin.
- 2) Todesco S., Gambari P.F.: Malattie Reumatiche-II edizione, Mc Graw-Hill, 1998.
- 3) Harvie J.N. et al.: Sacroiliitis in severe Psoriasis. Am. J. Roentgenol. 127,579-840, 1976.
- 4) Maldonado-Cocco et al.: Prevalence of Sacroiliitis and Ankylosing Spondylitis in psoriasis patients-J.Rheumatol.5,311-313, 1978.
- 5) Battistone M. et al: The prevalence of Sacroiliitis in Psoriatic Arthritis: new perspectives from a large, multicenter cohort. A department of veterans affairs cooperative study-Skelet.Radiol.28,196-201, 1999.
- 6) Bergfeldt L. et al.: Ankylosing Spondylitis: an important cause of severe disturbances of the cardiac conduction system. Prevalence among 223 pacemaker-treated men. Am.J.Med 73, 187-191,1982.
- 7) Prohaska E.: Isolated Sacroiliitis as monosymptomatic form of Ankylosing Spondylitis. A possible cause of chronic back-pain.Clin.Rheumatol.,3,33-37,1984.
- 8) Goie the HS et al. : Evaluation of diagnostic criteria for Ankylosing Spondylitis: a comparison of the Rome, New York and modified New York criteria in patients with a positive clinical history screening test for Ankylosing Spondylitis.Br.J.Rheumatol. 24. 242-249,1985.
- 9) Black E.R. et al.: Diagnostic strategies-Philadelphia, ACP,1999.
- 10) Isomaki H. et al.:HLAB27 and Artrytis.Ann.Clin.Res., 7,138-145,1975.
- 11) Goodman C.E. et al.:Ankylosing Spondylitis in women.Arch.Phys.Med.Rehabil.,61,167-170,1980.
- 12) Goldenberg K. et al.: Diagnostic Testing Handbook for clinical decision making-Chicago, Year Medical Publisher,1988.
- 13) Jajic Z. et al.:Clinical features of Ankylosing Spondylitis.Epidemiological Study. Reumatiz. Am., 45,13-80,1997.